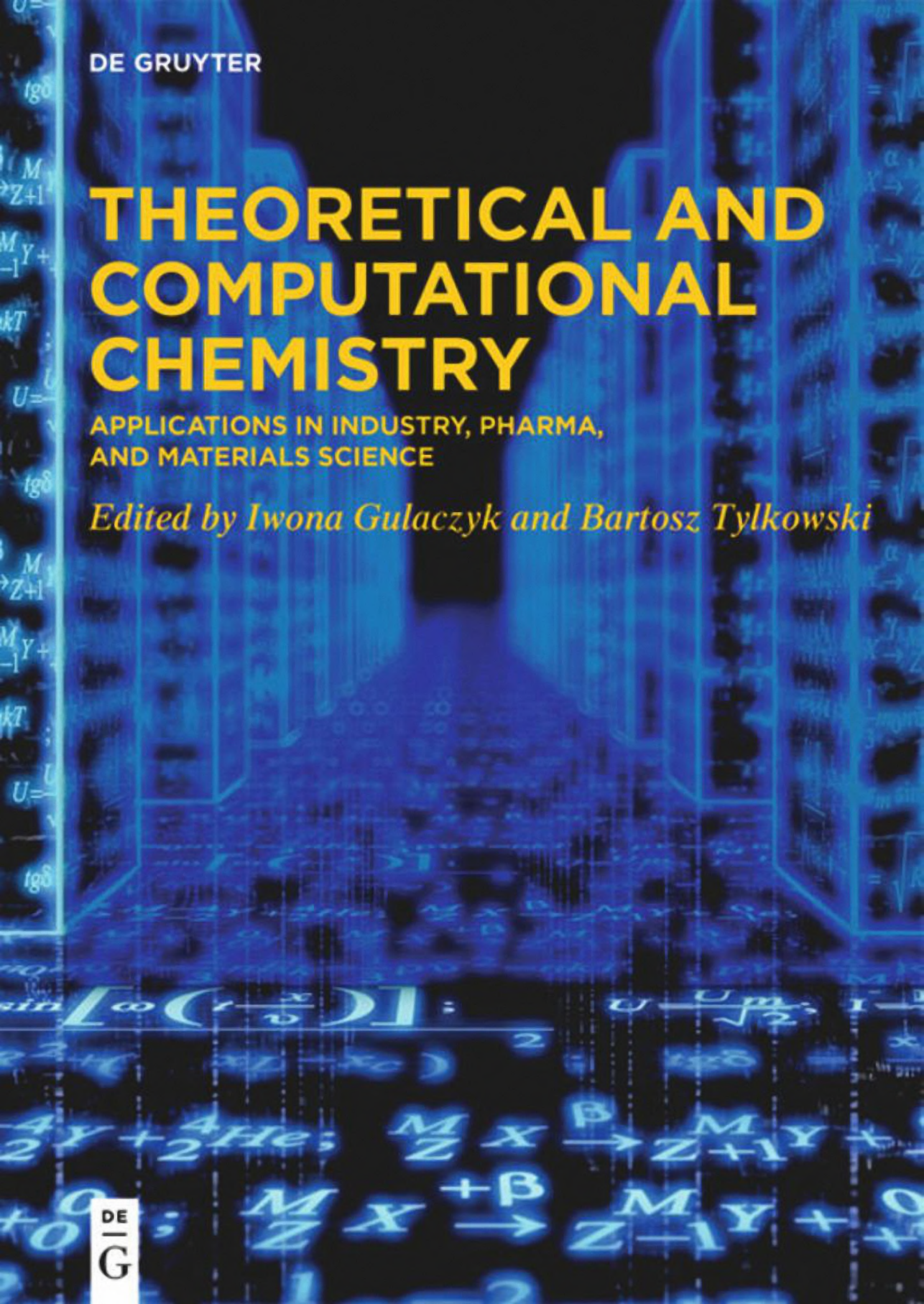
DE GRUYTER

THEORETICAL AND COMPUTATIONAL CHEMISTRY

APPLICATIONS IN INDUSTRY, PHARMA, AND MATERIALS SCIENCE

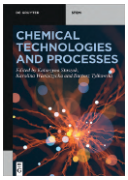
Edited by Iwona Gulaczyk and Bartosz Tylkowski



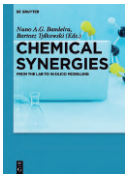
Iwona Gulaczyk, Bartosz Tylkowski (Eds.)

Theoretical and Computational Chemistry

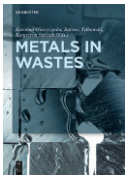
Also of interest



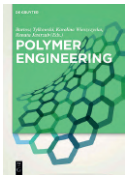
Chemical Technologies and Processes
Staszak, Wieszczycka, Tylkowski (Eds.), 2020
ISBN 978-3-11-065636-7, e-ISBN 978-3-11-065627-5



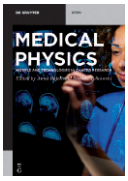
Chemical Synergies – From the Lab to In Silico Modelling
Bandeira, Tylkowski (Eds.), 2018
ISBN 978-3-11-048135-8, e-ISBN 978-3-11-048206-5



Metals in Wastes
Wieszczycka, Tylkowski, Staszak (Eds.), 2018
ISBN 978-3-11-054628-6, e-ISBN 978-3-11-054706-1



Polymer Engineering
Tylkowski, Wieszczycka, Jastrzab (Eds.), 2017
ISBN 978-3-11-046828-1, e-ISBN 978-3-11-046974-5



Medical Physics – Models and Technologies in Cancer Research
Bajek, Tylkowski (Eds.), 2021
ISBN 978-3-11-066229-0, e-ISBN 978-3-11-066230-6

Theoretical and Computational Chemistry

Edited by
Iwona Gulaczyk and Bartosz Tylkowski

DE GRUYTER

Editors
Dr. Iwona Gulaczyk
Adam Mickiewicz University in Poznań,
Faculty of Chemistry
ul. Uniwersytetu Poznańskiego 8,
61-614 Poznań
Poland
iwona.gulaczyk@amu.edu.pl

Dr. Bartosz Tylkowski
Eurecat, Centre Tecnològic de Catalunya
Marcel·lí Domingo s/n
43007, Tarragona
Spain
bartosz.tylkowski@eurecat.org

ISBN 978-3-11-067815-4
e-ISBN (PDF) 978-3-11-067821-5
e-ISBN (EPUB) 978-3-11-067828-4

Library of Congress Control Number: 2021933889

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck
Cover image: alengo/E+/Getty Images

www.degruyter.com

Preface

Modern research in the chemical sciences seeks not only to make useful molecules and materials but to understand, design, and control their properties. Theory is at the very center of this effort, providing the framework for an atomic and molecular level description of chemical structure and reactivity that forms the basis for interpreting experimental data and provides guidance toward new experimental directions. We decided to edit this book for those who see the potential benefit of using theoretical and computational chemistry but need practical insight into using the know-how. This book aims to review the state of the art and to provide the readers with a comprehensive and in-depth understanding of recent developments and innovation analysis of the subject. Drug design is an expensive and time-consuming process. Any method that allows reducing the time and the costs of the drug development project can have great practical value for the pharmaceutical industry. Chapter 1 describes computational methods for calculation of protein-ligand binding affinities in structure-based drug design. Chapters 2 emphasizes understanding (coupled) large amplitude motions, while Chapter 3 explains floppy molecules—their internal dynamics, spectroscopy, and applications. Chapter 4 critically discusses finding regarding the mechanisms of synthetic statins phototransformation. Artificial intelligence methods are the subjects of Chapter 5, while Chapter 6 deals with application of enzyme kinetics modelling in food, pharmaceutical and bioethanol industry. Computational approach to the study of morphological properties of polymer/fullerene blends in photovoltaics is detailed in Chapter 7, while modeling and assessment of the transfer effectiveness in integrated bioreactor with membrane separation are given in Chapter 8. We believe that the large number of references, to all significant topics mentioned, should make this book useful not only to undergraduates but also to graduate students and academic and industrial researchers.

Having an idea is one thing, turning it into a book is tough, however much satisfying!

We wish to acknowledge all contributing authors for making this book project a success.

Iwona Gulaczyk and Bartosz Tylkowski

https://doi.org/10.1515/9783110678215-202

Contents

List of contributing authors

Bartosz Tylkowski
Eurecat Technology Centre of Catalonia
Chemical Technology Unit
C/ Marcel·lí Domingo 2
43007 Tarragona, Spain

Daniela Dzhonova
Institute of Chemical Engineering
Bulgarian Academy of Sciences
Acad. G. Bonchev Str.
Bl. 103
1113 Sofia
Bulgaria

Donata Pluskota-Karwatka
Faculty of Chemistry
Adam Mickiewicz University in Poznań
ul. Uniwersytetu Poznańskiego 8,
61-614 Poznań
Poland

Francis M. Gaitho
Department of Physics
Masinde Muliro University of Science and Technology
P. O. Box 190- 50100
Kakamega- Kenya

Genene T. Mola
School of Chemistry and Physics
University of KwaZulu-Natal
Pietermaritzburg campus
Private Bag X01, Scottsville 3209
South Africa

Giuseppe Pellicane
Dipartimento di Scienze Biomediche
Odontoiatriche e delle Immagini
Morfologiche e Funzionali
Università degli Studi di Messina
98125 Messina
Italy

Ha Vinh Lam Nguyen
Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA)
CNRSUMR7583
Université Paris-Est Créteil, Université de Paris+
Institut Pierre Simon Laplace
61 avenue du Général de Gaulle
F-94010, Créteil
France

Irene Tsibranska
Institute of Chemical Engineering
Bulgarian Academy of Sciences
1113 Sofia
Bulgaria

Isabelle Kleiner
Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA)
CNRS UMR 7583,
Université Paris-Est Créteil
Université de Paris
Institut Pierre Simon Laplace
61 avenue du Général de Gaulle
F-94010, Créteil
France

Iwona Gulaczyk
Faculty of Chemistry
Adam Mickiewicz University in Poznań
ul. Uniwersytetu Poznańskiego 8,
61-614 Poznań
Poland

Josep M. Montornés
Eurecat Technology Centre of Catalonia
Chemical Technology Unit
C/ Marcel·lí Domingo 2
43007 Tarragona, Spain

https://doi.org/10.1515/9783110678215-204

Maciej Staszak
Institute of Chemical Technology and Engineering
Poznan University of Technology
ul. Berdychowo 4,
60-965 Poznań
Poland

Magdalena Olkiewicz
Eurecat Technology Centre of Catalonia
Chemical Technology Unit
C/ Marcel·lí Domingo 2, 43007 Tarragona
Spain

Marcin Hoffmann
Faculty of Chemistry
Adam Mickiewicz University in Poznań
ul. Uniwersytetu Poznańskiego 8,
61-614 Poznań
Poland

Marek Kręglewski
Faculty of Chemistry
Adam Mickiewicz University in Poznań
ul. Uniwersytetu Poznańskiego 8,
61-614 Poznań
Poland

Nadezhda Dermendzhieva
Institute of Chemical Engineering
Bulgarian Academy of Sciences
1113 Sofia
Bulgaria

Ricard Garcia-Valls
Eurecat Technology Centre of Catalonia
Chemical Technology Unit
C/ Marcel·lí Domingo 2
43007 Tarragona, Spain

Serafim Vlaev
Institute of Chemical Engineering
Bulgarian Academy of Sciences
1113 Sofia
Bulgaria

Stela Panyovska
Institute of Chemical Engineering
Bulgarian Academy of Sciences
1113 Sofia
Bulgaria

Zbigniew Dutkiewicz
Department of Chemical Technology of Drugs,
Poznan University of Medical Sciences
ul. Grunwaldzka 6,
60-780 Poznań
Poland

Zbigniew Dutkiewicz

1 Computational methods for calculation of protein-ligand binding affinities in structure-based drug design

Abstract: Drug design is an expensive and time-consuming process. Any method that allows reducing the time the costs of the drug development project can have great practical value for the pharmaceutical industry. In structure-based drug design, affinity prediction methods are of great importance. The majority of methods used to predict binding free energy in protein-ligand complexes use molecular mechanics methods. However, many limitations of these methods in describing interactions exist. An attempt to go beyond these limits is the application of quantum-mechanical description for all or only part of the analyzed system. However, the extensive use of quantum mechanical (QM) approaches in drug discovery is still a demanding challenge. This chapter briefly reviews selected methods used to calculate protein-ligand binding affinity applied in virtual screening (VS), rescoring of docked poses, and lead optimization stage, including QM methods based on molecular simulations.

Keywords: binding free energy, molecular docking, molecular dynamics, quantum mechanics, semi-empirical methods, structure-based drug design

1.1 Introduction

Drug design is a complicated, long-term, and very costly task [1]. On average, it takes 10 to 15 years from the start of work on a new drug to its launch [2], and costs reach up to $ 4 billion (for The Food and Drug Administration [FDA] approved drugs) [3]. The optimization of each lead compound selected as a hit in screening procedures costs about $ 150 million [3], [4].

About 44% of drug discovery projects fail at preclinical studies because of difficulties in identifying a ligand with desired properties (failure in ligand design) [5]. Therefore, techniques that lead the team to identify ligands with satisfactory properties more quickly and efficiently can potentially more than double the probability of successful discovery of a preclinical drug [6].

This article has previously been published in the journal Physical Sciences Reviews. Please cite as: Dutkiewicz, Z. Computational methods for calculation of protein-ligand binding affinities in structure-based drug design *Physical Sciences Reviews* [Online] 2021, 6 DOI: 10.1515/psr-2020-0034

https://doi.org/10.1515/9783110678215-001

Any method that allows to reduce the time of research and reduce the costs of the entire process can have great practical value. In computer-aided drug design (CADD), many methods are used to assess the affinity of a ligand for its biological target. Among the available techniques, methods that predict the binding mode of small molecules with biomolecules and calculate ligand-protein affinity are of particular importance.

CADD methods can be divided into two main groups: ligand-based drug design and structure-based drug design (SBDD). In SBDD methods, knowledge of the three-dimensional (3D) structure of the molecular target is required. The primary source of such structures is the Protein Data Bank (PDB). We use 3D structures of the target to design small molecules capable of binding strongly at the active site and modulating the biological function of the enzyme or receptor.

Affinity prediction methods that would be useful in the pharmaceutical industry should not only be accurate but also allow results to be obtained for many compounds within a reasonable time. In the early stages of drug design, when the affinity of tens or hundreds of thousands of compounds is estimated, it is only possible to use efficient, simplified methods, which provide a good trade-off between the needed level of accuracy and the required screening time. In the virtual screening (VS) of large ligand libraries, fast and approximate methods of estimating affinity are used.

Docking, as a method of VS, is successfully used to identify new hits from large libraries of chemical compounds and to predict their binding modes and affinity. However, VS methods are not suitable for predicting binding affinity with a high level of accuracy, thus not allowing for satisfactory correlations between experimental and predicted affinity. Only in subsequent stages of the project, when a small number of compounds are considered, more demanding and computationally intensive, but also more accurate methods of calculating affinity can be used.

The binding affinity depends on the balance between enthalpic and entropic contributions. Interactions between ligand and the residues in the binding site have a stabilizing effect [7]. Desolvation of binding partners, entropic loss during the formation of the protein-ligand complex, and conformational changes in the interacting partners are a source of unfavorable contributions to the binding [8]. Therefore binding affinity results from a small difference between two large values of enthalpy and entropy.

According to the equation $\Delta G = -RT\ln(K)$, the difference between a millimolar and a nanomolar inhibitor is less than 10 kcal/mol. Consequently, in order to be able to reliably predict whether a given inhibitor will be weak or potent, the theoretical method must correctly describe all components affecting binding affinity (such as intermolecular interactions, desolvation energies, changes in entropy and others) [2]. To increase the accuracy of binding affinity estimation, several free energy-based approaches have been proposed. They are based on MD or MC simulations, including end-state methods and so-called rigorous approaches, namely alchemical perturbation

methods, and physical pathway methods [9], [10], [11]. Fortunately, very often in drug design, it is sufficient to predict the relative affinity in a given series of ligands to find more potent compounds. Such calculations, thanks to error compensation, offer better agreement with experimental data than the prediction of absolute free energies of binding.

The vast majority of methods used to predict binding free energy for modeling protein-ligand interactions use classic force fields and molecular mechanics (MM) methods. However, many limitations of these methods are known; for example, conventional force fields cannot reliably describe the effects of polarization and charge transfer. Examples of interactions that are not sufficiently described by such force fields are also π-π, π-cation, and halogen bonds. In some cases, this may lead to an incorrect assessment of affinity and the inability to find the right binding mode. Also, systems containing metal cations are not correctly described by force field methods because of the difficulties in the parameterization of metal atoms and their ions. An attempt to go beyond these limits are new force fields taking into account polarization like Amber ff02pol, CHARMM Drude, NEMO, SIBFA, AMOEBA, ABEEMσπ, QMPFF and the application of quantum-mechanical description for all or only part of the analyzed system. Currently, we can see the growing importance of quantum chemical (electronic structure) methods in the study of protein-ligand interactions [12], [13], [14], [15], which results from the increase in available computing power and the development of quantum mechanical (QM) methods. However, the wide use of QM approaches in the drug discovery pipeline is still a demanding challenge.

This chapter briefly reviews selected methods used to calculate protein-ligand binding affinity applied in VS, rescoring of docked poses, and lead optimization stage, including QM methods based on molecular simulations.

1.1.1 Components of binding free energy

Experimental binding affinity can be defined by the use of equilibrium dissociation (K_d) or binding (K_b) constants. Under equilibrium conditions, the free energy of binding, ΔG is determined by the dissociation constant K_d of the protein-ligand, PL, complex (PL ⇄ P + L), as seen in the following equation:

$$\Delta G = \Delta G^0 - RT\ln K_d \tag{1.1}$$

where

$$K_d = \frac{[P][L]}{[PL]} = \frac{1}{K_b} \tag{1.2}$$

$$\Delta G = -RT\ln K_b = RT\ln K_d \tag{1.3}$$

R is the gas constant, T temperature in Kelvin, and ΔG^{O} is the change in free energy of reaction at standard conditions (all concentrations equal to 1 M, T = 298 K, pressure is 1 atm).

Many experimental methods have been developed to determine ΔG [16]. In isothermal titration calorimetry (ITC), we can directly measure thermodynamic energetics of binding. ITC allows for a determination of ΔG and its enthalpic contribution, ΔH. Stability shift assays allow determining the K_b constant. In turn, K_d constant can be determined by mobility shift assays. Techniques as equilibrium dialysis, analytical size exclusion chromatography, affinity selection chromatography, analytical ultracentrifugation, electrophoresis (capillary electrophoresis [CE], microchip electrophoresis [MCE], electrophoretic mobility shift assay [EMSA]), and microscale thermophoresis use for this purpose physical separation of ligand or protein from the protein-ligand complex. Among spectroscopic assays, special attention deserves surface plasmon resonance and biolayer interferometry due to their reliable binding affinity and kinetics data.

Binding free energy is a function of the state, which means that its value does not depend on the path leading from the substrates to the products. Therefore, it is possible to compare the ΔG value determined experimentally and calculated theoretically as the difference of free energy of reagents (protein and ligand) and free energy of the product:

$$\Delta G_{bind} = \Delta G_{PL} - (\Delta G_P + \Delta G_L) \tag{1.4}$$

where ΔG_{bind} is the free energy of binding, ΔG_{PL} the free energy of the protein-ligand complex and ΔG_P and ΔG_L the free energies of the protein and ligand, respectively.

The free energy of binding can be decomposed into various components corresponding to the specific contributions in the binding process. With the assumption, that these contributions to the free energy of binding are additive and can be calculated separately, by combining results of these calculations, we can obtain the value of ΔG_{bind}. In the following discussion on specific contributions to binding free energy, we will use "master equation" proposed by Ajay and Murcko [17] in the form:

$$\Delta G_{bind} = \Delta G_{int} + \Delta G_{solv} + \Delta G_{motion} + \Delta G_{conf} \tag{1.5}$$

The most obvious contribution to the binding affinity comes from the direct interactions between the binding partners. The first term, ΔG_{int}, is the binding free energy component resulting from the direct interaction between ligand and protein. This term is dominated by enthalpy contribution resulting from van der Waals interactions, electrostatic interactions, and hydrogen bonds responsible for the formation of the PL complex. For most MM methods, these contributions are additive, and therefore it is easy to determine the contributions to ΔG_{int} from individual atoms or groups of atoms. MM or QM can be used to calculate the energetic effects of this type of interaction. The QM methods are particularly important in the case

of PL complexes in which the ligand is strongly polarized, or the charge is transferred [18], [19], [20]. For almost all ligand-protein complexes, it is assumed that direct interactions increase binding affinity.

The free energy of solvation, ΔG_{solv}, takes into account the effect of solvent on the binding. When the ligand binds to the protein, both the ligand and the protein binding site usually get rid of the water molecules. Desolvation or loss of interaction with the solvent is an essential contribution to the binding free energy. For non-polar groups, due to the positive free energy for solvation [21], it is beneficial to dispose of the solvation layer and involve directly in van der Waals interactions. Non-polar contacts (hydrophobic interactions) are the driving force of the binding. In the case of polar groups, even though their desolvation is unfavorable, but it is beneficial to engage in electrostatic interactions or the formation of hydrogen bonds. However, the resultant contribution of polar groups to ΔG_{bind} depends on the mutual proportions in the value of desolvation energy and polar interactions in the complex. It can, therefore, be beneficial, disadvantageous, or neutral. The free energy of solvation also has an entropic component, ΔS_{solv}, which is associated with an increase in entropy of water molecules when they are removed from the receptor binding site and move to bulk water.

To calculate the free energy of solvation at the MM level, the Poisson-Boltzmann (PB) equation [22], [23] or the generalized Born (GB) method [24] can be used. In both approaches, the solvent is treated implicitly as a continuum with certain dielectric constant (ε). In this approach, the free energy of solvation is the sum of electrostatic (ΔG_{electr}) and nonpolar (ΔG_{n-p}) contributions:

$$\Delta G_{solv} = \Delta G_{electr} + \Delta G_{n-p} \tag{1.6}$$

The electrostatic part is derived from the solution of the PB/GB equation. The nonpolar part is calculated on the basis of appropriate parameters regarding the size (area) of the solvated surface and the type of atoms on it [25], [26]. ΔG_{n-p} term allows estimating the cost of forming a cavity in the solvent and the strength of interactions between the solute and solvent (dispersion and repulsion), considering usually the solvent-accessible surface area (SASA). In the QM approach to model solvation effects, the polarizable continuum model (PCM) [27], [28], [29] and the conductor-like screening model (COSMO) [30] are used.

When the protein and ligand form a complex, three degrees of rotational freedom and three degrees of translational freedom are lost and transformed to some extent into vibrational motions. The change in free energy resulting from this restriction on the free movement of the protein, ligand, and complex is designated ΔG_{motion}. Entropy and internal thermal energy of the system are calculated from the partition functions related to translation, rotation, and vibration. Calculations show that the loss of rotational and translational degrees of freedom during complex

formation is energetically unfavorable; however, it is partly compensated by the appearance of vibrational components (six new internal normal modes) [31].

The change in free energy resulting from conformational changes that occur during the formation of the complex, ΔG_{conf} depends on the conformation of the ligand and protein both before and after binding. ΔG_{conf} has an enthalpic and entropic component. The enthalpic part arises from a change in the internal energy of the ligand and protein during complex formation, which can be calculated using MM or quantum chemistry methods. Both the protein and ligand lose their conformational entropy during the formation of the complex. The conformational entropy of the ligand can be estimated based on the number of single bonds that are frozen during binding. The estimated energy of such a process is from 0.4 to 0.9 kcal/mol [32] for every single bond. Similarly, the loss of entropy change for side chains at the protein binding site can be estimated.

1.2 Computational chemistry methods in binding affinity calculations

1.2.1 Molecular mechanics

MM methods are widely used in calculations related to biomolecules. Due to the size of protein molecules, they are incomparably faster than QM, density functional theory (DFT) or semiempirical quantum mechanics (SQM) methods. MM treats a molecule as a set of rigid spheres (atoms) connected by springs (bonds). Individual types of atoms, depending on their hybridization and environment in the molecule, must be properly parameterized. MM allows calculating the potential energy of the molecule itself and the energies of intermolecular interactions owing to the presence of terms describing van der Waals and electrostatic interactions.

The most commonly used force fields in biomolecular simulations are CHARMM [33], AMBER [34], and GROMOS [35], [36]. A typical form of the potential energy function is as follows:

$$E = \sum_{i=1}^{N_{bonds}} k_{b,i}(r_i - r_{0,i})^2 + \sum_{i=1}^{N_{angles}} k_{a,i}(\theta_i - \theta_{0,i})^2 + \sum_{i=1}^{N_{dihedrals}} k_{\phi,i}[1 + \cos(n_i\phi_i + \delta_i)] + \sum_{i}^{N_{atoms}} \times \sum_{j\neq i}^{N_{atoms}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i}^{N_{atoms}}\sum_{j\neq i}^{N_{atoms}} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (1.7)$$

The first and the second term describe energy change during bond stretching and angle bending, where $k_{b,i}$ and $k_{a,i}$ are force constants for *ith* bond and *ith* angle, respectively. Distance between the two atoms is denoted as r_i, and $r_{0,i}$ is the equilibrium

bond length. Similarly, θ_i is the angle between three atoms and the $\theta_{0,i}$ is the equilibrium angle.

The third term accounts for the potential of dihedral angles (torsions), where $k_{\phi,i}$ is the force constant for the dihedral angle, n_i is the periodicity and δ_i is the phase shift.

The fourth and fifth terms account for the van der Waals interactions and the Coulombic interactions, respectively. Van der Waals interactions are described by Lennard-Jones potential between all pairs of atoms in the molecule. Electrostatic interaction energy depending on the partial atomic charges of two atoms q_i and q_j can be attractive or repulsive.

The last two terms allow calculating interaction energy between different molecules, for example, in protein-ligand complexes. Parameters in this empirical energy function (Eq. (1.7)), such as force constants, equilibrium values for bond lengths and angles, van der Waals parameters and atomic partial charges can be obtained by fitting to high-level quantum calculations.

1.2.2 Quantum-mechanics methods

Major changes in computing power and the availability of quantum-chemical computing software over the last two decades allowed the application of quantum mechanics to increasingly larger and more complex systems.

QM methods allow the accounting of the effects omitted in the description of ligand-protein complexes by MM methods. These include ligand polarization at the binding site, charge transfer, interactions with metals (difficult to parameterize in MM), or covalent binding of the ligand to the protein. Parameterization of ligands, which are usually new molecules, is also avoided. Parameters for proteins (amino acids) are very well determined in individual force fields, while the parameters for ligands are most often derived from CHARMM General force field [37], [38], [39], [40] or General AMBER Force Field [41], [42]. QM methods are based on the solution of the Schrödinger equation, which can be solved analytically only for the simplest one-electron system. Approximate solutions should be found for all other systems; hence a large number of such approximate quantum methods have been developed.

The Hartree-Fock (HF) method assumes that each electron interacts with the average field of all other electrons, which is a relatively crude approximation because it omits the electron movement correlations. The easiest way to correct these shortcomings can be the second-order Møller-Plesset perturbative method, which is also the simplest theoretical method that directly takes into account dispersion interactions. More accurate results are obtained by other methods, such as the coupled cluster method. Currently, the so-called gold standard of this type of method is the CCSD(T) method including a single, double, and triple excitations (treated perturbatively). This method usually gives results with an accuracy of approximately 1 kcal/mol.

On the other hand, HF calculations can be accelerated, up to 1000 times, by ignoring some expressions and replacing others with empirical parameters, which leads to semiempirical methods. Many SQM methods are based on the neglect of diatomic differential overlap (NDDO) approximation [43]. NDDO method ignoring a large number of two-electron integrals and assuming parameters for others, do not manage noncovalent interactions well. There are many methods of this type including AM1 [44], PM3 [45], RM1 [46], and PM6 [47] The semiempirical methods usually give rather poor energies for large structures, which results from omitting dispersion effects, as well as an inadequate description of hydrogen and halogen bonds. Therefore, several groups have developed corrections that include omitted or incorrectly described interactions; for example, DH [48], D2H [49], DH2X [50], and D3H4 [51]. It turns out that the PM6-D3H4 method can reproduce the interaction energies in test benchmarks with an error of less than 1 kcal/mol [51].

DFT focuses on electron density, which is a function of the three Cartesian coordinates. There is a one-to-one relationship between wave function and electron density, but the equation that needs to be solved to find the electron density for a given system is not exactly known. That is why a very large number of functionals were developed. Some of them contain the fraction of exchange energy from the HF method, and they are called hybrid functionals. In general, DFT methods are both faster and more accurate than the MP2 method. Later, there were also suggestions that the results of the HF and DFT methods could be improved by incorporating simple empirical corrections for neglected dispersion effects. The most commonly used corrections are DFT-D2 and DFT-D3, developed by Grimme [52], [53], [54]. Similar results also give other approaches [55], [56], [57].

Almost all quantum-mechanical methods use atomic basis functions in calculations. The quality of the results obtained depends on the quality of the basis set. In energy calculations, to achieve convergence, split-valence basis sets of triple or quadruple-zeta quality are needed. Additionally, diffuse functions are required for anions and proper consideration of dispersion. By definition, semi-empirical methods are based on a minimal basis set. One of the fundamental problems associated with a small atomic basis set is the fact that interacting molecules “borrow” functions from neighboring atoms belonging to another molecule. This effect is a particular problem when we calculate the interaction energy because it leads to an overestimation of the binding energy. This basis-set superposition error can be limited by performing additional calculations using the so-called counterpoise correction [58].

Quantum-mechanical calculations are very time consuming, and the time of calculations strongly depends on the size of the system. For example, the DFT, HF, MP2, and CCSD(T) methods exhibit an exponential dependence on the number of basic functions with exponents 3, 4, 5, and 7, respectively. In practice, single-point energy calculations now allow using semiempirical methods for systems with about 10,000 atoms, DFT methods for 2000 atoms, and CCSD(T) method for about 30 atoms [14].

1.2.2.1 QM/MM methods

In QM/MM methods, the ligand and sometimes an important part of the receptor involved in protein-ligand binding (surrounding residues) are treated by QM while the remainder of the system is treated by MM [59], [60] (Figure 1.1). Compared to full QM calculations on the whole system, QM/MM methods are computationally much more efficient. Contrary to MM force fields, which have been carefully parameterized for proteins, accurate parameters for novel compounds are not often available, so the QM model of the ligand can overcome these difficulties.

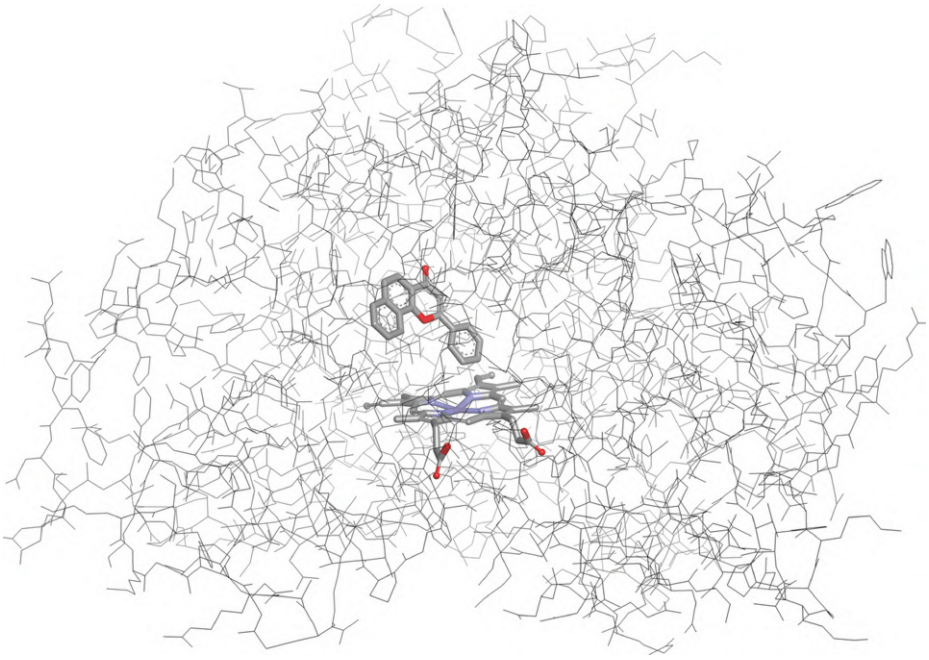


Figure 1.1: Partitioning of a system in QM/MM calculations. The QM region includes ligand (α-naphthoflavone, ANF) and heme (both as stick model) depicted for CYP1A2 complexed with ANF (PDB code: 2hi4).

One of QM/MM schemes is the ONIOM approach in which the molecular system is divided into an arbitrary number of layers, and different QM, SQM, and MM methods can be applied to different parts of the system [61].

In the combined QM/MM approach, the total energy is represented by the following equation:

$$E_{\text{total}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}} \tag{1.8}$$

where E_{QM} is the QM energy of the QM system, E_{MM} is the MM energy of the MM system, and $E_{\text{QM/MM}}$ is the interaction energy between the QM and MM systems. The

most common way to consider the interactions between the QM and MM systems is electrostatic embedding (EE) when the QM system is polarized by MM point charges. In a polarized embedding scheme, it is also possible to polarize QM and MM systems by each other, but this requires a polarizable MM force field [62], [63].

1.2.2.2 Linear-scaling methods

QM/MM methods encounter some problems, for example, where to set boundaries between QM and MM systems, how to deal with bonds that span both regions, and the mutual polarization of QM and MM systems. The solution to these problems would be to treat the entire system using QM methods.

To accelerate the most time-consuming parts of *ab initio* calculations (the Fock-matrix construction and its diagonalization), several linear-scaling approaches were developed [64], [65], [66], making possible the calculation of the electronic structure for the full system [67]. Most linear-scaling methods were developed for the semiempirical QM methods based on the NDDO approximation.

1.2.2.3 Fragmentation methods

Another method to speed up calculations is to divide the system into smaller fragments. In the first group of such methods used in the binding free energy calculations, the energy of the entire system is calculated, and any relative energies (binding energy) are obtained by subtracting the relevant terms. This approach is used in the fragment molecular orbital (FMO) method [68], [69], [70], [71]. In the FMO approach, various QM methods can be used, ranging from HF, MP2 MP3, MP4, to CCSD(T).

In the second group of methods, the interaction energies are calculated in dimers composed each time of the ligand and receptor fragment. This approach is used, for example, in molecular fractionation with conjugate caps (MFCC) method [72] and its derivatives, such as electrostatic embedding generalized molecular fractionation with conjugate caps (EE-GMFCC) method [73]. The polarizable multipole interaction with supermolecular pairs (PMISPs) method also involves MFCC calculations for the ligand and all chemical groups within 4 Å. It combines QM calculations with a polarizable multipole description of the many-body effects at MM level [74].

1.3 Virtual screening

In SBDD, the primary purpose of docking, as a method for virtual high-throughput screening, is to separate binders from non-binders in large compound libraries. The binding affinity is estimated here by simplified methods, which allow quickly evaluate thousands or millions of binding modes. These approximations often lead to a situation where there are many false positive results among molecules with the

highest scores (hits). Eliminating or reducing the number of such results is important before proceeding to the next stages of drug design, such as hit-to-lead optimization. For this purpose, re-scoring of docked poses by more advanced methods or use in VS more accurate QM or SQM scoring functions (SFs) can be applied.

1.3.1 Classical SFs

Docking in combination with scoring is widely used in VS of large molecular databases to identify new potential ligands, predict its binding mode, and estimate its binding affinity to macromolecular target [75], [76], [77]. Fast and simplified assessment methods used to evaluate ligand affinity in docking are called SFs. They are divided into four classes: force field-based, empirical, knowledge-based, and machine-learning–based ones [78].

In **force field–based SFs** for evaluation of binding affinity, mainly the noncovalent van der Waals and the electrostatic energy terms calculated at MM level are used. The additional term can be added for hydrogen bonding. The gas-phase potential energy is only one component of the free energy change during protein-ligand complex formation, so force field-based SFs were augmented by terms describing solvation energy change [79], [80]. The general functional form for this class of SFs can be written as:

$$\Delta G_{bind} = \Delta E_{vdW} + \Delta E_{elec} + \Delta E_{H-bond} + \Delta G_{solv} \tag{1.9}$$

Examples of force field-based SFs are GoldScore [81], MedusaScore [82], ICM [83] and LigandFit [84].

Empirical SFs estimate binding affinity by summing up the energetic contributions from particular physical events involved in the formation of the ligand-receptor complex:

$$\Delta G_{bind} = \sum_i w_i \cdot \Delta G_i \tag{1.10}$$

where ΔG_i are different energy terms, namely for hydrogen bonds, ionic and hydrophobic interactions, desolvation, as well as entropic effects. Weight coefficients w_i for these contributions are determined with the use of regression analysis by fitting calculated ΔG_{bind} to experimentally determined affinities for a series of PL complexes with known X-ray structures. Unfortunately, empirical SFs do not include less common interactions, as π-cation because they are not significant in the regression analysis. The first empirical SF (SCORE1) was developed by Böhm [85] in 1994. Other examples of this type of SFs are AutoDock [86], ChemScore [87], LigScore [88], and PLP [89].

The third type, **knowledge-based** SFs, uses pairwise statistical potentials between protein and ligand [78]:

$$A = \sum_{i}^{lig} \sum_{j}^{prot} \omega_{ij}(r) \tag{1.11}$$

Pairwise potentials, $\omega_{ij}(r)$, are derived from the occurrence frequency of atom pairs *i-j* in a structural database (PDB) using the inverse Boltzmann relation:

$$\omega_{ij}(r) = -k_B T \ln\left[g_{ij}(r)\right] = -k_B T \ln\left[\frac{\rho_{ij}(r)}{\rho_{ij}^{\star}}\right] \tag{1.12}$$

where k_B is the Boltzmann constant, T is the temperature in Kelvin, $\rho_{ij}(r)$ is the numeric density of atom pair *i–j* at distance *r*, and $\rho_{ij}^{\star}$ is the pair density in a reference state where interatomic interactions are zero. If specific interatomic distances occur more often than average distances, it indicates favorable contact between the given atom pair. On the other hand, if distances occur less frequently are likely to decrease affinity. Several knowledge-based SFs have been developed, such as SMoG [90], PMF [91], [92], [93], DrugScore [94], [95], [96], and IT-Score [97], [98], [99].

Machine-learning–based SFs are a relatively new type of methods, which uses approaches similar to quantitative structure-activity relationship analysis to obtain statistical models based on different descriptors that can compute binding scores. These methods, as empirical SFs, need a training set of PL complexes with known binding affinities and 3D structures. A few examples of machine learning-based SFs include NNScore [100], [101], RF-Score [102], [103], SFCscore^{RF} [104], and ID-Score [105].

1.3.2 QM scoring functions

In 2004 Raha and Merz [18] introduced a semiempirical SF based on the AM1 method with Amber FF96 force field dispersion term, PB implicit solvent model, and entropy estimates. In the SQM calculations, the divide-and-conquer (D&C) linear-scaling approach [106] was used. They studied 18 carbonic anhydrase inhibitors and five carboxypeptidase inhibitors obtaining a correlation of R^2 = 0.69 compared to the experimental binding free energies without fitting any of the contributions of the total score.

This approach was tested on the set of 57 protein-ligand complexes giving for Total Score correlation with experimental data R^2 = 0.48. When coefficients before each term in the binding energy were fit to the experimental ΔG, giving QMScore, R^2 raised to 0.53, and QMScore outperformed the other 11 tested SFs [107] for this set.

Another method based on a linear-scaling approach was introduced by Fanfrlík et al. [108] to rescore docked ligand poses. SQM SF was used to test the HIV-1 protease (PR) and its inhibitors, cyclin-dependent kinase 2 (CDK2) with sets of structurally diverse ligands, and casein kinase 2 (CK2) with halogenated ligands [108], [109],

[110], [111]. In this score, vacuum interaction energy was calculated by the corrected PM6 method (PM6-DH2 or PM6-DH2X). For the solvation free energy calculations, COSMO [30], while for the ligands steered molecular dynamics (SMD) [112] methods were used. The $-T\Delta S$ term was estimated either by a rotatable bond approach or by vibrational analysis at MM level.

This SQM SF improved docking results for a series of HIV-1 protease (HIV PR) inhibitors, allowing to identify correct binding poses and improve the ranking of the ligands. A good correlation to the experimental data was obtained (R^2 = 0.62) [108]. For less flexible CDK2 inhibitors, the best correlation with the experimental affinities was obtained when only SQM interaction energy and solvation terms were included in the final score [110]. Complexes of CK2 with halogenated inhibitors [111] were studied with the use of the PM6-DH2X method for interaction energy. However, deterioration of the correlation between calculated and experimental affinity was observed when the entropy term was included. MM method used for entropy calculations was not able to correctly describe halogen bonds.

Their SQM SF [113] was later simplified. The original equation included four terms:

$$\text{Score} = \Delta E_{int} + \Delta\Delta G_{solv} + \Delta G_{conf} - T\Delta S \tag{1.13}$$

where ΔE_{int} is the gas-phase interaction energy, $\Delta\Delta G_{solv}$ the change in solvation energy, ΔG_{conf} the change in conformational free energy, and $-T\Delta S$ entropy change upon ligand binding. In the SQM/COSMO energy filter [114], only the first two dominant terms were kept. ΔE_{int} is calculated at the PM6-D3H4X level, with corrections for dispersion and hydrogen-bonding and halogen-bonding interactions. To calculate solvation effects, $\Delta\Delta G_{solv}$, the implicit solvent model COSMO is used.

The performance of this filter was tested on four systems: acetylcholinesterase (AChE), TNF-α converting enzyme (TACE), aldose reductase (AR), and HIV PR. SQM/COSMO approach outperformed seven well-known empirical SFs and a physics-based AMBER/GB in discriminating binding-like poses from decoy poses presenting the lowest number of false-positive solutions. Only for TACE, which is a metalloprotein having Zn^{2+} coordinated by S^-, false-positive solutions were obtained. However, for achieving such high accuracy, a higher computational time is needed. The SQM/COSMO filter is about 100-times slower than the statistics- and knowledge-based SFs. Its accuracy and time requirements make it an excellent tool for the late stages of VS.

The same set of proteins (AChE, TACE, AR, and HIV PR) was tested by SQM/COSMO filter in which the PM6-D3H4X method was replaced by SCC-DFTB3-D3H4, a higher-level SQM method [115]. Calculation of ΔE_{int} at the DFTB3-D3H4 level significantly improved results for TACE metalloprotein, eliminating false-positive solutions.

Further validation of the two SQM/COSMO filters, based on PM6-D3H4X and DFTB3-D3H4X methods, was conducted on 17 pharmaceutically relevant protein-ligand complexes [116]. Results were compared with classical SFs (Glide XP, AutoDock4,

AutoDock Vina, and UCSF Dock). Both variants of the SQM/COSMO SF outperform standard SFs, as was shown by the low number of false positives.

1.4 Rescoring of docked ligands and lead optimization

1.4.1 Single-structure approaches

One of the simplest, although approximate, approach in binding affinity calculations is to use single structures of protein-ligand complexes. These structures can be obtained directly from docking or crystal structures; they can also be single snapshots from MD or MC simulations.

1.4.2 QM/MM methods

Different QM/MM methods have been used to estimate binding affinity [117], [118], [119], [120], [121], [122]. Several researcher groups calculated solvation effects by adding PB and SASA energy terms [123], [124], [125], [126], [127].

Approaches proposed by Merz [128] and Hobza [129] include all components required to obtain accurate energies. Hayik et al. [128] studied 23 metalloprotein-ligand complexes using QM/MM approach, similar to their QMScore [18]. Before the energy calculations, complexes were minimized in a vacuum. For QM systems, which included the ligand and residues within 5 Å of the Zn ion semiempirical AM1 method was used. Atoms outside of this region were treated with the AMBER ff99SB47 force field. Solvation effects, ΔG_{solv}, were calculated in the SQM/MM calculations with the PB method and SASA term in case of nonpolar interactions. For entropy change estimation normal mode analysis (NMA) of the QM system was performed. However, better correlation (R^2 = 0.64) with experimental data was obtained with use only the QM energy, instead of the full QM/MM (R^2 = 0.56).

Brahmkshatriya et al. [129] studied series of cyclin-dependent kinase (CDK2) inhibitors. They used a hybrid three-layer QM/MM approach (DFT-D/PM6-D3H4X/AMBER) as an extension of their all-protein linear-scaling SQM method [108]. The QM part, including the ligand and residues up to 4 Å distance, was treated with the RI-DFT-D method (BLYP/SVP for geometry optimization and TPSS/TZVP for the single-point energies). Complexes were optimized by QM/SQM/MM calculations, during which the MM part was kept frozen (to speed up calculations). The semiempirical PM6-D3H4X method was used for the part including residues within 8 Å of the ligand. For the remaining part of the protein AMBER force field with GB implicit solvent model was applied. Ligand solvation energy was calculated with a more demanding SMD model at the HF/6-31G* level. In calculations of ΔG_{bind} besides entropy estimate, calculated from the number of rotatable bonds in ligand hindered

during complex formation, deformation energies of the protein and the ligand were added. Inhibitor molecule from the crystal structure (PDB id: 2R3J) was the base for manual modifications and modeling series of 30 congeneric inhibitors. For the modified structures good correlation with experimental data was obtained (R^2 = 0.64), while for docked molecules, a satisfactory correlation was not found.

1.4.3 Linear-scaling methods

Many groups studying protein-ligand binding affinity with linear-scaling methods used the MOZYME approach [64] available in MOPAC package [130].

AM1 and a COSMO continuum-solvation model were employed in studies of six ligands binding to a 33-residue RNA aptamer and six ligands binding to trypsin [131]. Similar studies with MOZYME and COSMO were performed with the use of AM1, PM3, PM5, and PM6 (also with PM6-DH2) methods [132], [133], [134], [135].

1.4.4 Fragmentation methods

For the first time, Kitaura applied the FMO method for binding affinity calculations [136]. HF/STO-3G approach was used in studies with 11 estrogen receptor ligands. Despite the omission of dispersion, solvation, and entropy effects, a good correlation with experimental results was obtained (R^2 = 0.70).

The significantly improved approach was used by Kitaura in a study of FK506 binding protein [137]. The ligand geometry was optimized inside the binding site taking into account the residues within the 5.5 Å from the ligand, then the energy was calculated at MP2/6-31G* level of theory. Solvent effects were also included using the PB model (calculated at MM level) and SASA term.

In studies on 28 CDK2 inhibitors [138], considering in addition to FMO energy, solvation from the PB + SASA model (at MM level), and changes in ligand entropy, it was possible to obtain a very good correlation of results with the experiment (R^2 = 0.94).

Water molecules were considered explicitly in investigations of inhibitory dipeptides binding to thermolysin [139]. The MP2/6-31G method was used for the catalytic Zn ion, ligand and protein residues, as well as water molecules within 5 Å. The rest of the enzyme and a solvation shell was treated at the HF/6-31G level of theory. Entropy change was estimated from MM frequencies. This approach gave a good correlation to the experimental affinities (R^2 = 0.94). Nonetheless, absolute infinities were highly overestimated.

The accurate vacuum interaction energy of the insulin dimer with 4′-hydroxyacetanilide at the MP2/CBS level was calculated in the overlapping-multicenter ONIOM/FMO method [140].

Zhang et al. [72], [73] developed the MFCC approach to calculate ligand-binding affinities. They used this method in studies of biotin to streptavidin binding [141], binding of benzamidine to trypsin [142] and for the interaction of efavirenz with HIV-1 reverse transcriptase [143] employing mainly HF/3-21G method.

In calculations of the interaction energy for the binding of an inhibitor to neuraminidase, the MP2/6-311+G(2d, p) level of theory was used [144]. MP2 calculations with cc-pVDZ and cc-pVTZ basis were also employed to study the interaction between p53 and the oncoprotein MDM2 [145].

A variety of DFT methods [146], [147], [148], [149], [150], [151], often with dispersion corrections, were applied to calculate binding affinities for several different systems. However, unlike other MFCC calculations, not all protein was used here, but residues at a distance of 8–12 Å from the ligand, depending on the studied system.

1.4.5 Fully quantum mechanical

Ehrlich et al. [152] proposed a general, fully quantum mechanical scheme for the computation of protein-ligand affinities. The method was applied to study the binding of FXa and TYK2 proteins with 25 and 16 ligands, respectively. In their approach, the absolute value of ligand binding free energy is the sum of three components, the electronic interaction energy, the entropic term, and the change in the free energy of solvation. Calculations were performed on a subsystem with about 1000 atoms, including the ligand and neighboring binding site residues. The subsystem was fully optimized at the HF-3c level [153] of theory with an implicit solvent model (C-PCM) [154]. For the interaction energy, HF-3c or PBEh-3c DFT methods were used. Entropic contributions were calculated using a semiempirical DFTB3-D3 method [53], [155], [156], [157], and the free energy of solvation was calculated with COSMO-RS method [158], [159], [160]. Correlation between calculated and experimental binding free energies calculated at the HF-3C level, expressed by the Pearson correlation coefficient, is acceptable with 0.47 and 0.55 for FXa and TYK2, respectively. However, this approach is not yet applicable in an industrial setting due to residual errors and computational power needed for calculations.

1.4.6 End-point approaches

SFs discussed above and single-structure methods are focused exclusively on the bound state. The end-point approaches estimate binding free energy employing simulations at the bound state and also of the free ligand and the free receptor, thus for initial and final states (end-points). These methods are placed between single-structure approaches and rigorous free energy perturbations (FEPs) both in terms of accuracy and calculation time. Although they are too expensive to estimate binding affinities

for large compound libraries, they are commonly used in the pharmaceutical industry in the hit-to-lead and lead optimization phases of drug design [161].

1.4.6.1 QM/MM-Poisson-Boltzmann Surface Area methods

The MM Poisson-Boltzmann Surface Area (MM-PBSA) and MM Generalized Born Surface Area (MM-GBSA) methods combine the energies computed from MD simulations at the MM level with either PBSA or GBSA implicit solvent model [162], [163], [164].

MM-(PB/GB)SA makes use of MD simulations of the free ligand, free protein, and their complex with explicit solvent molecules. After the simulations, explicit solvent molecules are removed. Next, for each simulation frame, potential energies (MM) for ligand, protein, and complex, as well as their solvation energies with the use of the implicit solvent model (PB or GB) and SASA are calculated. Averaging over each trajectory allows calculating the changes in mean potential and solvation energy. In this approach general equation for binding free energy as the difference between the free energy of PL complex and free energies of free protein and unbound ligand:

$$\Delta G_{bind} = G_{PL} - G_P - G_L \tag{1.14}$$

can be decomposed into specific contributions:

$$\Delta G_{bind} = \Delta H - T\Delta S = \Delta E_{\mathrm{MM}} + \Delta G_{solv} - T\Delta S \tag{1.15}$$

in which

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{elec} + \Delta E_{vdW} \tag{1.16}$$

$$\Delta G_{solv} = \Delta G_{PB/GB} + \Delta G_{SA} \tag{1.17}$$

$$\Delta G_{SA} = \gamma \Delta SASA + b \tag{1.18}$$

where ΔE_{MM} are the changes in the MM energy, ΔG_{solv} solvation free energy and $-T\Delta S$ change in conformational entropy upon ligand binding. ΔE_{MM} includes the changes in the internal energies ΔE_{int} (from bonded terms: bond, angle and dihedral energies), electrostatic ΔE_{elec} and van der Waals interactions ΔE_{vdW}. ΔG_{solv} is the sum of the electrostatic solvation energy $\Delta G_{PB/GB}$, polar contribution to ΔG_{solv} calculated using either the PB or GB model and the nonpolar contribution $_{\Delta GSA}$ estimated using the SASA [165], [166]. The change in conformational entropy $-T\Delta S$ can be estimated through a normal-mode analysis [161], [167], [168] on a set of conformational snapshots of the free and bound molecules hence obtaining the average entropy change over these snapshots. However, this estimation can be neglected in studies of the relative binding free energies of similar ligands.

Replacing some terms obtained at the MM level in ΔE_{MM} and sometimes in ΔG_{solv} by values calculated with QM methods lead us to QM/MM-(PB/GB)SA approaches.

The QM/MM-PBSA approach was applied in studies of five bromobenzimidazole inhibitors binding to the kinase CK2 [168]. Fifty one snapshots from QM/MM molecular dynamics simulation were used in calculations of binding free energy. Ligands were treated by the AM1 method. Polar solvation energy was calculated from the PB equation end entropy term was omitted. This approach gave a good correlation with experimental data (R^2 = 0.69). Eight halogen-containing inhibitors of CDK2 were studied by a similar method, but with GB continuum solvation model. Unfortunately, due to the lack of halogen bonds corrections results of these calculations were poor, with a negative correlation [169].

Good results were obtained for five inhibitors of the urokinase plasminogen activator studied by QM/MM method [170]. With the use of the 18 snapshots from MD simulation at MM level and AM1 method for ligand and surrounding residues within 6 Å (R^2 = 0.68). A correlation was even better when structures were SQM minimized (R^2 = 0.96).

QM/MM method developed to estimate free energies for reactions in proteins [171] was also used to study the binding of six ruthenium-containing ligands to cathepsin B [117]. For the ligand and the side chain of the cysteine coordinating Ru ion, the TPSS/def2-SV(P) method was used. The estimate of ΔG calculated by QM/MM-PBSA method gave a decent correlation for fixed protein (R^2 = 0.59). The use of snapshots from MD simulations with a fixed QM region led to a significant improvement in results (R^2 = 0.91).

A similar approach was used for metalloenzymes [172], [173], [174]. For the ligand, metal ion, and selected residues, the B3LYP density functional with the 6-31G* or 6-31+G* basis set was applied. Entropy estimates were calculated from the number of rotatable bonds restricted during complex formation, for the solvation PBSA model was used. Accurate predictions for four inhibitors of CYP2A6 were obtained using this approach [174]. The QM system contained heme and the inhibitor molecule; the rest of the system was treated with MM method. Structures extracted from the 10 selected frames from the last 500 ps simulation trajectories were energy minimized. B3LYP/6-31G(d) level of theory was used for the QM system and Amber force field for the remaining atoms. After minimization, for each CYP2A6-inhibitor complex, the binding free energies were calculated with the use of the QM/MM-PBSA method. Good agreement between calculated and experimentally derived binding energies allowed to recreate the correct order of affinity for all inhibitors.

An MM/PBSA-like approach (PMISP method) was used to calculate binding free energies for seven ligands to avidin at the MP2/cc-pVTZ level [175]. Using for each ligand 10 snapshots from an MD simulation (at the MM level), entropies from MM/PBSA calculations, and the standard SASA term for nonpolar solvation, a moderate correlation was obtained (R^2 = 0.52). This result was worse than that obtained in the MM/PBSA method (R^2 = 0.96).

Also, FMO calculations have been performed on snapshots from MD simulations. MP2/cc-pVDZ method was used in studies of trisaccharide binding to lysozyme [176].

HF and MP2 methods with 6/31G* basis set were applied for three ligands binding to the DJ-1 protein [177]. Unfortunately, these approaches gave considerable uncertainty or positive binding energies.

Using the same approach as for QMScore [18], [107] Díaz et al. [178] developed the QM-PBSA method as a linear-scaling SQM variant of MM/PBSA. This method was used to calculate the binding affinity of benzylpenicillin and cephalosporin to the TEM-1 β-lactamase. The binding affinity calculations were performed on 50 snapshots from MD simulations (at the MM level) previously minimized by QM/MM. Next to SQM energy, solvation effects (PB and SASA term) and entropy (calculated by an MM NMA**)** were taken into account. According to those results, benzylpenicillin binds stronger to TEM-1 β-lactamase than cephalosporin.

Another linear-scaling approach, the MM/QM-COSMO method, was used to study the binding of five tetrapeptides to the Lck SH2 domain [179]. The PM3 method was applied for the full protein along with solvation effects calculated with the COSMO continuum model and SASA term plus entropy correction gained from MM NMA. Calculations were conducted on PM3 minimized 1000 snapshots from an MD simulation of the complex (at the MM level). After the optimization of the COSMO atomic radii, a remarkable good correlation with the average unsigned error of 0.7 kcal/mol was obtained for relative binding energies.

Similarly, using linear-scaling (mozyme keyword in MOPAC), the performance of AM1, RM1, and PM6 with the inclusion of dispersion and hydrogen-bond corrections was tested [180]. The method was used to calculate binding affinities of seven biotin analogs to avidin, nine inhibitors to blood-clotting factor Xa, and nine phenol derivatives to ferritin. Authors used a strict SQM version of MM/GBSA, replacing the $\Delta E_{elec}+\Delta E_{vdW}+\Delta G_{solv}$ terms by SQM calculations with COSMO solvation. The entropy was estimated by an NMA at the MM level. On average, the most promising results gave AM1-DH2 method, although better results were achieved in the standard MM/GBSA approach.

QM-PBSA approach with linear-scaling DFT-D and MM/PBSA calculations were applied to compute the binding free energy of the eight ligands to the T4 lysozyme double mutant L99A/M102Q, the protein which contains 2601 atoms [181]. In DFT calculations, the GGA exchange-correlation functional PBE with corrections accounting for dispersion was used. The change in entropy was calculated with MM NMA. Solvation energies for ligands were calculated using the SMD model [112] at the M05-2X/6-31G(d) level. By using 50 snapshots from MM MD simulations, the RMS error in binding free energies for QM-PBSA was equal to 2.7 kcal/mol, whereas for standard MM/PBSA, error increased to 4.0 kcal/mol.

QM/MM-GBSA method was used in the study on binding α- and β-anomers of monosaccharides to lectin RSL [182]. The QM system consisted of monosaccharide and seven ligand-interacting amino acid residues. In binding free energy calculations, 10 SQM methods and two DFTB Hamiltonians was combined with five GB models. Calculations were based on selected frames from 10 ns molecular dynamics

simulations at MM level. The entropy change was estimated using NMA at MM level. The PM6 method and its variants (PM6-D, PM6-DH+), as well as DFTB and SCC-DFTB methods, gave the highest correlation coefficient values ($R^2 \geq 0.96$). However, for VS, where thousands of compounds need to be calculated, the PM6 method is recommended as less computationally demanded than the DFTB/SCC-DFTB approach.

1.4.6.2 Linear interaction energy method

Linear interaction energy (LIE) method requires two MD simulations with explicitly included solvent molecules, one for the solvated ligand and the other for the solvated ligand-protein complex [183]. Then the trajectories obtained in this way are used for the calculations of the Boltzmann-averaged electrostatic and van der Waals interaction energies of the ligand with its environment.

$$\Delta G_{bind} = \alpha\left(\left\langle V^{vdW}_{lig-surr}\right\rangle_{bound} - \left\langle V^{vdW}_{lig-surr}\right\rangle_{free}\right) + \beta\left(\left\langle V^{el}_{lig-surr}\right\rangle_{bound} - \left\langle V^{el}_{lig-surr}\right\rangle_{free}\right) + \gamma \tag{1.19}$$

It denotes MD or MC averages of the van der Waals (*vdW*) and electrostatic (*el*) interaction energies between the ligand and its surrounding environment, either in solvent (*free*) or in the protein (*bound*). The parameters α and β of this equation are coefficients used to scale the nonpolar and polar binding energy contributions, respectively. Values of α and β and offset constant y are obtained by fitting calculated ΔG_{bind} to experimentally determined binding affinities for a training set of known ligands. Therefore, an experimental ΔG_{bind} should be available for at least several ligands used in the simulations.

LIE free energy calculations do not take into account the protein solvation energy, and also conformational entropy of the ligand is omitted. Thus, the best correlation of calculated ΔG_{bind} with experimental values is obtained for a series of ligands with a similar structure.

QM methods were applied in a parametrized LIE-type approach to study binding free energies of 28 hydroxamate inhibitors to zinc-dependent matrix metalloproteinase 9 (MMP-9) [184]. The inhibitors were docked to MMP-9 using FlexX. Protein-ligand complexes were minimized by QM/MM with the use of B3LYP hybrid functional for the QM part and OPLS-AA force field for remaining atoms. The QM system consisted of three His residues, one Glu residue, the ligand, and Zn ion. Next, the QM/MM optimized complexes were subjected to the 200 ps MD simulation to collect data and check their stability. However, for single point calculation of QM/MM interaction energies, time-averaged structures obtained after 5 ps simulations were used. Finally, three terms, QM/MM energies with SASA desolvation and constant term, were fitted to the experimental data. The FlexX scores gave no correlation ($R^2 = 0.04$). For QM/MM optimized structures, a correlation was better ($R^2 = 0.50$). However, the highest value

of the correlation coefficient was obtained for the QM/MM energy calculated for the time-averaged structures with the SASA term (R^2 = 0.90).

The same system was also studied with the use of a multimode approach [185]. Time-averaged structures for eight 25 ps intervals of 200 ps MD simulations of the complex and the free ligand (at MM level) were treated as different binding modes (mode 1, mode 2, … , mode 8). In contrast to the first approach [184], authors considered separately polar and nonpolar solvent-accessible surface obtaining additional parameter for fitting. Despite the model extension, the correlation with experimental affinities for the QM/MM-LIE was not improved (R^2 = 0.90).

Three parameter approach with only the SASA term for solvation [184] was also used in studies of MMP-3 protein, giving similar results [186] as described previously. The FlexX scores correlated poorly with experimental activity (R^2 = 0.06). The correlation was improved for QM/MM minimized structures (R^2 = 0.46), achieving the highest value (R^2 = 0.90) for the model with QM/MM energies of the time-averaged structures and SASA term. This QM/MM-LIE approach was able to capture subtle differences in binding affinities for the most active inhibitors binding to both MMP-9 and MMP-3.

Later the method was extended to consider multispecies, different tautomeric, and ionization states [187]. This approach was applied to 66 inhibitors (233 tautomer/ionization species) of protein kinase (MK2). Ligands were docked into MK2 binding site with the use of FlexiDock. Next, protein-ligand complexes were optimized by the hybrid QM/MM ONIOM method. In the QM system were included ligands, backbone atoms of three residues, and two other full residues (Thr206 and Asp207). The rest of the complex was defined as MM region and treated with Amber force field. QM system was minimized with B3LYP/6-31G(d, p) level of theory. Minimized structures were subjected to 1 ns MD simulations. Single point QM/MM energies for time-averaged structures were computed with the same methods as used for QM/MM minimization. FlexiDock scores resulted in no correlation, and the QM/MM minimized energies with SASA gave R^2 = 0.20. Finally, in the full QM/MM-LIE model, good agreement between calculated and experimental data was achieved (R^2 = 0.90). Neglecting different tautomeric and ionization states lowered correlation to R^2 = 0.66.

Another LIE-like method was developed for the study of 11 inhibitors of HIV-1 integrase [188]. The QM system (ligand) was treated with AM1 semiempirical method, protein, and water molecules represented MM part and were described by classical OPLS-AA force field. From 1ns QM/MM MD simulations, average interaction energies were calculated and compared with experimental data giving quite good correlation (R^2 = 0.82).

A similar approach based on QM/MM MD simulations and the use of AM1 method for QM system was applied for the binding study of eight cyclic (arginine-glycine-aspartate) peptides to $\alpha v\beta_3$ integrin$\alpha\beta$ [189]. QM system involved ligand only, and MM part was treated with CHARMM27 force field. For each complex, 500 frames from the production run were used in further analyses. A linear fit of QM/MM interaction energy correlated well with experimental affinity, which was expressed

by $\log IC_{50}$ (R^2 = 0.89). Even better results were obtained in the three-variable model with electrostatic, van der Waals, and SASA terms (R^2 = 0.94).

The four proteins of pharmaceutical relevance: BACE1, TYK2, HSP90, and PERK were also studied with the QM/MM-LIE method using 22, 16, 70, and 32 ligands, respectively [190]. The QM/MM energies were calculated for each of the 40 snapshots extracted from MD simulations at MM level. In binding free energy calculations, only ligands were treated by the semiempirical AM1 method. The remaining parts of the system were treated by OPLS force field. The good correlation between calculated binding energy and experimental affinity was obtained for each of four systems, 0.73, 0.71, 0.60, and 0.86, respectively.

1.4.7 FEP approaches

The FEP methods give the most accurate results but at the expense of computational time. The evaluation of the free energy of binding involves a large number of intermediate states, making these methods computationally very demanding, thus useful mainly at the lead optimization stage, where a small number of compounds need to be ranked. One of the most rigorous methods for computing (relative) binding free energies is alchemical FEP (AP) in which the binding free-energy difference between bound and unbound states is calculated with the use of statistical mechanics analysis on the trajectories obtained in simulations at MM level. These trajectories define a thermodynamic path, which connects two end states by unphysical (alchemical) intermediate states during the transformation of one ligand into the other while it is bound to protein as well as is free in the solution.

The most widely used alchemical methods are FEP [191], thermodynamic integration (TI) [192], and Bennett acceptance ratio (BAR) [193], [194]. The alchemical transformation of one state into the other is obtained with the use of a coupling parameter λ (λ = 0 for the initial state, and λ = 1 for final state). The final free-energy difference between the two end states is a sum of the alchemical transformations between states for which λ varies from 0 to 1.

Unlike TI and BAR, FEP, also called exponential averaging (EA) compute the relative difference in free-energy change between the bound and unbound state of the two ligands toward the same protein. In this method, alchemical transformations are used to transform one molecule into another structurally related one. To use alchemical methods, we need to know the ligand binding mode. TI or BAR can be applied to estimate the absolute binding free energy for one ligand toward protein or, in the case of FEP, calculate relative binding free energy for two or more structurally related ligands toward the same protein [10].

Besides alchemical transformation methods for absolute binding free energy calculations, physical pathway methods can be used. Pathway methods differ from alchemical transformations because they are aimed at reproducing the physical

binding pathway between unbound ligand and ligand bound to its protein target. These methods estimate the free energy of the ligand binding mechanism and also allow identifying the most energetically favorable bound state as the correct ligand binding mode. Commonly used methods from this group are umbrella sampling (US) [195], [196], SMD [197] and funnel metadynamics [198].

In FEP calculations relative binding free energies $\Delta\Delta G_{bind}$ between two ligands L_A and L_B are estimated with the use of a thermodynamic cycle (Figure 1.2), which relates $\Delta\Delta G_{bind}$ to the free energy of alchemically transforming L_A into L_B in a bound state, $\Delta G^{b}_{(A\to B)}$, and when they are free in solution, $\Delta G^{f}_{(A\to B)}$:

$$\Delta\Delta G^{(A\to B)}_{bind} = \Delta G^{(B)}_{bind} - \Delta G^{(A)}_{bind} = \Delta G^{b}_{(A\to B)} - \Delta G^{f}_{(A\to B)} \tag{1.20}$$

where $\Delta G^{b}_{(A\to B)}$, and $\Delta G^{f}_{(A\to B)}$ are estimated by FEP. The actual binding event is not sampled, only the protein-ligand complex in solution and the solvated ligands. FEP calculations usually are performed at the MM level of theory.

The reliability of the AP approaches depends on the similarity of the two ligands and also on their size [199]. Sometimes even for very similar ligands, the alchemical approach can fail in reproducing the experimental binding free energy difference. The reason for this may be the not enough accurate description of interactions by the MM force field [200]. Classical MM force fields used for simulations of biomolecules involve many approximations. They omit, for example, polarization effects, charge transfer and charge penetration. The force field employed in binding free energy estimations can be improved by a careful parametrization of the ligand [201] or by using a polarizable force field [202], [203].

The simplest way to improve results would be the direct use of quantum-mechanical methods in alchemical simulations. However, the extensive sampling at the QM level is not feasible due to the size of the systems, especially for a level of theory that would guarantee high accuracy. Nevertheless, improving the energy function by using QM or QM/MM calculations can be important for drug candidates possessing in their molecules halogen atoms or aromatic systems. Standard MM force fields poorly describe halogen bonds and π-π or π-cation interactions.

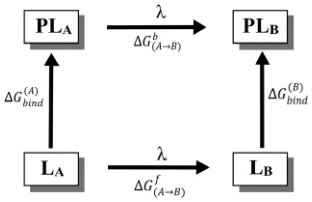


Figure 1.2: Thermodynamic cycle used in alchemical transformations.

QM/MM free energy estimate can be obtained by replacing the MM potential with QM/MM using direct alchemical QM/MM free-energy simulations. However, it is a computationally very expensive solution. Sampling with the QM/MM Hamiltonian can be avoided with the reference potential methods [204], [205], [206]. The free energy difference is estimated here by normal FEP simulation at the MM level, then corrections resulting from changing the model from reference (MM) to QM or QM/MM are evaluated [204], [205], [207], [208]. These additional FEP simulations are performed at the endpoints of the transformation (Figure 1.3), and perturbations need to converge in a single step (single-step EA). The QM or QM/MM free energy between the two states can be obtained as:

$$\Delta G^{QM}_{A\to B} = \Delta G^{MM}_{A\to B} - \Delta G^{(MM\to QM)}_{A} + \Delta G^{(MM\to QM)}_{B} \tag{1.21}$$

Another approach that enables the rigorous computation of free energies based on ensembles generated at MM level is the non-Boltzmann BAR (NBB) [209], [210]. The method uses the bias, as the difference between the approximate (MM energy) and true (QM or QM/MM energy) potentials for the same geometry and use this knowledge to reweight the free energy contributions to obtain the free energy for the unbiased (QM or QM/MM) potential. As a result of the NBB calculations, the free energy estimated at the QM level is obtained.

Only a few studies have been published with full FEP simulations at the QM/MM level. Direct QM/MM-based FEP calculations were used for relative solvation and binding free energies of the five AMP analogs complexed with human fructose-1,6-bisphosphatase (FBPase) [211]. Using AM1 method for the ligand and MM for the protein and solvent, relative binding affinities were reproduced with an error of less than 0.5 kcal/mol. The QM/MM-based FEP method required five times more CPU time than conventional FEP, but it allowed to achieve increased accuracy. The same approach was applied to other 22 FBPase inhibitors giving excellent results, with maximum errors of 0.5 kcal/mol for SQM and less than 1 kcal/mol for MM simulations [212].

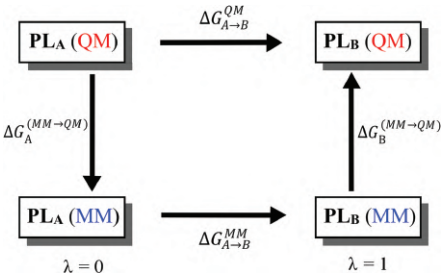


Figure 1.3: Thermodynamic cycle used in the reference-potential methods to obtain relative QM binding energies.

Binding free energies obtained in the FEP method based on MD simulations within hybrid QM/MM potentials were compared with the pathway method (US) for five ligands binding to HIV-1 reverse transcriptase [213]. The only ligand was treated with a semiempirical AM1 method; the rest of the system was described using the OPLS-AA force field (TIP3P model for water molecules). For each ligand, 100 ps MD simulations were used in FEP calculations giving better agreement with experimental data than pathway method. However, both attempts overestimated the value of binding free energy. A very similar approach with the FEP simulation at QM/MM level was also applied to five classical AChE inhibitors [214]. In this case, the use of QM was important due to the polarizable nature of the ligands, and calculated binding free energy reproduced the experimental order of inhibitory potency, giving a very good correlation to the experimental data ($R^2 = 0.96$).

The reference-potential method was applied to study the binding affinity of two inhibitors of COX-2 [215]. Ligands differed only in the single substituent. One ligand with -OH group was transformed into another with the $-CH_3$ group (celecoxib). FEP was performed at the MM level (Amber force field). For QM/MM postprocessing, only the end-points (from MM trajectories) were used, and 3000 QM/MM single point calculations were performed for $\lambda = 0$ and $\lambda = 1$. Calculations for ligand were performed at the B3LYP/6-31G(d, p) level of theory. Results obtained in the QM/MM approach were closer to experimental values than from MM FEP calculations.

Various approaches to QM/MM FEP have been developed, mainly using smaller host-guest systems, they are used and tested in blind-test SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) challenges. Binding free energy of nine cyclic carboxylate guest molecules to the octa-acid host was studied with FEP simulations at MM level and by DFT-D3 calculations [216]. Selected snapshots from FEP MM simulations were reweighted at the DFT-D level. For the ligands, a single-point TPSS-D3 calculation was performed with the def2-QZVP basis set. The def2-TZVP basis set was applied for the host and water molecules. Although three different approaches, which used QM energies such as single-step EA, NBB, and extrapolation corrections, reproduced experimental data worse than FEP calculations at the MM level.

A similar approach was used in studies of the relative binding affinity of two synthetic disaccharide ligands to galectin-3 [217]. The BLYP-D3/def2-SV(P) calculations were used for QM system, which included the ligand, protein groups, and water molecules within 6Å (748 or 744 atoms for the bound state and 527 or 523 for the free state). The rest of the system was treated by an accurate polarizable multipole description. Due to significant differences between potential energy surfaces at the MM and QM level, end-point corrections based on single-step EA and NBB do not converged properly. Only BAR calculations with QM energies, ignoring the differences in the structures sampled at MM and QM/MM potentials, gave comparable results as FEP simulations at MM level.

In a series of publications, the performance of different QM/MM-FEP approaches was also tested for smaller systems [218], [219], [220], [221]. In studies of binding free energies of guest molecules to the octa-acid host for the QM system semiempirical methods PM6-DH2X [218], PM6-DH+ [219], [220], [221] and dispersion-corrected density functional TPSS-D3 method [221] were applied.

1.5 Conclusions

This chapter has attempted to present the computational chemistry methods used in protein-ligand binding affinity predictions. In binding affinity calculations, many different QM methods are applicable, ranging from SQM approaches, through DFT, to advanced CCSD(T) calculations. Dispersion corrections to DFT and more accurate semi-empirical QM approaches have significantly improved the results of calculations for nonpolar interactions that play a very important role in protein-ligand complexes. Moreover, QM methods include effects not considered in classical force fields as a polarization of the ligand, charge transfer, metal-coordination bonds, and others. With the use of the QM approach, the parameterization of force fields for the novel compounds can also be avoided. Quantum mechanical methods do not always lead to improved calculated binding affinities. Often they give comparable or even worse results than classical MM approaches. However, continuous progress in QM methods gives prospects for quantum chemical approaches applied to binding affinity calculations to be widespread use in drug design, on the timescale required by pharmaceutical companies.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. DiMasi, JA, Grabowski, HG, Hansen, RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ* 2016;47:20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012.
2. Parenti, MD, Rastelli, G. Advances and applications of binding affinity prediction methods in drug discovery. *Biotechnol Adv* 2012;30:244–50. https://doi.org/10.1016/j.biotechadv.2011.08.003.
3. Paul, SM, Mytelka, DS, Dunwiddie, CT, Persinger, CC, Munos, BH, Lindborg, SR, et al.et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 2010;9:203–14. https://doi.org/10.1038/nrd3078.

4. Morgan, S, Grootendorst, P, Lexchin, J, Cunningham, C, Greyson, D. The cost of drug development: a systematic review. *Health Pol* 2011;100:4–17. https://doi.org/10.1016/j.healthpol.2010.12.002.
5. Cook, D, Brown, D, Alexander, R, March, R, Morgan, P, Satterthwaite, G, et al.et al. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov* 2014;13:419–31. https://doi.org/10.1038/nrd4309.
6. Abel, R, Manas, ES, Friesner, RA, Farid, RS, Wang, L. Modeling the value of predictive affinity scoring in preclinical drug discovery. *Curr Opin Struct Biol* 2018;52:103–10. https://doi.org/10.1016/j.sbi.2018.09.002.
7. Bissantz, C, Kuhn, B, Stahl, M. A medicinal chemist's guide to molecular interactions. *J Med Chem* 2010;53:5061–84. https://doi.org/10.1021/jm100112j.
8. Barril, X, Javier Luque, F. Molecular simulation methods in drug discovery: a prospective outlook. *J Comput Aided Mol Des* 2012;26:81–6. https://doi.org/10.1007/s10822-011-9506-1.
9. Montalvo-Acosta, JJ, Cecchini, M. Computational approaches to the chemical equilibrium constant in protein-ligand binding. *Mol Inform* 2016;35:555–67. https://doi.org/10.1002/minf.201600052.
10. Limongelli, V. Ligand binding free energy and kinetics calculation in 2020. *WIREs Comput Mol Sci* 2020;10:e1455. https://doi.org/10.1002/wcms.1455.
11. de Ruiter, A, Oostenbrink, C. Advances in the calculation of binding free energies. *Curr Opin Struct Biol* 2020;61:207–12. https://doi.org/10.1016/j.sbi.2020.01.016.
12. Shaw, KE, Woods, CJ, Mulholland, AJ. QM and QM/MM approaches to evaluating binding affinities. *Burger's Med Chem Drug Discov* 2010:725–52. https://doi.org/10.1002/0471266949.bmc143.
13. Mucs, D, Bryce, RA. The application of quantum mechanics in structure-based drug design. *Expet Opin Drug Discov* 2013;8:263–76. https://doi.org/10.1517/17460441.2013.752812.
14. Ryde, U, Söderhjelm, P. Ligand-binding affinity estimates supported by quantum-mechanical methods. *Chem Rev* 2016;116:5520–66. https://doi.org/10.1021/acs.chemrev.5b00630.
15. Cavasotto, CN, Adler, NS, Aucar, MG. Quantum chemical approaches in structure-based virtual screening and lead optimization. *Front Chem* 2018;6:1–7. https://doi.org/10.3389/fchem.2018.00188.
16. Kairys, V, Baranauskiene, L, Kazlauskiene, M, Matulis, D, Kazlauskas, E. Binding affinity in drug design: experimental and computational techniques. *Expet Opin Drug Discov* 2019;14:755–68. https://doi.org/10.1080/17460441.2019.1623202.
17. Ajay, Murcko, MA. Computational methods to predict binding free energy in ligand-receptor complexes. *J Med Chem* 1995;38:4953–67. https://doi.org/10.1021/jm00026a001.
18. Raha, K, Merz, KM. A quantum mechanics-based scoring function: study of zinc ion-mediated ligand binding. *J Am Chem Soc* 2004;126:1020–1. https://doi.org/10.1021/ja038496i.
19. Van Der Vaart, A, Merz, KM. The role of polarization and charge transfer in the solvation of biomolecules. *J Am Chem Soc* 1999;121:9182–90. https://doi.org/10.1021/ja9912325.
20. Garcia-Viloca, M, Truhlar, DG, Gao, J. Importance of substrate and cofactor polarization in the active site of dihydrofolate reductase. *J Mol Biol* 2003;327:549–60. https://doi.org/10.1016/s0022-2836(03)00123-2.
21. Ben-Naim, A, Marcus, Y. Solvation thermodynamics of nonionic solutes. *J Chem Phys* 1984;81:2016–27. https://doi.org/10.1063/1.447824.
22. Sharp, KA, Honig, B. Salt effects on nucleic acids. *Curr Opin Struct Biol* 1995;5:323–8. https://doi.org/10.1016/0959-440x(95)80093-x.
23. Warwicker, J, Watson, HC. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J Mol Biol* 1982;157:671–9. https://doi.org/10.1016/0022-2836(82)90505-8.

24. Clark Still, W, Tempczyk, A, Hawley, RC, Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* 1990;112:6127–9. https://doi.org/10.1021/ja00172a038.
25. Eisenberg, D, Mclachlan, AD. Solvation energy in protein folding and binding. *Nature* 1986;319:199–203. https://doi.org/10.1038/319199a0.
26. Gogonea, V, Merz, KM. Fully quantum mechanical description of proteins in solution. Combining linear scaling quantum mechanical methodologies with the Poisson-Boltzmann equation. *J Phys Chem A* 1999;103:5171–88. https://doi.org/10.1021/jp990266w.
27. Miertuš, S, Scrocco, E, Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects. *Chem Phys* 1981;55:117–29. https://doi.org/10.1016/0301-0104(81)85090-2.
28. Cossi, M, Tomasi, J, Cammi, R. Analytical expressions of the free energy derivatives for molecules in solution. Application to the geometry optimization. *Int J Quant Chem* 1995;56:695–702. https://doi.org/10.1002/qua.560560876.
29. Tomasi, J, Mennucci, B, Cammi, R. Quantum mechanical continuum solvation models. *Chem Rev* 2005;105:2999–3093. https://doi.org/10.1021/cr9904009.
30. Klamt, A, Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc Perkin Trans* 1993;2:799–805. https://doi.org/10.1039/p29930000799.
31. Schwarzl, SM, Tschopp, TB, Smith, JC, Fischer, S. Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction?. *J Comput Chem* 2002;23:1143–9. https://doi.org/10.1002/jcc.10112.
32. Searle, MS, Williams, DH. The cost of conformational order: entropy changes in molecular associations. *J Am Chem Soc* 1992;114:10690–7. https://doi.org/10.1021/ja00053a002.
33. Brooks, BR, Bruccoleri, RE, Olafson, BD, States, DJ, Swaminathan, S, Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 1983;4:187–217. https://doi.org/10.1002/jcc.540040211.
34. Cornell, WD, Cieplak, P, Bayly, CI, Gould, IR, Merz, KM, Ferguson, DM, et al.et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 1995;117:5179–97. https://doi.org/10.1021/ja00124a002.
35. Scott, WRP, Hünenberger, PH, Tironi, IG, Mark, AE, Billeter, SR, Fennen, J, et al.et al. The GROMOS biomolecular simulation program package. *J Phys Chem A* 1999;103:3596–607. https://doi.org/10.1021/jp984217f.
36. van Gunsteren, WF, Berendsen, HJC. *Groningen molecular simulation ({GROMOS}) library manual*. Groningen: Biomos; 1987.
37. Vanommeslaeghe, K, Hatcher, E, Acharya, C, Kundu, S, Zhong, S, Shim, J, et al.et al. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 2009;31:NA-NA. https://doi.org/10.1002/jcc.21367.
38. Vanommeslaeghe, K, MacKerell, AD. Automation of the CHARMM general force field (CGenFF) I: bond perception and atom typing. *J Chem Inf Model* 2012;52:3144–54. https://doi.org/10.1021/ci300363c.
39. Vanommeslaeghe, K, Raman, EP, MacKerell, AD. Automation of the CHARMM general force field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J Chem Inf Model* 2012;52:3155–68. https://doi.org/10.1021/ci3003649.
40. Yu, W, He, X, Vanommeslaeghe, K, MacKerell, AD. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J Comput Chem* 2012;33:2451–68. https://doi.org/10.1002/jcc.23067.

41. Wang, J, Wolf, RM, Caldwell, JW, Kollman, PA, Case, DA. Development and testing of a general amber force field. *J Comput Chem* 2004;25:1157–74. https://doi.org/10.1002/jcc.20035.
42. Wang, J, Wang, W, Kollman, PA, Case, DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* 2006;25:247–60. https://doi.org/10.1016/j.jmgm.2005.12.005.
43. Pople, JA, Santry, DP, Segal, GA. Approximate self-consistent molecular orbital theory. I. Invariant procedures. *J Chem Phys* 1965;43:S129–35. https://doi.org/10.1063/1.1701476.
44. Dewar, MJS, Zoebisch, EG, Healy, EF, Stewart, JJP. AM1: a new general purpose quantum mechanical molecular Model1. *J Am Chem Soc* 1985;107:3902–9. https://doi.org/10.1021/ja00299a024.
45. Stewart, JJP. Optimization of parameters for semiempirical methods I. Method. *J Comput Chem* 1989;10:209–20. https://doi.org/10.1002/jcc.540100208.
46. Rocha, GB, Freire, RO, Simas, AM, Stewart, JJPP. RM1: a reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J Comput Chem* 2006;27:1101–11. https://doi.org/10.1002/jcc.20425.
47. Stewart, JJP. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J Mol Model* 2007;13:1173–213. https://doi.org/10.1007/s00894-007-0233-4.
48. Řezáč, J, Fanfrlík, J, Salahub, D, Hobza, P. Semiempirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of noncovalent complexes. *J Chem Theor Comput* 2009;5:1749–60. https://doi.org/10.1021/ct9000922.
49. Korth, M, Pitoňák, M, Řezáč, J, Hobza, P. A transferable H-bonding correction for semiempirical quantum-chemical methods. *J Chem Theor Comput* 2010;6:344–52. https://doi.org/10.1021/ct900541n.
50. Řezáč, J, Hobza, P. A halogen-bonding correction for the semiempirical PM6 method. *Chem Phys Lett* 2011;506:286–9. https://doi.org/10.1016/j.cplett.2011.03.009.
51. Řezáč, J, Hobza, P. Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods. *J Chem Theor Comput* 2012;8:141–51. https://doi.org/10.1021/ct200751e.
52. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J Comput Chem* 2006;27:1787–99. https://doi.org/10.1002/jcc.20495.
53. Grimme, S, Antony, J, Ehrlich, S, Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 2010:132. https://doi.org/10.1063/1.3382344.
54. Grimme, S, Ehrlich, S, Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* 2011;32:1456–65. https://doi.org/10.1002/jcc.21759.
55. Risthaus, T, Grimme, S. Benchmarking of London dispersion-accounting density functional theory methods on very large molecular complexes. *J Chem Theor Comput* 2013;9:1580–91. https://doi.org/10.1021/ct301081n.
56. Burns, LA, Vázquez-Mayagoitia, Á, Sumpter, BG, Sherrill, CD. Density-functional approaches to noncovalent interactions: a comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals. *J Chem Phys* 2011:134. https://doi.org/10.1063/1.3545971.
57. Otero-De-La-Roza, A, Johnson, ER. Predicting energetics of supramolecular systems using the XDM dispersion model. *J Chem Theor Comput* 2015;11:4033–40. https://doi.org/10.1021/acs.jctc.5b00044.

58. Boys, SF, Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol Phys* 1970;19:553–66. https://doi.org/10.1080/00268977000101561.
59. Senn, HM, Thiel, W. QM/MM methods for biomolecular systems. *Angew Chem Int Ed* 2009;48:1198–229. https://doi.org/10.1002/anie.200802019.
60. Lin, H, Truhlar, DG. QM/MM: what have we learned, where are we, and where do we go from here?. *Theor. Chem. Acc.* 2007;117:185–99. https://doi.org/10.1007/s00214-006-0143-z.
61. Svensson, M, Humbel, S, Froese, RDJ, Matsubara, T, Sieber, S, Morokuma, K. ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)3)2 + H2 oxidative addition. *J Phys Chem* 1996;100:19357–63. https://doi.org/10.1021/jp962071j.
62. Poulsen, TD, Kongsted, J, Osted, A, Ogilby, PR, Mikkelsen, KV. The combined multiconfigurational self-consistent-field/molecular mechanics wave function approach. *J Chem Phys* 2001;115:2393–400. https://doi.org/10.1063/1.1374559.
63. Söderhjelm, P, Husberg, C, Strambi, A, Olivucci, M, Ryde, U. Protein influence on electronic spectra modeled by multipoles and polarizabilities. *J Chem Theor Comput* 2009;5:649–58. https://doi.org/10.1021/ct800459t.
64. Stewart, JJP. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *Int J Quant Chem* 1996;58:133–46. https://doi.org/10.1002/(sici)1097-461x(1996)58:2<133::aid-qua2>3.0.co;2-z.
65. Daniels, AD, Millam, JM, Scuseria, GE. Semiempirical methods with conjugate gradient density matrix search to replace diagonalization for molecular systems containing thousands of atoms. *J Chem Phys* 1997;107:425–31. https://doi.org/10.1063/1.474404.
66. Dixon, SL, Merz, KM. Fast, accurate semiempirical molecular orbital calculations for macromolecules. *J Chem Phys* 1997;107:879–93. https://doi.org/10.1063/1.474386.
67. Ochsenfeld, C, Kussmann, J, Lambrecht, DS. *Chapter 1 linear-scaling methods in quantum chemistry*. In: Kenny B. Lipkowitz, Thomas R. Cundari, editors. Reviews in computational chemistry. Hoboken: Wiley-VCH, John Wiley & Sons; 2006, vol 23.
68. Fedorov, DG, Kitaura, K. Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J Phys Chem A* 2007;111:6904–14. https://doi.org/10.1021/jp0716740.
69. Fedorov, DG, Nagata, T, Kitaura, K. Exploring chemistry with the fragment molecular orbital method. *Phys Chem Chem Phys* 2012;14:7562–77. https://doi.org/10.1039/c2cp23784a.
70. Fedorov, DG. The fragment molecular orbital method: theoretical development, implementation in gamess, and applications. *Wiley Interdiscip Rev Comput Mol Sci* 2017;7:1–17. https://doi.org/10.1002/wcms.1322.
71. Kitaura, K, Ikeo, E, Asada, T, Nakano, T, Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem Phys Lett* 1999;313:701–6. https://doi.org/10.1016/s0009-2614(99)00874-x.
72. Zhang, DW, Zhang, JZH. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J Chem Phys* 2003;119:3599–605. https://doi.org/10.1063/1.1591727.
73. He, X, Zhu, T, Wang, X, Liu, J, Zhang, JZH. Fragment quantum mechanical calculation of proteins and its applications. *Acc Chem Res* 2014;47:2748–57. https://doi.org/10.1021/ar500077t.
74. Söderhjelm, P, Ryde, U. How accurate can a force field become? A polarizable multipole model combined with fragment-wise quantum-mechanical calculations. *J Phys Chem A* 2009;113:617–27. https://doi.org/10.1021/jp8073514.

75. Huang, SY, Grinter, SZ, Zou, X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys* 2010;12:12899–908. https://doi.org/10.1039/c0cp00151a.
76. Danishuddin, M, Khan, AU. Structure based virtual screening to discover putative drug candidates: necessary considerations and successful case studies. *Methods* 2015;71:135–45. https://doi.org/10.1016/j.ymeth.2014.10.019.
77. Ferreira, LG, Dos Santos, RN, Oliva, G, Andricopulo, AD. Molecular docking and structure-based drug design strategies. *Molecules* 2015;20:13384–421. https://doi.org/10.3390/molecules200713384 .
78. Liu, J, Wang, R. Classification of current scoring functions. *J Chem Inf Model* 2015;55:475–82. https://doi.org/10.1021/ci500731a.
79. Gilson, MK, Given, JA, Head, MS. A new class of models for computing receptor-ligand binding affinities. *Chem Biol* 1997;4:87–92. https://doi.org/10.1016/s1074-5521(97)90251-9.
80. Zou, X, Sun, Y, Kuntz, ID. Inclusion of solvation in ligand binding free energy calculations using the generalized-born model. *J Am Chem Soc* 1999;121:8033–43. https://doi.org/10.1021/ja984102p.
81. Jones, G, Willett, P, Glen, RC, Leach, AR, Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 1997;267:727–48. https://doi.org/10.1006/jmbi.1996.0897.
82. Yin, S, Biedermannova, L, Vondrasek, J, Dokholyan, NV. MedusaScore: an accurate force field-based scoring function for virtual drug screening. *J Chem Inf Model* 2008;48:1656–62. https://doi.org/10.1021/ci8001167.
83 Abagyan, R, Totrov, M, Kuznetsov, D. ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* 1994;15:488–506. https://doi.org/10.1002/jcc.540150503.
84. Venkatachalam, CM, Jiang, X, Oldfield, T, Waldman, M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J Mol Graph Model* 2003;21:289–307. https://doi.org/10.1016/s1093-3263(02)00164-x.
85. Böhm, HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des* 1994;8:243–56.
86. Morris, GM, Goodsell, DS, Huey, R, Olson, AJ. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J Comput Aided Mol Des* 1996;10:293–304. https://doi.org/10.1007/bf00124499.
87. Eldridge, MD, Murray, CW, Auton, TR, Paolini, GV, Mee, RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 1997;11:425–45. https://doi.org/10.1023/a:1007996124545.
88. Krammer, A, Kirchhoff, PD, Jiang, X, Venkatachalam, CM, Waldman, M. LigScore: a novel scoring function for predicting binding affinities. *J Mol Graph Model* 2005;23:395–407. https://doi.org/10.1016/j.jmgm.2004.11.007.
89. Gehlhaar, DK, Verkhivker, GM, Rejto, PA, Sherman, CJ, Fogel, DR, Fogel, LJ, et al.et al. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem Biol* 1995;2:317–24. https://doi.org/10.1016/1074-5521(95)90050-0.
90. DeWitte, RS, Shakhnovich, EI. SMoG: de novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *J Am Chem Soc* 1996;118:11733–44. https://doi.org/10.1021/ja960751u.

91. Muegge, I, Martin, YC. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem* 1999;42:791–804. https://doi.org/10.1021/jm980536j.
92. Muegge, I. A knowledge-based scoring function for protein-ligand interactions: probing the reference state. *Perspect Drug Discov Des* 2000;20:99–114. https://doi.org/10.1023/A:1008729005958.
93. Muegge, I. Effect of ligand volume correction on PMF scoring. *J Comput Chem* 2001;22:418–25. https://doi.org/10.1002/1096-987x(200103)22:4<418::aid-jcc1012>3.0.co;2-3.
94. Gohlke, H, Hendlich, M, Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* 2000;295:337–56. https://doi.org/10.1006/jmbi.1999.3371.
95. Velec, HFG, Gohlke, H, Klebe, G. DrugScoreCSD-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J Med Chem* 2005;48:6296–303. https://doi.org/10.1021/jm050436v.
96. Neudert, G, Klebe, GDSX. A knowledge-based scoring function for the assessment of protein-ligand complexes. *J Chem Inf Model* 2011;51:2731–45. https://doi.org/10.1021/ci200274q.
97. Huang, SY, Zou, X. An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. *J Comput Chem* 2006;27:1866–75. https://doi.org/10.1002/jcc.20504.
98. Huang, SY, Zou, X. An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. *J Comput Chem* 2006;27:1876–82. https://doi.org/10.1002/jcc.20505.
99. Huang, SY, Zou, X. Inclusion of solvation and entropy in the knowledge-based scoring function for protein-ligand interactions. *J Chem Inf Model* 2010;50:262–73. https://doi.org/10.1021/ci9002987.
100. Durrant, JD, McCammon, JA. NNScore: a neural-network-based scoring function for the characterization of protein-ligand complexes. *J Chem Inf Model* 2010;50:1865–71. https://doi.org/10.1021/ci100244v.
101. Durrant, JD, McCammon, JA. NNScore 2.0: a neural-network receptor-ligand scoring function. *J Chem Inf Model* 2011;51:2897–903. https://doi.org/10.1021/ci2003889.
102. Ballester, PJ, Mitchell, JBO. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics* 2010;26:1169–75. https://doi.org/10.1093/bioinformatics/btq112.
103. Ballester, PJ, Schreyer, A, Blundell, TL. Does a more precise chemical description of protein-ligand complexes lead to more accurate prediction of binding affinity?. *J Chem Inf Model* 2014;54:944–55. https://doi.org/10.1021/ci500091r.
104. Zilian, D, Sotriffer, CA. SFCscoreRF: a random forest-based scoring function for improved affinity prediction of protein-ligand complexes. *J Chem Inf Model* 2013;53:1923–33. https://doi.org/10.1021/ci400120b.
105. Li, GB, Yang, LL, Wang, WJ, Li, LL, Yang, SY. ID-score: a new empirical scoring function based on a comprehensive set of descriptors related to protein-ligand interactions. *J Chem Inf Model* 2013;53:592–600. https://doi.org/10.1021/ci300493w.
106. Dixon, SL, Merz, KM. Semiempirical molecular orbital calculations with linear system size scaling. *J Chem Phys* 1996;104:6643–9. https://doi.org/10.1063/1.471382.
107. Raha, K, Merz, KM. Large-Scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. *J Med Chem* 2005;48:4558–75. https://doi.org/10.1021/jm048973n.

108. Fanfrlík, J, Bronowska, AK, Řezáč, J, Přenosil, O, Konvalinka, J, Hobza, P. A reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and H-bonding: HIV-1 protease with 22 ligands. *J Phys Chem B* 2010;114:12666–78. https://doi.org/10.1021/jp1032965.
109. Kolář, M, Fanfrlík, J, Hobza, P. Ligand conformational and solvation/desolvation free energy in protein-ligand complex formation. *J Phys Chem B* 2011;115:4718–24. https://doi.org/10.1021/jp2010265.
110. Dobeš, P, Fanfrlík, J, Řezáč, J, Otyepka, M, Hobza, P. Transferable scoring function based on semiempirical quantum mechanical PM6-DH2 method: CDK2 with 15 structurally diverse inhibitors. *J Comput Aided Mol Des* 2011;25:223–35. https://doi.org/10.1007/s10822-011-9413-5.
111. Dobeš, P, Řezáč, J, Fanfrlík, J, Otyepka, M, Hobza, P. Semiempirical quantum mechanical method PM6-DH2X describes the geometry and energetics of CK2-inhibitor complexes involving halogen bonds well, while the empirical potential fails. *J Phys Chem B* 2011;115:8581–9. https://doi.org/10.1021/jp202149z.
112. Marenich, AV, Cramer, CJ, Truhlar, DG. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J Phys Chem B* 2009;113:6378–96. https://doi.org/10.1021/jp810292n.
113. Lepšík, M, Řezáč, J, Kolář, M, Pecina, A, Hobza, P, Fanfrlík, J. The semiempirical quantum mechanical scoring function for in silico drug design. *ChemPlusChem* 2013;78:921–31. https://doi.org/10.1002/cplu.201300199.
114. Pecina, A, Meier, R, Fanfrlík, J, Lepšík, M, Řezáč, J, Hobza, P, et al.et al. The SQM/COSMO filter: reliable native pose identification based on the quantum-mechanical description of protein-ligand interactions and implicit COSMO solvation. *Chem Commun* 2016;52:3312–5. https://doi.org/10.1039/c5cc09499b.
115. Pecina, A, Haldar, S, Fanfrlík, J, Meier, R, Řezáč, J, Lepšík, M, et al.et al. SQM/COSMO scoring function at the DFTB3-D3H4 level: unique identification of native protein-ligand poses. *J Chem Inf Model* 2017;57:127–32. https://doi.org/10.1021/acs.jcim.6b00513.
116. Ajani, H, Pecina, A, Eyrilmez, SM, Fanfrlík, J, Haldar, S, Řezáč, J, et al.et al. Superior performance of the SQM/COSMO scoring functions in native pose recognition of diverse protein-ligand complexes in cognate docking. *ACS Omega* 2017;2:4022–9. https://doi.org/10.1021/acsomega.7b00503.
117. Ciancetta, A, Genheden, S, Ryde, U. A QM/MM study of the binding of RAPTA ligands to cathepsin B. *J Comput Aided Mol Des* 2011;25:729–42. https://doi.org/10.1007/s10822-011-9448-7.
118. Li, Y, Yang, Y, He, P, Yang, Q. QM/MM study of epitope peptides binding to HLA-A*0201: the roles of anchor residues and water. *Chem Biol Drug Des* 2009;74:611–8. https://doi.org/10.1111/j.1747-0285.2009.00896.x.
119. Li, Q, Gusarov, S, Evoy, S, Kovalenko, A. Electronic structure, binding energy, and solvation structure of the streptavidin-biotin supramolecular complex: ONIOM and 3D-RISM study. *J Phys Chem B* 2009;113:9958–67. https://doi.org/10.1021/jp902668c.
120. Burger, SK, Thompson, DC, Ayers, PW. Quantum mechanics/molecular mechanics strategies for docking pose refinement: distinguishing between binders and decoys in cytochrome c peroxidase. *J Chem Inf Model* 2011;51:93–101. https://doi.org/10.1021/ci100329z.
121. Gleeson, MP, Gleeson, D. QM/MM as a tool in fragment based drug discovery. A cross-docking, rescoring study of kinase inhibitors. *J Chem Inf Model* 2009;49:1437–48. https://doi.org/10.1021/ci900022h.

122. Gleeson, MP, Burton, NA, Hillier, IH. Prediction of the potency of inhibitors of adenosine deaminase by QM/MM calculations. *Chem Commun* 2003;3:2180–1. https://doi.org/10.1039/b305790a.
123. Tian, F, Yang, L, Lv, F, Luo, X, Pan, Y. Why OppA protein can bind sequence-independent peptides? A combination of QM/MM, PB/SA, and structure-based QSAR analyses. *Amino Acids* 2011;40:493–503. https://doi.org/10.1007/s00726-010-0661-9.
124. Guo, X, He, D, Huang, L, Liu, L, Liu, L, Yang, H. Strain energy in enzyme-substrate binding: an energetic insight into the flexibility versus rigidity of enzyme active site. *Comput Theor Chem* 2012;995:17–23. https://doi.org/10.1016/j.comptc.2012.06.017.
125. Guo, X, He, D, Liu, L, Kuang, R, Liu, L. Use of QM/MM scheme to reproduce macromolecule-small molecule noncovalent binding energy. *Comput Theor Chem* 2012;991:134–40. https://doi.org/10.1016/j.comptc.2012.04.010.
126. Ai, X, Sun, Y, Wang, H, Lu, S. A systematic profile of clinical inhibitors responsive to EGFR somatic amino acid mutations in lung cancer: implication for the molecular mechanism of drug resistance and sensitivity. *Amino Acids* 2014;46:1635–48. https://doi.org/10.1007/s00726-014-1716-0.
127. Yang, L, Mo, X, Yang, H, Dai, H, Tan, F. Testing the sensitivities of noncognate inhibitors to varicella zoster virus thymidine kinase: implications for postherpetic neuralgia therapy with existing agents. *J Mol Model* 2014;20:1–10. https://doi.org/10.1007/s00894-014-2321-6.
128. Hayik, SA, Dunbrack, R, Merz, KM. Mixed quantum mechanics/molecular mechanics scoring function to predict protein-ligand binding affinity. *J Chem Theor Comput* 2010;6:3079–91. https://doi.org/10.1021/ct100315g.
129. Brahmkshatriya, PS, Dobes, P, Fanfrlik, J, Rezac, J, Paruch, K, Bronowska, A, et al.et al. Quantum mechanical scoring: structural and energetic insights into cyclin-dependent kinase 2 inhibition by pyrazolo[1,5-a]pyrimidines. *Curr Comput Aided Drug Des* 2013;9:118–29. https://doi.org/10.2174/157340913804998784.
130. Stewart, JJP. *Stewart computational chemistry. Colorado Springs. CO USA. MOPAC2016*. Avaiable from: http://OpenMOPAC.net. [accessed 14 February 2020].
131. Vasilyev, V, Bliznyuk, A. Application of semiempirical quantum chemical methods as a scoring function in docking. *Theor Chem Acc* 2004;112:313–7. https://doi.org/10.1007/s00214-004-0589-9.
132. Pichierri, F. A quantum mechanical study on phosphotyrosyl peptide binding to the SH2 domain of p56lck tyrosine kinase with insights into the biochemistry of intracellular signal transduction events. *Biophys Chem* 2004;109:295–304. https://doi.org/10.1016/j.bpc.2003.12.006.
133. Ohno, K, Wada, M, Saito, S, Inoue, Y, Sakurai, M. Quantum chemical study on the affinity maturation of 48G7 antibody. *J Mol Struct Theochem* 2005;722:203–11. https://doi.org/10.1016/j.theochem.2004.11.061.
134. Li, J, Reynolds, CH. A quantum mechanical approach to ligand binding — calculation of ligand–protein binding affinities for stromelysin-1 (MMP-3) inhibitors. *Can J Chem* 2009;87:1480–4. https://doi.org/10.1139/v09-129.
135. Kamel, K, Kolinski, A. Assessment of the free binding energy of 1,25-dihydroxyvitamin D3 and its analogs with the human VDR receptor model. *Acta Biochim Pol* 2012;59:653–60. https://doi.org/10.18388/abp.2012_2106.
136. Fukuzawa, K, Kitaura, K, Uebayasi, M, Nakata, K, Kaminuma, T, Nakano, T. Ab initio quantum mechanical study of the binding energies of human estrogen receptor α with its ligands: an application of fragment molecular orbital method. *J Comput Chem* 2005;26:1–10. https://doi.org/10.1002/jcc.20130.

137. Nakanishi, I, Fedorov, DG, Kitaura, K. Molecular recognition mechanism of FK506 binding protein: an all-electron fragment molecular orbital study. *Proteins Struct Funct Bioinforma* 2007;68:145–58. https://doi.org/10.1002/prot.21389.
138. Mazanetz, MP, Ichihara, O, Law, RJ, Whittaker, M. Prediction of cyclin-dependent kinase 2 inhibitor potency using the fragment molecular orbital method. *J Cheminf* 2011;3:1–15. https://doi.org/10.1186/1758-2946-3-2.
139. Dedachi, K, Hirakawa, T, Fujita, S, Khan, MTH, Sylte, I, Kurita, N. Specific interactions and binding free energies between thermolysin and dipeptides: molecular simulations combined with Ab initio molecular orbital and classical vibrational analysis. *J Comput Chem* 2011;32:3047–57. https://doi.org/10.1002/jcc.21887.
140. Asada, N, Fedorov, DG, Kitaura, K, Nakanishi, I, Merz, KM. An efficient method to evaluate intermolecular interaction energies in large systems using overlapping multicenter ONIOM and the fragment molecular orbital method. *J Phys Chem Lett* 2012;3:2604–10. https://doi.org/10.1021/jz3010688.
141. Zhang, DW, Xiang, Y, Zhang, JZH. New advance in computational chemistry: full quantum mechanical ab initio computation of Streptavidin–Biotin interaction energy. *J Phys Chem B* 2003;107:12039–41. https://doi.org/10.1021/jp0359081.
142. Zhang, DW, Xiang, Y, Gao, AM, Zhang, JZH. Quantum mechanical map for protein-ligand binding with application to β-trypsin/benzamidine complex. *J Chem Phys* 2004;120:1145–8. https://doi.org/10.1063/1.1639152.
143. Mei, Y, He, X, Xiang, Y, Zhang, DW, Zhang, JZH. Quantum study of mutational effect in binding of efavirenz to HIV-1 RT. *Proteins Struct Funct Bioinforma* 2005;59:489–95. https://doi.org/10.1002/prot.20455.
144. Bettens, RPA, Lee, AM. On the accurate reproduction of ab initio interaction energies between an enzyme and substrate. *Chem Phys Lett* 2007;449:341–6. https://doi.org/10.1016/j.cplett.2007.10.073.
145. Ding, Y, Mei, Y, Zhang, JZH. Quantum mechanical studies of residue-specific hydrophobic interactions in p53-MDM2 binding. *J Phys Chem B* 2008;112:11396–401. https://doi.org/10.1021/jp8015886.
146. Da Costa, RF, Freire, VN, Bezerra, EM, Cavada, BS, Caetano, EWS, De Lima Filho, JL, et al. Explaining statin inhibition effectiveness of HMG-CoA reductase by quantum biochemistry computations. *Phys Chem Chem Phys* 2012;14:1389–98. https://doi.org/10.1039/c1cp22824b.
147. Martins, ACV, De Lima-Neto, P, Barroso-Neto, IL, Cavada, BS, Freire, VN, Caetano, EWS. An ab initio explanation of the activation and antagonism strength of an AMPA-sensitive glutamate receptor. *RSC Adv* 2013;3:14988–92. https://doi.org/10.1039/c3ra42149j.
148. Rodrigues, CRF, Oliveira, JIN, Fulco, UL, Albuquerque, EL, Moura, RM, Caetano, EWS, et al. et al. Quantum biochemistry study of the T3-785 tropocollagen triple-helical structure. *Chem Phys Lett* 2013;559:88–93. https://doi.org/10.1016/j.cplett.2012.12.061.
149. Da Silva Ribeiro, TC, Da Costa, RF, Bezerra, EM, Freire, VN, Lyra, ML, Manzoni, V. The quantum biophysics of the isoniazid adduct NADH binding to its InhA reductase target. *New J Chem* 2014;38:2946–57. https://doi.org/10.1039/c3nj01453c.
150. Zanatta, G, Nunes, G, Bezerra, EM, Da Costa, RF, Martins, A, Caetano, EWS, et al.et al. Antipsychotic haloperidol binding to the human dopamine D3 receptor: beyond docking through QM/MM refinement toward the design of improved schizophrenia medicines. *ACS Chem Neurosci* 2014;5:1041–54. https://doi.org/10.1021/cn500111e.
151. Dantas, DS, Oliveira, JIN, Lima Neto, JX, Da Costa, RF, Bezerra, EM, Freire, VN, et al.et al. Quantum molecular modelling of ibuprofen bound to human serum albumin. *RSC Adv* 2015;5:49439–50. https://doi.org/10.1039/c5ra04395f.

152. Ehrlich, S, Göller, AH, Grimme, S. Towards full quantum-mechanics-based protein-ligand binding affinities. *ChemPhysChem* 2017;18:898–905. https://doi.org/10.1002/cphc.201700082.
153. Sure, R, Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J Comput Chem* 2013;34:1672–85. https://doi.org/10.1002/jcc.23317.
154. Barone, V, Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J Phys Chem A* 1998;102:1995–2001. https://doi.org/10.1021/jp9716997.
155. Aradi, B, Hourahine, B, Frauenheim, T. DFTB+ a sparse matrix-based implementation of the DFTB method. *J Phys Chem A* 2007;11:5678–84. https://doi.org/10.1021/jp070186p.
156. Gaus, M, Cui, Q, Elstner, M. DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J Chem Theor Comput* 2011;7:931–48. https://doi.org/10.1021/ct100684s.
157. Brandenburg, JG, Grimme, S. Accurate modeling of organic molecular crystals by dispersion-corrected density functional tight binding (DFTB). *J Phys Chem Lett* 2014;5:1785–9. https://doi.org/10.1021/jz500755u.
158. Klamt, A. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *J Phys Chem* 1995;99:2224–35. https://doi.org/10.1021/j100007a062.
159. Eckert, F, Klamt, A. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J* 2002;48:369–85. https://doi.org/10.1002/aic.690480220.
160. Klamt, A. The COSMO and COSMO-RS solvation models. *WIREs Comput Mol Sci* 2011;1:699–709. https://doi.org/10.1002/wcms.56.
161. Wang, E, Sun, H, Wang, J, Wang, Z, Liu, H, Zhang, JZH, et al.et al. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem Rev* 2019;119:9478–508. https://doi.org/10.1021/acs.chemrev.9b00055.
162. Srinivasan, J, Cheatham, TE, Cieplak, P, Kollman, PA, Case, DA. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J Am Chem Soc* 1998;120:9401–9. https://doi.org/10.1021/ja981844+.
163. Kollman, PA, Massova, I, Reyes, C, Kuhn, B, Huo, S, Chong, L, et al.et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 2000;33:889–97. https://doi.org/10.1021/ar000033j.
164. Kuhn, B, Kollman, PA. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J Med Chem* 2000;43:3786–91. https://doi.org/10.1021/jm000241h.
165. Gilson, MK, Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins Struct Funct Bioinforma* 1988;4:7–18. https://doi.org/10.1002/prot.340040104.
166. Wang, J, Hou, T, Xu, X. Recent advances in free energy calculations with a combination of molecular mechanics and continuum models. *Curr Comput Aided Drug Des* 2006;2:287–306. https://doi.org/10.2174/157340906778226454.
167. Gohlke, H, Case, DA. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J Comput Chem* 2004;25:238–50. https://doi.org/10.1002/jcc.10379.
168. Retegan, M, Milet, A, Jamet, H. Exploring the binding of inhibitors derived from tetrabromobenzimidazole to the CK2 protein using a QM/MM-PB/SA approach. *J Chem Inf Model* 2009;49:963–71. https://doi.org/10.1021/ci8004435.

169. Ibrahim, MAA. Performance assessment of semiempirical molecular orbital methods in describing halogen bonding: quantum mechanical and quantum mechanical/molecular mechanical-molecular dynamics study. *J Chem Inf Model* 2011;51:2549–59. https://doi.org/10.1021/ci2002582.
170. Barbault, F, Maurel, F. Is inhibition process better described with MD(QM/MM) simulations? The case of urokinase type plasminogen activator inhibitors. *J Comput Chem* 2012;33:607–16. https://doi.org/10.1002/jcc.21983.
171. Kaukonen, M, Söderhjelm, P, Heimdal, J, Ryde, U. QM/MM-PBSA method to estimate free energies for reactions in proteins. *J Phys Chem B* 2008;112:12537–48. https://doi.org/10.1021/jp802648k.
172. Lu, H, Goren, AC, Zhan, CG. Characterization of the structures of phosphodiesterase 10 binding with adenosine 3′,5′-monophosphate and guanosine 3′,5′- monophosphate by hybrid quantum mechanical/molecular mechanical calculations. *J Phys Chem B* 2010;114:7022–8. https://doi.org/10.1021/jp911527y.
173. Chen, X, Zhao, X, Xiong, Y, Liu, J, Zhan, CG. Fundamental reaction pathway and free energy profile for hydrolysis of intracellular second messenger adenosine 3′,5′-cyclic monophosphate (cAMP) catalyzed by phosphodiesterase-4. *J Phys Chem B* 2011;115:12208–19. https://doi.org/10.1021/jp205509w.
174. Lu, H, Huang, X, Abdulhameed, MDM, Zhan, CG. Binding free energies for nicotine analogs inhibiting cytochrome P450 2A6 by a combined use of molecular dynamics simulations and QM/MM-PBSA calculations. *Bioorg Med Chem* 2014;22:2149–56. https://doi.org/10.1016/j.bmc.2014.02.037.
175. Söderhjelm, P, Kongsted, J, Ryde, U. Ligand affinities estimated by quantum chemical calculations. *J Chem Theor Comput* 2010;6:1726–37. https://doi.org/10.1021/ct9006986.
176. Ishikawa, T, Burri, RR, Kamatari, YO, Sakuraba, S, Matubayasi, N, Kitao, A, et al.et al. A theoretical study of the two binding modes between lysozyme and tri-NAG with an explicit solvent model based on the fragment molecular orbital method. *Phys Chem Chem Phys* 2013;15:3646–54. https://doi.org/10.1039/c3cp42761g.
177. Shigemitsu, Y. Quantum chemical study on molecular-level affinity of DJ-1 binding compounds. *Int J Quant Chem* 2013;113:574–9. https://doi.org/10.1002/qua.24132.
178. Díaz, N, Suárez, D, Merz, KM, Sordo, TL. Molecular dynamics simulations of the TEM-1 β-lactamase complexed with cephalothin. *J Med Chem* 2005;48:780–91. https://doi.org/10.1021/jm0493663.
179. Anisimov, VM, Cavasotto, CN. Quantum mechanical binding free energy calculation for phosphopeptide inhibitors of the Lck SH2 domain. *J Comput Chem* 2011;32:2254–63. https://doi.org/10.1002/jcc.21808.
180. Mikulskis, P, Genheden, S, Wichmann, K, Ryde, U. A semiempirical approach to ligand-binding affinities: dependence on the Hamiltonian and corrections. *J Comput Chem* 2012;33:1179–89. https://doi.org/10.1002/jcc.22949.
181. Fox, SJ, Dziedzic, J, Fox, T, Tautermann, CS, Skylaris, CK. Density functional theory calculations on entire proteins for free energies of binding: application to a model polar binding site. *Proteins Struct Funct Bioinforma* 2014;82:3335–46. https://doi.org/10.1002/prot.24686.
182. Mishra, SK, Koča, J. Assessing the performance of MM/PBSA, MM/GBSA, and QM-MM/GBSA approaches on protein/carbohydrate complexes: effect of implicit solvent models, QM methods, and entropic contributions. *J Phys Chem B* 2018;122:8113–21. https://doi.org/10.1021/acs.jpcb.8b03655.

183. Åqvist, J, Medina, C, Samuelsson, J-E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng Des Sel* 1994;7:385–91. https://doi.org/10.1093/protein/7.3.385.
184. Khandelwal, A, Lukacova, V, Comez, D, Kroll, DM, Raha, S, Balaz, S. A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J Med Chem* 2005;48:5437–47. https://doi.org/10.1021/jm049050v.
185. Khandelwal, A, Balaz, S. Improved estimation of ligand-macromolecule binding affinities by linear response approach using a combination of multi-mode MD simulation and QM/MM methods. *J Comput Aided Mol Des* 2007;21:131–7. https://doi.org/10.1007/s10822-007-9104-4.
186. Khandelwal, A, Balaz, S. QM/MM linear response method distinguishes ligand affinities for closely related metalloproteins. *Proteins Struct Funct Genet* 2007;69:326–39. https://doi.org/10.1002/prot.21500.
187. Natesan, S, Subramaniam, R, Bergeron, C, Balaz, S. Binding affinity prediction for ligands and receptors forming tautomers and ionization species: inhibition of mitogen-activated protein kinase-activated protein kinase 2 (MK2). *J Med Chem* 2012;55:2035–47. https://doi.org/10.1021/jm201217q.
188. Alves, CN, Martí, S, Castillo, R, Andrés, J, Moliner, V, Tuñón, I, et al.et al. A quantum mechanics/molecular mechanics study of the protein-ligand interaction for inhibitors of HIV-1 integrase. *Chem Eur J* 2007;13:7715–24. https://doi.org/10.1002/chem.200700040.
189. Xiang, M, Lin, Y, He, G, Chen, L, Yang, M, Yang, S, et al.et al. Correlation between biological activity and binding energy in systems of integrin with cyclic RGD-containing binders: a QM/MM molecular dynamics study. *J Mol Model* 2012;18:4917–27. https://doi.org/10.1007/s00894-012-1487-z.
190. Frush, EH, Sekharan, S, Keinan, S. Silico prediction of ligand binding energies in multiple therapeutic targets and diverse ligand sets – a case study on BACE1, TYK2, HSP90, and PERK proteins. *J Phys Chem B* 2017;121:8142–8. https://doi.org/10.1021/acs.jpcb.7b07224.
191. Zwanzig, RW. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J Chem Phys* 1954;22:1420–6. https://doi.org/10.1063/1.1740409.
192. Kirkwood, JG. Statistical mechanics of fluid mixtures. *J Chem Phys* 1935;3:300–13. https://doi.org/10.1063/1.1749657.
193. Bennett, CH. Efficient estimation of free energy differences from Monte Carlo data. *J Comput Phys* 1976;22:245–68. https://doi.org/10.1016/0021-9991(76)90078-4.
194. Shirts, MR, Chodera, JD. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* 2008;129:124105. https://doi.org/10.1063/1.2978177.
195. Souaille, M, Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput Phys Commun* 2001;135:40–57. https://doi.org/10.1016/s0010-4655(00)00215-0.
196. Torrie, GM, Valleau, JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J Comput Phys* 1977;23:187–99. https://doi.org/10.1016/0021-9991(77)90121-8.
197. Izrailev, S, Stepaniants, S, Balsera, M, Oono, Y, Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys J* 1997;72:1568–81. https://doi.org/10.1016/s0006-3495(97)78804-0.
198. Limongelli, V, Bonomi, M, Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proc Natl Acad Sci U S A* 2013;110:6358–63. https://doi.org/10.1073/pnas.1303186110.

199. Merz, KM. Limits of free energy computation for protein-ligand interactions. *J Chem Theor Comput* 2010;6:1769–76. https://doi.org/10.1021/ct100102q.
200. Pearlman, DA. Evaluating the molecular mechanics Poisson-Boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP Kinase. *J Med Chem* 2005;48:7796–807. https://doi.org/10.1021/jm050306m.
201. Muddana, HS, Gilson, MK. Prediction of SAMPL3 host-guest binding affinities: evaluating the accuracy of generalized force-fields. *J Comput Aided Mol Des* 2012;26:517–25. https://doi.org/10.1007/s10822-012-9544-3.
202. Jiao, D, Golubkov, PA, Darden, TA, Ren, P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc Natl Acad Sci U S A* 2008;105:6290–5. https://doi.org/10.1073/pnas.0711686105.
203. Zhang, J, Yang, W, Piquemal, JP, Ren, P. Modeling structural coordination and ligand binding in zinc proteins with a polarizable potential. *J Chem Theor Comput* 2012;8:1314–24. https://doi.org/10.1021/ct200812y.
204. Muller, RP, Warshel, A. Ab initio calculations of free energy barriers for chemical reactions in solution. *J Phys Chem* 1995;99:17516–24. https://doi.org/10.1021/j100049a009.
205. Iftimie, R, Salahub, D, Wei, D, Schofield, J. Using a classical potential as an efficient importance function for sampling from an ab initio potential. *J Chem Phys* 2000;113:4852–62. https://doi.org/10.1063/1.1289534.
206. Wood, RH, Yezdimer, EM, Sakane, S, Barriocanal, JA, Doren, DJ. Free energies of solvation with quantum mechanical interaction energies from classical mechanical simulations. *J Chem Phys* 1999;110:1329–37. https://doi.org/10.1063/1.478009.
207. Luzhkov, V, Warshel, A. Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies. *J Comput Chem* 1992;13:199–213. https://doi.org/10.1002/jcc.540130212.
208. Bandyopadhyay, P. Accelerating quantum mechanical/molecular mechanical sampling using pure molecular mechanical potential as an importance function: the case of effective fragment potential. *J Chem Phys* 2005;122:091102. https://doi.org/10.1063/1.1861890.
209. König, G, Boresch, S. Non-Boltzmann sampling and Bennett's acceptance ratio method: how to profit from bending the rules. *J Comput Chem* 2011;32:1082–90. https://doi.org/10.1002/jcc.21687.
210. König, G, Hudson, PS, Boresch, S, Woodcock, HL. Multiscale free energy simulations: an efficient method for connecting classical MD simulations to QM or QM/MM free energies using non-Boltzmann Bennett reweighting schemes. *J Chem Theor Comput* 2014;10:1406–19. https://doi.org/10.1021/ct401118k.
211. Reddy, MR, Erion, MD. Relative binding affinities of fructose-1,6-bisphosphatase inhibitors calculated using a quantum mechanics-based free energy perturbation method. *J Am Chem Soc* 2007;129:9296–7. https://doi.org/10.1021/ja072905j.
212. Rathore, RS, Reddy, RN, Kondapi, AK, Reddanna, P, Reddy, MR. Use of quantum mechanics/molecular mechanics-based FEP method for calculating relative binding affinities of FBPase inhibitors for type-2 diabetes. *Theor Chem Acc* 2012;131:1–10. https://doi.org/10.1007/s00214-012-1096-z.
213. Świderek, K, Martí, S, Moliner, V. Theoretical studies of HIV-1 reverse transcriptase inhibition. *Phys Chem Chem Phys* 2012;14:12614–24. https://doi.org/10.1039/C2CP40953D.
214. Nascimento, ÉCM, Oliva, M, Świderek, K, Martins, JBL, Andrés, J. Binding analysis of some classical acetylcholinesterase inhibitors: insights for a rational design using free energy perturbation method calculations with QM/MM MD simulations. *J Chem Inf Model* 2017;57:958–76. https://doi.org/10.1021/acs.jcim.7b00037.

215. Beierlein, FR, Michel, J, Essex, JW. A Simple QM/MM approach for capturing polarization effects in protein-ligand binding free energy calculations. *J Phys Chem B* 2011;115:4911–26. https://doi.org/10.1021/jp109054j.
216. Mikulskis, P, Cioloboc, D, Andrejić, M, Khare, S, Brorsson, J, Genheden, S, et al.et al. Free-energy perturbation and quantum mechanical study of SAMPL4 octa-acid host-guest binding energies. *J Comput Aided Mol Des* 2014;28:375–400. https://doi.org/10.1007/s10822-014-9739-x.
217. Genheden, S, Ryde, U, Söderhjelm, P. Binding affinities by alchemical perturbation using QM/MM with a large QM system and polarizable MM model. *J Comput Chem* 2015;36:2114–24. https://doi.org/10.1002/jcc.24048.
218. Olsson, MA, Söderhjelm, P, Ryde, U. Converging ligand-binding free energies obtained with free-energy perturbations at the quantum mechanical level. *J Comput Chem* 2016;37:1589–600. https://doi.org/10.1002/jcc.24375.
219. Olsson, MA, Ryde, U. Comparison of QM/MM methods to obtain ligand-binding free energies. *J Chem Theor Comput* 2017;13:2245–53. https://doi.org/10.1021/acs.jctc.6b01217.
220. Steinmann, C, Olsson, MA, Ryde, U. Relative ligand-binding free energies calculated from multiple short QM/MM MD simulations. *J Chem Theor Comput* 2018;14:3228–37. https://doi.org/10.1021/acs.jctc.8b00081.
221. Caldararu, O, Olsson, MA, Misini Ignjatović, M, Wang, M, Ryde, U. Binding free energies in the SAMPL6 octa-acid host–guest challenge calculated with MM and QM methods. *J Comput Aided Mol Des* 2018;32:1027–46. https://doi.org/10.1007/s10822-018-0158-2.

Ha Vinh Lam Nguyen and Isabelle Kleiner

2 Understanding (coupled) large amplitude motions: the interplay of microwave spectroscopy, spectral modeling, and quantum chemistry

Abstract: A large variety of molecules contain large amplitude motions (LAMs), *inter alia* internal rotation and inversion tunneling, resulting in tunneling splittings in their rotational spectrum. We will present the modern strategy to study LAMs using a combination of molecular jet Fourier transform microwave spectroscopy, spectral modeling, and quantum chemical calculations to characterize such systems by the analysis of their rotational spectra. This interplay is particularly successful in decoding complex spectra revealing LAMs and providing reference data for fundamental physics, astrochemistry, atmospheric/environmental chemistry and analytics, or fundamental researches in physical chemistry. Addressing experimental key aspects, a brief presentation on the two most popular types of state-of-the-art Fourier transform microwave spectrometer technology, i.e., pulsed supersonic jet expansion–based spectrometers employing narrow-band pulse or broad-band chirp excitation, will be given first. Secondly, the use of quantum chemistry as a supporting tool for rotational spectroscopy will be discussed with emphasis on conformational analysis. Several computer codes for fitting rotational spectra exhibiting fine structure arising from LAMs are discussed with their advantages and drawbacks. Furthermore, a number of examples will provide an overview on the wealth of information that can be drawn from the rotational spectra, leading to new insights into the molecular structure and dynamics. The focus will be on the interpretation of potential barriers and how LAMs can act as sensors within molecules to help us understand the molecular behavior in the laboratory and nature.

Keywords: conformational analysis, large amplitude motions, microwave spectroscopy, molecular jet Fourier transform, rotational spectroscopy, spectral fitting.

2.1 Introduction

Since the jet-based Fourier transform microwave (FTMW) spectroscopy technique was introduced by Balle and Flygare in 1979 [1, 2], it rapidly became popular. The complexity of molecules that could be studied strongly increased with this method. In

This article has previously been published in the journal Physical Sciences Reviews. Please cite as: Nguyen, H. V. L., Kleiner, I. Understanding (coupled) large amplitude motions: the interplay of microwave spectroscopy, spectral modeling, and quantum chemistry *Physical Sciences Reviews* [Online] 2021, 6 DOI: 10.1515/psr-2020-0037

https://doi.org/10.1515/9783110678215-002

combination with newly developed jet sources, it shows a great potential toward even larger and more dynamic systems, transient molecules, or weakly bounded complexes with applications in research fields as diverse as, e.g., molecular biology, astrophysics, or environmental sciences.

Rotational spectra have been always known to be related directly to molecular structures because the most important geometry parameters they deliver are the rotational constants, which reflect the mass distribution of atoms in the molecules. Information on bond lengths and angles can be routinely obtained. However, microwave spectroscopy is today far more advanced and offers much more than just the molecular structures. It addresses various fundamental and applied problems in molecular physics, physical chemistry, and related fields such as astrophysics, atmospheric sciences, or biochemistry. It answers challenging questions on conformational landscapes, characters of the chemical bond, electronic surrounding of a given nucleus, charge transfer, and internal dynamics.

The power of microwave spectroscopy in decoding molecular structures is very impressive compared to other spectroscopic techniques because pure rotational transitions can be observed directly within the lowest vibrational state. Furthermore, in combination with supersonic expansion where the rotational temperature is decreased almost to the absolute zero point (from 0.5 to 2 K), only the lowest rotational levels are populated. This leads to a tremendous simplification of the spectra. Although important chemical and spectroscopic information could still be obtained with static gas cells, many problems could only be solved using supersonic jets. At the same time, with the capability to observe spectra of even larger molecules, microwave spectroscopy is emphasizing and consolidating the key role in yielding accurate information on various physical and chemical objectives. Atmospheric compounds, interstellar species, and biomolecules are targeted in many laboratory studies, which provide information on properties of molecules, clusters, and radicals, as well as offer applications directed toward atmospheric sensing or radio astronomy.

The two classical books known by almost all microwave spectroscopists are those written by Townes and Schawlow [3] and Gordy and Cook [4]. The fundamental theoretical background of rotational spectroscopy is referred to those books and will not be repeated here.

The effects of intramolecular dynamics cause all rotational energy levels to split into several levels due to tunneling effects. Two prototypes of such large amplitude motions (LAMs) are methyl internal rotation and inversion tunneling involving a double minimum potential. These LAMs sometimes complicate the spectra that much that their assignments were inhibited. To model the experimental spectra, theoretical tools including correct Hamiltonian need to be developed. The theory on LAMs occurring in rotational spectra is described in detail by Wollrab [5] and Lister et al. [6] or in reviews, e.g., by Lin and Swalen [7]. Their group theoretical treatment using the permutation inversion group can be found in a book written by Bunker and Jensen [8].

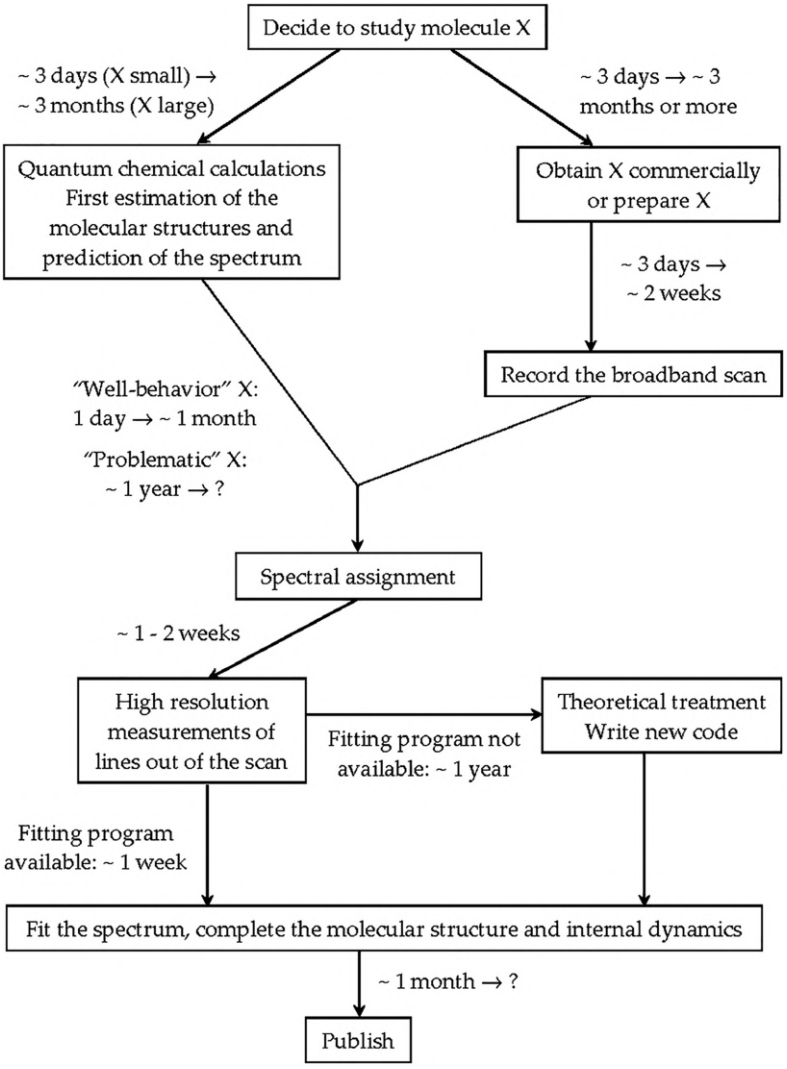
Several computer programs are available to analyze the rotational spectra. They are often developed by microwave spectroscopists during the investigations on their particular molecular systems and then generalized for other molecules with similar molecular symmetry and internal dynamics. Nevertheless, in a number of cases where no program is available to treat the LAM problems occurring in the molecules of interest, specialized codes are needed. They are either extended from an already existing program or written newly. In those cases, group theoretical treatment is required in addition. The rotational spectra of many molecules can be fitted by the program *SPFIT/SPCAT* written by Pickett [9], often used to produce line frequencies and intensities for molecules in the astrophysical databases such as the Jet Propulsion Laboratory (JPL) [10] or "the Cologne Database for Molecular Spectroscopy" (CDMS) [11]. Many nonrigid molecules are also available at other astrophysical line lists and websites such as Splatalogue [12], the Observed Interstellar Molecular Microwave Transitions NIST database [13], and the Toyama Microwave Atlas [14].

Many programs for rotational spectroscopy dealing with LAMs are available at the "Programs for ROtational SPEctroscopy" (PROSPE) website managed by Kisiel [15]. Among them, the program *XIAM* written by Hartwig and Dreizler [16] can treat the internal rotation effects of up to three methyl internal rotors. Other codes exist which contain higher order perturbation terms such as the one written by Ohashi and Hougen [17] for methylamine-like molecules, by Hougen et al. [18] and Hougen and Kleiner [19] to deal with a number of one-top and two-top molecules [20] with C_s and C_1 symmetry, by Groner [21] to deal with one or two internal tops, or by Ilyushin et al. [22] to treat one-top molecules with C_s and two-top molecules with C_{2v} symmetry [23]. Kleiner and Ilyushin use a global treatment including all the torsional levels for a given vibrational state, an approach which is particularly successful for low barriers or excited torsional states [18–20, 22, 23]. Internal rotations with intermediate to high torsional barriers can be handled by the *IAMCALC* program intergrating in Pickett's *SPFIT/SPCAT* suite [10] and by the spectral fitting program *JB95* provided by Plusquellic [24]. A few models for complicated interactions arising in individual molecules have been also developed based on group theoretical approaches, e.g., studies on the rotatory wagging coupled motions in 2-methylmalonaldehyde, a methylamine-like molecule, by Kleiner and Hougen [25] using the hybrid model.

Density functional theory (DFT) and *ab initio* calculations implemented in computational programs such as *Gaussian* [26] or *GAMESS* [27] are commonly used today in the spectroscopic community to predict i.e., the potential energy surfaces (PES) to determine the conformational preferences, the molecular equilibrium structures, and electric field gradients. Such calculations yield starting values of spectroscopic parameters including barrier heights for different types of LAMs to support the experimental work.

In the present review, the theoretical or experimental background required to analyze the rotational spectrum of a molecule containing one of more LAMs will not be given in detail, as they were presented in a number of books, papers, and review

papers [3, 8]. Instead, this review will summarize the modern strategy to analyze microwave spectra, how to achieve fits with standard deviations within experimental accuracy, and serve as a guideline to study LAM systems, as outlined in Scheme 2.1. Furthermore, it will give an overview on how to use the information coming out from the spectrum to bring new insights on the molecular gas phase structures and dynamics. The focus will be on the interpretation of potential barriers and the use of methyl internal rotors as sensors for the molecular structures.



Scheme 2.1: The general procedure for microwave spectroscopic investigations on a molecular system.

2.2 Spectrometer technology

With the unique ability to provide simultaneously high sensitivity and high resolution, FTMW spectroscopy based on pulsed supersonic jet expansions has become popular in the last decades. This technique is a powerful tool for studying isolated molecules in the gas phase. The most attractive advantages of integrating supersonic jet expansion in the FTMW spectrometers is the significant simplification of observed spectra. Furthermore, the signal to noise ratio is increased due to statistical depopulation of the excited states in the molecules because of the effective cooling of the sample to extremely low rotational temperature (0.5–2 K). In this review, the two most popular molecular jet FTMW spectrometer techniques will be briefly presented. More details can be found in the study by Grabow [28].

2.2.1 Resonator-based molecular jet FTMW (2 – 40 GHz) spectrometers: *high resolution and sensitivity but time consuming for survey spectra*

In a molecular sample, the electric dipole interaction occurs while exposing to a standing wave field of the microwave radiation which propagates in a "transverse electric magnetic"-mode of a Fabry-Pérot-type resonator formed by two spherical mirrors, typically made of aluminum, with equal curvature at a distance d. Today, resonator-based molecular jet FTMW spectrometers in the frequency range below 26 GHz are commonly used in many microwave laboratories and have shown their superior sensitivity and resolving power [29, 40]. For medium-sized or large biomolecules or volatile compounds important for atmospheric chemistry and biology (≥15 atoms), this frequency range is suitable because of the small rotational constants. Therefore, lower J and K rotational transitions, which can be assigned more easily, fall in this range. Spectrometers from 26 to 40 GHz are fairly rare but extremely useful for studying small molecules (<15 atoms) abundant in the Earth's atmosphere or in astrophysical objects such as molecular clouds or planetary atmospheres [41].

A diluted gas mixture of about 1% sample seeded in a rare gas at a total pressure in the range of 50–200 kPa is expanded into the vacuum chamber which contains the resonators. The simplest form of the nozzle consists of a circular orifice with a diameter of 0.5–2.0 mm. The exit channel of the nozzle of 2 mm length is widened conically to 4 mm. The signals appear as doublets due to the Doppler effect in the case of a coaxial arrangement between the resonator and the molecular beam (Coaxially Oriented Beam Resonator Arrangement [COBRA] type). The typical measurement accuracy is about 2 kHz. Figure 2.1 depicts a typical broadband scan and a spectrum recorded at high resolution using a COBRA-type molecular jet FTMW spectrometer [33].

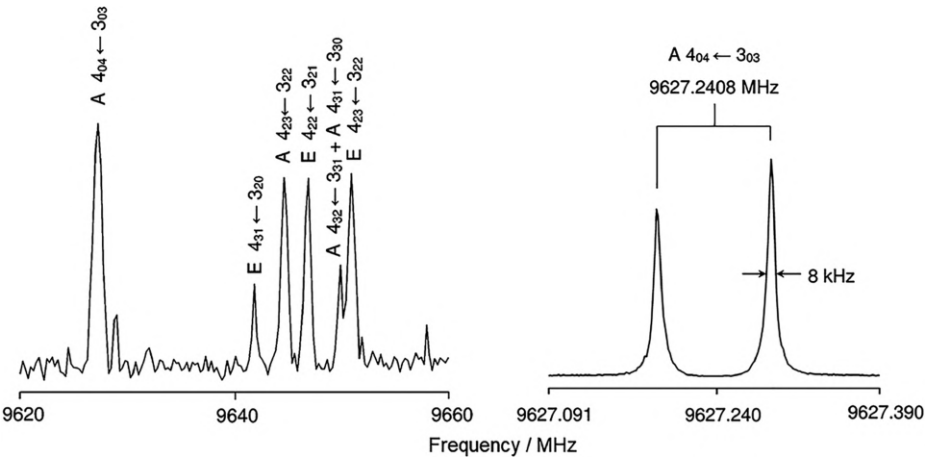


Figure 2.1: A typical broadband scan (left hand side) and a spectrum recorded at high resolution (right hand side) of allyl acetate using a Coaxially Oriented Beam Resonator Arrangement (COBRA)-type molecular jet FTMW spectrometer [33]. The broadband scan was recorded by overlapping spectra at a step size of 0.25 MHz, where only the line positions are present. The linewidth in the high resolution spectrum of the $4_{04} \leftarrow 3_{03}$ rotational transition, A torsional species, is 8 kHz. The line frequency can be determined with a measurement accuracy of about 0.8 kHz [42].

2.2.2 Chirped-pulse molecular jet FTMW spectrometers: *how to reduce the time requirements for survey spectra?*

The classical versions of resonator-based molecular jet FTMW spectrometers have an unrivaled resolution but suffer from the time requirement to acquire survey spectra (scans) because the resonator has to be tuned mechanically for every frequency element at rather narrow steps of less than 0.25 MHz ($<10^{-5}$ cm^{-1}), see, for example, the figure caption of Figure 2.1 [42]. An approach to overcome this problem was developed a decade ago using the chirped-pulse (CP) method, which relies on a very short but powerful frequency ramp signal with a band width of 1 GHz (3333 times 10^{-5} cm^{-1}) or even more, and thus can reduce the time requirements for scans dramatically [43].

The fast passage excitation arises when the frequency of an electromagnetic field is swept through a molecular resonance in a time which is much shorter than the relaxation time. Despite this very short resonance time, the population of the states changes remarkably, which leads to a detectable oscillating macroscopic polarization. Although traditional narrow-band resonator-based spectrometers can provide better resolution of the spectra of a target molecule, CP-FTMW spectrometers provide unparalleled speed for highly complex broadband spectra arising from mixtures of reaction products, isomers, or weakly bound molecular clusters. In many studies where optimizations of the molecular jet are required, e.g., van der Waals complexes or electric discharges to produce ions or radicals, the use of a CP machine is indispensable. It is

necessary to monitor a wide spectral range in real time in order to recognize changes in the spectrum while the strength of discharge and other parameters are varied. Furthermore, the line intensities of CP spectra are more reliable than those of a resonator-based apparatus. This is an important feature for assignments of the spectra and for conformational analysis.

With all of these advantages, CP-FTMW is the spectrometer of choice for many scientists worldwide and has been equipped in almost all microwave spectroscopic labs [44–59].

2.3 Quantum chemical calculations

The combination of microwave spectroscopy and quantum chemical calculations is a powerful tool to determine the conformational landscape of medium-sized and large molecules, as well as LAM parameters. Studies where the microwave heavy atom structures were determined experimentally, such as those on methyltetrahydrofuran [60], 2-ethylfuran [61], and 2- and 3-nitrobenzonitrile [62], have proven that the molecular geometries can be calculated sufficiently well. In the present review, we constraint the scope of quantum chemical calculations on geometry optimizations. They are crucial and serve as a supporting tool for the assignment of microwave spectra by delivering predicted values to start the assignment [63]. If the molecules are large, it is often not possible to obtain r_s geometries from the experiments, as not all isotopologues can be observed, but geometry information can be indirectly obtained from the rotational constants, which can be compared with those of the theoretical results. Like Townes and Schawlow [3] and Gordy and Cook [4] for microwave spectroscopists, details on quantum chemistry are referred to the two books written by Cramer [64] and Jensen [65]. The present review will only address the use of quantum chemistry as a supporting tool for rotational spectroscopy.

The relation between the positions of all nuclei in a molecule and the corresponding potential of the molecule can be described by the PES. Each energy minimum on the PES represents a stable conformer because any change in geometry parameters such as bond angles or lengths would lead to a conformation with a higher energy. The absolute minimum is the energetically lowest, i.e., the most favorable, conformation of the molecule. Transition states are also present on the maxima of the PES as saddle points. Figure 2.2 illustrates the PES of phenetole depending on the rotations of the phenyl ring and the ethyl group as an example [66]. This PES describes completely the conformational landscape of phenetole.

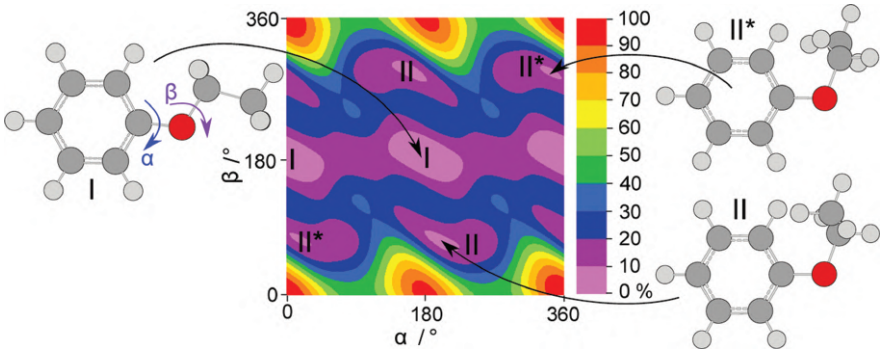


Figure 2.2: The potential energy surface (PES) of phenetole calculated at the MP2/6-311++G(d,p) level of theory obtained by rotating the phenyl ring (dihedral angle α) and the ethyl group (dihedral angle β) [66]. The numbers in the color code indicate the energy (in per cent) relative to the energetic minimum E_{min} = −385.0444525 Hartree (0%) and the energetic maximum E_{max} = −385.02509670 Hartree (100%). Two energy minima (I and II) are found on this PES. II* is the enantiomer of II.

2.3.1 Geometry optimizations: *how to start?*

To calculate the stable conformers of a molecule, geometry optimizations are carried out on different starting geometries computed in quantum chemical programs like, for example, the commercially available *Gaussian* [26] or the freeware *GAMESS* [27] where the total energy is minimized by varying the atom positions in small steps until convergence.

As stated in Scheme 2.1, theoretical investigations can be started immediately after deciding to study a system of interest (called molecule X in Scheme 2.1), even before the experimental spectrum is recorded. However, several points should be carefully thought before running the first calculation. The first point is to decide which level of theory is most suitable to treat molecule X and how accurate the results will be. The choice of course depends on the available experimental data.

2.3.2 Method choice: *be careful. Discrepancy!*

Different methods are available to optimize the geometry of a molecule and find the stable conformations. The two methods most commonly used with a reasonable ratio between calculation time and efficiency ratio are probably the DFT method using the B3LYP functional [67], [68] and the MP2 method [69]. They mainly differ in their optimized quantity. While it is the electron density for DFT, the wave function is optimized by MP2. If the molecule belongs to a systematic investigation on a molecular class, a certain level of theory is often preferred, which has yielded more

reliable results in previous studies [70]. Otherwise, MP2 is usually recommended and represents one of the most favorable methods in the microwave spectroscopic community because of the reasonable rate between accuracy and required computational time. However, an advantage of the semiempirical B3LYP method over the pure *ab initio* MP2 method is that it is much faster and more cost efficient. However, a disadvantage of the B3LYP functional is its underestimation of intramolecular interactions. As a consequence, energy minima may be overlooked during the conformational analysis [71]. This can be improved by using Grimme's dispersion correction [72, 73], which turned out to be quite helpful for large molecules containing internal rotation. The case study on methyl jasmonate and zingerone, two substances produced by plants, has shown that a wrong choice of the theoretical method can lead to misinterpretation of the experimental results (see Figure 2.3) [74].

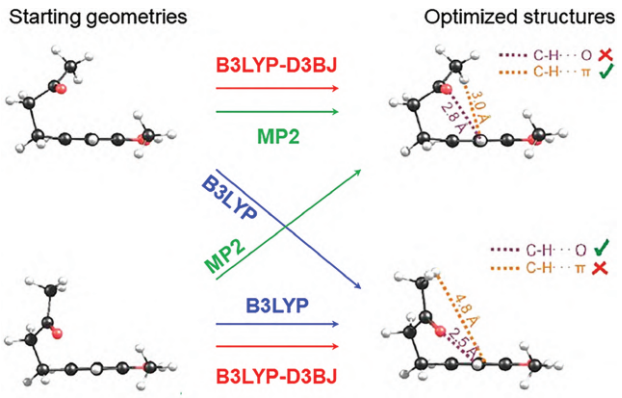


Figure 2.3: Schematic representation of the prototype behavior of theoretical methods in the conformational casuistry of zingerone. For details see the study by Uriarte et al. [74].

The discrepancy between B3LYP and MP2 found in methyl jasmonate and zingerone is not unusual and can also be observed in smaller molecules. An example is the case of diethyl ketone, where MP2 predicts a tilt angle of about 10° for each of the ethyl group and the corresponding symmetry is C_2, while B3LYP calculates a C_{2v} symmetry with all heavy atoms located on the symmetry plane (see Figure 2.4) [75]. There are also more accurate methods, for example, higher orders of perturbation theory (MP4) [76] or the golden standard in quantum chemistry, coupled cluster calculations [77], but they are much more expensive than the B3LYP and MP2 methods. Therefore, a conformational analysis at a level higher than B3LYP or MP2 is not recommended. However, it might be interesting to reoptimize the geometries of the lowest energy conformers obtained from the PES at a higher level to predict more accurate structures.

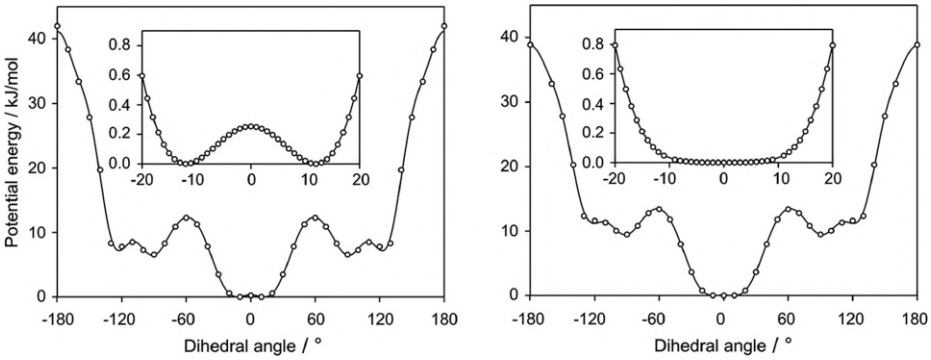


Figure 2.4: Left hand side: a potential energy curve of diethyl ketone calculated at the MP2/6-311++G(d,p) level of theory. Right hand side: the same potential curve calculated at the B3LYP/6-311++G(d,p) level of theory. For details see the study by Nguyen and Stahl [75].

2.3.3 Basis set choice: *make a lot of tests...*

The second point is to choose a suitable basis set. The Pople valence triple-zeta basis set 6-311++G(d,p) is often used in combination with the MP2 and B3LYP methods because the valence shell electrons are described well and the diffuse functions (++) and the polarization function (d,p) are included [78]. A set of polarizing d-functions is used on all atoms heavier than helium, and a set of p-functions on all hydrogen and helium atoms is included. Dunning correlation consistent basis sets, e.g., aug-cc-pVDZ and aug-cc-pVTZ are also often chosen [79]. It is obvious that the same level of theory should be used throughout the conformational analysis for a reasonable comparison.

For each energy minimum, harmonic frequency calculations have to be performed to verify whether an optimized geometry on the PES is an energy minimum or a saddle point (transition state). Zero-point corrections on the equilibrium energy can also be obtained which yield the energy of the system in the vibrational ground state.

The accuracy of calculations with different methods or basis sets depends on the system of interest X. Therefore, the method/basis set combination should be chosen carefully before running the calculations. A level often used in the spectroscopic community is MP2/6-311++G(d,p), which has provided good starting values in investigations of many small- to medium-sized molecules [80–82]. The geometries and consequently the predicted rotational constants usually agree well with the experimental values. However, the B3LYP/6-311++G(d,p) level of theory turns out to be more reliable in some cases such as the three isomers of mono-methylanisoles [83–85]. For some esters like ethyl valerate [86] or molecules containing aromatic rings such as coumarin [87], quinolone [88], and isoquinoline [88], MP2/6-31G(d,p) turns out to be a "magic level" which calculates rotational constants accidentally closest to the experimental values and has eased significantly the assignment process.

It is therefore recommended for any new class of molecules to first test which combination of methods and basis set is appropriate and then apply it for all following investigations in the series.

Finally, it should be mentioned that the predicted rotational constants refer to the equilibrium structure, and the experimental rotational constants obtained from the microwave spectrum are those of the vibrational ground state. Since the deviations are often only about 1%, the theoretical B_e values are usually compared directly with the experimental B_0 results [89]. Though this comparison is not physically meaningful, theoretical B_e values offer cost-efficient calculations with sufficient accuracy for spectral assignment purposes. Theoretical B_0 rotational constants at the vibrational ground state can be provided by anharmonic frequency calculations, which are sometimes important, for example, to determine the semiexperimental equilibrium r_e^{SE} structure [90]. However, for the assignment of microwave spectra, these calculations are firstly very time consuming and secondly often do not predict rotational constants in much better agreement to the experimental values due to error compensation.

2.3.4 Estimation of the torsional barriers: *still challenging*

Barriers hindering the LAMs can be obtained from potential energy scans where the LAM coordinates are varied in a certain grid while all other geometry parameters are allowed to relax. Though geometry optimizations yield reliable results in many cases, calculating energies still remains a challenge for today's quantum chemistry with even growing computational capacity. Especially, if the energy differences are small, it is difficult to predict the conformers of lowest energy or the barriers of LAMs with a sufficient accuracy to guide the spectral assignment [91]. Although for quantum chemistry, an accuracy of 1 kJ·mol^{-1} (approximately 84 cm^{-1}) states energy calculations with high quality, it is far from accurate regarding the experimental requirements. By a change of only one cm^{-1} in the potential barrier, the predicted frequencies can differ hundreds of MHz from the experimental values. However, the order of magnitude is often given correctly and serves as a good orientation for studying the effects of LAMs in the spectrum.

2.4 A small historical perspective on large amplitude motions

One of the first recorded rotational spectra were those of ammonia with its inversion tunneling motion of the nitrogen atom through the plane spanned by the three hydrogen atoms [92]. Ammonia thus exists in two energetically equivalent geometries, which are connected by a tunneling path with the shape of a double minimum potential [4].

The spectra of many molecular systems are dominated by the tunneling splittings of their LAMs. In some cases, two or more LAMs simultaneously generate tunneling splittings, sometimes with Coriolis couplings between them. There are several types of LAMs, but the two most frequent are the internal rotation, typically of methyl groups, and inversion tunneling. The present review does not consider the puckering LAMs of saturated rings and pseudorotation, as occurring, for example, in tetrahydrofuran [93]. Furthermore, the investigations are limited to isolated molecules and do not consider weakly bound complexes, where complex LAMs also often occur, such as in the methyl glycidate-water complex [94].

2.4.1 Internal rotation

Internal rotation is a LAM where an internal rotor rotates with respect to the rest of the molecule, called the frame. The internal rotor can be symmetric, e.g., a methyl group, or asymmetric, e.g., OH, SH, NH_2 groups. Depending on the symmetry of the rotor and the frame, the number of equivalent minima of the torsional potential can be different.

2.4.1.1 Symmetric internal rotor

A threefold potential is present if the internal rotor is a group with C_3 symmetry like a methyl group attached to an asymmetric frame, as described in Equation (2.1) [4]:

$$V(\alpha) = \frac{1}{2}V_3(1-\cos\ 3\alpha) + \frac{1}{2}V_6(1-\cos\ 6\alpha) + \frac{1}{2}V_9(1-\cos\ 9\alpha)... \qquad (2.1)$$

The height of this potential, i.e., the V_3 value, can be quite different, from essentially free rotation (almost 0 cm^{-1}) to over 1000 cm^{-1}.

The first internal rotors studied in the microwave domain were often molecules detected in the astrophysical medium. The history of internal rotation is thus closely related to the history of astrophysical detections. *Methanol*, a very important molecule in chemistry and industry, is one of the simplest molecules undergoing hindered internal rotation with a barrier height of about 380 cm^{-1}, and was therefore studied since a long time by numerous experimental and theoretical investigations, for example, by Koehler and Dennison [95] already in 1940. The structure of methanol has been determined in 1951 by Hughes et al. [96]; the millimeter wave spectrum was investigated by Lees and Baker [97]. The studies were extended by De Lucia et al. [98] in 1989 and then continued by Xu et al. [99]. Xu, Lees and their collaborators contributed extensively to the knowledge of both microwave and infrared spectrum of methanol and its isotopologues.

The barrier to internal rotation of 398(14) cm^{-1} of another astrophysical molecule, *acetaldehyde*, was estimated by Lin and Kilb [100] in 1956. The spectral analysis was completed later by Bauder and Günthard [101], Liang et al. [102], and Maes

et al. [103]. *Methyl formate*, a further important astrophysical molecule, was studied in the microwave region in 1959 by Curl [104], who reported a barrier to methyl internal rotation of 416(14) cm^{-1}. Later on, Plummer et al. [105], Demaison et al. [106], Oesterling et al. [107], and Oka et al. [108] improved the spectral analysis to great accuracy. *Acetic acid*, a structural isomer of methyl formate, also shows internal rotation [109], [110] where the methyl torsional barrier of 174 cm^{-1} was determined in 1957 by Tabor [111] and later improved to be 168.16(17) cm^{-1} by Krischer and Saegebarth [112]. A review of molecules with internal rotation is available in the study by Kleiner [113].

2.4.1.2 Asymmetric internal rotor

An example among many asymmetric internal rotors is the primary amino group $-NH_2$. The typical example is ethyl amine for which Fischer and Botskor [114] reported on the spectrum of the *trans* conformer in 1982 and two years later also on the *gauche* conformer [115]. Another interesting example is water, which can also acts as an asymmetric internal rotor, causing a V_2 energy potential, as observed in a number of complexes like nitric acid-water [116] or water-carbon oxide [117], but we will not consider it further in this review.

2.4.1.3 Why is internal rotation important?

The smaller the barrier to internal rotation, the larger the splittings in the spectrum [4] and the more complicated the spectrum analysis. Why should we care about the laboratory analysis of those spectra? One reason is to understand the molecular structure itself which can be used to study chemical and biochemical intrinsical properties. With increasing sensitivity in atmospheric detection, internal rotors have been frequently observed and play an even more important role in chemistry of the Earth atmosphere though they are present only in traces as volatile organic compounds. Another reason is that knowledge about internal rotation is essential for the detection of molecules in the interstellar medium. More than 200 molecules have been found in the circumstellar shells or interstellar medium. A number of them show methyl internal rotation with observable splittings. The classical example is methanol, one of the simplest molecules showing methyl internal rotation with a barrier height of about 380 cm^{-1}, which was detected in Orion-A by Lovas et al. [118] in 1976. About at the same time, Churchwell and Winnewisser [119] found methyl formate in Sgr B2 with the A-E doublet of the $1_{10} \leftarrow 1_{11}$ transition. Acetaldehyde was also found in Sgr B2 [120] and then in two other clouds, TMC-1 and L134N [121]. Acetic acid was first detected in the interstellar medium by Mehringer et al. [122]. Many internal rotor molecules were observed in Orion-KL such as dimethyl ether [123]. Larger molecules like acetone [124] and ethyl methyl ether [125] have been detected as well. Most of the identifications were enabled by the interplay between laboratory studies of rotational spectra and observations of interstellar surveys in the microwave, submillimeter

wave, and millimeter wave frequency ranges. The laboratory investigations yield reliable data such as line lists and line intensities needed for the interpretation of interstellar surveys.

Only a few years ago, 215 unblended lines as well as 163 lines moderately blended with other species of *methyl acetate* with its five torsional splittings arising from two inequivalent methyl rotors were detected in the Orion cloud [126]. The detection is fully secured taking into account (i) the large number of detected lines and (ii) the fact that the systematic pattern of the lines arising from the different internal rotation states is always present (Figure 2.5). This success was due to the laboratory work and knowledge on the molecular transitions occurring in the microwave spectra of a two-top internal rotor.

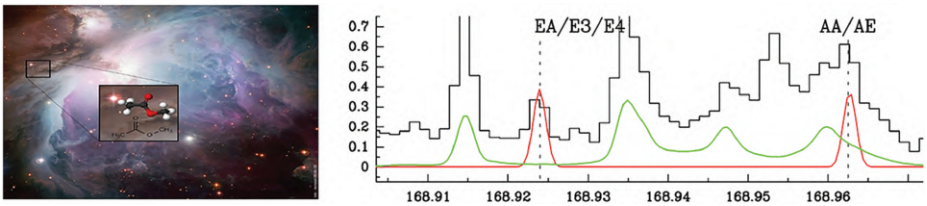


Figure 2.5: Selected lines (in red) of methyl acetate, CH_3COOCH_3, at 2 mm toward Orion-IRc2. Lines from five different torsional states are identified. The continuous green line corresponds to all lines already modeled from the study by Tercero et al. [126].

2.4.1.4 Understanding two-top molecules toward astrophysical detections

All rotational lines of a one-top molecule split into an A and an E torsional component. As illustrated in Figure 2.6, in the case of *two nonequivalent* internal rotors, all A components split into the so-called AA-AE doublets, and all E components into the EA-EE-EE* triplets, according to the notation by Dreizler based on the direct product $C_{3v}^{(1)} \otimes C_{3v}^{(2)}$ [127]. In the spectra of molecules with *two equivalent rotors*, AA-AE-EE-EE* quartets arise instead of quintets due to the degeneracy of the AE and EA torsional components (see Figure 2.6). Filled circles in Figure 2.6 indicate the nonrotating state of the rotors and round arrows symbolize the two rotating states. If the molecular symmetry of the rest of the molecule excluding the methyl groups (so-called the "frame") is C_s, new labeling schemes using the semidirect products have been introduced for both, the two inequivalent and two equivalent rotor cases. The torsional species are denoted as (00), (01), (10), (11), and (12) [128], where (01) = (10) in equivalent rotor cases [129]. The molecular symmetry groups are G_{36} and G_{18}, respectively. Laboratory studies on the smallest acetate, methyl acetate, first by Sheridan et al. [130], then by Tudorie et al. [20] and Nguyen et al. [131], and the understanding of molecular energy levels have eventually provided reliable line lists and line intensities for the successful search and unambiguous detection of this molecule in the Orion cloud, as shown in Figure 2.5 [126].

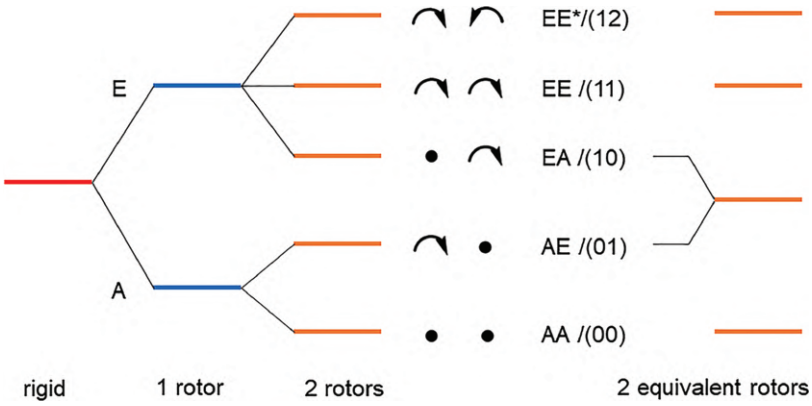


Figure 2.6: Schematic splittings due to internal rotation in the rotational spectrum of molecules with one rotor, two nonequivalent rotors, and two equivalent rotors.

There are also some other investigations on two inequivalent top C_s molecules such as 2-acetyl-5-methyl-furan [132], ethyl methyl ketone [133], methyl propionate [134], dimethylbenzaldehydes [135], dimethylanisoles [128, 136, 137], and 4-methylacetophenone [138], but the number is still scarce. To our knowledge, so far only one molecule with two inequivalent tops and C_1 symmetry for the frame, isopropenyl acetate, was reported in the literature [139]. The spectra show splittings arising from a low barrier internal rotation of the acetyl methyl group (135.3498(38) cm^{-1}) and a high barrier internal rotation of the isopropenyl methyl group (711.7(73) cm^{-1}).

For two equivalent methyl internal rotors, only a handful of molecules like acetone [140], dimethyl ether [141], dimethyl diselenide [142], diethyl ketone [75], 2,5-dimethylfuran [143], 2,5-dimethylthiophene [129], and dimethyl sulfide [144] were reported in the literature.

2.4.1.5 Beyond two internal rotors

Considerably fewer molecules with more than two methyl internal rotors such as trimethyl silyl iodide $(\mathbf{CH_3})_3$SiI [145] and similar systems [146] and *N*,*N*-dimethylacetamide [147] have been studied. The number of methyl tops is three in all cases. In *N*,*N*-dimethylacetamide, interactions between the methyl internal rotations and the overall rotation were characterized by Coriolis-like coupling parameters, which allowed the determination of the torsional barrier heights to be 677 cm^{-1} for the C-methyl top, 237 cm^{-1} for the *trans*-*N*-methyl top, and 183 cm^{-1} for the *cis*-*N*-methyl top [147].

2.4.2 Inversion tunneling

An well-known inversion process is the case of ammonia, NH_3, with a double minimum potential tunneling pathway [92]. Nevertheless, this effect is not only found there but also in larger and more complex molecules.

In an investigation on *ethylene diamine*, $NH_2–CH_2–CH_2–NH_2$, Marstokk and Møllendal [148] reported two conformers, in both of which tunneling LAMs occurred when the amino groups interchanged their donor and acceptor roles. Separately fitting the (+) and (–) energy levels yielded quite different inversion splittings of 1.564(66) MHz and 86.356(21) MHz for the two conformers. After theoretical improvements, a global fit could be achieved by Merke and Coudert [149]. *Hydrazine*, $NH_2–NH_2$, is another molecule possessing two amino groups, where three LAMs simultaneously occur, namely an internal rotation about the N–N bond and the inversion motions asiring from both amino groups. The rotational spectrum of hydranzine was first assigned by Kasuja [150] and then improved by Tsunekawa et al. [151] as well as by Kreglewski et al. [152] in the submillimeter wave range and in the infrared range [153], [154]. Inversion tunneling was also observed in almost all primary amines such as aniline $C_6H_5–NH_2$ [155].

Tunneling motion is also linked to many complexes like dimers as the weak van der Waal bonds are usually quite "floppy". In a theoretical study, Ohashi and Hougen [156] suggested 25 tunneling motions which might occur in methanol dimers in addition to the internal rotations of the two methyl groups. The group theoretical results had been confirmed by the experiments where Lovas et al. [157] revealed in the microwave spectrum 15 different tunneling states of the *a*-type *R*-branch $K_a = 0$ rotational transitions. While a considerable number of studies on molecules with internal rotations are available, investigations on inversion tunneling are scarce. Furthermore, this effect is often combined with internal rotation(s). An example is the inversion of the two protons in the amino group found in primary amines, which is accompanied by the internal rotation of the entire amino group. Only in a few molecules like planar secondary amines such as dimethyl amine, $CH_3–\mathbf{NH}–CH_3$ [158], ethyl methyl amine, $CH_3–CH_2–\mathbf{NH}–CH_3$ [159], and diethyl amine, $CH_3–CH_2–\mathbf{NH}–CH_2–CH_3$ [160], the inversion tunneling of the proton is not coupled with internal rotations (though small splittings from methyl torsions can be resolved). In phenyl formate, the tunneling of the phenyl ring also represents the only LAM of this molecule [161].

2.4.3 Interaction of internal rotation(s) with tunneling motion(s): *from rotation wagging to hydrogen transfer*

One of the classical example is the coupling between the methyl internal rotation and the inversion motion of the amino group in *methyl amine*, where the back-and-forth motion of the amino group triggers a 60° corrective rotation of the methyl

group, thereby resulting in six energetically equivalent molecular frameworks. The coupling of inversion, vibration, and torsion states of methyl amine has been studied by Gulaczyk and Kreglewski [162] and Gulaczyk et al. [163] in the infrared spectral range and more recently the excited torsional states in the far-infrared range [164].

Another interesting example where a methyl internal rotation interacts with another LAM, the proton tunneling, is *2-methylmalonaldehyde*. Similar to the case of malonaldehyde, the hydrogen transfer in 2-methylmalonaldehyde is accompanied by tautomeric rearrangements of single and double bonds [165], but the additional consequence is a rotation of the methyl group by 60°. Chou and Hougen [166] developed a tunneling-rotation Hamiltonian based on the G^{12}_{m} molecular symmetry group to perform global fits of several isotopologues with root-mean-square deviations close to measurement accuracy [167].

The microwave spectrum of *pinacolone* also caused surprises with splittings arising from the acetyl methyl group in A-E doublets coupled with the oscillation LAM of the *tert*-butyl group with splittings into the $v_t = 0$ and 1 states [168]. The same problems occur in the spectrum of acetanilide [169], where the internal rotation of the acetyl methyl group interacts with the tunneling motion of the phenyl ring, which is tilted out of the acetyl plane and a double minimum potential is present. Probably, there are more cases where such coupled LAMs are observed, but the number of publications on those systems is still very limited. According to Scheme 2.1 in the Introduction section, such "problematic" molecules require extensive measurements and spectral analysis, as well as very often new program code to fit their spectra. All of these usually inhibit the way to publication. On the other hand, they are ideal systems for the development of new group theory, as well as for testing the Hamiltonian and theoretical models.

2.5 Spectral modeling

2.5.1 Global fits of rotational spectra with LAMs: *the way to achieve standard deviations within experimental accuracy*

Treating the microwave spectrum of a rigid rotor is a classical approach where a rigid rotor Hamiltonian H_r supplemented by centrifugal distortion terms H_{CD} is sufficient for a fit of high quality. Many programs are available to treat rigid rotor spectra, considering even higher J and K transitions, as well as several vibrational excited states. The programs most frequently used in the high resolution spectroscopic community for this type of molecules are probably *SPFIT/SPCAT* [9], *XIAM* in its rigid rotor mode [16], *JB95* [24], *AUTOFIT* [170], etc. A number of softwares for rigid or nonrigid molecules are available with examples of input/output files at the PROSPE website [15].

The microwave spectrum of a molecule undergoing methyl internal rotation features splittings of each rotational transition into torsional components and can

no longer be treated using a rigid rotor model. In comparison to the large number of rigid rotor programs, only a few programs have been developed to deal with the effects of internal rotations.

2.5.1.1 The *SPFIT/SPCAT (IAMCALC)* package

A fitting and prediction program very convenient and familiar to many spectroscopists is the *SPFIT/SPCAT* package of Pickett with addition of the front-end program *IAMCALC* [10], as described for the case of propane, a two-top molecule with high torsional barriers [171]. The Mathieu function was used to generate a set of parameters connecting torsional sublevels (v_t = 0, 1, 2, A, E). The theoretical model takes into account interactions between the lowest torsional states. This package is also often used in the astrophysical communities, as well as for the JPL [10] and CDMS databases [11].

2.5.1.2 The *XIAM* code

Another program popularly used to model rotational spectra with spittings arising from up to three symmetric internal rotors, typically methyl groups, is *XIAM*, written by Hartwig and Dreizler [16]. Geometry parameters like the rotational and centrifugal distortion constants as well as internal rotation parameters such as the V_3 and higher potential terms, the angles between the internal rotor axes with the principal axes, the moment of inertia of the internal rotors, some top-top potential, and kinetic coupling terms can be fitted. In addition, *XIAM* uses a first-order approximation to treat the hyperfine structures arising from nuclear quadrupole interactions of one nucleus. The fit quality is often quite satisfactory in the cases of nuclei with small quadrupole moments such as ^{14}N.

The rigid frame-rigid top Hamiltonian for an asymmetric molecule with a methyl top in the principal axis system can be written as follows:

$$H = AP_a^2 + BP_b^2 + CP_c^2 + F\left(p_\alpha - \rho_a P_a - \rho_b P_b - \rho_c P_c\right)^2 + V(\alpha) \tag{2.2}$$

where P_g ($g = a, b, c$) are the components of the total rotational angular momentum and p_α is the internal rotation angular momentum associating with the torsion angle α. The relations between the rotational constants A, B, C of the molecule, the rotational constant F of the internal rotor, and the ρ_g components of the ρ vector to the principal moments of inertia I_g of the molecule and to the moment of inertia of the top I_α can be expressed as follows:

$$A = \frac{\hbar^2}{2I_a}, B = \frac{\hbar^2}{2I_b}, C = \frac{\hbar^2}{2I_c}, F = \frac{\hbar^2}{2rI_\alpha}, \rho_g = \frac{\lambda_g I_\alpha}{I_g}, r = 1 - \sum_g \lambda_g^2 \frac{I_\alpha}{I_g} \tag{2.3}$$

where $\lambda_g = \cos < (i, g)$ are the direction cosines of the internal rotation axis i of the top.

The program *XIAM* uses a combined axis method where the internal rotation problem is first set up in the principal axis system. For each individual internal rotor, the Hamiltonian matrix is transformed into a rho axis system to eliminate the

Coriolis coupling terms which occur in Equation (2.2). In the rho axis system, the eigenvalues are calculated in the product basis of planar rotor functions for the torsion and symmetric top functions for the overall rotation. Subsequently, the eigenvalue matrix is transformed back to the principal axis system.

The *XIAM* code is very user-friendly and offers a reasonable compromise of accuracy and speed of the calculations because of suitable basis transformations and matrix factorization. Therefore, it rapidly became one of the programs most frequently used to fit rotational spectra of many molecules where internal rotations take place. However, a known weakness of *XIAM* is the treatment of methyl torsions with low barrier heights, for example, in 3-pentyn-1-ol [172], vinyl acetate [173], and *N*-ethylacetamide [174], where in the fits standard deviations within measurement accuracy cannot be achieved.

2.5.1.3 The *BELGI* code

The program *BELGI*, initially written by Kleiner et al. [18], is also popularly used to deal with internal rotation problems. The Hamiltonian is written in the Rho Axis Method (RAM) as follows:

$$H_{RAM} = H_r + H_{CD} + H_T + H_{int} \tag{2.4}$$

where H_r is the rotational Hamiltonian, H_{CD} is the centrifugal distortion Hamiltonian, H_T is the torsional Hamiltonian, and H_{int} consists of higher order torsional-rotational interaction terms:

$$\begin{aligned} H_R &= A_{RAM}P_a^2 + B_{RAM}P_b^2 + C_{RAM}P_c^2 + D_{ab}(P_aP_b + P_bP_a), \\ H_T &= F(p_\alpha - \rho P_a)^2 + V(\alpha). \end{aligned} \tag{2.5}$$

D_{ab} is an out-of-plane parameter arising from the use of a nonprincipal system. The relation between the rotational constants A, B, C in the principal axis system is given in Equation (2.2), and the constants in the rho axis system in Equation (2.5) are obtained by diagonalizing the 3 × 3 matrix of the RAM rotational constants [175].

There are two versions of *BELGI* for molecules with a C_s frame symmetry, *BELGI-C_s* for one internal rotor [18] and *BELGI-C_s-2Tops* for two internal rotors of C_{3v} symmetry [20], as well as *BELGI-C_1* for one rotor and a C_1 frame symmetry [19]. *BELGI-C_s* was tested extensively on acetaldehyde [176, 177]. Later, spectra of other molecules like acetic acid [178, 179] and ^{13}C-methyl formate ($H^{13}COO–CH_3$) [180] were treated with this program as well. The *BELGI-C_S-2Tops* code was applied to methyl acetate [20, 131], ethyl methyl ketone [133], methyl propionate [134], and dimethylbenzaldehyde [135]. Unlike *XIAM*, *BELGI-C_s* and *BELGI-C_1* use only the RAM method.

The program *BELGI* in its recently developed hyperfine versions *BELGI-C_s-hyperfine* and *BELGI-C_1-hyperfine* [181], as well as *BELGI-C_s-2Tops-hyperfine* [182], includes weak nuclear quadrupole coupling using a first-order perturbation approximation

and has proved its predictive power for a number of molecules such as *N-tert*-butylacetamide (C_s, one-top) [181], *N*-ethylacetamide (C_1, one-top) [174, 181], 3-nitrotoluene (C_s, one-top) [183], and 4,5-dimethylthiazole [182].

2.5.1.4 The *RAM36* code

Similar approaches are utilized by the program *RAM36*, written by Ilyushin et al. [22], currently existing in a version for fitting a V_6 potential with its application on e.g., toluene [22], as well as 3,5- and 2,6-difluorotoluene [184]. The combination of ^{14}N quadrupole coupling and methyl internal rotation can be handled with the program *RAM36* but so far only for molecules with C_s symmetry, for example, in the case of nitromethane with a V_6 potential [185] or *N*-methylformamide with a V_3 potential [186]. *RAM36* is very fast and uses a procedure which allows the user to choose various higher order terms and to achieve fits for very high J values.

2.5.1.5 The *ERHAM* code

A further program often applied to treat rotational spectra of molecules with up to two internal rotors and $J_{max} = 120$ is *ERHAM*, written by Groner [21]. The C_{3v} symmetry of internal rotors restricted in *XIAM* and *BELGI* is not applied for *ERHAM*. In molecules with one rotor or two nonequivalent rotors, the frame symmetry can be C_s or C_1, as well as C_2, C_{2v}, or C_s for two equivalent rotor cases. The physical meaning of the fitted parameters is not as clear as in *XIAM*, *BELGI*, or *RAM36*, because *ERHAM* sets up and solves an effective rotational Hamiltonian. For example, the rotational barrier is not a fit parameter and needs to be extracted. On the other hand, *ERHAM* is very fast and therefore advantageous in fitting large data sets. The fit quality is usually quite satisfactory with standard deviation close to experimental accuracy. Dimethyl ether was the first molecule fitted by *ERHAM*, first by its author [187] and then by Endres et al. [188] Another two-top molecule with equivalent rotors, acetone, was also extensively studied with *ERHAM* [140]. A large number of one-top molecules have been treated, such as methyl carbamate [189], pyruvic acid [190], methyl formate [191], pyruvonitrile [192], methyl isobutyl ketone [193], as well as molecules with two inequivalent rotors like isopropenyl acetate [139].

2.5.1.6 The *PAM-C_{2v}-2tops* code

A further program written explicitly for molecules with two equivalent methyl rotors and a C_{2v} symmetry at equilibrium is *PAM-C_{2v}-2tops*, written by Ilyushin and Hougen [23]. *PAM-C_{2v}-2tops* was applied first to acetone [23, 194], then to 2,5-dimethylfuran [143]. It uses a two-dimensional potential function and can fit rotational transitions of different torsional states simultaneously. Under group theoretical considerations, the program is based on the G_{36} permutation-inversion group. *PAM-C_{2v}-2tops* utilizes the principal axis method and applies a two-step diagonalization procedure. The two-top

torsion-rotational Hamiltonian matrix under the full G_{36} permutation-inversion group is split into four submatrices under the G_9 permutation-inversion group, corresponding to the (00), (01), (11), and (12) symmetry species, for blockwise diagonalization. These (00), (01), (11), and (12) blocks represent the $(A_1 \otimes A_2 \otimes A_3 \otimes A_4)$, G, $(E_3 \otimes E_4)$, and $(E_1 \otimes E_2)$ symmetry species in G_{36}, respectively.

2.5.2 Separate fits of LAM species: *quick check of the assignments*

With the use of supersonic expansion, torsional symmetry species in the vibrational ground state ($v_t = 0$) can be observed. The excited states are no longer populated because of the very low rotational temperature of the jet. Consequently, it is often not possible to determine the contributions of higher order terms (V_6, V_9 ...) in the potential function of Equation (2.1). Simultaneously fitting the potential terms and moments of inertia of the internal rotors often fails. It is also difficult to determine the potential or kinetic coupling terms in the cases of coupled internal rotors. Checking the assignments and finding new lines become very difficult and time-consuming tasks using a global fit. For these purposes, fitting the different symmetry species separately and treating them as individual species without any coupling with each other turn out to be helpful. This is a method which is known to work usually well for the A species where a semirigid rotor Hamiltonian supplemented by centrifugal distortion corrections, $H = H_r + H_{CD}$, is sufficient to obtain a fit with standard deviation close to the measurement accuracy, even for molecules where the barriers to methyl internal rotation are lower than 10 cm^{-1}, such as 3-pentyn-1-ol (fit A in Table 2.2 of the study by Eibl et al. [172]).

By adding effective terms in the Hamiltonian, taking into account the E species lines and using a model including a number of interactions within the torsional bath, a standard deviation close to measurement accuracy of 2 kHz can be achieved. In the above mentioned example on 3-pentyn-1-ol, the global fit was performed with the program *BELGI-C_1* with a standard deviation of 1.5 kHz [172]. However, such global fits only work with relatively well-assigned data sets. The initial assignment of torsional excited species (E species for one-top molecules, (01), (10), (11), and (12) for two-top molecules, and so on) is often performed with the program *XIAM*. This procedure is very time consuming and uncertain because of the large standard deviation of the *XIAM* fit, especially for low barrier cases. Furthermore, with more than two rotors, no effective Hamiltonian programs are available to check the assignment. Checking the assignments with combination difference loops is not always possible.

2.5.2.1 Test case 1: *a very low barrier (10 cm^{-1}) with C_1 symmetry*

The question to be addressed is how to get a good prediction for all torsional species to ease the assignment purpose. As mentioned, this works well for the torsional ground state (A or (00) species and so on) which can be fitted separately with $H = H_r + H_{CD}$. However, such separate fit can be done for any other symmetry species if angular momentum operators of odd power, e.g., P_a , P_b, P^2P_a, are included in the Hamiltonian. A computer code, called "Separately Fitting Large Amplitude Motion Species (*SFLAMS*)", was written for this purpose [138]. The main part of the Hamiltonian implemented in *SFLAMS* has the following form:

$$H = AP_a^2 + BP_b^2 + CP_c^2 + Q_a p_\alpha P_a + Q_b p_\alpha P_b + Q_c p_\alpha P_c \tag{2.6}$$

where all terms are of even order. If the $Q_a p_\alpha$, $Q_b p_\alpha$, and $Q_c p_\alpha$ terms are substituted by the expectation value $\langle E|p_\alpha|E\rangle$ of p_α and become Q'_a, Q'_b, Q'_c, the power of some terms in the Hamiltonian becomes odd. The expectation value is obviously zero for the A species. A similar approach was applied in the program *JB95* [24] and a program written by Ohashi et al. [195] to fit separately the five torsional species of *N*-methylacetamide. Under the time-reversal operation, the odd power terms of the angular momentum components P_a, P_b, and P_c change sign. The Q'_a, Q'_b, Q'_c coefficients enclose numerical expectation values of an odd power of the torsional angular momentum operators of the methyl group (p_α), making the torsion-rotational Hamiltonian effective. Nevertheless, p_α would also change sign under time reversal as the odd power P_a, P_b, and P_c operators would do if they were included explicitly in the Hamiltonian. The cancelation of sign changes makes the Hamiltonian invariant as it should be.

Coming back to the example on 3-pentyn-1-ol mentioned above: while the program *SFLAMS* was applied, the standard deviation of 2.3 kHz was obtained for 51 A species lines by fitting only the three rotational and five quartic distortion constants. A similar standard deviation of 3.2 kHz could be obtained for 36 E species lines by adding the q, r, and s parameters, which are Q'_a, Q'_b, Q'_c, respectively, in Equation (2.6) and some higher odd order parameters such as q_J, q_{JK}, r_K etc. (see Table 2.1).

2.5.2.2 Test case 2: *a four-top molecule*

The use of *SFLAMS* is best demonstrated in the case of 2,3,4,5-tetramethylthiophene, a molecule featuring four methyl rotors, two of which are in close proximity to the sulfur atom and are equivalent. The other two are also equivalent (see Figure 2.7). Group theoretical treatment has shown that there are 25 torsional species for this molecule, among them 13 species could be assigned [196]. A typical spectrum is illustrated in the left hand-side of Figure 2.8. No global fit could be made due to the lack of suitable program. Separate two-top and three-top fits were carried out with the program *XIAM*, but the results of all those fits are not satisfactory as they do not reach the experimental accuracy [196].

Table 2.1: Molecular parameters of the A and the E species of 3-pentyn-1-ol obtained with the program *SFLAMS*.

3-Pentyn-1-ol A			3-Pentyn-1-ol E			
Par.[a]	**Unit**	**Values**[b]	**Par.**[a]	**Unit**	**Values**[b]	**Operator**[c]
A	GHz	8.45793216(76)	A	GHz	8.39348(24)	P_a^2
B	GHz	1.68514419(72)	B	GHz	1.6843810(23)	P_b^2
C	GHz	1.46706317(70)	C	GHz	1.4674993(45)	P_c^2
D_J	kHz	0.8730(32)	D_{JK}	kHz	44.16(67)	P^4
D_{JK}	kHz	−4.524(19)	D_K	MHz	−0.680(66)	$P^2P_a^2$
D_K	kHz	56.48(18)	d_1	MHz	−0.09318(62)	P_a^4
d_1	kHz	0.20637(51)	q	GHz	14.834780(97)	P_a
d_2	kHz	5.57(34)	q_J	kHz	−20.97(67)	P^2P_a
N^d		51	q_{JK}	kHz	−1.900(19)	$P^2P_a^3$
rms^e	kHz	2.3	q_K	MHz	−1.53(22)	P_a^3
			r	GHz	1.20349(10)	$P_+ + P_-$
			r_J	kHz	−8.10(13)	$P^2 (P_+ + P_-)$
			r_K	MHz	1.093(78)	$\{P_a^2, (P_+ + P_-)\}$
			r_{KK}	MHz	−0.2506(51)	$\{P_a^4, (P_+ + P_-)\}$
			s	GHz	0.192863(23)	$P_- - P_+$
			N^d		36	
			rms^e	kHz	3.2	

[a]Watson's S reduction and I^r representation were used.
[b]Standard errors in parentheses are in the units of the last digit.
[c]The product of the parameter and operator from a given row yields the term actually used in the Hamiltonian.
[d]Number of line.
[e]Root-mean-square deviation of the fit.

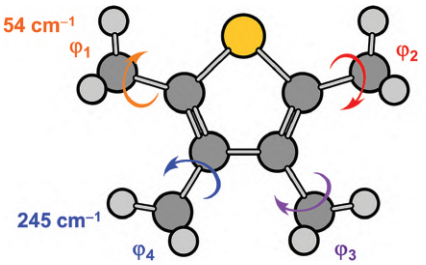


Figure 2.7: The four-top molecule 2,3,4,5-tetramethylthiophene. The two upper tops φ_1 and φ_2 as well as the two bottom tops φ_3 and φ_4 are equivalent [196]. The preliminary torsional barriers are given.

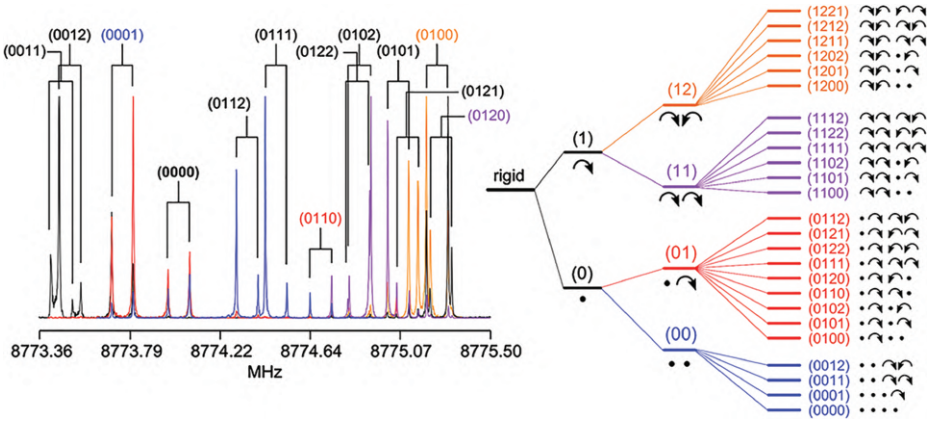


Figure 2.8: Left hand side: Splittings due to internal rotation in the rotational spectrum of 2,3,4,5-tetramethylthiophene with two pairs of equivalent rotors. Right hand side: The 25 allowed *b*-type transitions using a scheme similar to that given in Figure 2.6 [196].

Using *SFLAMS*, the torsional species are fitted well separately with standard deviations close to measurement accuracy. From the *q* and *r* parameters, the symmetry species could be unambiguously assigned. For example, the *q* and *r* values of the (0101) species should be the sum of the *q* and *r* values of the (0001) and (0100) species, respectively (see Table 2.2). A small deviation occurs because this sum rule does not take into account the higher order terms connecting to *q* and *r*.

Table 2.2: The rotational constants, *q*, *r* parameters, and the root-mean-square deviations of some assigned species of 2,3,4,5-tetramethylthiophene obtained with the program *SFLAMS* [138].

Per.	Unit	(0000)	(0001)	(0011)	(0012)	(0100)
A	GHz	1.831412151	1.831174074	1.830935893	1.830936122	1.816106857
B	GHz	1.386904417	1.386632349	1.386359251	1.386359726	1.385709333
C	GHz	0.800078070	0.800078483	0.800078536	0.800078238	0.800110164
q	MHz	0.0	19.6	0.0	39.3	1224.3
r	MHz	0.0	22.4	44.9	0.8	285.9
rms	kHz	2.6	2.7	1.8	1.4	1.2
Per.	**Unit**	**(0101)**	**(0110)**	**(0111)**	**(0120)**	
A	GHz	1.815876903	1.815838209	1.815602766	1.815863307	
B	GHz	1.385426931	1.385433503	1.385150624	1.385429503	
C	GHz	0.800110910	0.800111933	0.800110473	0.800110210	
q	MHz	1244.9	1205.4	1226.1	1246.3	
r	MHz	309.9	308.3	332.4	262.6	
rms	kHz	4.6	2.8	1.5	2.9	

SFLAMS and all effective programs fitting separately the torsional states yield very precise predictions for spectral assignment purposes. Extracting the barrier heights from the q, r, and (if allowed by symmetry) s parameters or vice versa is possible. However, it is currently performed manually and only for certain one-rotor cases. The use of global approaches taking all states into account simultaneously and allowing a direct determination of the potential barrier is thus complementary to this so-called "local" approach.

2.6 Variety of large amplitude motions in molecules and their applications

This section gives an overview of different molecular classes, each of which features an individual character on LAMs. The information deduced from the interpretation of the microwave spectra is discussed, mainly the barrier heights but also their applications on structure determinations.

2.6.1 Challenges in internal rotation problems: *some examples*

2.6.1.1 Torsional barriers in acetates: *low (100–150 cm^{-1}) and predictable*

The barriers to internal rotation of the acetyl methyl group in acetates (acetic acid esters), CH_3-COO-R, have been thoroughly studied. Almost all acetates investigated so far can be divided into three classes, as illustrated in Figure 2.9. Class I encloses *α,β-saturated acetates*, where the barrier to internal rotation is usually around 100 cm^{-1}. Examples are *n*-alkyl acetate from methyl acetate [131] up to *n*-hexyl acetate [175, 197–200]. Even in the cases of isopropyl acetate [201] and isoamyl acetate [202] where the alkyl chain is branched, or when a double bond is present at the end of the alkyl chain as in allyl acetate [42], the barrier height seems to be almost unaffected.

Class II comprises *α,β-unsaturated acetates* where the C=C double bond is located in the COO plane. The torsional barrier increases from 100 to 150 cm^{-1} due to extended conjugation over the C=C bond. Vinyl acetate [173] and the *E* and the *Z* isomer of butadienyl acetate [203] are two representatives in this class. The two molecules have proven that when the number of conjugated double bonds is augmented, there is no further increase in the barrier height of the acetyl methyl torsion. The configuration of the conjugated double bond system does not influence the barrier significantly, either, since the value of this parameter remains almost the same in the *E* and *Z* isomers of butadienyl acetate [203].

Finally, class III contains *α,β-unsaturated acetates* where the double bond is not located in the COO plane. Since the molecule is not planar, the conjugation is less effective than that in the molecules in class II, and the torsional barrier is

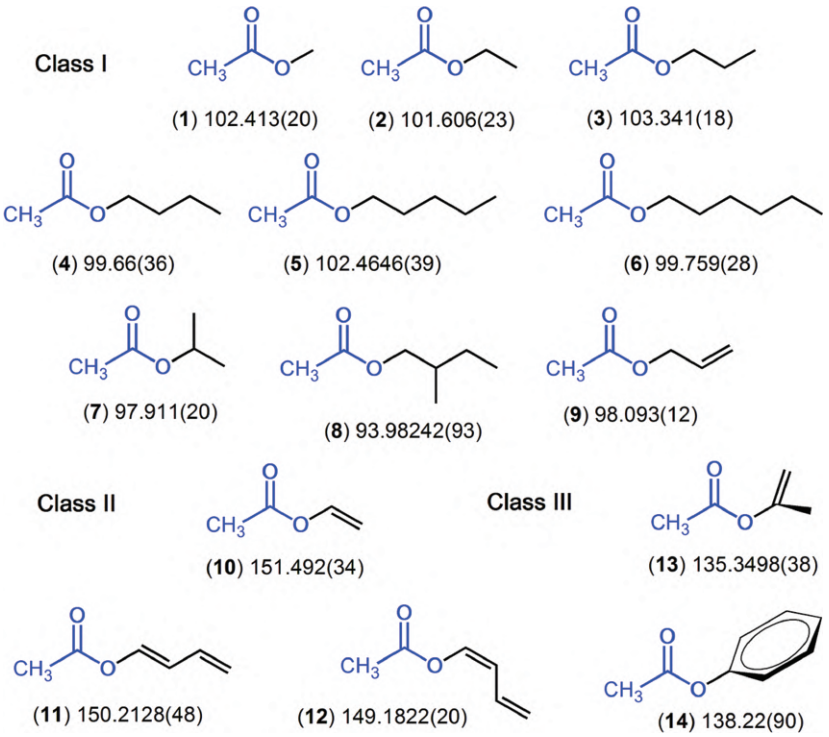


Figure 2.9: Torsional barriers of the acetyl methyl group in acetates (in cm^{-1}). Class I: (**1**) Methyl acetate [131], (**2**) ethyl acetate [175], (**3**) *n*-propyl acetate [197], (**4**) *n*-butyl acetate [198], (**5**) *n*-pentyl acetate [199], (**6**) *n*-hexyl acetate [200], (**7**) isopropyl acetate [201], (**8**) isoamyl acetate [202], and (**9**) allyl acetate [42]. Class II: (**10**) Vinyl acetate [173], (**11**) *E*-butadienyl acetate [203], and (**12**) *Z*-butadienyl acetate [203]. Class III: (**13**) Isopropenyl acetate [139] and (**14**) preliminary results of phenyl acetate [204].

around 135 cm^{-1}. Two typical examples of class III are isopropenyl acetate [139] and phenyl acetate [204].

2.6.1.2 Torsional barriers in acetamides: yet *unpredictable*

The barrier to internal rotation of a methyl group varies strongly but generally could be divided in three classes. Molecules with torsional barriers higher than 600 cm^{-1} belong to the first class with ethane as the classical example (~1000 cm^{-1}) [205]. The second class comprises molecules with intermediate barrier heights from about 200 to 600 cm^{-1}. A vast number of one- or two-top molecules like *o*-methylanisole [83], 2,5-dimethylfuran [143], and methyl methacrylate [206] fall in this class. Splittings arising from the internal rotation(s) are typically up to some hundreds of MHz. In most cases, taking into account the V_3 term and the internal rotor position(s) is sufficient to

model the rotational spectrum with high accuracy. Finally, the last class contains molecules where the methyl group(s) undergo(es) internal rotation(s) with barrier heights lower than 200 cm^{-1}, with a subclass of very low barriers (say < 30 cm^{-1}). The torsional splittings can be in the order of several GHz, making the assignment of the rotational spectra challenging. In many cases, high order terms are required in the Hamiltonian to model the spectra to experimental accuracy. In the previous section on acetates, a classification can be reliably predicted (see Figure 2.9). This is not yet the case of acetamides, where a summary is given in Figure 2.10.

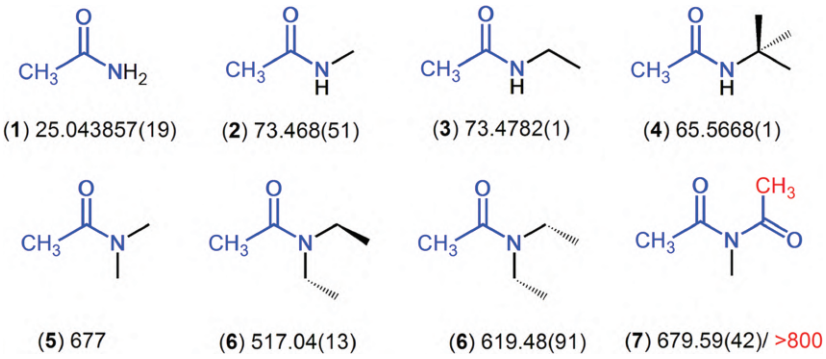


Figure 2.10: Torsional barriers of the acetyl methyl group in acetamides (in cm^{-1}). (**1**) Acetamide [207], (**2**) *N*-methylacetamide [195], (**3**) *N*-ethylacetamide [174], (**4**) *N*-*tert*-butylacetamide [181], (**5**) *N*,*N*-dimethylacetamide [147], (**6**) two conformers of *N*,*N*-diethylacetamide [91], and (**7**) *N*-methyldiacetamide [208].

We now focus on the the acetyl methyl group in Figure 2.10. This group in acetamide has a remarkably low barrier of 25 cm^{-1} [207]. Probably, the electronic configuration from the amide bond with two hydrogen atoms attached to the nitrogen atom influences the V_3 potential of the methyl group on the other side of the carbonyl group. If one of these hydrogen atoms are substituted, secondary acetamides are obtained, where the barrier is higher than the value found for acetamide, but still low, and varies between 65 and 75 cm^{-1}, depending on the type of the substituent. While the acetyl methyl group in *N*-*tert*-butylacetamide has a barrier to internal rotation of approximately 65.6 cm^{-1} [181], the respective values of *N*-methylacetamide [195] and *N*-ethylacetamide [174] are both about 73.5 cm^{-1}. Currently, there is no conclusive explanation for this difference.

It is quite impressive that the substitution of one hydrogen atom leads to a difference of approximately 50 cm^{-1} in barrier height between acetamide and the secondary acetamides. If both hydrogen atoms in acetamide are substituted, the effect is even greater as the barrier height increases by an order of magnitude. For all *tert*-acetamides, the barriers are high, in the order of 500–800 cm^{-1} [91, 147, 208]. Because the substituents are well-separated from the acetyl methyl group by the

amide moiety (C=O)N, the reason is probably electronic rather than steric effects. Obviously, the electronic situation is quite special in amides, where information on the electron density can be transferred more easily through the amide bond to the acetyl methyl rotor.

2.6.1.3 Essentially free internal rotation of the propynyl methyl group: *very low (<10 cm^{-1}), very challenging*

Many torsional barriers were accurately determined by using microwave spectroscopy, but only a few studies concern molecules where the barriers are lower than 30 cm^{-1}. Acetamide, mentioned previously in Section 2.6.1.2, represents a typical example where the barrier height of the threefold potential is 25 cm^{-1} [207]. Surprisingly, the number of very low barrier investigations does not increase much in the modern ages of microwave spectroscopy. Probably, the reason is the challenge in assigning and fitting rotational spectra of molecules with very low torsional barriers, which are in addition hard to calculate by quantum chemistry. As a consequence of the small number of available studies, the barrier heights cannot be predicted by chemical intuition, either. In some cases such as *trans*-methyl nitrite [209] or *m*-fluorotoluene [210], the comparable values of the threefold and sixfold potential terms complicated the spectral analysis.

The barrier to internal rotation of the two methyl groups in ethane is high, at around 1000 cm^{-1} [205]. But if a C≡C triple bond is inserted as a spacer to separate the methyl groups, as in 2-butyne (dimethylacetylene), CH_3–C≡C–CH_3, it is assumed that the molecule exhibits two essentially free methyl internal rotations [211]. Though dimethylacetylene cannot be investigated by microwave spectroscopy due to the lack of a permanent dipole moment, the basic concepts of this chemical bonding suggest that a methyl group connected to an acetylene fragment CH_3–C≡C–R (called the propynyl methyl group) features an extremely low torsional barrier (V_3 < 10 cm^{-1}). This assumption was confirmed by dimethylacetylene-d_3 (molecule (**1**) in Figure 2.11) [212] and methylsilylacetylene (**2**) [213], two molecules related to dimethylacetylene possessing a barrier to methyl internal rotation of 5.62(16) and 3.77(70) cm^{-1}, respectively. Very low V_3 potentials of 1.00900(42) and 2.20(12) cm^{-1} were observed for 2-butynoic acid (**9**) (when R = COOH) [214] and tetrolyl fluoride (**8**) (R = COF) [215], respectively.

The torsional barrier of the propynyl methyl group increases to 6.93(9) cm^{-1} in 2-butynol (**4**) (R = CH_2OH) [217]. Currently, the largest value for this group was observed in 1-chloro-2-butyne (**3**) (R = CH_2Cl), for which Solwijk and van Eijck [216] reported a V_3 potential of 10.05(9) cm^{-1}. Three further alkynols, 3-pentyn-1-ol (**5**) [172], 3-pentyn-2-ol (**6**) [218], and 4-hexyn-3-ol (**7**) [219], were studied to find out the effects of the alkyl length on the torsional barrier of the propynyl methyl group. From the results summarized in Figure 2.11, we found a trend that the barrier height decreases slightly by longer alkyl chain. The barrier height of 2-butynol (**4**) [217] does not fit in this trend, but the authors have reported a lot of fitting difficulties.

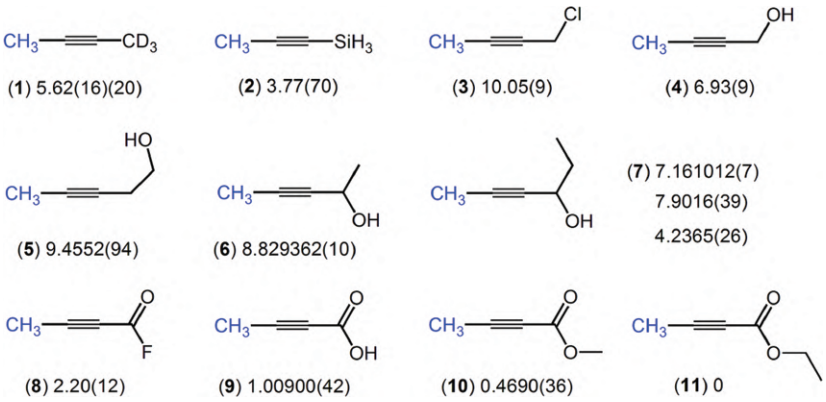


Figure 2.11: Torsional barriers of the propynyl methyl group (in cm^{-1}). (**1**) Dimethylacetylene-d_3 [212], (**2**) methylsilylacetylene [213], (**3**) 1-chloro-2-butyne [216], (**4**) 2-butynol [217], (**5**) 3-pentyn-1-ol [172], (**6**) 3-pentyn-2-ol [218], (**7**) 4-hexyn-3-ol [219], (**8**) tetrolyl fluoride [215], (**9**) 2-butynoic acid [214], (**10**) methyl-2-butyonate [218], and (**11**) ethyl-2-butyonate [218].

Therefore, the barrier of 6.93(9) cm^{-1} might not be as exact as the barriers observed for 3-pentyn-1-ol (**5**), 3-pentyn-2-ol (**6**), and 4-hexyn-3-ol (**7**).

The microwave spectrum of 2-butynoic acid has given the lowest barrier height ever analyzed with a RAM Hamiltonian (1.00900(42) cm^{-1}) [214]. The questions are as follows: Is it still possible to lower this barrier? Can we deduce how structural changes affect the methyl torsional barrier? From the study on the alkynols mentioned above, where a decreasing trend by longer alkyl chain was observed, two molecules were investigated where the COOH group in 2-butynoic acid was lengthened to a methoxylate ($COO–CH_3$) (**10**) and an ethoxylate group ($COO–CH_2–CH_3$) (**11**). The barrier of 0.4690 (36) cm^{-1} found in methyl-2-butyonate (**10**) [218] provides a new record in low V_3 potential, proving that the methyl group is a good sensor of the molecular structure. For the two conformers of ethyl-2-butyonate (**11**), the internal rotation of the propynyl methyl group is quasi free. A V_3 potential could not be fitted and therefore was fixed to zero in fits with root-mean-square deviation within the measurement accuracy [218].

2.6.2 Sensing the molecular conformations of natural substances by internal rotors

2.6.2.1 Acetyl methyl torsion in honey bee pheromones: is the barrier height 180 or 240 cm^{-1}?

Previous sections have shown that the barriers to methyl internal rotation can be linked to functional groups and other structural characteristics of the molecules, making the methyl group a spectroscopic “detector” of the structure. This is quite important if we want to understand eventually the structure and intrinsic properties

of larger and more complicated molecules, such as those present in natural substances produced by the plants or biomolecules.

A series of saturated methyl alkyl ketones, naturally present in the honey bee pheromone cocktail, were investigated, starting from ethyl methyl ketone (butan-2-one) to octan-2-one. Except for ethyl methyl ketone, at least two conformers are present in the microwave spectra. One of the two features a C_s structure, where all heavy atoms are located on a symmetry plane, and the other one exhibits C_1 symmetry, where the *y*-carbon atom of the alkyl chain is bent in a nearly synclinal position. While different conformations possess almost the same barrier height in the cases of *n*-alkyl acetates (see *Section 2.6.1.1*), this is not the situation found for methyl alkyl ketones where a scheme with two different classes can be drawn, as summarized in Figure 2.12.

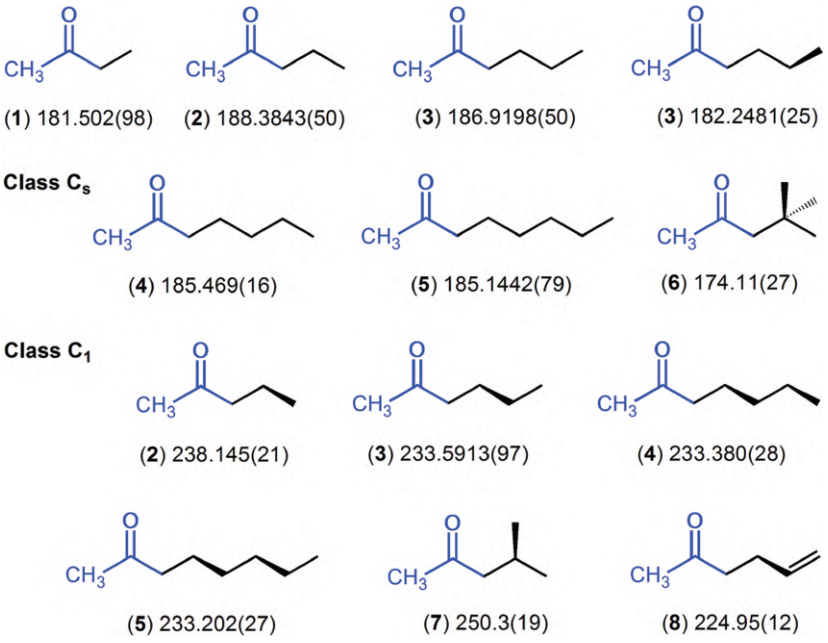


Figure 2.12: Torsional barriers of the acetyl methyl group in methyl alkyl ketones (in cm^{-1}). (**1**) Ethyl methyl ketone [133], (**2**) pentan-2-one [220], (**3**) hexan-2-one [221], (**4**) heptan-2-one [222], (**5**) octan-2-one [223], (**6**) methyl neopentyl ketone [224], (**7**) methyl isobutyl ketone [193], and allyl acetone [225].

The first class comprises of ketones with a C_s *structure* where the barrier height is always around 180 cm^{-1}. Examples are ethyl methyl ketone [133], the C_s conformer of pentan-2-one [220], hexan-2-one [221], heptan-2-one [222], and octan-2-one [223]. Even, if the alkyl chain is branched, as in the case of methyl neopentyl ketone, the barrier height seems to be almost unaffected [224].

The second class contains molecules with C_1 *symmetry* and a higher torsional barrier of approximately 240 cm^{-1}. The C_1 conformer of pentan-2-one [220], heptan-2-one [222], octan-2-one [223], two C_1 conformers of hexan-2-one [221], allyl acetone [225], and methyl isobutyl ketone [193] belong to this class. Within each class, the same trend as found for the propynyl group containing molecules (Section 2.6.1.3) also applies, i.e., a slight decrease of the barrier height by longer alkyl chain [223]. This trend is more pronounced in the C_s than in the C_1 series (see Figure 2.12).

2.6.2.2 Conformational determination by internal rotor in lavender oil

Linalool, illustrated in Figure 2.13, is an acyclic monoterpene. With its agreeable, floral, and refreshing scent, linalool is an important component in several essential oils of plants e.g., lavender (40%), coriander (70%), and ho-leaf (80%). The large size and open chain of linalool offer a rich conformational landscape with hundreds of possible conformers [226]. Among them, finally only one is observed in the jet-cooled microwave spectrum, which is shown in the lower trace of Figure 2.14. Compared to the spectrum in the same spectral range of the natural product "essence de lavendin", purchased in the Provence, France (see the upper trace of Figure 2.14), almost all lines of the linalool spectrum are present with very similar intensity in the spectrum of lavender oil. This shows that high resolution spectroscopy can be applied to study the chemical compounds used in olfaction as an analytic tool and to determine the compounds that play an important role for the scent of a perfume.

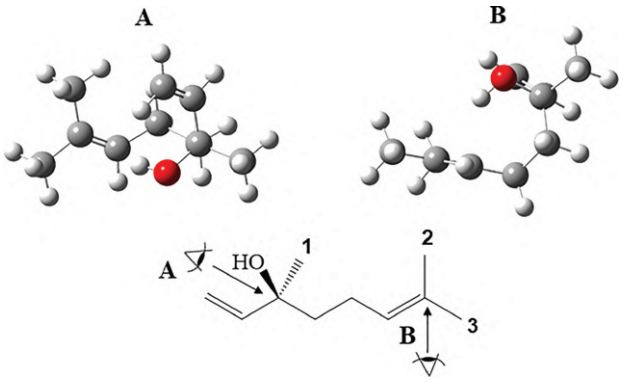


Figure 2.13: The globular geometry of the lowest energy conformer of linalool in two different views (A and B) corresponding to the respective Newman projections given in the upper trace. The three methyl groups are labeled with numbers; two of them show resolvable torsional splittings in the experimental spectrum.

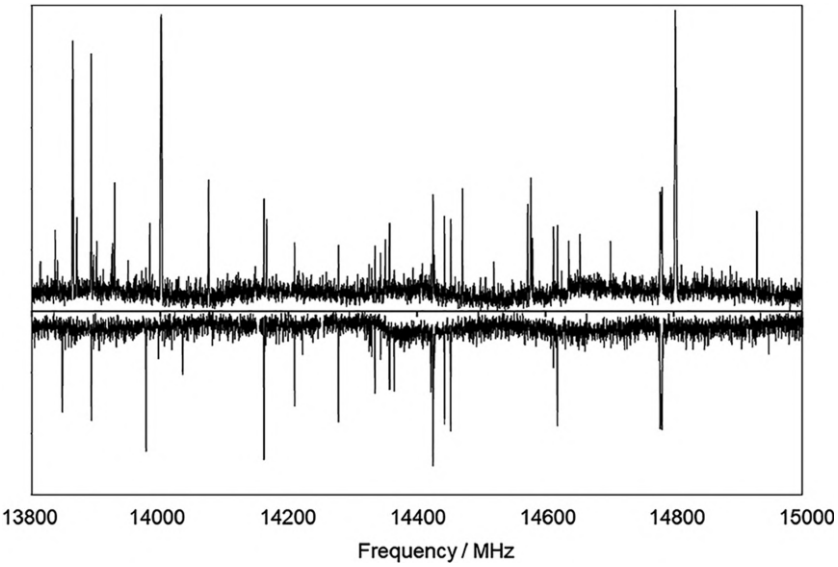


Figure 2.14: Upper trace: the spectrum of essence de lavendin (lavender oil). Lower trace: the linalool spectrum. The intensities are given in a logarithmic scale and in arbitrary units.

A standard opinion is that a comparison of the experimental rotational constants and those calculated by quantum chemistry is sufficient to validate the molecular structure. Nevertheless, in the case of linalool, due to the large number of conformers with low energy, the calculated rotational constants of three conformers close in energy are very similar. All of them are in good agreement with the experimentally deduced constants. In order to identify the observed conformer, the angles between the methyl internal rotor axis and the principal axes provided decisive information, since those of only one conformer (that given in Figure 2.13) agree with the experimental ones. Therefore, the internal rotation analysis is crucial for structure determinations in large molecules.

2.6.3 Coupled internal rotations

Coupled internal rotations are reported much more rarely. The present section will give an example on a systematic investigation of six isomers of dimethylanisole given in Figure 2.15 [204].

2.6.3.1 From no trouble...

The only member in the family which does not cause fitting troubles is 3,4-dimethylanisole (molecule (**5**) in Figure 2.15) [128]. It is also the only isomer which exists as

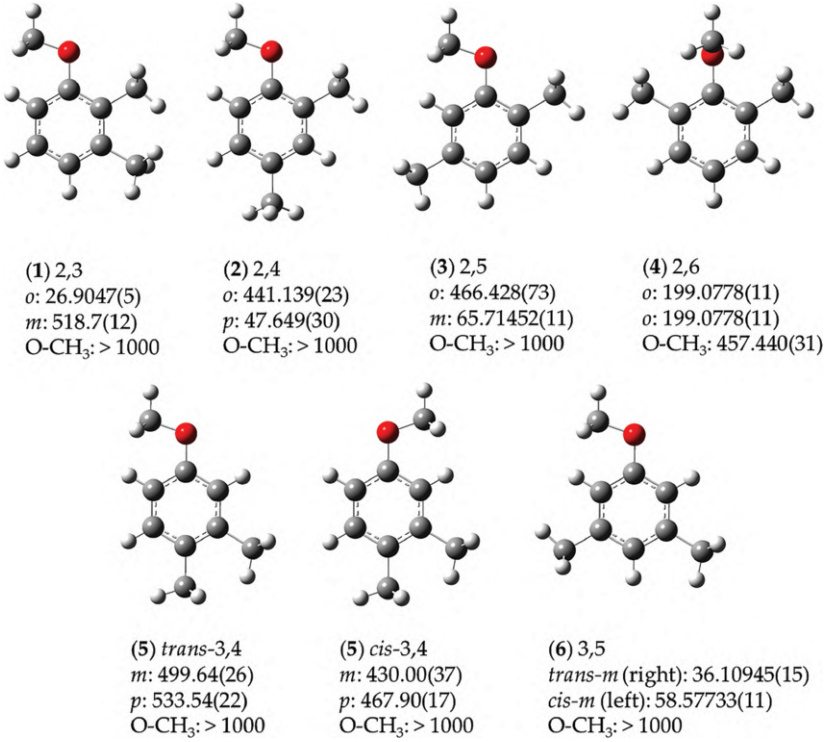


Figure 2.15: Torsional barriers of the methyl groups in six isomers of dimethylanisole (in cm^{-1}). (**1**) 2,3-Dimethylanisole [136], (**2**) 2,4-dimethylanisole [137], (**3**) 2,5-dimethylanisole [204], (**4**) 2,6-dimethylanisole [204], (**5**) 3,4-dimethylanisole [128], and (**6**) 3,5-dimethylanisole [204]. For each compound, "*o*", "*m*", or "*p*" corresponds to the barrier height of the methyl group in the *ortho*-, *meta*-, or *para*-position relative to the $-OCH_3$ methoxy group, respectively. "$-OCH_3$" corresponds to the barrier height of the methoxy methyl group, which is only determinable in 2,6-dimethylanisole (**4**).

two conformers in the microwave spectrum, called *syn* and *anti*, with different orientations of the methoxy group. The torsional barrier of the methyl group in the methoxy moiety $-OCH_3$ (called the methoxy methyl group) is high enough that its internal rotation splittings are not observable with the resolution of the experimental setup. This is in agreement with the study on anisole, which can be treated as a rigid rotor [227]. The coupled LAMs in 3,4-dimethylanisole is therefore a two-top problem, where the methyl groups at the *meta*- and *para*-position relative to the methoxy group are the two internal rotors. Because they are in close proximity, steric hindrance causes an intermediate barrier between 400 and 550 cm^{-1} for both methyl tops which can be modeled well without any top-top coupling term in the Hamiltonian.

2.6.3.2 ...to some troubles

Two other isomers, 2,4- (**2**) [137] and 2,5-dimethylanisole (**3**) [204], feature an intermediate and a low barrier height. Also in these cases, splittings from the internal rotation of the methoxy methyl group are not observable. The immediate V_3 potential of about 450 cm^{-1} found for the *o*-methyl group is similar in both molecules and probably also arises from steric hindrance from the neighboring methoxy group. The methyl rotor at the *para* or *meta* position of 2,4- or 2,5-dimethylanisole, respectively, possesses a much lower barrier height of less than 70 cm^{-1} (see Figure 2.15) and is thus a challenge for the fitting process. The top-top coupling term multiplying $(1 - \cos\alpha_1)(1 - \cos\alpha_2)$ and higher order effective parameters are required in the Hamiltonian to achieve standard deviation close to measurement accuracy [137, 204]. The spectral assignments are quite extensive because every frequency has to be verified by combination difference loops to assure a correct assignment.

2.6.3.3 ...and a lot of troubles

Since the 3,4-isomer is a "no-trouble-case" due to steric hindrance of the two neighboring methyl groups, which increases the barrier heights to intermediate values, one tends to assume that 2,3-dimethylanisole (**1**), also featuring two neighboring methyl groups, would not cause any fitting troubles, either. This assumption finally turns out to be wrong. In 2,3-dimethylanisole (**1**), while the V_3 potential of 519 cm^{-1} of the *m*-methyl group is intermediate and can be treated well, the surprisingly low barrier to internal rotation of about 27 cm^{-1} of the methyl group at the *ortho*-position cannot be captured correctly with the *XIAM* program [136]. Assuming that a methoxy group and a methyl group are similar, then the *o*-methyl group, featuring a C_{3v} symmetry, would experience potentials on a frame with C_{2v} symmetry, similar to the cases of toluene [22] or nitromethane [185], where the V_3 contribution of the potential vanishes and only a small V_6 term exists. In the case of 2,3-dimethylanisole, the frame symmetry is out of balance from C_{2v}, resulting in a small V_3 term of about 27 cm^{-1}. This very low V_3 potential term is challenging for the spectral analysis. The situation gets even worse in the case of 3,5-dimethylanisole (**6**), where both ring methyl groups undergo internal rotation with low torsional barriers [204].

2.6.3.4 But also some nice surprise: *lowering the torsional barrier by sterical hindrance*

Sterical hindrance is known to increase the barrier to methyl internal rotation because the methyl torsion is not only hindered quantum mechanically but also mechanically if a bulky group is present in close proximity of the methyl group. However, there are exceptions.

In all five isomers of dimethylanisole mentioned previously, slittings due to the internal rotation of the methoxy methyl group have never been observed. This is in agreement with the experimental results of anisole, as well as the high torsional

barriers of over 1000 cm^{-1} predicted by quantum chemistry. All these isomers have a common structural feature: the methoxy group is always located on the phenyl ring plane, which is at the same time the symmetry plane.

The situation changes in 2,6-dimethylanisole (**4**): because of steric hindrance caused by the two methyl groups occupying both *ortho*-positions of the phenyl ring, the methoxy group located between them is forced to tilt out of the phenyl plane by 90°. The molecular symmetry is still C_s, but the symmetry plane is now perpendicular to the plane containing the phenyl ring, and the two ring methyl groups are equivalent [204] (see Figure 2.15). This extraordinary orientation of the methoxy methyl group in 2,6-dimethylanisole has decreased its torsional barrier height from over 1000 to 457 cm^{-1}, and torsional splittings become resolvable. Therefore, this isomer surprisingly presents a three-top problem with a total of 10 torsional species. The potential barriers of the two equivalent methyl groups are approximately 200 cm^{-1} [204].

2.6.4 Inversion tunneling

As mentioned in *Section 2.4* on Large amplitude motions, much less investigations have been reported on inversion tunneling motions than on internal rotations. In addition, inversion tunneling is often combined with internal rotation. Only in very few molecules like *planar secondary amines*, the inversion tunneling of the proton at the nitrogen atom is not accompanied by internal rotation. This section will focus on a series of such molecules, starting from dimethyl amine [158], then ethyl methyl amine [159], and diethyl amine [160].

The *c*-component of the dipole moment changes sign, while the *b*-component retains its sign during the proton inversion (see Figure 2.16). Therefore, tunneling splittings are observed in the spectrum for all *c*-type transitions, which are twice the separation between the lowest symmetric and antisymmetric inversion energy levels. The values of these splittings are given in Figure 2.17. A trend can be easily recognized that the splittings decrease in larger molecules.

Phenyl formate also features a pure inversion tunneling motion, which is the tunneling of the phenyl ring [161]. At the beginning, phenyl formate was expected as a normal rigid-rotor molecule, but a rigid-rotor model has failed completely to reproduce its microwave spectra with a root-mean-square deviation of 3 MHz while the measurement accuracy was 2 kHz. Quantum chemical calculations have hinted that some state other than the ground state is populated in the molecular jet, as a consequence of the ring tunneling quantum effect. This low-lying $v_t = 1$ tunneling state is calculated to lie at 48.24 GHz (1.6 cm^{-1}) above the $v_t = 0$ ground state, corresponding to tunneling splittings in the order of about 100 GHz for all *c*-type transitions. For a comparison, the largest splittings observed for the secondary amines illustrated in Figure 2.18 are 2.6459 GHz, as found in dimethyl amine [158]. The splittings arising for

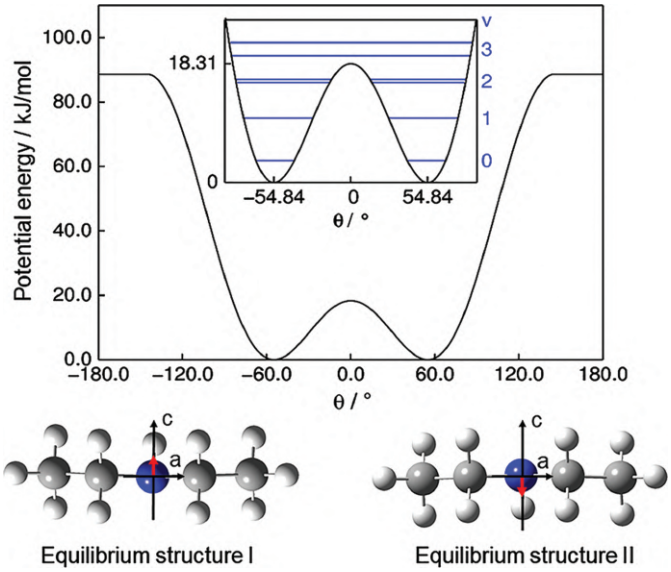
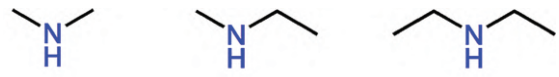


Figure 2.16: The potential energy curve describing the inversion tunneling at the nitrogen atom of diethyl amine. θ is the angle of the NH bond against the NCC plane. Inset: Horizontal lines indicate the lowest torsional energy levels v_t = 0, 1, 2, 3, which are doubly degenerated. The two versions of the equilibrium geometry given in the lower trace correspond to the two energy minima of the nitrogen tunneling process. The two equilibrium structures on the lower figure show that the dipole moment, indicated as a red arrow, changes sign in the *c*-direction upon proton tunneling.

b-type transitions due to Coriolis interaction are up to 100 MHz, three order of magnitude larger than those found in the the secondary amines, as depicted in Figure 2.18.



(**1**) 2645.9 (**2**) 1981.0 (**3**) 1521.5

Figure 2.17: Tunneling splittings of *c*-type transitions in secondary amines (in MHz). (**1**) Dimethyl amine [158], (**2**) ethyl methyl amine [159], and (**3**) diethyl amine [160].

2.7 Conclusions

A large variety of molecular systems with applications in diverse research fields from molecular biology, astrophysics to environmental sciences contain LAMs. Studying them at the molecular level is an extremely lively field. The focus of the present review was on two types of LAMs which frequently occur: internal rotation

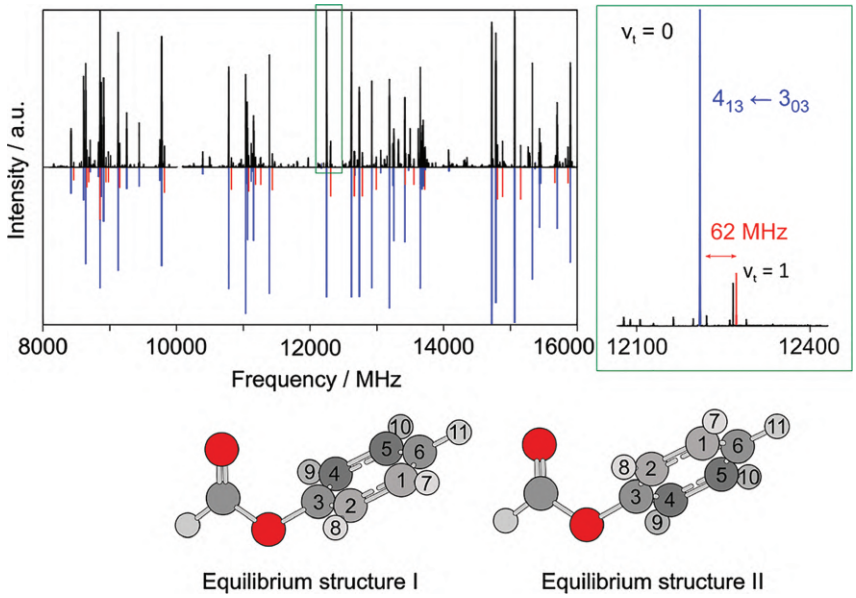


Figure 2.18: Upper figure, left hand side: A portion from 8000 to 16000 MHz of the survey spectrum of phenyl formate (upper trace) compared to the theoretical spectrum (lower trace) predicted using the program *SPFIT/SPCAT*. Transitions of the $v_t = 0$ ground state are marked in blue, those of the $v_t = 1$ excited state in red. Right hand side: The frequency range from 12100 to 12400 MHz in an enlarged scale, capturing the $v_t = 0$ and $v_t = 1$ components of the $4_{13} \leftarrow 3_{03}$ transition which are separated by 62 MHz due to Coriolis interactions. Lower figure: Two versions of the equilibrium geometry corresponding to the two energy minima of the ring tunneling process.

and inversion tunneling. The combination of molecular jet Fourier transform microwave spectroscopy, spectral modeling, and quantum chemical calculations is particularly successful in decoding complex spectra with LAMs and providing reference data for astrophysical research, atmospheric chemistry, or general applications in physical chemistry.

The two most popular types of state-of-the-art Fourier transform microwave spectrometer technology based on pulsed supersonic jet expansions are resonator-based and chirped-pulse spectrometers. The former version is commonly used in many microwave laboratories and has shown its superior sensitivity and resolving power but suffers from the time requirement to acquire survey spectra because the resonator has to be tuned mechanically for every frequency element at rather narrow steps of less than 0.25 MHz. The latter version relies on a very short but powerful frequency ramp signal with a band width of 1 GHz and thus provides unparalleled speed for scans with more reliable line intensities.

The use of quantum chemistry using programs such as *Gaussian* or *GAMESS* as a supporting tool for rotational spectroscopy is becoming very popular. The most common methods used by many microwave spectroscopic labs is the DFT method using

the B3LYP functional and the MP2 method, along with the Pople valence triple-zeta basis set 6-311++G(d,p). Conformational analysis performed by calculating potential energy surfaces and geometry optimizations is sufficiently accurate to start the spectral assignment. In the contrary, predicting the barriers of LAMs, corresponding to calculations of energy, is not yet sufficiently accurate to satisfy experimental requirements and needs experimental values for benchmark calculations.

Several computer codes for global fits of rotational spectra with splittings arising from LAMs have been developed along the last four or five decades, such as, not exhaustively *SPFIT/SPCAT*, *IAMCALC*, *JB95*, *XIAM*, *BELGI*, *RAM36*, *PAM-C_{2v}-2tops*, and ERHAM, just to mention some of them. A local approach such as in the *SFLAMS* code for separately fitting the LAM species can ease the assignments of the microwave spectra, especially when problems with a large number of internal rotors (> 3) and low torsional barriers have to be conquered, and it is still challenging to model the splittings in a global approach. Some "problematic", yet unsolved cases remain where the internal rotation(s) interact(s) with tunneling motion(s) or with other small amplitude vibrations. Theoretical developments and new codes are needed for those cases.

Some examples on chemical aspects and comparison of barrier heights in systematic investigations on acetates, amides, propynyl group containing molecules, ketones, and dimethylanisoles as well as the inversion tunneling in secondary amines and phenyl formate have given information about how LAMs can help us to understand molecular structures and conformations in nature. Applications of LAMs on conformational determinations are only at their beginning on natural substances and biomolecules where they can provide unique "sensors" of the molecular structure. There is no doubt that many future investigations toward this direction will be performed.

Acknowledgements: Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: This research was funded by the Agence Nationale de la Recherche, ANR-18-CE29-0011.
Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. Balle, TJ, Campbell, EJ, Keenan, MR, Flygare, WH. A new method for observing the rotational spectra of weak molecular complexes: KrHCl. *Chem J Phys* 1979;71:2723. https://doi.org/10.1063/1.438631.
2. Balle, TJ, Flygare, WH. Fabry-Perot cavity pulsed Fourier transform microwave spectrometer with a pulsed nozzle particle source *Rev Sci Instrum* 1981;52:33. https://doi.org/10.1063/1.1136443.

3. Townes, CH, Schawlow, AL. *Microwave spectroscopy*. New York: McGraw-Hill; 1955.
4. Gordy, W, Cook, LR. Microwave molecular spectra. In: Weissberger, A, editor, *Techniques of chemistry, vol. XVIII*, 3rd ed. New York: John Wiley & Sons Inc.; 1984.
5. Wollrab, JE. *Rotational spectra and molecular structure*. New York: Academic Press; 1967.
6. Lister, DG, Macdonald, JN, Owen, NL. *Internal rotation and inversion: an introduction to large amplitude motions in molecules*. New York: Academic Press; 1978.
7. Lin, CC, Swalen, JD. Internal rotation and microwave spectroscopy. *Rev Mod Phys* 1959;31:841. https://doi.org/10.1103/revmodphys.31.841.
8. Bunker, PR, Jensen, P. *Molecular symmetry and spectroscopy*. Ottawa: NRC Research Press; 1998.
9. Pickett, HM. The fitting and prediction of vibration-rotation spectra with spin interactions. *J Mol Spectrosc* 1991;148:371. https://doi.org/10.1016/0022-2852(91)90393-o.
10. Pickett, HM, Poynter, RL, Cohen, EA, Delitsky, ML, Pearson, JC, Müller, HSP. Submillimeter, millimeter, and microwave spectral line catalog. *J Quant Spectrosc Rad Transfer* 1998;60:883. https://doi.org/10.1016/s0022-4073(98)00091-0.
11. Endres, CP, Schlemmer, S, Schilke, P, Stutzki, J, Müller, HSP. The cologne database for molecular spectroscopy, CDMS, in the virtual atomic and molecular data centre, VAMDC. *J Mol Spectrosc* 2016;327:95. https://doi.org/10.1016/j.jms.2016.03.005.
12. Markwick-Kemper, AJ, Remijan, AJ. The splatalogue (spectral line catalogue) and calibase (calibration source database). *Bull Am Astron Soc* 2006;38:130.
13. Lovas, FJ. NIST recommended rest frequencies for observed interstellar molecular microwave transitions National Institute of Standards and Technology; 2009. https://doi.org/10.18434/T4JP4Q.
14. Toyama microwave Atlas for spectroscopists and astronomers; 2017. Available from: http://www.sci.u-toyama.ac.jp/phys/4ken/atlas/.
15. Available from: http://info.ifpan.edu.pl/ kisiel/prospe.htm.
16. Hartwig, H, Dreizler, H. *The microwave spectrum of trans-2,3-dimethyloxirane in torsional excited states. Z Naturforsch* 1996;51a:923.
17. Ohashi, N, Hougen, JT. Approximate selection rules for |K| = 1-0 and 1-1 tunneling-rotation transitions in the methanol dimer. *J Mol Spectrosc* 2000;203:170. https://doi.org/10.1006/jmsp.2000.8153.
18. Hougen, JT, Kleiner, I, Godefroid, M. Selection rules and intensity calculations for a Cs asymmetric top molecule containing a methyl group internal rotor. *J Mol Spectrosc* 1994;163:559. https://doi.org/10.1006/jmsp.1994.1047.
19. Kleiner, I, Hougen, JT. Rho-axis-method Hamiltonian for molecules having one methyl rotor and C1 point-group symmetry at equilibrium. *J Chem Phys* 2003;119:5505. https://doi.org/10.1063/1.1599354.
20. Tudorie, M, Kleiner, I, Hougen, JT, Melandri, S, Sutikdja, LW, Stahl, W. A fitting program for molecules with two inequivalent methyl tops and a plane of symmetry at equilibrium: application to new microwave and millimeter-wave measurements of methyl acetate. *J Mol Spectrosc* 2011;269:211. https://doi.org/10.1016/j.jms.2011.07.005.
21. Groner, P. Effective rotational Hamiltonian for molecules with two periodic large-amplitude motions. *J Chem Phys* 1997;107:4483. https://doi.org/10.1063/1.474810.
22. Ilyushin, VV, Kisiel, Z, Pszczółkowski, L, Mäder, H, Hougen, JT. A new torsion-rotation fitting program for molecules with a sixfold barrier: application to the microwave spectrum of toluene. *J Mol Spectrosc* 2010;259:26. https://doi.org/10.1016/j.jms.2009.10.005.
23. Ilyushin, VV, Hougen, JT. A fitting program for molecules with two equivalent methyl tops and C2v point-group symmetry at equilibrium: application to existing microwave, millimeter, and

sub-millimeter wave measurements of acetone. *J Mol Spectrosc* 2013;289:41. https://doi.org/10.1016/j.jms.2013.05.012.
24. Plusquellic, DF, Suenram, RD, Mate, B, Jensen, JO, Samuels, AC. The conformational structures and dipole moments of ethyl sulfide in the gas phase. *J Chem Phys* 2001;115:3057. https://doi.org/10.1063/1.1385527.
25. Kleiner, I, Hougen, JT. A hybrid program for fitting rotationally resolved spectra of floppy molecules with one large-amplitude rotatory motion and one large-amplitude oscillatory motion. *J Phys Chem* 2015;119:10664. https://doi.org/10.1021/acs.jpca.5b08437.
26. Frisch, MJ, Trucks, GW, Schlegel, HB, Scuseria, GE, Robb, MA, Cheeseman, JR, et al. *Gaussian 16, Revision B.01, Inc.*, Wallingford CT; 2016.
27. Schmidt, MW, Baldridge, KK, Boatz, JA, Elbert, ST, Gordon, MS, Jensen, JJ, et al. General atomic and molecular electronic structure system. *J Comput Chem* 1993;14:1347. https://doi.org/10.1002/jcc.540141112.
28. Grabow, J-U. Fourier transform microwave spectroscopy measurement and instrumentation. In: Quack, M., Merkt, F., editors. *Handbook of high-resolution spectroscopy*, Vol. 2. Chichester, UK: Wiley; 2011. Chapter 1.
29. Grabow, J-U, Palmer, ES, McCarthy, MC, Thaddeus, P. Supersonic-jet cryogenic-resonator coaxially oriented beam-resonator arrangement Fourier transform microwave spectrometer. *Rev Sci Instrum* 2005;093106.
30. Alonso, JL, Pérez, C, Sanz, ME, López, JC, Blanco, S. Seven conformers of L-threonine in the gas phase: a LA-MB-FTMW study. *Phys Chem Chem Phys* 2009;11:617. https://doi.org/10.1039/b810940k.
31. Cocinero, EJ, Lesarri, A, Écija, P, Grabow, J-U, Fernández, JA, Castaño, F. Conformational equilibria in vanillin and ethylvanillin. *Phys Chem Chem Phys* 2010;12:12486. https://doi.org/10.1039/c0cp00585a.
32. Grabow, J-U. *Habilitationschrift*. Hannover University; 1994.
33. Grabow, J-U, Stahl, W, Dreizler, H. A multioctave coaxially oriented beam-resonator arrangement Fourier-transform microwave spectrometer. *Rev Sci Instrum* 1996;67:4072. https://doi.org/10.1063/1.1147553.
34. Avilés Moreno, JR, Demaison, J, Huet, TR. Conformational flexibility in hydrated sugars: the glycolaldehyde–water complex. *J Am Chem Soc* 2006;128:10467. https://doi.org/10.1021/ja062312t.
35. Xu, Y, Jäger, W. Evidence for heavy atom large amplitude motions in RG-cyclopropane van der Waals complexes (RG=Ne, Ar, Kr) from rotation-tunneling spectroscopy. *J Chem Phys* 1997;106:7968. https://doi.org/10.1063/1.473808.
36. Sedo, G, van Wijngaarden, J. Fourier transform microwave spectra of a “new” isomer of OCS-CO2. *J Chem Phys* 2009;131:044303. https://doi.org/10.1063/1.3186756.
37. Leung, HO, Gangwani, D, Grabow, J-U. Nuclear quadrupole hyperfine structure in the microwave spectrum of Ar-N2O. *J Mol Spectrosc* 1997;184:106. https://doi.org/10.1006/jmsp.1997.7293.
38. Batten, RC, Cole, GC, Legon, AC. Rotational spectroscopy of a weak complex of thiirane and ethyne: the identification and properties of a highly nonlinear S H-C hydrogen bond. *J Chem Phys* 2003;119:7903. https://doi.org/10.1063/1.1607318.
39. Hight Walker, AR, Chen, W, Novick, SE, Bean, BD, Marshall, MD. Determination of the structure of HBr OCS. *J Chem Phys* 1995;102:7298. https://doi.org/10.1063/1.469041.
40. Newby, JJ, Serafin, MM, Peebles, RA, Peebles, SA. Rotational spectrum, structure and modeling of the OCS-CS2 van der Waals dimer. *Phys Chem Chem Phys* 2005;7:487. https://doi.org/10.1039/b414897e.

41. Merke, I, Stahl, W, Dreizler, H. A molecular beam Fourier transform microwave spectrometer in the range 26.5 to 40 GHz. Tests of performance and analysis of the D- and 14N-hyperfine structure of methylcyanide-d1. *Z Naturforsch* 1994;49a:490.
42. Nguyen, HVL, Mouhib, H, Stahl, W, Kleiner, I. The microwave spectrum of allyl acetate. *Mol Phys* 2010;108:763. https://doi.org/10.1080/00268971003645354.
43. Pajski, JJ, Logan, MD, Douglass, KO, Brown, GG, Suenram, RD, Dian, BC, et al. Chirped-pulse Fourier Transform microwave spectroscopy: a new technique for rapid identification of chemical agents. *Int J high Speed Electron* 2008;18:31. https://doi.org/10.1142/s0129156408005114.
44. Loru, D, Bermúdez, MA, Sanz, ME. Structure of fenchone by broadband rotational spectroscopy. *J Chem Phys* 2016;145:074311. https://doi.org/10.1063/1.4961018.
45. Schmitz, D, Shubert, VA, Betz, T, Schnell, M. Multi-resonance effects within a single chirp in broadband rotational spectroscopy: the rapid adiabatic passage regime for benzonitrile. *J Mol Spectrosc* 2012;280:77. https://doi.org/10.1016/j.jms.2012.08.001.
46. Neill, JL, Shipman, ST, Alvarez-Valtierra, L, Lesarri, A, Kisiel, Z, Pate, BH. Rotational spectroscopy of iodobenzene and iodobenzene-neon with a direct digital 2–8 GHz chirped-pulse Fourier transform microwave spectrometer. *J Mol Spectrosc* 2011;269:21. https://doi.org/10.1016/j.jms.2011.04.016.
47. Marshall, FE, Gillcrist, DJ, Persinger, TD, Jaeger, S, Hurley, CC, Shreve, NE, et al. The CP-FTMW spectrum of bromoperfluoroacetone. *J Mol Spectrosc* 2016;328:59. https://doi.org/10.1016/j.jms.2016.07.014.
48. Hernandez-Castillo, AO, Abeysekera, C, Hays, BM, Zwier, TS. Broadband multi-resonant strong field coherence breaking as a tool for single isomer microwave spectroscopy. *J Chem Phys* 2016;145:114203. https://doi.org/10.1063/1.4962505.
49. Finneran, IA, Holland, DB, Carroll, PB, Blake, GA. A direct digital synthesis chirped pulse Fourier transform microwave spectrometer. *Rev Sci Instrum* 2013;84;083104. https://doi.org/10.1063/1.4818137.
50. Crabtree, KN, Martin-Drumel, MA, Brown, GG, Gaster, SA, Hall, TM, McCarthy, MC. Microwave spectral taxonomy: a semi-automated combination of chirped-pulse and cavity Fourier-transform microwave spectroscopy. *J Chem Phys* 2016;144:124201. https://doi.org/10.1063/1.4944072.
51. Mata, S, Peña, I, Cabezas, C, López, JC, Alonso, JL. A broadband Fourier-transform microwave spectrometer with laser ablation source: the rotational spectrum of nicotinic acid. *J Mol Spectrosc* 2012;280:91. https://doi.org/10.1016/j.jms.2012.08.004.
52. Uriarte, I, Écija, P, Spada, L, Zabalza, E, Lesarri, A, Basterretxea, FJ, et al. Potential energy surface of fluoroxene: experiment and theory. *Phys Chem Chem Phys* 2016;18:3966. https://doi.org/10.1039/c5cp06522d.
53. Grabow, J-U, Mata, S, Alonso, JL, Peña, I, Blanco, S, López, JC, et al. Rapid probe of the nicotine spectra by high-resolution rotational spectroscopy. *Phys Chem Chem Phys* 2011;13:21063. https://doi.org/10.1039/c1cp22197c.
54. Dempster, SP, Sukhorukov, O, Lei, Q-Y, Jäger, W. Rotational spectrum of 1,1,1-trifluoro-2-butanone using chirped-pulse Fourier transform microwave spectroscopy. *J Chem Phys* 2012;137:174303. https://doi.org/10.1063/1.4762862.
55. Evangelisti, L, Sedo, G, van Wijngaarden, J. Rotational spectrum of 1,1,1-Trifluoro-2-butanone using chirped-pulse Fourier transform microwave spectroscopy. *J Phys Chem* 2011;115:685. https://doi.org/10.1021/jp1089905.
56. Marshall, MD, Leung, HO, Scheetz, BQ, Thaler, JE, Muenter, JS. A chirped pulse Fourier transform microwave study of the refrigerant alternative 2,3,3,3-tetrafluoropropene. *J Mol Spectrosc* 2011;266:37. https://doi.org/10.1016/j.jms.2011.02.005.

57. Stephens, SL, Walker, NR. Determination of nuclear spin-rotation coupling constants in CF3I by chirped-pulse Fourier-transform microwave spectroscopy. *J Mol Spectrosc* 2010;263:27. https://doi.org/10.1016/j.jms.2010.06.007.
58. Grubbs, GSII, Dewberry, CT, Etchison, KC, Kerr, KE, Cooke, SA. A search accelerated correct intensity Fourier transform microwave spectrometer with pulsed laser ablation source. *Rev Sci Instrum* 2007;78:096106. https://doi.org/10.1063/1.2786022.
59. Obenchain, DA, Elliott, AA, Steber, AL, Peebles, RA, Peebles, SA, Wurrey, CJ, et al. Rotational spectrum of three conformers of 3,3-difluoropentane: construction of a 480 MHz bandwidth chirped-pulse Fourier-transform microwave spectrometer. *J Mol Spectrosc* 2010;261:35. https://doi.org/10.1016/j.jms.2010.03.002.
60. Van, V, Stahl, W, Nguyen, HVL. The heavy atom microwave structure of 2-methyltetrahydrofuran. *J Mol Struct* 2016;1123:24. https://doi.org/10.1016/j.molstruc.2016.05.078.
61. Nguyen, HVL. The heavy atom substitution and semi-experimental equilibrium structures of 2-ethylfuran obtained by microwave spectroscopy. *J Mol Struct* 2020;1208:127909. https://doi.org/10.1016/j.molstruc.2020.127909.
62. Graneek, JB, Bailey, WC, Schnell, M. Electron-withdrawing effects on the molecular structure of 2- and 3-nitrobenzonitrile revealed by broadband rotational spectroscopy and their comparison with 4-nitrobenzonitrile. *Phys Chem Chem Phys* 2018;20:22210. https://doi.org/10.1039/c8cp01539b.
63. Mouhib, H. Understanding the structure and dynamic of odorants in the gas phase using a combination of microwave spectroscopy and quantum chemical calculations. *J Phys B At Mol Opt Phys* 2014;47:143001. https://doi.org/10.1088/0953-4075/47/14/143001.
64. Cramer, CJ. *Essentials of computational chemisty: theories and models*, 2nd ed.Chichester Wiley; 2004.
65. Jensen, F. *Introduction to computational chemistry*. Chichester: Wiley; 2007.
66. Ferres, L, Stahl, W, Nguyen, HVL. The molecular structure of phenetole studied by microwave spectroscopy and quantum chemical calculations. *Mol Phys* 2016;114:2788. https://doi.org/10.1080/00268976.2016.1177217.
67. Becke, AD. Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 1993;98:5648. https://doi.org/10.1063/1.464913.
68. Lee, C, Yang, W, Parr, RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 1988;37:785. https://doi.org/10.1103/physrevb.37.785.
69. Møller, C, Plesset, MS. Note on an approximation treatment for many-electron systems. *Phys Rev* 1934;46:618.
70. Nair, KPR, Herbers, S, Grabow, J-U. Structure and methyl torsion of halogenated toluenes: rotational spectrum of 3,4-difluorotoluene. *J Mol Spectrosc* 2019;355:19. https://doi.org/10.1016/j.jms.2018.11.007.
71. Karpfen, A, Choi, CH, Kertesz, M. Single-bond torsional potentials in conjugated systems: a comparison of ab initio and density functional results. *J Phys Chem* 1997;101:7426. https://doi.org/10.1021/jp971606l.
72. Grimme, S, Antony, J, Ehrlich, S, Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 2010;132:154104. https://doi.org/10.1063/1.3382344.
73. Grimme, S, Ehrlich, S, Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* 2011;32:1456. https://doi.org/10.1002/jcc.21759.
74. Uriarte, I, Insausti, A, Cocinero, EJ, Jabri, A, Kleiner, I, Mouhib, H, et al. Competing dispersive interactions: from small energy differences to large structural effects in methyl jasmonate and zingerone. *J Phys Chem Lett* 2018;9:5906. https://doi.org/10.1021/acs.jpclett.8b02339.

75. Nguyen, HVL, Stahl, W. The rotational spectrum of diethyl ketone. *ChemPhysChem* 2011;12:1900. https://doi.org/10.1002/cphc.201001021.
76. Krishnan, R, Pople, JA. Approximate fourth-order perturbation theory of the electron correlation energy. *Int J Quant Chem* 1978;14:91. https://doi.org/10.1002/qua.560140109.
77. Bartlett, RJ, Musial, M. Coupled-cluster theory in quantum chemistry.*Rev Mod Phys* 2007;79:291. https://doi.org/10.1103/revmodphys.79.291.
78. Pople, JA, Head-Gordon, M, Fox, DJ, Raghavachari, K, Curtiss, LA. Gaussian-1 theory: a general procedure for prediction of molecular energies. *J Chem Phys* 1989;90:5622. https://doi.org/10.1063/1.456415.
79. Dunning, THJr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J Chem Phys* 1989;90:1007.
80. Mouhib, H, Van, V, Stahl, W. Sulfur-containing flavors: gas phase structures of dihydro-2-methyl-3-thiophenone. *J Phys Chem* 2013;117:6652. https://doi.org/10.1021/jp4041748.
81. Pérez, C, Muckle, MT, Zaleski, DP, Seifert, NA, Temelso, B, Shields, GC, et al. Structures of cage, prism, and book isomers of water hexamer from broadband rotational spectroscopy. *Science* 2012;336:897. https://doi.org/10.1126/science.1220574.
82. Van, V, Stahl, W, Schwell, M, Nguyen, HVL. Gas-phase conformations of 2-methyl-1,3-dithiolane investigated by microwave spectroscopy. *J Mol Struct* 2018;1156:348. https://doi.org/10.1016/j.molstruc.2017.11.084.
83. Ferres, L, Mouhib, H, Stahl, W, Nguyen, HVL. Methyl internal rotation in the microwave spectrum of o-methyl anisole. *ChemPhysChem* 2017;18:1855. https://doi.org/10.1002/cphc.201700276.
84. Ferres, L, Stahl, W, Nguyen, HVL. Conformational effects on the torsional barriers in m-methylanisole studied by microwave spectroscopy. *J Chem Phys* 2018;148:124304. https://doi.org/10.1063/1.5016273.
85. Ferres, L, Stahl, W, Kleiner, I, Nguyen, HVL. The effect of internal rotation in p-methyl anisole studied by microwave spectroscopy. *J Mol Spectrosc* 2018;343:44. https://doi.org/10.1016/j.jms.2017.09.008.
86. Mouhib, H, Stahl, W. Conformational analysis of green apple flavour: the gas-phase structure of ethyl valerate validated by microwave spectroscopy. *ChemPhysChem* 2012;13:1297. https://doi.org/10.1002/cphc.201100932.
87. Nguyen, HVL, Grabow, J-U. The scent of Maibowle – π electron localization in coumarin from its microwave-determined structure. *ChemPhysChem* 2020;21:1243. https://doi.org/10.1002/cphc.202000234.
88. Kisiel, Z, Desyatnyk, O, Pszczółkowski, L, Charnley, SB, Ehrenfreund, P. Rotational spectra of quinoline and of isoquinoline: spectroscopic constants and electric dipole moments. *J Mol Spectrosc* 2003;217:115. https://doi.org/10.1016/s0022-2852(02)00020-6.
89. Grimme, S, Steinmetz, M. Effects of London dispersion correction in density functional theory on the structures of organic molecules in the gas phase. *Phys Chem Chem Phys* 2013;15:16031. https://doi.org/10.1039/c3cp52293h.
90. Demaison, J, Császár, AG, Margulès, LD, Rudolph, HD. Equilibrium structures of heterocyclic molecules with large principal axis rotations upon isotopic substitution. *J Phys Chem* 2011;115:14078. https://doi.org/10.1021/jp2063595.
91. Kannengießer, R, Klahm, S, Nguyen, HVL, Lüchow, A, Stahl, W. The effects of methyl internal rotation and 14N quadrupole coupling in the microwave spectra of two conformers of N,N-diethylacetamide. *J Chem Phys* 2014;141:204308. https://doi.org/10.1063/1.4901980.
92. Cleeton, CE, Williams, NH. Movies and conduct. *Phys Rev* 1934;45:234. https://doi.org/10.1103/physrev.45.234.

93. Meyer, R, López, JC, Alonso, JL, Melandri, S, Favero, PG, Caminati, W. Pseudorotation pathway and equilibrium structure from the rotational spectrum of jet-cooled tetrahydrofuran. *J Chem Phys* 1999;111:7871. https://doi.org/10.1063/1.480122.
94. Gall, JTA, Thomas, J, Xie, F, Wang, Z, Jäger, W, Xu, Y. Rotational spectroscopy of the methyl glycidate-water complex: conformation and water and methyl rotor tunnelling motions. *Phys Chem Chem Phys* 2017;19:29508. https://doi.org/10.1039/c7cp05464e.
95. Koehler, JS, Dennison, DM. Hindered rotation in methyl alcohol. *Phys Rev* 1940;57:1006. https://doi.org/10.1103/physrev.57.1006.
96. Hughes, RH, Good, WE, Coles, DK. Microwave spectrum of methyl alcohol. *Phys Rev* 1951;84:418. https://doi.org/10.1103/physrev.84.418.
97. Lees, RM, Baker, JG. Torsion-vibration-rotation interactions in methanol. I. Millimeter wave spectrum. *J Chem Phys* 1968;48:5299. https://doi.org/10.1063/1.1668221.
98. De Lucia, FC, Herbst, E, Anderson, T, Helminger, P. The analysis of the rotational spectrum of methanol to microwave accuracy. *J Mol Spectrosc* 1989;134:395. https://doi.org/10.1016/0022-2852(89)90325-1.
99. Xu, L-H, Fisher, J, Lees, RM, Shi, HY, Hougen, JT, Pearson, JC, et al. Torsion-rotation global analysis of the first three torsional states (vt = 0, 1, 2) and terahertz database for methanol. *J Mol Spectrosc* 2008;251:305. https://doi.org/10.1016/j.jms.2008.03.017.
100. Lin, CC, Kilb, RW. Microwave spectrum and internal barrier of acetaldehyde. *J Chem Phys* 1956;24:631. https://doi.org/10.1063/1.1742578.
101. Bauder, A, Günthard, HH. Internal rotation in acetaldehyde. *J Mol Spectrosc* 1976;60:290. https://doi.org/10.1016/0022-2852(76)90133-8.
102. Liang, W, Baker, JG, Herbst, E, Booker, RA, De Lucia, FC. The millimeter-wave spectrum of acetaldehyde in its two lowest torsional states. *J Mol Spectrosc* 1986;120:298. https://doi.org/10.1016/0022-2852(86)90006-8.
103. Maes, H, Wlodarczak, G, Boucher, D, Demaison, J. The millimeter- and submillimeter-wave spectra of acetaldehyde: centrifugal distortion and internal rotation analysis. *Z. Naturforsch* 1987;42a:97.
104. Curl, RF. Microwave spectrum, barrier to internal rotation, and structure of methyl formate. *J Chem Phys* 1959;30:1529. https://doi.org/10.1063/1.1730232.
105. Plummer, GM, Blake, GA, Herbst, E, De Lucia, FC. The millimeter and submillimeter laboratory spectrum of methyl formate in its ground symmetric torsional state. *Astrophys J Suppl* 1984;55:633. https://doi.org/10.1086/190972.
106. Demaison, J, Boucher, D, Dubrulle, A, Van Eijck, BP. Millimeter wave spectrum of methyl formate. *J Mol Spectrosc* 1983;102:260. https://doi.org/10.1016/0022-2852(83)90242-4.
107. Oesterling, LC, Albert, S, De Lucia, FC, Sastry, KVLN, Herbst, E. The millimeter- and submillimeter-wave spectrum of methyl formate (HCOOCH3). *Astrophys J* 1999;521:255. https://doi.org/10.1086/307543.
108. Karakawa, Y, Oka, K, Odashima, H, Takagi, K, Tsunekawa, S. The microwave spectrum of methyl formate (HCOOCH3) in the frequency range from 7 to 200 GHz. *J Mol Spectrosc* 2001;210:196.
109. Van Eijck, BP, Van Ophensden, J, Van Schaik, MMM, Van Zoeren, E. Acetic acid: microwave spectra, internal rotation and substitution structure. *J Mol Spectrosc* 1981;86:465. https://doi.org/10.1016/0022-2852(81)90294-0.
110. Demaison, J, Dubrulle, A, Boucher, D, Burie, J, van Eijck, BP. Millimeter wave spectrum of acetic acid. *J Mol Spectrosc* 1982;94:211. https://doi.org/10.1016/0022-2852(82)90306-x.
111. Tabor, WJ. Microwave spectrum and barrier to internal rotation of acetic acid. *J Chem Phys* 1957;27:974. https://doi.org/10.1063/1.1743896.

112. Krischer, LC, Saegebarth, E. Microwave spectrum of acetic acid, CH3COOH and CD3COOH. *J Chem Phys* 1971;54:4553.
113. Kleiner, I. Spectroscopy of interstellar internal rotors: an important tool for investigating interstellar chemistry. *ACS Earth Space Chem* 2019;3:1812. https://doi.org/10.1021/acsearthspacechem.9b00079.
114. Fischer, E, Botskor, I. The microwave spectrum of trans-ethylamine. *J Mol Spectrosc* 1982;91:116. https://doi.org/10.1016/0022-2852(82)90035-2.
115. Fischer, E, Botskor, I. The microwave spectrum of gauche-ethylamine. *J Mol Spectrosc* 1984;104:226. https://doi.org/10.1016/0022-2852(84)90117-6.
116. Canagaratna, M, Phillips, JA, Ott, ME, Leopold, KR. The nitric acid–water complex: microwave spectrum, structure, and tunneling. *J Phys Chem* 1998;102:1489. https://doi.org/10.1021/jp980033p.
117. Columberg, G, Bauder, A, Heineking, N, Stahl, W, Makarewicz, J. Internal rotation effects and hyperfine structure in the rotational spectrum of a water-carbon dioxide complex. *Mol Phys* 1998;93:215. https://doi.org/10.1080/00268979809482205.
118. Lovas, FJ, Johnson, DR, Buhl, D, Snyder, LE. Millimeter emission lines in Orion A. *Astrophys J* 1976;209:770. https://doi.org/10.1086/154774.
119. Churchwell, E, Winnewisser, WG. Observations of methyl formate in the galactic center. *Astron Astrophys* 1975;45:229.
120. Bell, MB, Matthews, HE, Feldman, PA. Observations of microwave transitions of A-state acetaldehyde in Sgr B2. *Astron Astrophys* 1983;127:420.
121. Matthews, HE, Friberg, P, Irvine, WM. The detection of acetaldehyde in cold dust clouds. *Astron Astrophys* 1985;290:609. https://doi.org/10.1086/163018.
122. Mehringer, DM, Snyder, LE, Miao, Y, Lovas, FJ. Detection and confirmation of interstellar acetic acid. *Astrophys J Lett* 1997;480:71. https://doi.org/10.1086/310612.
123. Lee, CW, Cho, SH, Lee, SM. A spectral line survey from 138.3 to 150.7 GHz toward Orion-KL. *Astrophys J* 2001;551:333. https://doi.org/10.1086/320062.
124. Combes, F, Gerin, M, Wootten, A, Wlodarczak, G, Clausset, F, Encrenaz, PJ. Acetone in interstellar space. *Astron Astrophys* 1987;180:13.
125. Fuchs, GW, Fuchs, U, Giesen, TF, Wyrowski, F. Trans-ethyl methyl ether in space - a new look at a complex molecule in selected hot core regions. *Astron Astrophys* 2005;444:521. https://doi.org/10.1051/0004-6361:20053599.
126. Tercero, B, Kleiner, I, Cercicharo, J, Nguyen, HVL, López, A, Munoz Caro, GM. Discovery of methyl acetate and gauche ethyl formate in Orion. *Astrophys J* 2013;770:L13. https://doi.org/10.1088/2041-8205/770/1/l13.
127. Dreizler, H. Gruppentheoretische Betrachtungen zum Mikrowellenspektrum von Molekülen enthaltend zwei behindert drehbare Methylgruppen mit verschiedenen Kohlenstoffisotopen. *Z. Naturforsch* 1961;16a:1354.
128. Ferres, L, Cheung, J, Stahl, W, Nguyen, HVL. Conformational effect on the large amplitude motions of 3,4-dimethylanisole explored by microwave spectroscopy. *J Phys Chem* 2019;123:3497. https://doi.org/10.1021/acs.jpca.9b00029.
129. Van, V, Stahl, W, Nguyen, HVL. Two equivalent methyl internal rotations in 2,5-dimethylthiophene investigated by microwave spectroscopy. *Phys Chem Chem Phys* 2015;17:32111. https://doi.org/10.1039/c5cp03513a.
130. Sheridan, J, Bossert, W, Bauder, A. Internal rotation of molecules with two inequivalent methyl groups: The microwave spectrum of methyl acetate. *J Mol Spectrosc* 1980;80:1. https://doi.org/10.1016/0022-2852(80)90265-9.

131. Nguyen, HVL, Kleiner, I, Shipman, ST, Mae, Y, Hirose, K, Hatanaka, S, et al. Extension of the measurement, assignment, and fit of the rotational spectrum of the two-top molecule methyl acetate. *J Mol Spectrosc* 2014;299:17. https://doi.org/10.1016/j.jms.2014.03.012.
132. Van, V, Stahl, W, Nguyen, HVL. The structure and torsional dynamics of two methyl groups in 2-acetyl-5-methylfuran as observed by microwave spectroscopy. *ChemPhysChem* 2016;17:3223. https://doi.org/10.1002/cphc.201600757.
133. Nguyen, HVL, Van, V, Stahl, W, Kleiner, I. The effects of two internal rotations in the microwave spectrum of ethyl methyl ketone. *J Chem Phys* 2014;140:214303. https://doi.org/10.1063/1.4878412.
134. Nguyen, HVL, Stahl, W, Kleiner, I. Structure and rotational dynamics of methyl propionate studied by microwave spectroscopy. *Mol Phys* 2012;110:2035. https://doi.org/10.1080/00268976.2012.683884.
135. Tudorie, M, Kleiner, I, Jahn, M, Grabow, J-U, Goubet, M, Pirali, O. Coupled large amplitude motions: a case study of the dimethylbenzaldehyde isomers. *J Phys Chem A* 2013;117:13636. https://doi.org/10.1021/jp407603y.
136. Ferres, L, Truong, K-N, Stahl, W, Nguyen, HVL. Interplay between microwave spectroscopy and X-ray diffraction: the molecular structure and large amplitude motions of 2,3-dimethylanisole. *ChemPhysChem* 2018;19:1781. https://doi.org/10.1002/cphc.201800115.
137. Ferres, L, Stahl, W, Nguyen, HVL. Low torsional barrier challenges in the microwave spectrum of 2,4-dimethylanisole. *J Chem Phys* 2019;151:104310. https://doi.org/10.1063/1.5116304.
138. Herbers, S, Fritz, SM, Mishra, P, Nguyen, HVL, Zwier, TS. Local and global approaches to treat the torsional barriers of 4-methylacetophenone using microwave spectroscopy. *J Chem Phys* 2020;152:074301. https://doi.org/10.1063/1.5142401.
139. Nguyen, HVL, Stahl, W. The microwave spectrum of isopropenyl acetate - an asymmetric molecule with two internal rotors. *J Mol Spectrosc* 2010;264:120. https://doi.org/10.1016/j.jms.2010.10.002.
140. Groner, P, Albert, S, Herbst, E, De Lucia, FC, Lovas, FJ, Drouin, BJ, et al. Acetone: laboratory assignments and predictions through 620 GHz for the vibrational-torsional ground state. *Astrophys J* 2002;142:145. https://doi.org/10.1086/341221.
141. Neustock, W, Guarnieri, A, Demaison, J, Wlodarczak, G. The millimeter and submillimeter-wave spectrum of dimethylether. *Z Naturforsch* 1990;45a:702.
142. Groner, P, Gillies, CW, Gillies, JZ, Zhang, Y, Block, E. Microwave structural studies of organoselenium compounds 1. Microwave spectra, molecular structure, and methyl barrier to internal rotation of dimethyl diselenide. *J Mol Spectrosc* 2004;226:169. https://doi.org/10.1016/j.jms.2004.04.001.
143. Van, V, Bruckhuisen, J, Stahl, W, Ilyushin, V, Nguyen, HVL. The torsional barriers of two equivalent methyl internal rotations in 2,5-dimethylfuran investigated by microwave spectroscopy. *J Mol Spectrosc* 2018;343:121. https://doi.org/10.1016/j.jms.2017.11.007.
144. Jabri, A, Van, V, Nguyen, HVL, Mouhib, H, Kwabia-Tchana, F, Manceron, L, et al. Laboratory microwave, millimeter wave and far-infrared spectra of dimethyl sulfide. *Astron Astrophys* 2016;589:A127. https://doi.org/10.1051/0004-6361/201628074.
145. Merke, I, Lüchow, A, Stahl, W. Internal rotation, quadrupole coupling and structure of (CH3)3SiI studied by microwave spectroscopy and ab-initio calculations. *J Mol Struct* 2006;780-781:295. https://doi.org/10.1016/j.molstruc.2005.07.011.
146. Schnell, M, Hougen, JT, Grabow, J-U. Towards the complete analysis of the rotational spectrum of (CH3)3SnCl. *J Mol Spectrosc* 2008;251:38. https://doi.org/10.1016/j.jms.2008.01.007.

147. Fujitake, M, Kubota, Y, Ohashi, N. Fourier transform microwave spectroscopy of N,N-dimethylacetamide. *J Mol Spectrosc* 2006;236:97. https://doi.org/10.1016/j.jms.2005.12.013.
148. Marstokk, KM, Møllendal, H. Microwave spectrum, conformational equilibrium, intramolecular hydrogen bonding, inversion tunnelling, dipole moments and centrifugal distortion of ethylenediamine. *J Mol Struct* 1978;49:221. https://doi.org/10.1016/0022-2860(78)87259-7.
149. Merke, I, Coudert, LH. Microwave spectrum, tunneling motions, and quadrupole coupling hyperfine structure of ethylene diamine. *J Mol Spectrosc* 2006;237:174. https://doi.org/10.1016/j.jms.2006.03.013.
150. Kasuja, T, Kojima, T. Internal motions of hydrazine. *J Phys Soc Japan* 1963;18:364.
151. Tsunekawa, S, Kojima, T, Hougen, JT. Analysis of the microwave spectrum of hydrazine. *J Mol Spectrosc* 1982;95:133. https://doi.org/10.1016/0022-2852(82)90243-0.
152. Kreglewski, M, Cosléou, J, Wlodarczak, G. Rotational spectrum of hydrazine in the submillimeter range. *J Mol Spectrosc* 2002;216:501.
153. Gulaczyk, I, Pyka, J, Kręglewsk, M. The effective Hamiltonian for the coupling between inversion-torsion states of hydrazine. *J Mol Spectrosc* 2007;241:75. https://doi.org/10.1016/j.jms.2006.11.001.
154. Gulaczyk, I, Kręglewski, M. The symmetric amino-wagging band of hydrazine: assignment and analysis. *J Mol Spectrosc* 2008;249:73. https://doi.org/10.1016/j.jms.2008.02.012.
155. Kleibömer, B, Sutter, DH. The vibrational state dependence of the 14N quadrupole coupling tensor in aniline. A microwave Fourier-transform study combind with semirigid bender calculations. *Z Naturforsch* 1988;43a:561.
156. Ohashi, N, Hougen, JT. Analysis and global fit of tunneling splittings in the K = 0 a-type microwave spectrum of the methanol dimer. *J Mol Spectrosc* 1995;170:493. https://doi.org/10.1006/jmsp.1995.1087.
157. Lovas, FJ, Belov, SP, Tretyakov, MY, Stahl, W, Suenram, RD. The a-type K = 0 microwave spectrum of the methanol dimer. *J Mol Spectrosc* 1995;170:478. https://doi.org/10.1006/jmsp.1995.1086.
158. Wollrab, JE, Laurie, VW. Microwave spectrum of dimethylamine. *J Chem Phys* 1968;48:5058. https://doi.org/10.1063/1.1668177.
159. Penn, RE, Boggs, JE. Rotation-inversion spectrum of methylaminoethane. *J Mol Spectrosc* 1973;47:340. https://doi.org/10.1016/0022-2852(73)90017-9.
160. Nguyen, HVL, Stahl, W. The effects of nitrogen inversion tunneling, methyl internal rotation, and 14N quadrupole coupling observed in the rotational spectrum of diethyl amine. *J Chem Phys* 2011;135:024310.
161. Ferres, L, Mouhib, H, Stahl, W, Schwell, M, Nguyen, HVL. Molecular structure and ring tunneling of phenyl formate as observed by microwave spectroscopy and quantum chemistry. *J Mol Spectrosc* 2017;337:59. https://doi.org/10.1016/j.jms.2017.04.017.
162. Gulaczyk, I, Kręglewski, M. The effective Hamiltonian for the coupling between inversion-torsion states in methylamine. *J Mol Spectrosc* 2009;256:86. https://doi.org/10.1016/j.jms.2009.02.013.
163. Gulaczyk, I, Kręglewski, M, Horneman, V-M. The C-N stretching band of methylamine. *J Mol Spectrosc* 2011;270:70. https://doi.org/10.1016/j.jms.2011.09.003.
164. Gulaczyk, I, Kręglewski, M. Second torsional state of methylamine from high resolution IR spectra. *J Quant Spectrosc Radiat Transf* 2018;217:321. https://doi.org/10.1016/j.jqsrt.2018.06.008.
165. Baughcum, SL, Smith, Z, Wilson, EB, Duerst, RW. Microwave spectroscopic study of malonaldehyde. 3. Vibration-rotation interaction and one-dimensional model for proton tunneling. *J Am Chem Soc* 1984;106:2260. https://doi.org/10.1021/ja00320a007.

166. Chou, Y-C, Hougen, JT. Two-dimensional tunneling Hamiltonian treatment of the microwave spectrum of 2-methylmalonaldehyde. *J Chem Phys* 2006;124:074319. https://doi.org/10.1063/1.2162545.
167. Sanders, ND. Vibration-torsion interaction in the microwave spectrum of internally hydrogen-bonded methylmalonaldehyde. *J Mol Spectrosc* 1981;86:27. https://doi.org/10.1016/0022-2852(81)90103-x.
168. Zhao, Y, Nguyen, HVL, Stahl, W, Hougen, JT. Unusual internal rotation coupling in the microwave spectrum of pinacolone *J Mol Spectrosc* 2015;318:91. https://doi.org/10.1016/j.jms.2015.10.005.
169. Cabezas, C, Varela, M, Caminati, W, Mata, S, López, JC, Alonso, JL. The two conformers of acetanilide unraveled using LA-MB-FTMW spectroscopy. *J Mol Spectrosc* 2011;268:42. https://doi.org/10.1016/j.jms.2011.03.033.
170. Seifert, NA, Finneran, IA, Perez, C, Zaleski, DP, Neill, JL, Steber, AL, et al. AUTOFIT, an automated fitting tool for broadband rotational spectra, and applications to 1-hexanal. *J Mol Spectrosc* 2015;312:13. https://doi.org/10.1016/j.jms.2015.02.003.
171. Drouin, BJ, Pearson, JC, Walters, A, Lattanzi, V. THz measurements of propane. *J Mol Spectrosc* 2006;240:227. https://doi.org/10.1016/j.jms.2006.10.007.
172. Eibl, K, Kannengießer, R, Stahl, W, Nguyen, HVL, Kleiner, I. Low barrier methyl rotation in 3-pentyn-1-ol as observed by microwave spectroscopy. *Mol Phys* 2016;114:3483. https://doi.org/10.1080/00268976.2016.1239034.
173. Nguyen, HVL, Jabri, A, Van, V, Stahl, W. Methyl internal rotation in the microwave spectrum of vinyl acetate. *J Phys Chem* 2014;118:12130. https://doi.org/10.1021/jp5075829.
174. Kannengießer, R, Lach, MJ, Stahl, W, Nguyen, HVL. Acetyl methyl torsion in N-ethylacetamide: a challenge for microwave spectroscopy and quantum chemistry. *ChemPhysChem* 2015;16:1906. https://doi.org/10.1002/cphc.201500087.
175. Jelisavac, D, Cortés-Gómez, DC, Nguyen, HVL, Sutikdja, LW, Stahl, W, Kleiner, I. The microwave spectrum of the trans conformer of ethyl acetate. *J Mol Spectrosc* 2009;257:111. https://doi.org/10.1016/j.jms.2009.07.002.
176. Kleiner, I, Lovas, FJ, Godefroid, M. Microwave spectra of molecules of astrophysical interest. XXIII. Acetaldehyde. *J Phys Chem* 1996;25:1113. https://doi.org/10.1063/1.555983.
177. Kleiner, I, Hougen, JT, Grabow, J-U, Belov, SP, Tretyakov, MY, Cosleou, J. The third and fourth torsional states of acetaldehyde. *J Mol Spectrosc* 1996;179:41. https://doi.org/10.1006/jmsp.1996.0182.
178. Ilyushin, VV, Alekseev, EA, Dyubko, SF, Podnos, SV, Kleiner, I, Margulès, L, et al. The ground and first excited torsional states of acetic acid. *J Mol Spectrosc* 2001;205:286. https://doi.org/10.1006/jmsp.2000.8270.
179. Ilyushin, VV, Alekseev, EA, Dyubko, SF, Kleiner, I. The second torsional state of acetic acid. *J Mol Spectrosc* 2003;220:170. https://doi.org/10.1016/s0022-2852(03)00073-0.
180. Carvajal, M, Margulès, L, Tercero, B, Demyk, K, Kleiner, I, Guillemin, JC, et al. Rotational spectrum of 13C2-methyl formate (HCOO13CH3) and detection of the two 13C-methyl formate in Orion. *Astron Astrophys* 2009;500:1109. https://doi.org/10.1051/0004-6361/200811456.
181. Kannengießer, R, Stahl, W, Nguyen, HVL, Kleiner, I. 14N nuclear quadrupole coupling and methyl internal rotation in N-tert-butylacetamide as observed by microwave spectroscopy. *J Phys Chem* 2016;120:3992. https://doi.org/10.1021/acs.jpca.6b02111.
182. Van, V, Nguyen, T, Stahl, W, Nguyen, HVL, Kleiner, I. Coupled large amplitude motions: the effects of two methyl internal rotations and 14N quadrupole coupling in 4,5-dimethylthiazole investigated by microwave spectroscopy. *J Mol Struct* 2020;1207:127787. https://doi.org/10.1016/j.molstruc.2020.127787.

183. Roucou, A, Kleiner, I, Goubet, M, Bteich, S, Mouret, G, Bocquet, R, et al. Towards the detection of explosive taggants: microwave and millimetre-wave gas-phase spectroscopies of 3-nitrotoluene. *ChemPhysChem* 2018;19:1056. https://doi.org/10.1002/cphc.201701266.
184. Nair, KPR, Jahn, MK, Lesarri, A, Ilyushin, VV, Grabow, J-U. Six-fold-symmetry internal rotation in toluenes: the low barrier challenge of 2,6-and 3,5-difluorotoluene. *Phys Chem Chem Phys* 2015;17:26463.
185. Ilyushin, V. Millimeter wave spectrum of nitromethane. *J Mol Spectrosc* 2018;345:64. https://doi.org/10.1016/j.jms.2017.12.005.
186. Belloche, A, Meshcheryakov, AA, Garrod, RT, Ilyushin, VV, Alekseev, EA, Motiyenko, RA, et al. Rotational spectroscopy, tentative interstellar detection, and chemical modeling of N-methylformamide. *Astron Astrophys* 2017;601:A49. https://doi.org/10.1051/0004-6361/201629724.
187. Groner, P, Albert, S, Herbst, E, De Lucia, FC. Dimethyl ether: laboratory assignments and predictions through 600 GHz. *Astrophys J* 1998;500:1059. https://doi.org/10.1086/305757.
188. Endres, CP, Drouin, BJ, Pearson, JC, Müller, HSP, Lewen, F, Schlemmer, S, et al. Dimethyl ether: laboratory spectra up to 2.1 THz. *Astron Astrophys* 2009;504:635. https://doi.org/10.1051/0004-6361/200912409.
189. Groner, P, Winnewisser, M, Medvedev, IR, De Lucia, FC, Herbst, E, Sastry, KVLN. The millimeter- and submillimeter-wave spectrum of methyl carbamate [CH3OC(:O)NH2]. *Astrophys J Suppl Ser* 2007;169:28. https://doi.org/10.1086/511133.
190. Kisiel, Z, Pszczólkowski, L, Bialkowska-Jaworska, E, Charnley, SB. The millimeter wave rotational spectrum of pyruvic acid. *J Mol Spectrosc* 2007;241:220. https://doi.org/10.1016/j.jms.2006.12.011.
191. Maeda, A, De Lucia, FC, Herbst, E. Submillimeter-wave spectra of H12COOCH3 and H13COOCH3 in excited CH3 torsional states. *J Mol Spectrosc* 2008;251:293. https://doi.org/10.1016/j.jms.2008.03.014.
192. Krasnicki, A, Pszczólkowski, L, Kisiel, Z. Analysis of the rotational spectrum of pyruvonitrile up to 324 GHz. *J Mol Spectrosc* 2010;260:57.
193. Zhao, Y, Stahl, W, Nguyen, HVL. Ketone physics - structure, conformations, and dynamics of methyl isobutyl ketone explored by microwave spectroscopy and quantum chemical calculations. *Chem Phys Lett* 2012;545:9. https://doi.org/10.1016/j.cplett.2012.07.009.
194. Armieieva, IA, Ilyushin, VV, Alekseev, EA, Dorovskaya, OA, Margulès, L, Motiyenko, RA. Millimeter wave spectroscopy of the ground, first and second excited torsional states of acetone. *Radio Phys Radio Astron* 2016;21:37. https://doi.org/10.15407/rpra21.01.037.
195. Ohashi, N, Hougen, JT, Suenram, R, Lovas, FJ, Kawashima, Y, Fujitake, M, et al. Analysis and fit of the Fourier-transform microwave spectrum of the two-top molecule N-methylacetamide. *J Mol Spectrosc* 2004;227:28. https://doi.org/10.1016/j.jms.2004.04.017.
196. Van, V. *RWTH Aachen University*. Dissertation; 2017.
197. Sutikdja, LW, Stahl, W, Sironneau, V, Nguyen, HVL, Kleiner, I. Structure and internal dynamics of n-propyl acetate studied by microwave spectroscopy and quantum chemistry. *Chem Phys Lett* 2016;663:145. https://doi.org/10.1016/j.cplett.2016.09.062.
198. Attig, T, Sutikdja, LW, Kannengießer, R, Kleiner, I, Stahl, W. The microwave spectrum of n-butyl acetate. *J Mol Spectrosc* 2013;284–285:8. https://doi.org/10.1016/j.jms.2013.02.003.
199. Attig, T, Kannengießer, R, Kleiner, I, Stahl, W. Conformational analysis of n-pentyl acetate using microwave spectroscopy. *J Mol Spectrosc* 2013;290:24. https://doi.org/10.1016/j.jms.2013.07.001.
200. Attig, T, Kannengießer, R, Kleiner, I, Stahl, W. The microwave spectrum of n-hexyl acetate and structural aspects of n-alkyl acetates. *J Mol Spectrosc* 2014;298:47. https://doi.org/10.1016/j.jms.2014.02.008.

201. Mouhib, H, Jelisavac, D, Stahl, W, Wang, R, Kalf, I, Englert, U. The conformation of odorants in different states of aggregation: a joint venture in microwave spectroscopy and X-ray diffraction. *ChemPhysChem* 2011;12:761. https://doi.org/10.1002/cphc.201000986.
202. Sutikdja, LW, Jelisavac, D, Stahl, W, Kleiner, I. Structural studies on banana oil, isoamyl acetate, by means of microwave spectroscopy and quantum chemical calculations. *Mol Phys* 2012;110:2883. https://doi.org/10.1080/00268976.2012.679630.
203. Jabri, A, Van, V, Nguyen, HVL, Stahl, W, Kleiner, I. Probing the methyl torsional barriers of the E and Z isomers of butadienyl acetate by microwave spectroscopy . *ChemPhysChem* 2016;17:2660. https://doi.org/10.1002/cphc.201600265.
204. Ferres, L. *RWTH Aachen University*. Dissertation; 2019.
205. Pitzer, RM. The barrier to internal rotation in ethane. *Acc Chem Res* 1983;16:207. https://doi.org/10.1021/ar00090a004.
206. Herbers, S, Wachsmuth, D, Obenchain, DA, Grabow, J-U. Rotational characterization of methyl methacrylate: Internal dynamics and structure determination. *J Mol Spectrosc* 2018;343:96. https://doi.org/10.1016/j.jms.2017.10.006.
207. Ilyushin, VV, Alekseev, EA, Dyubko, SF, Kleiner, I, Hougen, JT. Ground and first excited torsional states of acetamide. *J Mol Spectrosc* 2004;227:115. https://doi.org/10.1016/j.jms.2004.05.014.
208. Kannengießer, R. *RWTH Aachen University*. Dissertation; 2017.
209. Turner, PH, Corkill, MJ, Cox, AP. Microwave spectra and structures of cis- and trans-methyl nitrite. Methyl barrier in trans-methyl nitrite. *J Phys Chem* 1979;83:1473. https://doi.org/10.1021/j100474a023.
210. Herbers, S, Nair, KPR, Nguyen, HVL, Grabow, J-U. The structure and low-barrier methyl torsion of 3-fluorotoluene. *Spectrochim Acta A* 2020;242:118709.
211. Bunker, PR, Longuet-Higgins, HC. The infra-red spectrum of dimethylacetylene and the torsional barrier. *Proc R Soc A* 1964;280:340.
212. Nakagawa, J, Hayashi, M, Endo, Y, Saito, S, Hirota, E. Microwave spectrum and internal rotation of 2-butyne-1,1,1-d3 (dimethylacetylene), CH3C≡CCD3. *J Chem Phys* 1984;80:5922. https://doi.org/10.1063/1.446697.
213. Nakagawa, J, Yamada, K, Bestera, M, Winnerwisser, G. Millimeter-wave spectrum and internal rotation of 1-silylpropyne, CH3C≡CSiH3. *J Mol Spectrosc* 1985;110:74. https://doi.org/10.1016/0022-2852(85)90213-9.
214. Ilyushin, V, Rizzato, R, Evangelisti, L, Feng, G, Maris, A, Melandri, S, et al. Almost free methyl top internal rotation: Rotational spectrum of 2-butynoic acid. *J Mol Spectrosc* 2011;267:186. https://doi.org/10.1016/j.jms.2011.03.028.
215. Hensel, KD, Gerry, MCL. Microwave spectrum of tetrolyl fluoride. *J Chem Soc Faraday Trans* 1994;90:3023. https://doi.org/10.1039/ft9949003023.
216. Stolwijk, VM, van Eijck, BP. Microwave spectrum and barrier to internal rotation of 1-chloro-2-butyne. *J Mol Spectrosc* 1987;124:92. https://doi.org/10.1016/0022-2852(87)90124-x.
217. Subramanian, R, Novick, SE, Bohn, RK. Torsional analysis of 2-butynol. *J Mol Spectrosc* 2003;222:57. https://doi.org/10.1016/s0022-2852(03)00170-x.
218. Eibl, K. *RWTH Aachen University*. Dissertation; 2019.
219. Eibl, K, Stahl, W, Kleiner, I, Nguyen, HVL. Conformational effect on the almost free internal rotation in 4-hexyn-3-ol studied by microwave spectroscopy and quantum chemistry. *J Chem Phys* 2018;149:144306. https://doi.org/10.1063/1.5044542.
220. Andresen, M, Kleiner, I, Schwell, M, Stahl, W, Nguyen, HVL. Acetyl methyl torsion in Pentan-2-one as observed by microwave spectroscopy. *J Phys Chem* 2018;122:7071. https://doi.org/10.1021/acs.jpca.8b06183.

221. Andresen, M, Kleiner, I, Schwell, M, Stahl, W, Nguyen, HVL. Sensing the molecular structures of hexan-2-one by internal rotation and microwave spectroscopy. *ChemPhysChem* 2019;20:2063.
222. Andresen, M, Kleiner, I, Schwell, M, Stahl, W, Nguyen, HVL. Microwave spectrum and internal rotations of heptan-2-one: a pheromone in the gas phase. *J Phys Chem* 2020;124:1353. https://doi.org/10.1021/acs.jpca.9b11577.
223. Andresen, M, Schöngen, D, Kleiner, I, Schwell, M, Stahl, W, Nguyen, HVL. Internal rotation of the acetyl methyl group in methyl alkyl ketones: the microwave spectrum of octan-2-one. *ChemPhysChem* 2020;21. https://doi.org/10.1002/cphc.202000522. In press.
224. Zhao, Y, Jin, J, Stahl, W, Kleiner, I. The microwave spectrum of methyl neopentyl ketone. *J Mol Spectrosc* 2012;281:4. https://doi.org/10.1016/j.jms.2012.09.002.
225. Tulimat, L, Mouhib, H, Kleiner, I, Stahl, W. The microwave spectrum of allyl acetone. *J Mol Spectroc* 2015;312:46. https://doi.org/10.1016/j.jms.2015.03.006.
226. Nguyen, HVL, Mouhib, H, Klahm, S, Stahl, W, Kleiner, I. A touch of lavender: gas-phase structure and dynamics of the monoterpene linalool validated by microwave spectroscopy. *Phys Chem Chem Phys* 2013;15:10012. https://doi.org/10.1039/c3cp50271f.
227. Reinhold, B, Finneran, IA, Shipman, ST. Room temperature chirped-pulse Fourier transform microwave spectroscopy of anisole. *J Mol Spectrosc* 2011;270:89. https://doi.org/10.1016/j.jms.2011.10.002.

Iwona Gulaczyk and Marek Kręglewski

3 Floppy molecules—their internal dynamics, spectroscopy and applications

Abstract: Floppy molecules can be defined as molecules performing large amplitude vibrations (LAVs). There are different types of LAVs among which the most common are inversion and internal rotation. Molecules with LAVs have been of great interest for a very long time since their dynamic, geometry and molecular spectra were very often considered as a challenge. In the review, we present an outline of the history and development of various theoretical approaches concerning molecules with LAVs. Different types of LAVs are described with the emphasis on inversion tunneling (wagging) and internal rotation (torsion). Furthermore, strategies for building explicit and effective Hamiltonians are given and explained in detail using a hydrazine molecule, which is an exemplary molecule performing three LAVs—two inversions and one internal rotation. Since floppy molecules play a significant role in numerous areas as chemistry, pharmacy, astrophysics, biology, agriculture etc., we also provide an overview of their applications.

Keywords: a hydrazine molecule, effective Hamiltonian, explicit Hamiltonian, high resolution spectroscopy, large amplitude vibrations

3.1 Introduction

The study of rotational and vibration–rotational spectra of isolated molecules in a gas phase is of fundamental importance for determining the geometry of the molecule in various vibrational states, the shape of the multidimensional potential surface, the magnitude of the dipole moment, detection at distant locations etc. Moreover, these analyses help to generate accurate models for *ab initio* quantum mechanical calculations, and many of the results have also interdisciplinary character. For instance, precise spectroscopic standards are necessary for astrophysical research, automatic control of atmospheric pollutants, or stimulation of chemical reactions through laser beams. Both theoretical and experimental studies on the spectroscopy of molecules with large amplitude vibrations (LAVs) have special significance in this research. It is very important to find the appropriate methods of describing these molecules for kinetics of chemical reactions, the study of molecules

This article has previously been published in the journal Physical Sciences Reviews. Please cite as: Gulaczyk, I., Kręglewski. M. Floppy molecules—their internal dynamics, spectroscopy and applications *Physical Sciences Reviews* [Online] 2021, 6. DOI: 10.1515/psr-2020-0035

https://doi.org/10.1515/9783110678215-003

vibrationally excited by high temperature, determining the structure of weakly bound Van der Waals complexes, or the structure of molecules in excited electronic states. However, none of the above-mentioned goals can be achieved without an appropriate theory that will allow a correct interpretation of the vibration–rotational spectra of the molecules of interest.

For more than 70 years molecular spectroscopy along with other methods has been able to characterize the structures and shapes of molecules, both rigid and nonrigid. The nonrigid molecules play a significant role in such varied areas as biological activity, energy transfer, or chemical reactivity. The development of theory [group theoretical methods and potential energy surfaces (PES)] was essential to understand the unique and often very complex spectra of many so called floppy molecules.

Classically, molecules can be treated as almost rigid structures rotating in space. Their vibrational motion causes structural deformations, but these displacements are rather small in comparison to the bond lengths and angles in a molecule. In such a case, one can assume that vibrations of a molecule are only weakly coupled with its rotation, thus the standard way of treating the rotation–vibration interaction is by using an effective rotational Hamiltonian, which is obtained by a perturbation theory. For molecules that are "floppy", because perform LAVs, nuclear displacements caused by LAV can be the size of linear dimensions of the equilibrium structure, thus vibration and rotation cannot be treated separately, and the classic approach for the rotation–vibration data fails. Instead, special formalisms or methods have to be implemented to describe dynamics of such molecules. For instance, the LAVs have to be described by curvilinear coordinates, which make difficult to generate quantum-mechanical Hamiltonian. The couplings between LAVs and rotation are very strong, what is reflected in a strong coupling between the rotational and vibrational moment of inertia. Although small amplitude vibrations theory is not much older than the LAV theory, the latter was developed much later and for the first time was described in a systematic way in a work by Hougen, Bunker and Johns (HBJ) [1]. This method will be described in more details in the next paragraph.

3.2 Large amplitude vibrations (LAVs)

In the first part of this chapter, we shall introduce the reader to a handful of information about different theoretical approaches for treating molecules with LAV and show how these theories and models developed over the years. In the second part, examples of various types of large amplitude motions are presented with particular emphasis on inversion and internal rotation.

First studies of systematic theoretical description of LAV were introduced for triatomic molecules. There are two main methods that have been extensively applied to molecules in which there are several LAVs. These are: (i) the HBJ method

[1] and its extensions to the multivibration case; (ii) the slow tunneling formalism developed by Hougen [2], which became the effective tool for the interpretation of high-resolution spectra of many molecules.

3.2.1 Theories involving LAVs

First studies on vibration–rotation spectroscopy of polyatomic molecules were published in the 1930s of the previous century. In 1936, Wilson and Howard [3] presented a general vibration–rotation theory of molecules on the assumption that the amplitude of vibrations is small. Soon Darling and Dennison derived a correct quantum-mechanical approach [4], and in 1951, Nielsen [5] developed formulas that enabled phenomenological interpretation of vibration–rotational spectra. Nevertheless, all those theories failed when it came to analysis of spectra of molecules with LAVs. There were two main reasons for that. First, the assumption that vibrations are not coupled strongly with rotation, and second, that expansion of a potential function can be given as a power series. In molecules with LAVs, vibrations cannot be treated separately from rotation and power series expansions converge very slowly. Simultaneously with a general theory on vibrations and rotation, studies appeared where rotation was coupled with LAVs, but other vibrations were neglected. Classic examples of that were studies of inversion in ammonia [6], internal rotation in ethane [7] and methanol [8]. Of course, the approximation of the potential energy with a power series for this type of internal motion of molecules was subject to very large errors.

Nevertheless, for many years, no general theory has been proposed to analyze rotation–vibration spectra of molecules that perform large amplitude motions. It did not mean that the spectra of such molecules were not analyzed at all. Spectroscopists used a Wilson's method [9] to interpret vibrational spectra. For analysis of bands corresponding to excited LAVs or analysis of rotational spectra, the approximation of semi rigid molecule was used. This method assumed that a molecule performs either a LAV only or rotates. The other vibrations were neglected. It turned out that this assumption was not correct, since LAV is strongly coupled to overall rotation of a molecule, much more than other vibrations, and separately it is neither possible to explain a spectrum of LAV nor rotational spectrum of a molecule performing LAV. The first work devoted to the vibration–rotation problem in a symmetric quasilinear three-atom molecule was the one by Thorson and Nakagawa [10]. A disadvantage of their method was that that the vibration–rotational Hamiltonian did not include terms responsible for rotation around axes perpendicular to the molecular skeleton and did not take into account the dependence of a reduced mass for a bending mode on a coordinate of this vibration. It should be underlined at this point that studies of three-atom molecules performing a large amplitude bending vibration led later to the HBJ approach in 1970 (HBJ) [1]. Another attempt was made by Freed and Lombardi [11], who developed the complete Hamiltonian for any three-atom molecule with LAV. In their approach a

molecular coordinate system coinciding with the principal axes was applied instead of a system that minimizes vibrational angular momentum. Consequently, some of the Coriolis terms were too high. Kirtman [12], [13] analyzed couplings between harmonic vibrations and internal rotation as well as with rotation of a whole molecule implementing a perturbation theory. Thanks to the all mentioned above approaches, it was possible to explain many spectral features of the molecules with LAVs. Unfortunately, all those methods considered just one LAV. A general LAV theory was presented in 1968 by Meyer and Günthard [14]. In the Hamiltonian, each vibration could be defined as a LAV. Two molecular coordination systems were presented, i.e., principal axis system (PAS) and internal motion axis system (IMAS). In the first one, the deviation moments were minimized, whereas in the second Coriolis interaction terms. This model was too general, thus its application to LAVs was confined. Similar objections were directed to the models of Quade [15] and Louck [16].

All the above examples of different approaches showed clearly that it was extremely difficult to create the vibration–rotational theory for molecules performing LAVs, which were general and precise enough at the same time. In 1970, HBJ presented a model based on a three-atomic molecule with a large amplitude bending mode [1]. This work was a sort of compilation and an expansion of earlier theories by Hougen [17], [18], [19] and Bunker [20, [21] concerning internal rotation in dimethylacetylene. A fundamental assumption of this theory is based on an appropriate choice of coordinates in a molecular coordination system and a transformation of these coordinates into laboratory system. In other words, this theory applied two types of internal coordinates: curvilinear coordinates to describe LAVs of a molecule and linear coordinates for other vibrations, and the coupling between rotation and LAVs was treated explicitly. Thus, the vibration–rotational Hamiltonian in the HBJ approach treated explicitly LAVs, SAVs and rotation. Eq. (3.1) shows the transformation of the system of laboratory, r_i, to molecular coordinates:

$$\boldsymbol{R}_{\mathbf{i}} = \boldsymbol{R} + S^{-1}(\chi, \Theta, \phi)[\boldsymbol{a}_{\mathbf{i}}(\rho) + d_{\mathbf{i}}] \tag{3.1}$$

where $R_{\mathbf{i}}$ is represented by a 3 × 1 column vector containing the laboratory-fixed Cartesian coordinates of the *i*th atom. $\boldsymbol{R}$ is a 3 × 1 column vector containing the laboratory-fixed Cartesian coordinates of the center of mass of the nuclei. $\boldsymbol{S}(\chi, \theta, \phi)$ represents a 3 × 3 direction cosine matrix, which transforms from laboratory-fixed (X, Y, Z) to molecule-fixed (x, y, z) Cartesian coordinates and which is given as a function of rotational variables (the Eulerian angles χ, θ, ϕ). $a_{\mathbf{i}}(\rho)$ is a 3 × 1 column vector that contains the molecule-fixed Cartesian coordinates of a reference position of the *i*th atom, and $\mathbf{d_i}$ is a 3 × 1 column vector containing the molecule-fixed Cartesian coordinates of the vibrational displacement of the *i*th atom from its reference position and is used to describe the small amplitude vibrations only.

Since the above equation contains seven redundant coordinates on the right-hand side, seven constraint equations must be given. This can be done in different

ways. Some constraints will be more appropriate than others because they will eliminate or minimize certain quantities in the kinetic energy operator. The problem of choosing the coordinate system and minimizing the coupling in the Hamiltonian has been the subject of many considerations up to now. The study by Hougen et al. [1], apart from Eckart's conditions, which situate a center of molecular system in the center of a molecule, proposed the following conditions:

a) Eckart condition that locates the origin of the molecular system in the center of mass

$$\sum_i m_i \boldsymbol{d_i} = \mathbf{0} \tag{3.2}$$

b) Eckart's condition that eliminates a coupling between SAVs and rotation

$$\sum_i m_i \boldsymbol{a_i} \times \boldsymbol{d_i} = \mathbf{0} \tag{3.3}$$

c) Sayvetz's condition that minimizes a coupling between SAVs and LAVs

$$\sum_i m_i \, \partial \, \boldsymbol{a_i}/\partial\rho \cdot \boldsymbol{d_i} = 0 \tag{3.4}$$

These conditions were generated and fulfilled for a triatomic molecule. Eckart's conditions define the orientation of the axes of the molecular coordinate system such that the Coriolis terms for the selected internal coordinates become zero at some point. Thus, Eckart's axes do not eliminate Coriolis terms in the entire range of variables, only locally.

Since LAV generates an angular momentum, it is desirable that the coupling with the rotational angular momentum is minimized. This can be obtained through modifying the position of the molecule in a molecular axis system in a function of ρ using the Eckart condition for LAV:

$$\sum_i m_i \boldsymbol{a_i} \times \partial \boldsymbol{a_i}/\partial\rho = 0 \tag{3.5}$$

The HBJ theory proposed the rotation–vibrational Hamiltonian for a triatomic molecule by expanding its terms into a Taylor series of linear normal coordinates representing small vibrations. A similar Hamiltonian for the NH_3 molecule, in which the inversion was treated as LAV, was derived by Papoušek [22].

In the HBJ theory, an important role plays the rigid model, which is obtained by neglecting small amplitude vibrations and which is treated as a zero-order approximation. The rigid model can be then improved by applying the perturbation theory, yielding the effective nonrigid Hamiltonian. The effective Hamiltonian makes it possible to calculate the bending-rotational energy levels in a particular vibrational state.

The HBJ method could be then successfully applied to different types of molecules: linear, bent, symmetric tops, asymmetric tops and to various types of LAVs [23], [24], [25], [26], [27], [28], [29], [30], [31], [32]. Different versions of a nonrigid

bending Hamiltonian were presented for a triatomic molecule by Bunker et al. [33], [34], [35], [36], [37], [38], [39] and by Jensen [40], and a very similar Hamiltonian for inversion proposed Špirko et al. [28], [41], [42]. Applicability of the HBJ method was extensively tested on other molecules with LAVs, more complex than just three-atom systems and of lower symmetry. For instance, the generalized HBJ method was applied to describe a coupling between internal rotation and inversion in methylamine [43]. The vibration–inversion–torsion–rotational Hamiltonian was derived, taking into account two LAVs: inversion and torsion (rigid and semirigid model). Redundant coordinates were removed using conditions presented by Eqs. (3.2)–(3.4). Due to the C_{3v} symmetry of the methyl group, the elements of the inverted tensor of moment of inertia were functions of the inversion variable only. This made it possible to solve the problem in two stages: to solve a pure inversion problem by the numerical Numerov–Cooley integration method, and then the full equation by the variational method. Rovibrational transitions were calculated, and a two-dimensional potential function was determined by fitting to experimental data. Strong interactions between LAVs were observed too.

In one of other works a rotational structure was calculated in excited electronic state of formaldehyde [44]. The HBJ approach was also applied to molecules with two LAVs, which belonged to quasisymmetric top molecules (QSTs) [45], [46], [47]. Dynamic of molecules, which have quasilinear skeletons with one or two tops of C_{3v} symmetry attached, has not been understandable for a very long time. The correct Hamiltonian for such molecules was presented for the first time in 1983 [48]. The subject of the study was a disilane ether molecule, $(SiH_3)_2O$, with three LAVs: two torsions and a bending. It turned out that spectra of QST molecules can be interpreted properly only when couplings between rotation, internal rotation and bending vibration are treated explicitly. The Hamiltonian developed could be applied successfully to molecules with a bent skeleton and with one or two symmetric tops and to quasilinear molecules [49], [50]. The approach used in those works was similar to the one for methylamine [43].

Another proposal of the nonrigid bending Hamiltonian, in which both large and small amplitude vibrations are described by curvilinear coordinates, was given by Quade [15], [51], [52]. He proposed a transformation to a special coordinate system (*R* and *T* transformation) eliminating Coriolis coupling terms and obtaining an effective one-dimensional Hamiltonian for LAVs.

Szalay [53] showed how to minimize the coupling between LAV and rotation for a single LAV. Asymmetric vibrations of the ring (small amplitude vibrations) were neglected. Therefore, only one dynamic variable (LAV) was used to calculate the elements of the tensor of inertia, freezing all other vibrations (a rigid model) and eliminating all Coriolis terms in the entire range of this variable.

An interesting, completely numerical approach to the problem of LAVs was proposed by Pyka et al. (PFM theory) [54]. All vibrations were described by curvilinear coordinates. On the basis of adiabatic separation of small and LAVs, an effective

method for solving the vibration–rotational problem in a disilocyclobutane molecule was presented. Coordinate transformations enabling the elimination of kinetic coupling between large and small vibrations were also discussed in this study.

The problem of minimizing the coupling between rotation and vibration in a triatomic molecule (HDO) was raised in the study by Makarewicz and Łodyga [55], where the self-consisted field method SCF [56], [57] was used to the approximate separation of stretching vibrations, bending and rotation. Unlike other studies, the axes that eliminate Coriolis terms (called internal axes) were not assumed to be optimal axes. To describe the molecule, the Jacobi coordinates [58] and a moving coordinate system were used, and a direction of this system was determined by the following condition:

$$\vartheta_2 = -A\vartheta \tag{3.6}$$

where ϑ is an internal axis and a parameter A describes orientation of a molecular system (varies from 0 to 1). For $A = 0$, the axis is parallel to the bond r_1, whereas for $A = 1$ it is parallel to r_2, and for $A = 0.36$, for internal axis system. Results showed that the molecular coordinate system that eliminates the Coriolis coupling (internal axis system) minimizes the errors of calculated transition energies. The molecular axes have a different orientation in each excited state of stretching vibrations.

3.2.2 Types of large amplitude vibrations

There are various types of Large Amplitude Vibrations (LAVs). In general, LAVs are vibrations, which cannot be described approximately with harmonic potentials. The PESs of these vibrations often have two or more minima and distances between energy levels are usually very small. The LAVs are usually of the lowest-frequency vibrations in a molecule. In MW spectroscopy, LAVs can be displayed through the splitting of rotational lines, in IR and Raman spectra by hot bands appearance as well as many combination bands, and in low energy Raman spectra or far infrared (FIR) by well-developed absorption bands corresponding to excited LAVs. Some information about LAVs can be also provided by NMR analysis (especially bending modes of aromatic rings and pseudorotation) or by electron diffraction.

How do we define “floppy” molecules? For many molecules, such as ammonia NH_3 or the water dimer $(H_2O)_2$, it is not possible to describe their dynamics on the basis of the equilibrium structure, because they undergo various so called tunneling motions such as the umbrella motion in NH_3 or internal rotation of methyl group in methylamine or in ethyl radical. Such molecules with large amplitude motions are called nonrigid or simply “floppy”. Thus, “floppy” molecules are those which perform one or more LAVs. Such molecules cannot be described by harmonic potentials, their PESs have at least two minima or are very broad and LAVs are strongly coupled with rotation of a molecule. It should be underlined that the analysis of high-resolution spectra of nonrigid molecules is often a challenge. The

spectra are usually very tangled and complicated since the rovibronic transitions are additionally split by tunneling motions. In order to fit such spectra, very often the Hamiltonian has to be either adapted or invented.

3.2.2.1 Inversion

In both, theory and experiment, the most common LAV is inversion and the best known molecular inversion is the "umbrella-like" bending vibration in ammonia molecule, NH_3 [6], [59], [60], [61]. Ammonia molecule has a pyramidal shape, where LAV pushes nitrogen atom through the plane formed by three hydrogen atoms. A potential of inversion is usually described by a symmetric or asymmetric function with two minima, which means that a molecule can occur in two configurations. First models did not take into account the change of the reduced mass during inversion and used the value of the reduced mass calculated for the equilibrium configuration instead [62]. Originally, the inversion coordinate in ammonia was defined as a distance between nitrogen atom from a plane of three hydrogen atoms, but later a curvilinear coordinate was introduced. This coordinate described the inversion motion as a simultaneous change of three HNH angles [61], [62], [63] or as a change of an angle between 3-fold axis and one of the bonds [25], [27], [28]. Figure 3.1 presents the ammonia molecule with the inversion coordinate defined using the last definition.

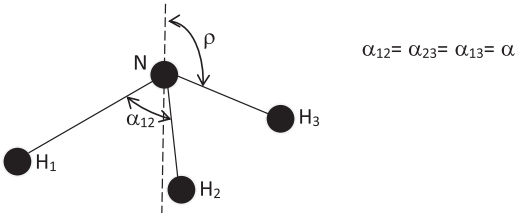


Figure 3.1: Inversion coordinate, ρ, in ammonia.

There are many other molecules performing inversion like NH_2D, ND_3 or H_2CO in the excited electronic state [64]. In the latter, the inversion coordinate is defined as an angle between the HCH plane and a CO bond, what makes the HCH angle constant during inversion motion [24]. Since the inversion barrier in ammonia is relatively high, the splitting of the lowest energy levels is very small in comparison to the formaldehyde molecule, where this splitting is comparable with higher energy levels. Inversion occurs also in more complex molecules like methylamine [43], [65], methylhydrazine [66] or hydrazine [67]. As far as a hydrazine molecule is concerned, inversion problem will be presented and discussed in more detail in the following parts.

Proton tunneling may be similar to inversion-like vibration. Malonaldehyde, 2-methylmalonaldehyde or tropolone are exemplary molecules, where this kind of

LAV appears. Figure 3.2 illustrates a 2-methylmalonaldehyde molecule with a definition of its proton tunneling coordinate [68].

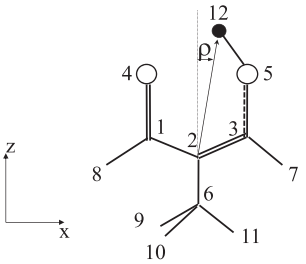


Figure 3.2: Definition of a proton tunneling coordinate, ρ, in 2-methylmalonaldehyde.

3.2.2.2 Internal rotation (torsion)

Internal rotation (torsion) can be described as a rotation of one part of the molecule (top) around a single bond with respect to the rest of the molecule (frame). The barriers hindering this rotation can vary within a very wide range, from zero (free rotation) to a few thousand cm^{-1}. A hindered internal rotation is a torsional vibration. Molecules performing internal rotation can be classified according to a symmetry of a rotor, which can be symmetric or asymmetric, and to a number of tops. For one top, a potential function describing internal rotation can be presented by the following expression:

$$V_n(\tau) = \sum\nolimits_m V_m/2[1 - \cos(mn\tau)] \tag{3.7}$$

where τ is a torsional angle, V_m are potential constants and n denotes multiplicity of the top axis. The expansion of the function depends on symmetry of the top, whether it is symmetric or asymmetric.

One can distinguish molecules with one top, which is symmetric along with symmetric frame (e.g., CH_3CH_3), symmetric top with asymmetric frame (e.g., CH_3OH, CH_3CHO) or asymmetric top and frame (e.g., $C_6H_5CH = CH_2$). Of course, there are molecules with two tops, for instance $(CH_3)_2X$-type molecules like $(CH_3)_2S$, $(CH_3)_2CH_2$ or even with more tops: e.g., $(CH_3)_3N$, $(CH_3)_2NSiH_3$. The most known and developed theories and methods for spectral analysis are these for molecules with a single top. Spectra analyses of molecules with more tops are much more complex, since torsional motions are usually strongly coupled. There were many review papers devoted to internal rotation [69], [70], [71]. We would like to mention a very interesting work by Tsuboi et al. [72], where in order to analyze a torsional band of ethylamine, a system of two coupled tops had to be implemented. This approach allowed to determine successfully an energy difference between trans and gauche conformers. Flaud et al. [73] focused on H_2O_2, several torsional bands centers were determined using the torsional Hamiltonian: $B_{YY}J_Y^2 + V_2(\gamma)$, where V_2 is a potential given by Eq. (3.7) and y is a

dihedral angle between two H–O–O planes. Another work [74] presented an interesting problem of three internal rotations in a $CH_2(OH)COOCH_3$ molecule. The properties of three-dimensional system were derived on the basis of the results for one- and two-dimensional systems calculated by a method proposed by Meyer [75].

There are numerous molecules performing both LAVs: inversion and torsion. One of such molecules is 5-methyltropolone, in which proton tunneling occurs along with internal rotation of a methyl top (Figure 3.3). Other examples of molecules performing both inversion and torsion motions are hydrazine and methylamine. In hydrazine, N_2H_4, there are three large amplitude motions altogether: two inversion motions of two equivalent amino groups, NH_2, and a torsion around the N–N axis. Methylamine, CH_3NH_2, has two LAVs, an inversion of the amino group and a torsion of a methyl group.

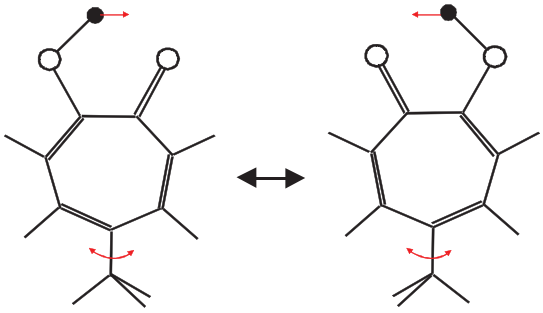


Figure 3.3: Proton tunneling and internal rotation in 5-methyltropolone.

3.2.2.3 Bending vibration of quasilinear molecules

A bending vibration of quasilinear molecules (e.g., CH_2, HCNO) is another type of LAV. The bending vibration of quasilinear molecules is usually described by a cylindrical potential with a minimum for a nonlinear configuration. This type of LAV possesses two degrees of freedom, i.e., it is described by two perpendicular Cartesian coordinates, and is doubly degenerate (Figure 3.4a). They can be transformed to radial coordinates, α and φ, where α describes a bending angle, whereas φ a free rotation of a molecule around its axis (Figure 3.4b). There are molecules which, apart from bending vibration, perform also other vibrations, especially the internal rotation of the CH_3 group. They are called QST with a general formula MH_3NCX. In case of these molecules, the main problem was to determine whether the four-atom MNCX skeleton was linear or bent on the nitrogen atom. It turned out that due to the low stiffness of the skeleton, these molecules cannot be considered neither as symmetric top molecules (which requires a linear skeleton), nor as asymmetric top molecules (with a permanently bent skeleton). Another group of molecules with low-frequency bending vibration of a heavy skeleton are molecules with two symmetric tops, for instance $(SiH_3)_2O$ [48], [76], or quasilinear polyatomic molecules like C_3O_2 [77], [78].

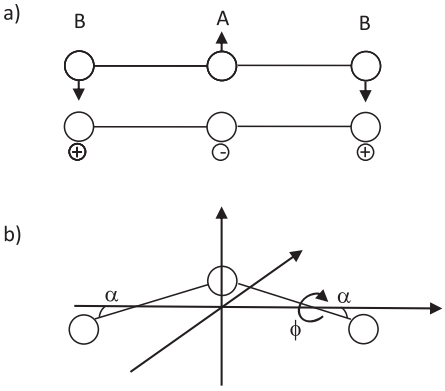


Figure 3.4: Two types of coordinates in degenerate bending vibration of linear molecule AB_2: a) linear Cartesian coordinates and b) curvilinear radial coordinates.

3.2.2.4 Ring puckering

Apart from inversion and torsion, which are the most common, there are also other types of LAVs. One of them is ring puckering, which is defined as an out-of plane bending vibration appearing in many ring molecules with small number of atoms (four-, five- or six-membered rings). Ring puckering produces two forms of a molecule and is described by symmetric or asymmetric potential with two minima. Figure 3.5 presents ring puckering motions in cyclobutane and cyclopentane. This type of vibration occurs in four-membered ring molecules, in five-membered with one double bond in a ring or in six-membered with two double bonds in a ring. In 1984, Mills [79] showed that a separation of LAVs from other vibrations is not always a good approximation. In the following years, several papers appeared that were devoted to studying couplings between different vibrations in ring molecules. For instance, Egawa et al. [80], [81] determined a two-dimensional potential function taking into account a coupling between a ring-bending vibration and a ring-scissoring. Tecklenberg and Laane [82] studied ring-bending and ring-twisting in asymmetric six-membered rings, whereas Laane analyzed four- and five-ring molecules [83]. Study of bending vibration in a molecule $(CH_2)_3NH$ [84], in which a hydrogen atom next to nitrogen can be either in axial or equatorial position, showed

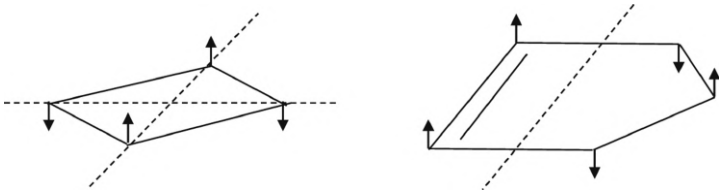


Figure 3.5: Ring puckering in cyclobutane and cyclopentane.

that the same FIR data can be interpreted using a potential with two minima and of small asymmetry, or with potential with one minimum but of a high asymmetry. A conclusion can be drawn that as far as molecules performing LAVs are concerned, one should be careful with potential calculations based on vibrational data only.

3.2.2.5 Pseudorotation in a five-membered ring

In 1947, Kilpatrick et al. [85] studied the thermodynamic properties of cyclopentane and in order to explain its abnormally high entropy, they had to assume a new type of vibration. It was an out-of-plane bending vibration of the ring in which the atoms moved the way that the phase of the bending vibration rotated around the ring. This type of motion, resulting from the superposition of two vibrations in which the atoms' movements are perpendicular to the plane of the ring, has been called pseudorotation. The vibration components are shown in Figure 3.6. Torsion in a five-membered ring makes the A-axis of the ring bending to connect sequentially different pairs of atoms. Pseudorotation can be free (potenial equals zero) or hindered.

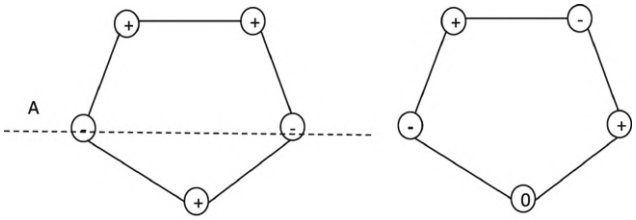


Figure 3.6: *Z*-axis shifts of the ring atoms for two out-of-plane vibrations R_1 and R_2.

3.2.2.6 Berry pseudorotation

LAVs may appear also in molecules of MF_5 type with trigonal bipyramid structure of D_{3h} symmetry, e.g., PF_5. This vibration is based on the intramolecular exchange of axial and equatorial fluorine atoms (Figure 3.7). The possibility of this exchange was first noticed by Berry [86], [87], and vibration is called Berry pseudorotation. Figure 3.7 presents a schematic view of a Berry pseudorotation. Since atoms 5 and 6, according to the numbering adopted in Figure 3.7, can tilt both in the *x* and *y* directions, the vibration is doubly degenerate, and can be described using the same formalism as for pseudorotation.

In this review, we present our experience with one "floppy" molecule: hydrazine. In the following paragraphs, we will focus therefore on the explicit and effective Hamiltonians built for this molecule. Before we come to this, we would like to remind briefly permutation–inversion (PI) group theory, which is essential for high-resolution molecular spectroscopy.

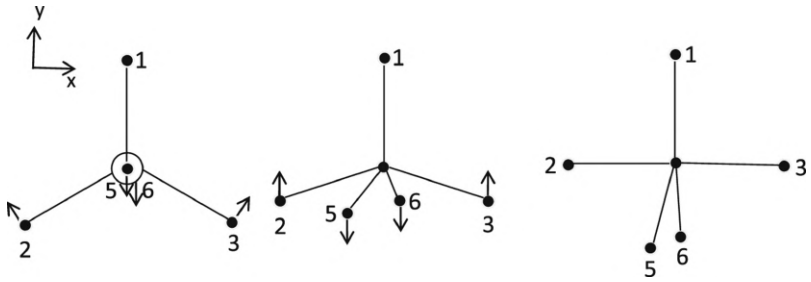


Figure 3.7: A schematic view of the equatorial-axial exchange during a Berry pseudorotation in a XY_5 type of molecule.

3.3 Permutation–inversion group theory

For spectra assignment of nonrigid molecules and for their selection rules development, it is necessary to apply a group theory. A group theory is used in quantum mechanics to determine selection rules, to assign symmetry to energy levels, or to calculate the overall shape of tunneling splitting in molecules with the LAVs. The size of tunneling splittings depends on the barrier height of a particular LAV, which may be calculated from *ab initio* methods. Having the barrier heights calculated, one can predict the LAV-vibration–rotation levels and finally the spectrum of a molecule.

There are two group theories used in high-resolution molecular spectroscopy: the point group theory and the PI group theory. The molecular point group symmetry is used for rigid molecules with one equilibrium configuration only, where no tunneling is observed between configurations. Thanks to a group theory, spectra can be analyzed by labeling the energy levels with irreducible representations of the corresponding symmetry group [88], [89], [90].

For nonrigid molecules, a point group symmetry is not sufficient, since such molecules have more than one equilibrium configuration. Instead, PI group is applied to describe their symmetries, thus the complete nuclear permutation inversion (CNPI) group is created [91], [92]. The invariance group of the molecular Hamiltonian is a basis on which the PI group theory was developed. In the PI groups, the symmetry elements consists of the permutations of identical nuclei and inversion in the center of mass [90].

Let us recall quickly how the PI groups work. For this, we take the ammonia molecule as an example. As already mentioned in a previous paragraph, the ammonia molecule performs so called umbrella motion. Its equilibrium structure is described by C_{3v} point-group symmetry, but the inversion motion transition state is described by D_{3h} symmetry. If the symmetry of ammonia was to be confined to the point-group symmetry, it would be problematic by which symmetry group its energy levels should be named, C_{3v} or D_{3h}. A PI symmetry group provides a solution for this situation and helps to label the energy levels explicitly. Thus, a PI group for

NH_3 consists of 12 symmetry elements: E, (12), (13), (23), (123), (132), E*, (12)*, (13)*, (23)*, (123)* and (132)*. E is the identity element, which does not change anything, symmetry operation (12) denotes the replacement of hydrogen 1 by hydrogen 2, similarly symmetry operations (13) or (23). The operation (123) is the cyclic permutation of three atoms and should be understood in the following way: hydrogen 1 is replaced by hydrogen 2, whereas hydrogen 2 is replaced by hydrogen 3, and finally hydrogen 3 is replaced by hydrogen 1. The asterisk describes the inversion in the center of mass. In ammonia, there are six permutation operations and six PI operations (all the operations with an asterisk).

In 1963, Longuet-Higgins introduced the Molecular Symmetry (MS) group as a group of permutation and permutation-inversion (PI) operations of all particles and spins, subject to the condition of "feasibility" [91]. The concept of the MS (or PI) group was presented using the examples of the following molecules: CH_3BF_2 [93], $B(CH_3)_3$, CH_3CH_3 [94], [95] and NH_2NH_2 [96]. Before the PI group was invented, many molecules with internal rotation and other LAVs had been studied like NH_3, H_2O_2, CH_3OH, CH_3NO_2 [71 and the references therein]. What is interesting that Longuet-Higgins did not mention any of these works in his approach. The first who connected the PI group with one of these studies was Watson [97], [98], [99].

Feasibility is essential for PI group theory considerations [90], [91]. It may happen that some group operations are not feasible due to the resolution of certain molecular spectra and therefore they are not important in their analysis. Let us explain this fact using again ammonia molecule along with phosphine, NH_3 and PH_3, respectively. Both molecules are described by C_{3v} molecular symmetry and CNPI group of PH_3 is similar to NH_3. It was already mentioned that in order to describe the symmetry of ammonia properly one needs to use the CNPI group. It consists of 12 elements and six irreducible representations ($A_1^{'}, A_1^{''}, A_2^{'}, A_2^{''}, E^{'}, E^{''}$). For the phosphine molecule, there are only three irreducible representations instead of six ($A_1^{'} + A_2^{''}, A_2^{''} + A_2^{'}, E^{'} + E^{''}$). This difference can be explained by the fact that for phosphine tunneling between two frameworks was not resolved experimentally and all the symmetry operations responsible for inversion motion can be left out and treated as unfeasible.

In some cases, the CNPI group is the same as the MS group, for instance in water molecule (a rigid molecule). All the CNPI group operations are feasible. The MS group is a subgroup of the CNPI group, which includes only feasible operations. The unfeasible operations are essential in molecules that tunnel through energy barrier. The point group and the CNPI group may be isomorphic. Isomorphism means that the elements of one group are mapped onto the elements of another group in such a way that the multiplication laws of both groups are not infringed. Examples of isomorphic point groups are: D_6, C_{6v}, D_{3h} and D_{3d}. Although these point groups are isomorphic, their symmetry operators do not affect the functions of the vibrational and rotational coordinates in the same way.

Since only one particular nonrigid molecule is considered in this review in detail, which is a hydrazine (N_2H_4) molecule, in further discussions we will focus only

on the symmetry of this molecule. If a reader is interested more in history and wide applications of the PI group theory, we recommend an article by Groner on MS group for LAVs [100], in which not only historical context is very interesting and widely presented, but also nomenclature and notation is fully described as well as an alternate labeling for molecular group is introduced.

3.4 Rovibrational Hamiltonian for a floppy molecule

If a full rovibrational problem, which takes into account small amplitude vibrations, LAVs and rotation, can be reduced to a smaller system, the obtained Hamiltonian is called effective. Thus, if SAVs are removed, we gain the effective Hamiltonian for LAVs and rotation. The examples of such reduction are studies presented in literature studies [33], [34], [35], [36], [37], [38], [39], [40], where for three-atom molecules, the original HBJ Hamiltonian was reduced to an effective rotational-bending Hamiltonian including terms from small vibrations (nonrigid model). If we go on with this reduction, the effective rotational Hamiltonian can be obtained, although it is not always possible due to strong couplings between LAVs and rotation. In order to get the effective Hamiltonian for LAVs, a separation of coordinates in the Schrödinger equation should be done. In the mentioned studies, the following methods for separation of variables can be distinguished: Van-Vleck's transformation (used in the HBJ approach, in its extensions and Quade's method [15]), adiabatic method (applied in PFM approach [54]) and self-consistent field method [55], [56]. As a result of the reduction, small vibration parameters appear in the effective Hamiltonian. Thus, the obtained Hamiltonian is effective for a LAV in each vibrational state of other vibrations in a form of a parametric equation that is easy to apply to the analysis of experimental data or theoretical calculations.

If a molecule with small vibrations only is considered and undergone the reduction, then in effect, the effective rotational Hamiltonian of Watson is obtained [101]. This Hamiltonian (Watsonian) contains higher powers of J, and its coefficients depend on parameters of small vibrations. For higher terms of the Watsonian, this relation is not explicitly derived because is too complicated. However, the Hamiltonian can be used to fit the experimental data by treating the Hamiltonian constants as the parameters of the fit. Obviously, the information about small vibrations is contained in the obtained parameters, but cannot be extracted.

A similar idea for solving a rovibrational problem was used by Hougen for a hydrazine molecule [2]. It was shown there how to obtain an effective rotational Hamiltonian for molecules performing LAVs. In the theory of slow tunneling through barriers, the full wave function is built from local functions, which overlap and give the tunneling parameters. The effective parameters that are obtained by fitting to spectroscopic data depend on the parameters of LAVs, but this information cannot be extracted explicitly. They only determine the splittings in the

spectrum. It should be mentioned at this point that Groner [102] succeeded in getting the dependence of such parameters on the parameters of LAVs, but only for a simple case with one LAV. Although there is still a possibility of extending it to two LAVs.

First time, a tunneling formalism was applied by Dalton [103]. It was used to analyze a MW spectrum of PF_5, where a Berry pseudorotation occurs. However, Hougen's formalism became the basis for spectra interpretation of many systems, like dimers in a gas phase, Van der Waals complexes and also molecules performing LAVs. The approach was applied to H_2O_2 [104], methylamine CH_3NH_2 [105] and CH_3NHD [106]. A formalism derived for methylamine was adapted to acetaldehyde, in which inversion was blocked [107]. An analysis of $(HF)_2$ [108] was performed assuming that LAVs take place in a plane. An extension of the internal axis method was proposed by Hougen in case of a tunneling over high barriers. This theory was used for water dimer [109], [110], ammonia dimer [111] as well as methanol-water system [112]. The latter is pretty complex, since two molecules are bonded via hydrogen bond, around which they can rotate, whereas a donor can be exchanged, i.e., once it can be methanol and then water. In addition, an internal rotation can take place in methanol, thus there are from 5 to 7 LAMs possible in this system.

The Hougen's tunneling formalism was extensively used in analyses of many spectra of complex systems. The main advantage of the formalism is that the problem can be solved easily and a relatively large number of fitting parameters is obtained, which describe the spectral splittings as a result of tunneling between configurations. This makes a spectrum analysis very precise. As a disadvantage, one can take a situation, where the minima corresponding to equivalent configurations are not deep enough. In addition the fitting parameters have no clear physical meaning.

In general, the first step to develop rovibrational Hamiltonian (either explicit or effective Hamiltonian) for a floppy molecule, is to describe its symmetry using the PI group. As mentioned in Section 3.3, there are two group theories used in high-resolution molecular spectroscopy, i.e., the point group theory and the PI group theory. The point group symmetry is used for rigid molecules, i.e., the molecules with one equilibrium configuration and is extensively described in many spectroscopic text books [88], [89], [90], [113]. As for the PI group, it is used to describe a symmetry of nonrigid molecules since the point group symmetry cannot be applied to molecules with more than one equilibrium configuration. In the PI groups, the symmetry elements consist of the permutations of identical nuclei with or without inversion [90]. Once internal rotation or inversion is feasible in a molecule, the PI theory must be applied. Table 3.1 presents all the required steps that are needed to build a rovibrational Hamiltonian.

After the symmetry of a molecule has been determined, in the second step of building a rovibrational Hamiltonian of both types (explicit and effective), small and large coordinates must be defined. It is known that it is not correct to use linear coordinates for description of LAVs. No single linear coordinate can correctly describe

Table 3.1: Instructions for derivation of explicit and effective Hamiltonians.

Explicit rovibrational Hamiltonian	Effective Hamiltonian
Permutation–inversion group of a molecule	
Definition of small and large amplitude coordinates	
Differences between linear and curvilinear coordinates	
Transformations of coordinates in the PI group	Transformations of configurations in the PI group
Hamiltonian in vibrational and rotational coordinates	Effective group-theoretical Hamiltonian
Symmetrized rovibrational basis functions	Symmetrized linear combinations of configurations
Calculation of the Hamiltonian matrix elements	
Labels (assignment of quantum numbers) on eigenvalues	
Results: geometry and potential function	Results: rovibrational and tunneling parameters

large changes in a configuration of a molecule (such as inversion or internal rotation) without involving large changes in the values of other linear coordinates at the same time. On the other hand, it is much easier to develop a quantum-mechanical Hamiltonian using linear coordinates, since the elements of metric tensor are not the functions of coordinates. It is not a case with curvilinear coordinates. Thus, one should use a minimum number of curvilinear coordinates for LAVs description, and other vibrations should be described with linear coordinates.

Speaking technically, a set of mathematical equations describes the correspondence between coordinates in the laboratory fixed Cartesian axis system and molecular coordinates. The equations define the molecule-fixed coordinates, their transformation properties under the symmetry operations are clear, the symmetry of basis set functions can be determined, and finally rovibrational Hamiltonian operators with appropriate symmetry can be created.

The subsequent three operations in developing rovibrational Hamiltonian are different for explicit and effective Hamiltonian. As far as explicit Hamiltonian is concerned, transformation of coordinates in the PI group is conducted whereas for effective Hamiltonian transformation of configurations. In effect, either Hamiltonian in vibrational and rotational coordinates is obtained, if it is developed explicitly, or effective group-theoretical Hamiltonian. Then the symmetrized rovibrational basis functions and linear combinations of configurations are built for explicit and effective Hamiltonians, respectively. In the next step matrix elements are calculated for both types of Hamiltonians, and the eigenvalues are properly labeled. The final results are different for explicit and effective Hamiltonians, since for the first one geometry and potential function is an explicit outcome, whereas for effective Hamiltonian rovibrational and tunneling parameters.

It is worth mentioning at this point that Jon Hougen presented similar strategies and recipes for advanced applications of PI groups to molecules with large amplitude motions in a very accessible and pedagogical way [114]. In this article, there are clear instructions divided into particular steps for the PI operations applications, or for determining the correct PI group. The Hougen's strategies concern a rotational problem only.

Summarizing, there are different ways to analyze rovibrational structure of molecules performing LAVs of various type like inversion, torsion or ring puckering. Some of them start from potential surfaces for LAVs, the other apply effective Hamiltonians, where tunneling splittings play a main role. Whatever method is chosen, one should solve the following problems. First of all, a symmetry of a molecule with LAVs should be defined by the PI molecular group. Then, the internal coordinates should be chosen and transformed in the symmetry group. Hamiltonian should be derived in the chosen coordinates and afterward the Hamiltonian matrix elements are calculated in a symmetrized basis set. As for the methods starting from geometry and potential surface for LAVs, internal dynamic of molecules are better described, whereas fitting results to experimental data are worse.

In the following sections, we would like to present the application of the steps gathered in Table 3.1 to one floppy molecule, hydrazine.

3.5 Hydrazine molecule

The hydrazine molecule, N_2H_4, performs three LAVs: inversion (an umbrella motion) of the amino group, $-NH_2$, at each end of the molecule and an internal rotation (torsion) of the two amino groups about the N–N bond. As Table 3.1 suggests, in order to develop either explicit or effective Hamiltonian for hydrazine, one should find a PI group for a molecule. Thus, the equilibrium structure of hydrazine [115] belongs to the point group C_2, which contains only two symmetry operations. The PI group for hydrazine evolved gradually from a group of order 8 [96], [116], containing all nuclear permutation operations which do not destroy the original $N_{[3]}H_{[1]}H_{[2]}$ and $N_{[4]}H_{[5]}H_{[6]}$ amino groupings (numbering as in Figure 3.8) in the molecule, through order 16 and up to 32. Longuet-Higgins introduced a group of order 16 for hydrazine [91], containing the eight permutations and the eight PI operations which do not destroy the original amino groupings. Finally, a group of order 32 was introduced by Papoušek et al. [117] and Merer and Watson [118] what was necessary to describe properly symmetry properties of rotational and vibrational functions. It is this double group of the Longuet-Higgins PI group $G_{16}^{(2)}$ that is used for hydrazine. The double PI group is generated by the operations a, b, c and d, where d operator is just responsible for a double group of the Longuet-Higgins PI group G_{16}. The PI group symmetry generators of the $G_{16}^{(2)}$ group are defined by the following PI operations:

$$\begin{aligned} \boldsymbol{a} &= (34)(1526)^{*} \\ \boldsymbol{b} &= (34)(15)(26) \\ \boldsymbol{c} &= E^{*} \\ \boldsymbol{d} &= \text{double group} \end{aligned} \tag{3.8}$$

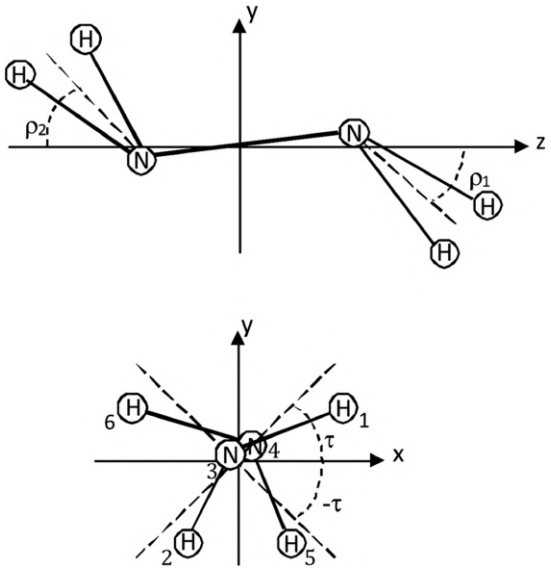


Figure 3.8: A schematic view of hydrazine in the molecular axis system with large amplitude motion coordinates: two inversion coordinates, ρ_1 and ρ_2, and an internal rotation coordinate, τ.

Concluding, the 16-element group is a subgroup of $G_{16}^{(2)}$, the pure permutation group G_8 is a subgroup of G_{16} and the equilibrium geometry point group C_2 is a subgroup of G_8. The character table of the group $G_{16}^{(2)}$ is presented in Table 3.2. In a ground electronic state, the rotation–inversion–torsion functions belong to one of 10 first species and some rotation or inversion–torsion functions belong to the last four species.

There are three large amplitude coordinates in hydrazine, i.e., ρ_1, ρ_2 and τ. The first two describe the inversion motions of two amino groups and the third one describes internal rotation of the two amino groups around the N–N bond. Figure 3.8 presents a schematic view of hydrazine with definitions of these three LAVs. Inversion is presented as a motion around a local mass center of the NH_2 group. The upper view shows a location of the molecular coordination system for N_2H_4 as well as a definition of inversion coordinates, ρ_1 and ρ_2. These curvilinear coordinates are defined as the angles between the HNH angle bisector and the axis passing through the local centers of mass of both amino groups. The lower view of Figure 3.8 presents an equilibrium configuration of a molecule along with

Table 3.2: The character table of a $G_{16}^{(2)}$ group.

	e	d	a^2 d	a^2 d	c cd	a^2c a^2cd	b a^2b	bd a^2bd	bc a^2bc	a^2bc bcd	a a^3 ad a^3d	ac a^3c acd a^3cd	ab a^3b abd a^3bd	abc a^3bc abcd a^3bcd	Nuclear spin weights
A_{1g}^{+}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6
A_{2u}^{+}	1	1	1	1	1	1	−1	−1	−1	−1	1	1	−1	−1	36
A_{1u}^{-}	1	1	1	1	−1	−1	1	1	−1	−1	1	−1	1	−1	45
A_{2g}^{-}	1	1	1	1	−1	−1	−1	−1	1	1	1	−1	−1	1	3
B_{1g}^{+}	1	1	1	1	1	1	1	1	1	1	−1	−1	−1	−1	45
B_{2u}^{+}	1	1	1	1	1	1	−1	−1	−1	−1	−1	−1	1	1	3
B_{1u}^{-}	1	1	1	1	−1	−1	1	1	−1	−1	−1	1	−1	1	6
B_{2g}^{-}	1	1	1	1	−1	−1	−1	−1	1	1	−1	1	1	−1	36
E^{+}	2	2	−2	−2	2	−2	0	0	0	0	0	0	0	0	27
E^{-}	2	2	−2	−2	−2	2	0	0	0	0	0	0	0	0	27
E_1	2	−2	2	−2	0	0	2	−2	0	0	0	0	0	0	–
E_2	2	−2	2	−2	0	0	−2	2	0	0	0	0	0	0	–
E_g	2	−2	−2	2	0	0	0	0	2	−2	0	0	0	0	–
E_u	2	−2	−2	2	0	0	0	0	−2	2	0	0	0	0	–

a definition of the torsional coordinate, τ, which is an angle between x axis of a molecular system and a plane of a straightened amino group NH_2 ($\rho_1 = \rho_2 = 0$). This torsional coordinate is measured counterclockwise to the upper right hydrogen atom. The angle τ is a half of a torsional angle. The molecular axis system shown in Figure 3.8 is chosen assuming the following issues: its origin is in the center of mass of the molecule, the z axis of the system is collinear with the axis passing through the centers of mass of two amino groups, atoms 1 and 5 have a positive value of coordinate x for $\tau = 0$, and finally, for $\rho_1 = \rho_2 = 0$, the atoms 1 and 6 have positive values of coordinate y while atoms 2 and 5 are negative, on condition that τ has small positive value.

In order to define a set of vibration–rotation coordinates for hydrazine in terms of the laboratory-fixed Cartesian coordinates of the component N and H atoms, the extension of Eq. (3.1) is considered:

$$\boldsymbol{R_i} = \boldsymbol{R} + \boldsymbol{S}^{-1}(\chi, \theta, \varphi)[\boldsymbol{a_i}(\tau, \rho_1, \rho_2) + \boldsymbol{d_i}] i = 1, \ldots, 6 \tag{3.9}$$

where the reference configuration $a_i(\tau, \rho_1, \rho_2)$ is a function of three curvilinear coordinates.

So far, instructions presented above are the same for constructing both explicit and effective Hamiltonian of hydrazine. The next step is going to be different for explicit and effective Hamiltonians, so they will be described separately.

3.5.1 Explicit rovibrational Hamiltonian for hydrazine

For the explicit Hamiltonian, the molecular coordinates are transformed using the PI symmetry group. Table 3.3 presents transformations of rovibrational coordinates.

Table 3.3: Transformations of rovibrational coordinates for hydrazine within the PI group $G_{16}^{(2)}$.

Transformations of rotational and LAV coordinates				
E	***a***	***b***	***c***	***d***
χ	$\chi + \pi/2$	$-\chi$	$-\chi$	$\chi + \pi$
θ	θ	$\pi - \theta$	$\pi - \theta$	θ
φ	φ	$\varphi + \pi$	$\varphi + \pi$	φ
τ	$-\tau + \pi/2$	τ	$-\tau + \pi$	$\tau + \pi$
ρ_1	ρ_2	$-\rho_2$	$-\rho_1$	ρ_1
ρ_2	$-\rho_1$	$-\rho_1$	$-\rho_2$	ρ_2

The rotation–inversion–torsion Hamiltonian for hydrazine, where small amplitude vibrations are neglected, can be written as:

$$H_{\mathrm{rit}} = \frac{1}{2}\mu^{\frac{1}{4}}\sum_{\alpha\beta} J_\alpha \mu_{\alpha\beta}\mu^{-\frac{1}{2}} J_\beta \mu^{\frac{1}{4}} + V(\tau,\rho_1,\rho_2)$$
$$\alpha,\beta = x,y,z,\tau,\rho_1,\rho_2 \tag{3.10}$$

J_x, J_y, J_z are components of the total angular momentum operator and

$$J_\eta = -\frac{i\hbar\partial}{\partial\eta}, \text{ for } \eta = \tau,\rho_1,\rho_2 \tag{3.11}$$

are operators of the torsion and inversion momenta. $\mu_{\alpha\beta}$ are elements of the inverse of the moment of inertia tensor defined below in Eq. (3.17).

V(τ, ρ_1, ρ_2) is a three-dimensional inversion–torsion potential function, which is written in the form:

$$V(\tau,\rho_1,\rho_2) = V_i(\rho_1) + V_i(\rho_2) + V_{ii}(\rho_1,\rho_2) + V_t(\tau) + V_{it}(\tau,\rho_1,\rho_2) \tag{3.12}$$

Each term of this function is expanded into a Fourier series:

$$V_i(\rho_k) = a_0 + a_1\cos\rho_k + a_2\cos 2\rho_k + a_3\cos 3\rho_k + a_4\cos 4\rho_k \; k = 1,2 \tag{3.13}$$

$$V_{ii}(\rho_1,\rho_2) = V_{22}\cos 2\rho_1\cos 2\rho_2 \tag{3.14}$$

$$V_t(\tau) = V_4\cos 4\tau + V_8\cos 8\tau \tag{3.15}$$

$$V_{it}(\tau,\rho_1,\rho_2) = V_{222}\sin 2\rho_1 \sin 2\rho_2 \cos 2\tau \tag{3.16}$$

$V_i(\rho_k)$ is a periodic function with two symmetric minima, which describes "single" inversion, i.e., the inversion motion of one amino group in a molecule (Eq. [3.13]). Eq. (3.14) presents coupling between two inversions, $V_{ii}(\rho_1,\rho_2)$ The "pure" torsion is described by a periodic function $V_t(\tau)$ in Eq. (3.15), and finally, a function $V_{it}(\tau,\rho_1,\rho_2)$ expresses coupling between torsion and two inversion motions, respectively (Eq. [3.16]). The form of the components of the potential energy is a result of the symmetry properties of trigonometric functions within $G_{16}^{(2)}$ group [118].

In order to obtain the rotation–torsion energy levels and the wavefunctions, one has to solve the Hamiltonian in Eq. (3.10). It is not trivial, since the elements of the tensor $\boldsymbol{\mu} = \boldsymbol{I}^{-1}$ are functions of large amplitude coordinates, τ,ρ_1,ρ_2 The tensor $\boldsymbol{\mu}$ is given by Eq. (3.17), in which three parts of the matrix can be distinguished. The first part stands for rotation and is characterized by elements I_{xx}, I_{xy} ... up to I_{zz}. The next group of elements corresponds to torsion and is presented by all the inertia elements including τ in their indices, i.e., $I_{\tau x}$, $I_{\tau y}$, etc. The last part of elements relates to inversion and these elements have indices including ρ, i.e., $I_{x\rho_1}, I_{x\rho_2,}, I_{y\rho_1}$ etc.

$$\mu = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} & I_{x\tau} & I_{x\rho_1} & I_{x\rho_2} \\ I_{yx} & I_{yy} & I_{yz} & I_{y\tau} & I_{y\rho_1} & I_{y\rho_2} \\ I_{zx} & I_{zy} & I_{zz} & I_{z\tau} & I_{z\rho_1} & I_{z\rho_2} \\ I_{\tau x} & I_{\tau y} & I_{\tau z} & I_{\tau\tau} & I_{\tau\rho_1} & I_{\tau\rho_2} \\ I_{\rho_1 x} & I_{\rho_1 y} & I_{\rho_1 z} & I_{\rho_1\tau} & I_{\rho_1\rho_2} & I_{\rho_1\rho_2} \\ I_{\rho_2 x} & I_{\rho_2 y} & I_{\rho_2 z} & I_{\rho_2\tau} & I_{\rho_2\rho_1} & I_{\rho_2\rho_2} \end{bmatrix}^{-1} \tag{3.17}$$

The Hamiltonian must be fully symmetric, thus each element of the tensor $\boldsymbol{\mu}$ belongs to the same symmetry species as the corresponding operator. These elements can be expanded into a Fourier series [118].

A set of basis functions for the rotation–inversion–torsion energies consists of products of inversion, torsion and rotational symmetry functions [117], [118]. In a zeroth approximation, a purely torsional Hamiltonian can be expressed in the following form:

$$H_t = \frac{1}{2}\mu_{\tau\tau}J_\tau^2 + V_t(\tau) \tag{3.18}$$

where the potential function is given by Eq. (3.15). The Schrödinger equation with the Hamiltonian presented by Eq. (3.18) is a Mathieu equation, which can be solved analytically. The symmetrized basis functions can be presented by:

$$\Psi_t = |m\rangle = \sum_m c_m \left[N_m\left(e^{im\tau} \pm e^{-im\tau}\right)\right] \tag{3.19}$$

where c_m is the expansion coefficient and N_m is a normalization factor. If a "pure" torsional function is considered, it has four equivalent minima, and hence the calculated torsional energy levels have the structure of quartets.

Similarly, the zero-order Hamiltonian for pure inversion can be described as:

$$H_i = \frac{1}{2}\mu_{\rho_1\rho_1}J_{\rho_1}^2 + V_i(\rho_1) \tag{3.20}$$

where the potential function is presented by Eq. (3.13) and the Schrödinger equation with above Hamiltonian can be solved in the same way as the torsional equation. The inversion function basis set is given by:

$$\begin{aligned} \Psi_{i1} = |v_1\rangle &= \sum_n c_n \left[N_m\left(e^{in\rho_1} \pm e^{-in\rho_1}\right)\right] \\ \Psi_{i2} = |v_2\rangle &= \sum_n c_n \left[N_m\left(e^{in\rho_2} \pm e^{-in\rho_2}\right)\right] \end{aligned} \tag{3.21}$$

Finally, the rotational function basis set has a form:

$$\Psi_r = |J, K\rangle \tag{3.22}$$

and the rotational Hamiltonian is presented by the following expression:

$$H_r = \frac{1}{4}\left(\mu_{xx} + \mu_{yy}\right)\left(J^2 + J_z^2\right) + \frac{1}{2}\mu_{zz}J_z^2 \tag{3.23}$$

Therefore, the inversion–torsion–rotational basis functions can be written in the following form:

$$\Psi_{rii} = \Psi_r\Psi_t\Psi_{i1}\Psi_{i2} = |J,k\rangle|m\rangle|v_1\rangle|v_2\rangle \tag{3.24}$$

where J, k, m, v_1, v_2 are the appropriate quantum numbers.

The basis functions from Eq. (3.24) are used to build the functions of a given symmetry that belongs to the irreducible representations of the $G_{16}^{(2)}$ group [118]. In Table 3.4, symmetrized rotational, torsional and inversion functions are gathered.

Table 3.4: Symmetrized rotational, torsional and inversion functions for hydrazine.

Function	Notation	Expression	Symmetry species
Rotational	$\|J,K^{\pm}\rangle$	$N_K\left(\|J,K\rangle \pm (-1)^J\|J,K\rangle\right)$	$A_{1g}^+, A_{2g}^-, B_{1g}^+, B_{2g}^-, E_g$
Torsional	$\|m^{\pm}\rangle$	$N_m(\|m\rangle \pm \|-m\rangle)$	$A_{1g}^+, A_{1u}^-, B_{1g}^+, B_{1u}^-, E_1$
Inversion	$\|v_1, v_2^{\pm}\rangle$	$N_u(\|u_1(1), u_2(2)\rangle \pm \|u_2(1), u_1(2)\rangle)$	$A_{1g}^+, A_{2u}^+, B_{1g}^+, B_{2u}^+, E^-$

Specific form of rotation–inversion–torsion symmetry functions depends on parity of quantum numbers and symmetry species to which they belong. These functions are presented and described in detail in the study by Łodyga et al. [118]. The rotation–inversion–torsion energy levels E_{rit} are obtained by diagonalization of the Hamiltonian matrix. Each rotation–inversion–torsion matrix element is a sum of the matrix elements that correspond to the terms of the Fourier expansion in large amplitude coordinates of an appropriate element $\mu_{\alpha\beta}$.

Three large amplitude motions in a hydrazine molecule give rise to a multiple splitting of each rotational level into sublevels. Each energy level is labeled according to an irreducible representation of the double PI group $G_{16}^{(2)}$. In the ground state, the largest splitting is caused by inversion [119]. Two inversions give rise to triplets. Subsequent splitting is caused by torsion. Concluding, each rotational level splits into six subleveles for $K = 0$ (two of A symmetry, two of B symmetry and two degenerate of E symmetry) or 12 sublevels for $K > 0$ (four of A symmetry, four of B symmetry and four of E symmetry).

The rotation–inversion–torsion Hamiltonian was used to reproduce the observed splitting in the spectrum for J = 0 [118]. The obtained parameters of potential function were fitted to experimental data, which correspond to inversion–torsional energy levels as well as to the torsional–inversion splitting presented in Table 3.5. Figure 3.9 shows a scheme of inversion–torsion energy levels for hydrazine.

Table 3.5: Experimental data of hydrazine, energy levels and torsional and inversion splitting.

Inv-tors level	Energy [cm^{-1}]	Torsional splitting [cm^{-1}]	Inversion splitting [cm^{-1}]
GS	0	0.00019	0.535
v_7	376.402	0.0608	0.266
$2v_7$	670.349	6.6896	0.689
$3v_7$	860.138	87.96	0.411
v_6	795.137	0.510	8.864
v_{12}	937.156	0.1178	13.735

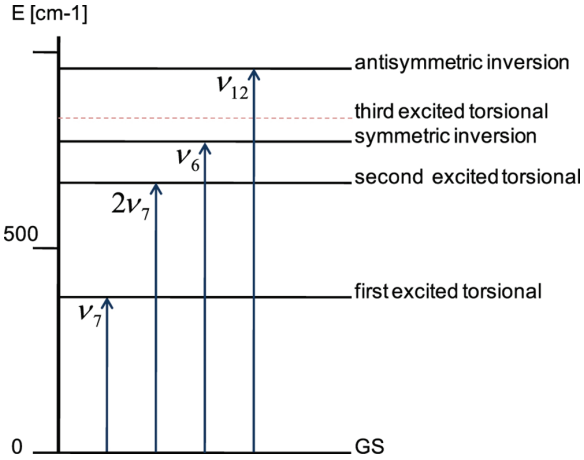


Figure 3.9: Inversion–torsion energy levels for hydrazine.

The properties of the potential can be demonstrated by presenting some of its 1D and 2D plots. Figure 3.10 presents the 1D torsional potential function for N_2H_4 constructed on the basis of the following expression:

$$V_t(\tau) = V_2 \sin 2\rho_1^{eq} sin 2\rho_2^{eq} \cos 2\tau + V_4 \cos 4\tau + V_8 \cos 8\tau \tag{3.25}$$

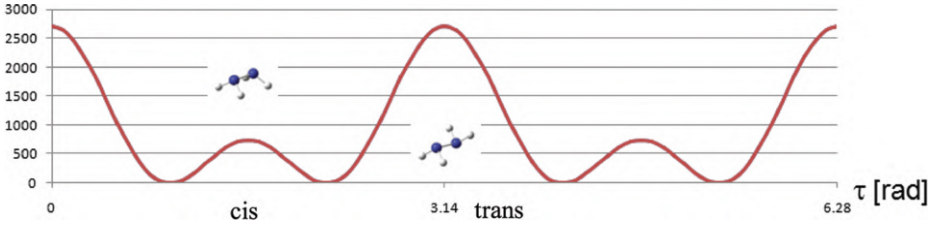


Figure 3.10: The 1D torsion potential of hydrazine.

This 1D potential was calculated by fixing the wagging coordinates at their equilibrium values which is $\rho_1^{eq} = \rho_2^{eq} = 49.8$

Figure 3.11 shows how the local minima are connected through the saddle points in a 2D space of the inversion coordinates (ρ_1, ρ_2). The plots were generated using the potential expression:

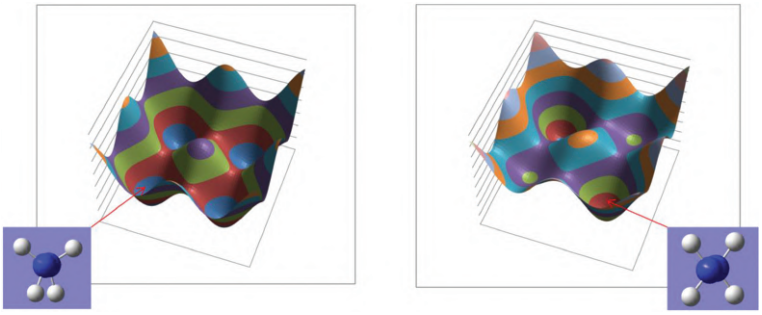


Figure 3.11: The 2D inversion–inversion potential of hydrazine.

$$V_i(\rho_1,\rho_2) = a_0 + \sum_{k=1}^{2}(a_1 cos\rho_k + a_2 cos2\rho_k + a_3 cos3\rho_k + a_4 cos4\rho_k) + V_{22} cos2\rho_1 cos2\rho_2 + V_{222} sin2\rho_1 sin2\rho_2 cos2\tau^{eq} \tag{3.26}$$

The plots were drawn for $\tau^{eq} = 45°$ (on the left-hand side) and for $\tau^{eq} = 0°$ (on the right-hand side).

It is worth mentioning at this point that Łodyga and Makarewicz [120] applied high-level *ab initio* methods to describe spectroscopic properties of hydrazine like geometries, anharmonic vibrations and torsion-wagging (TW) multiplets. To describe the splitting pattern caused by tunneling in TW states, the 3D PES for the large amplitude TW modes was built and characterized. This 3D PES was then used to construct numerically a Hamiltonian, which was used to solve the vibrational problem for TW coupled motion. The obtained frequencies were very precise and helped in assignment problems mentioned earlier in the literature studies. The energy barriers to the tunneling of the symmetric and antisymmetric inversion vibrations agreed well with the empirical values. In Table 3.6, the obtained potential function parameters are collected in terms of torsional and inversion barriers or angles.

3.5.2 Effective rovibrational Hamiltonian for hydrazine

In this section, the effective Hamiltonian for hydrazine will be explained, which is a very helpful tool in spectral analysis. Since its parameters do not always have an obvious physical meaning, it is rather not possible to use them for description of

Table 3.6: Potential function parameters obtained from *ab initio* methods for hydrazine [120].

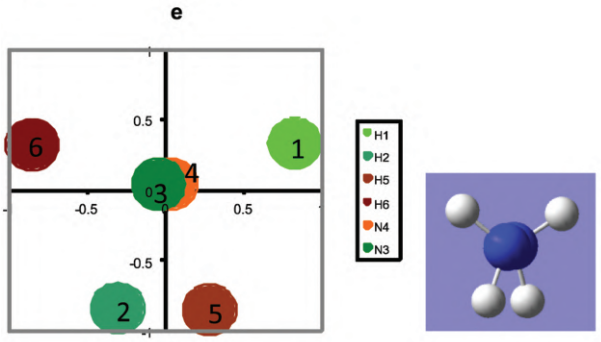
Torsional barrier (trans)	**744 cm^{-1}**
Torsional barrier (cis)	2706 cm^{-1}
Inversion barrier (one NH_2 group)	1997 cm^{-1}
Inversion barrier (both NH_2 groups)	3454 cm^{-1}
Inversion barrier in cis configuration	5480 cm^{-1}
Torsional equilibrium angle	45.1°
Inversion equilibrium angle	54.7°

interactions and dynamics of molecular modes. Such information can be obtained from spectra using an explicit Hamiltonian, discussed in the previous section, with a PES, which is expressed by empirical parameters.

In order to derive an effective Hamiltonian for a molecule like hydrazine, the first three steps presented in Table 3.1 are the same as for the explicit Hamiltonian. Hydrazine in its equilibrium configuration can appear in eight nonsuperimposable conformations. There are three large amplitude motions in the molecule (one internal rotation about the N–N bond and inversion of the NH_2 group at either end), each having two equivalent equilibrium positions, thus 2^3 gives eight frameworks. Figure 3.12 shows an initial conformation of hydrazine together with a conformation obtained after one exemplary transformation is applied. This is transformation a corresponding to the PI operation (34) (1526)* presented in its matrix form. The same effect can be achieved through the large amplitude coordinate transformations ($\tau \rightarrow -\tau + \pi/2$, $\rho_1 \rightarrow \rho_2$, $\rho_2 \rightarrow -\rho_1$). In Figure 3.13, all eight conformations are displayed. The conformations cannot be superimposed through a rotation in space. The first conformation can be transformed into another one through a proper change of LAV coordinates τ, ρ_1, ρ_2. The slow tunneling model was presented and described in detail by Hougen [2], thus we will only summarize its main features in order to give a reader a basic idea.

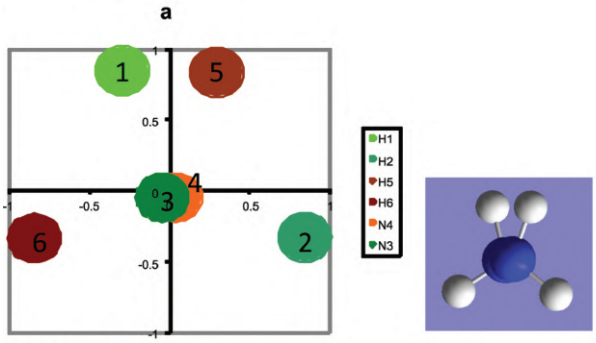
In the slow tunneling model, the assumption is made that a molecule spends most of its time vibrating in the vicinity of one of these eight frameworks presented in Figure 3.13, and it tunnels from one conformation to another from time to time only. This assumption is important since it requires that the splittings caused by tunneling motion are small in comparison to differences between vibrational energy levels in equilibrium conformation. For hydrazine, it is true in its ground state and in the first excited torsional state. Once the number of frameworks is determined for a given molecule (eight for hydrazine), one should define the vibration–rotation energy functions for one of the conformations. For hydrazine two types of vibrational functions can appear: symmetric or antisymmetric with respect to the rotation around the C_2 axis which is equivalent to the a^2b symmetry operation of the PI group. These two functions are denoted as A or B type, respectively. The vibrational function for the ground state is of A type and for the first excited torsional state of B type.

Initial configuration



	X	Y	Z
H1	0.8340	0.3270	1.1360
H2	-0.2977	-0.8449	1.1360
N3	-0.0386	0.0373	0.6944
H5	0.2882	-0.8541	-1.1228
H6	-0.8435	0.3178	-1.1228
N4	0.0400	0.0386	-0.6963

After transformation



(34)(1526)*

a

0	0	0	-1	0	0
0	0	0	0	-1	0
0	0	0	0	0	-1
0	-1	0	0	0	0
-1	0	0	0	0	0
0	0	-1	0	0	0

	X	Y	Z
H1	-0.2882	0.8541	1.1228
H2	0.8435	-0.3178	1.1228
N3	-0.0400	-0.0386	0.6963
H5	0.2977	0.8449	-1.1360
H6	-0.8340	-0.3270	-1.1360
N4	0.0386	-0.0373	-0.6944

Figure 3.12: Definition of the localized (slightly distorted) conformation ($\rho_1 = 44°$, $\rho_2 = 46°$) for hydrazine.

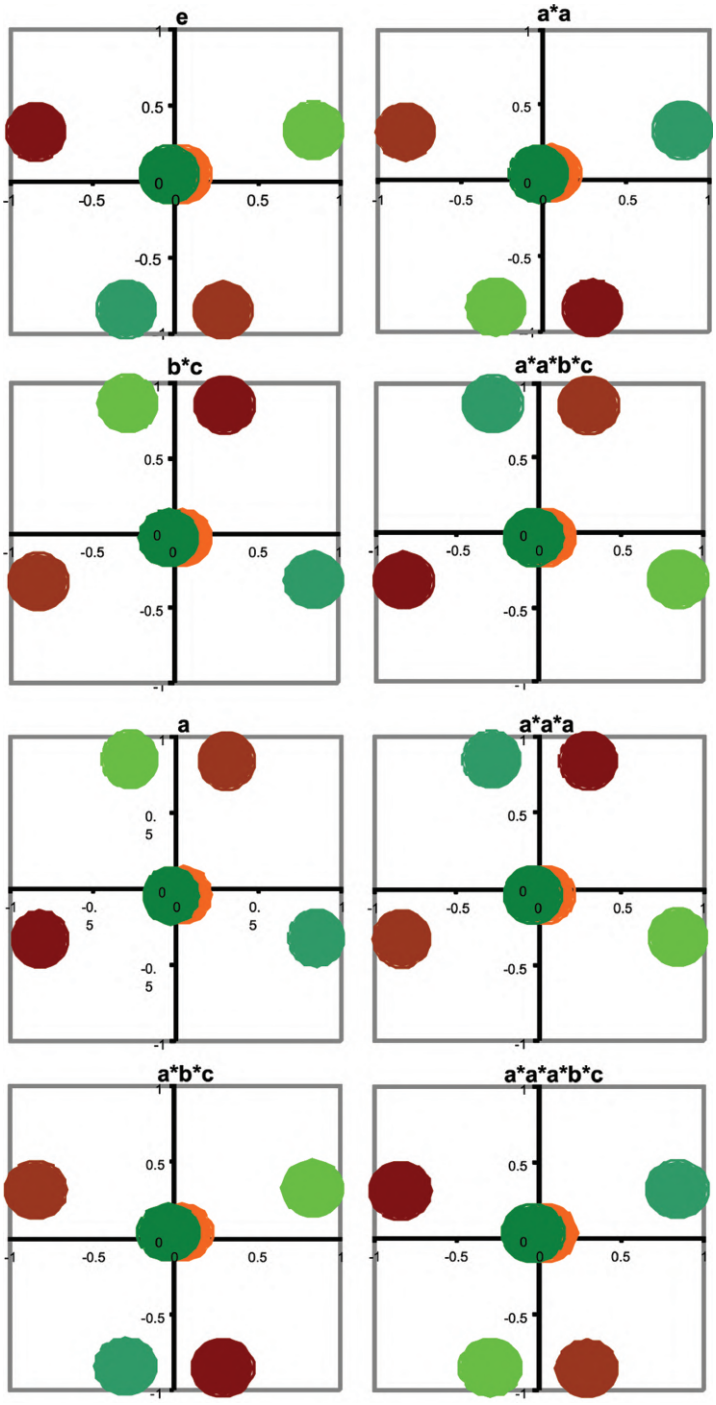


Figure 3.13: The eight nonsuperimposable equilibrium frameworks of hydrazine.

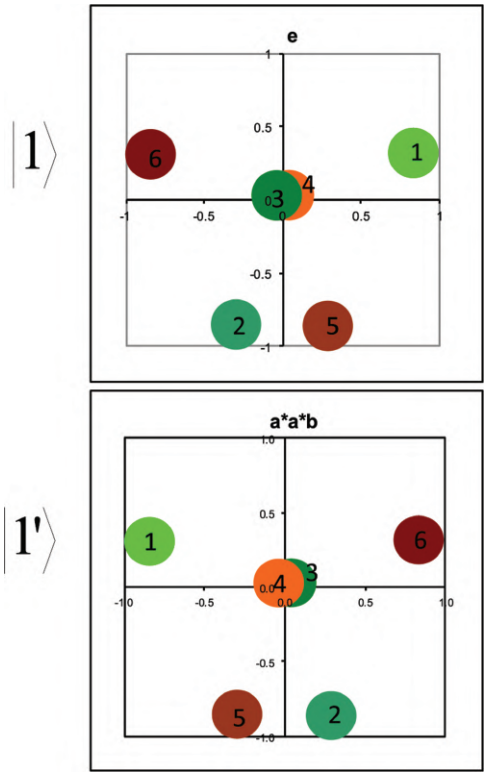


Figure 3.14: Exemplary transformation of two inversion–torsional functions in hydrazine.

Since the basis functions are defined as products of vibrational and rotational functions, it is necessary to analyze both functions separately. The effect of the *d* symmetry operator on the whole vibration–rotation function is simply an identity operation, but the rotational and torsional parts behave differently. The *d* operation always rotates the molecule by π in the molecular axis systems and every conformation has its superimposable partner rotated by π which are denoted as A and A′ or B and B′, respectively Figure 3.14. Finally, we can conclude that for the ground state we shall deal with 16 frameworks of A and A′ symmetry and for the antisymmetric inversion state with 16 frameworks of B and B′ symmetry.

Since the diagonal matrix elements for nonprimed and primed functions are identical the basis functions are built as combinations of both functions $\left(|1\rangle + |1'\rangle\right)$ or $\left(|1\rangle - |1'\rangle\right)$, where $|1\rangle$ is nonprimed function localized in the conformation 1. In this method, the explicit form of inversion–torsion functions is not important, only their symmetry properties matter. Combinations of inversion–torsion functions localized in eight nonsuperimposable configurations belong to the following symmetry species: $A_{1g}^{+}, A_{1u}^{-}, B_{1g}^{+}, B_{1u}^{-}, E^{+}, E^{-}, E_g, E_u, E_1$. The rotational functions are of $A_{1g}^{+}, A_{2g}^{-}, B_{1g}^{+}, B_{2g}^{-}, E_g$

symmetry. Thus, the rotation–inversion–torsional functions are of symmetry: $A_{1g}^{+}, A_{2g}^{-}, B_{1g}^{+}, B_{2g}^{-}, A_{1u}^{-}, A_{2u}^{+}, B_{1u}^{-}, B_{2u}^{+}, E^{+}, E^{-}$. Below in Eq. (3.27), an example of a symmetrized function is given for J = even and K = odd of A_{1g}^{+} symmetry:

$$\begin{aligned} |A_{1g}^{+}(K \geq 0, J)\rangle &= \frac{1}{\sqrt{2}} \left\{ |E_g^{(+)}\rangle_{tw} |J, K^{(-)}\rangle + (-1)^{\frac{(K-1)}{2}} i |E_g^{(-)}\rangle_{tw} |J, K^{(+)}\rangle \right\} |E_g^{(+)}\rangle \\ &= \frac{1}{2}[|1\rangle - |2\rangle + |3\rangle - |4\rangle] |E_g^{(-)}\rangle = \frac{1}{2}[|5\rangle - |6\rangle - |7\rangle + |8\rangle] \end{aligned} \tag{3.27}$$

where $|E_g^{(+)}\rangle$ and $|E_g^{(-)}\rangle$ are vibrational, and $|J, K^{(+)}\rangle$ and $|J, K^{(-)}\rangle$ are rotational functions.

A phenomenological rotational Hamiltonian operator, H, is built in a way that it contains symmetry of permitted products of purely rotational operators and coefficients, i.e., functions of the large and small vibrational coordinates. The rovibrational effective Hamiltonian is presented in the following form:

$$\begin{aligned} \mathrm{H} = \sum_{n=1}^{8} H_n = \sum_{n=1}^{8} & h_n + h_{nj} J^2 + h_{nk} J_z^2 + h_{njj} J^4 + h_{njk} J^2 J_z^2 + h_{nkk} J_z^4 + \\ & f_n \left(J_+^2 + J_-^2\right) + g_n \left[i\left(J_+^2 - J_-^2\right)\right] + q_n J_z + \\ & d_n \left(J_+^4 + J_-^4\right) + p_n \left[i\left(J_+^4 - J_-^4\right)\right] + \\ & \left[r_{n+} J_+ + r_{n-} J_-\right] + \left[s_{n+} \left(J_z J_+ + J_+ J_z\right) + s_{n-} \left(J_z J_- + J_- J_z\right)\right] \end{aligned} \tag{3.28}$$

The sum runs over eight nonsuperimposable configurations of hydrazine.

The subscript n indicates the number of a framework which interacts with the framework 1. Thus, matrix elements of the Hamiltonian in Eq. (3.28), can be written in the form $\langle 1|H|n\rangle$, where (n = 1) corresponds to no tunneling motion, i.e., to the effects within a single framework 1. The (n = 2) term corresponds to inversion of both NH_2 groups, (n = 3) to internal rotation, (n = 4) to inversion of both amino groups and internal rotation, (n = 5) to inversion of one amino group, (n = 6) to inversion of the other amino group, (n = 7) to inversion of one of the amino group and internal rotation, and (n = 8) to inversion the other amino group and internal rotation. According to symmetry, processes (n = 5) and (n = 6) are energetically equivalent, so are (n = 7) and (n = 8). In the end, only matrix elements with n = 1, 2, 3, 4, 5 and 7 should be taken into account explicitly.

The quantities from Eq. (3.28) h_n, h_{nj}, h_{nk}, f_n, g_n, q_n, d_n etc. represent numerical constants for a vibrational state of hydrazine. The quantities J and K are the usual total angular momentum quantum numbers and their projection along the a axis of the prolate near symmetric top. The Hamiltonian matrix is built from products of rotational and inversion–torsion function. Eq. (3.29) gives an example:

$$\langle inv.tors|q_n|inv.tors\rangle \langle rot|\hat{J}_z|rot\rangle \tag{3.29}$$

The symmetry allows to reduce all configuration interactions to basic <1|n> (n = 2, 3, 4, 5, 7) interactions and to determine, which of these interactions are zero for every term in the effective Hamiltonian.

To conclude, the formalism of Hougen [2] was originally developed to account for the observed microwave spectra of the ground state of hydrazine [119]. A few years later, this theory was used to assign and fit the spectrum of the torsional band of N_2H_4 [121]. It was a test whether the effective Hamiltonian would be also suitable for a state where the torsional splitting is much higher than in the ground state. The fitting results were satisfactory as well as in the second torsional state [122]. The effective Hamiltonian did not succeed in a global analysis of the antisymmetric amino-wagging band of hydrazine (ν_{12}) [123], although the inversion tunneling frequency was much smaller than the antisymmetric vibrational frequency and it could be treated as a high barrier case. It seemed that failure of a global fit was caused by perturbing states located close to the ν_{12} state. The antisymmetric band of hydrazine was reanalyzed a few years later and the effective parameters were calculated separately for each value of K′ using the Hougen-Ohashi phenomenological Hamiltonian [2], [124]. The same Hamiltonian was used to fit the observed transitions in the N–N stretching band (ν_5) of hydrazine [125]. As a result, the effective parameters were obtained as well as the torsional and inversion splittings. The most intense bands in the IR spectrum of hydrazine between 700 and 1200 cm^{-1} were assigned to wagging motions of two NH_2 groups [126]. The weaker band appearing at lower frequency was assigned as symmetric wagging band (ν_6) and a stronger doublet at 933–966 cm^{-1} as antisymmetric wagging band (ν_{12}) [127]. The analysis of the symmetric amino-wagging band (ν_6) revealed a strong Fermi-type perturbation between this band and the third excited torsional band ($3\nu_7$) [128]. It turned out that the structure of tunneling splittings was different from the one that was originally assumed for an unperturbed vibrational state. The best candidate for a perturber of the ν_6 state would be the ν_{12} state, since it is located very close to the symmetric wagging and even overlaps partly. According to the results obtained in work [129], where the effective Hamiltonian for coupling between inversion–torsion states of hydrazine was studied, only states of the same vibrational symmetry can interact with each other through anharmonic resonances. The matrix elements for A–A- and A–B-type interactions were derived, and it was shown that most of the resulting matrix elements for A-B-type coupling vanish. Thus, the best perturbing state was concluded the third excited torsional state ($3\nu_7$), which is symmetric with respect to the rotation about the C_2 axis of hydrazine and can interact with the symmetric wagging state. This is an A–A type of interaction [129] and therefore a Fermi-type coupling is permitted. One can look back at Figure 3.9, where all three mentioned states are presented. Although at the beginning, the third excited torsional state seemed to be located not close enough in energy to the ν_6 state, the torsional splitting is so strong that its lower part of energy levels can take part in the interaction (a dotted line in Figure 3.9).

As it turned out later, the $3\nu_7$ state was calculated not that far away from the ν_6 state. Unfortunately, no experimental data were available for the $3\nu_7$ and its position as well as structure had to be estimated using 3D PES for hydrazine [118]. Applying a rotation–torsion–inversion Hamiltonian, it was possible to reproduce the inversion–torsional splitting for the $J = 0$ sublevels. Parameters of the inversion–torsion potential function have been fitted to the band centers and splittings for the ground, first and second torsional, and symmetric and antisymmetric inversion states of hydrazine (Table 7 in the study by Łodyga et al. [118]). The geometric parameters were taken from the microwave data, so the predicted position for the perturber was at 830 cm^{-1} and for the symmetric inversion state at 780 cm^{-1}. The theoretical results obtained from the 3D model were used as a starting point for the subsequent analysis of the dyad symmetric inversion/third excited. The analysis of the symmetric inversion state is a nice example of how the explicit and effective Hamiltonians complement each other.

3.6 Floppy molecules applications

Hydrazine is mainly used as a foaming agent in the preparation of polymer foams, but is also applied as a precursor to polymerization catalysts, pharmaceuticals and agrochemicals. Besides, it is used in various rocket fuels as well as to prepare gas precursors for air-bags. Hydrazine is also used as a long-term storable propellant on board space vehicles, such as the NASA Dawn probe to Ceres and Vesta. The F-16 fighter jet, NASA Space Shuttle and U-2 spy plane use hydrazine to fuel their emergency power units [130]. Hydrazine's role in pharmacy is a precursor to many pharmaceutical compounds. It involves conversion of hydrazine to heterocyclic rings such as pyrazoles or pyridazines. Hydrazine compounds combined with other agricultural chemicals such as insecticides, fungicides or herbicides can be effective in agriculture.

Many floppy molecules have astrophysical importance. A list of such molecules is very long, thus we would like to mention just some of them like ammonia, methanol, ethylene, formamide, acetylaldehyde etc. One of such molecules is methylamine, CH_3NH_2, which is the simplest primary alkylamine. It is considered as a precursor of the simplest amino acid glycine. It has been detected as a constituent in the interstellar medium for the first time in 1974 at 3.5 cm [131] and at 3 cm [132] bands. Interstellar methylamine was also detected in a spiral galaxy with a high redshift of 0.89 [133] and in cometary samples of the Stardust mission [134]. The analysis of the molecular dynamics of the interstellar molecules appeared to be necessary for understanding the frequencies and intensities of the observed spectra in the laboratory and in interstellar space.

Apart from molecules described above, there is a group of large molecules with conjugated π-systems that play an important role in natural processes and can be

used in different materials of technological usage [135], [136], [137], [138], [139], [140]. The structure of these molecules is determined mainly by a balance between the electronic forces governing the π-system and the steric forces between nonbonded atoms in the molecule. To such molecules belong, for instance, biphenyl, benzophenone or *cis*-stilbene. If the molecules are electronically excited, the forces responsible for π-system change, and one can notice the difference between the structures of the ground state and the excited state. The vibrations caused by electronic excitation are usually considered to be slow and can lead to structural isomerization. For instance, once *cis*-stilbene molecule gets relaxed to the ground electronic state, two parallel photoisomerization reactions take place in its lowest excited singlet state [141], [142], [143].

Most recent and spectacular application of floppy molecules considers them as candidates for achieving optoelectronic molecular devices without skeletal rearrangement or bond breaking [144]. For this purpose, specific floppy molecules are used, usually the molecules that consist of two benzene or benzene-like rings, and which rotate relative to each other. In the paper by Baldea [144], a mechanism is proposed for the photoinduced switching of such molecules from a nonplanar conformation to the planar one with neither skeletal rearrangement nor bond breaking or significant molecular length changes. Since the planar conformation has a higher conductance than the nonplanar one, it acts as ON state of the molecular switch, whereas the nonplanar as OFF state of the molecular switch. To such molecules belong azobenzene or diarylethene. For diarylethene, for instance, the open and most stable as well as less-conjugated conformation corresponds to a low conductance and it can be considered as the OFF state of the molecule. Once the molecule is exposed to ultraviolet radiation, it is driven from this OFF state to a less stable closed structure, which is more conjugated and more conducting (ON state of the switch). Shining with visible light can toggle back the molecule to the insulating open state [145]. A similar switching between ON and OFF states was observed in azobenzene [146], [147], [148], [149], [150], where once the molecule is illuminated, it can switch reversibly from *cis* to *trans* conformations. Ultraviolet radiation toggles the more stable and more conjugated *trans* isomer into the *cis* one, which is less stable and less conjugated. Then shining with visible light toggles back the *cis* isomer into the *trans*.

Some molecules adsorbed at interfaces may exhibit unique behavior that can be very different from molecules in bulk situations [151]. This uniqueness considered in the context of molecular electronics, can be explained by the fact that in contrast to the fixed density in bulk cases, molecules in a monolayer self-assembled (SAM) on a substrate electrode have a density that can be varied and controlled. SAMs are based on floppy molecules consisting of two rings that can easily rotate relative to each other. Thus, their molecular conformation can be tuned by varying the density of the adsorbate.

To conclude, we believe that although dynamics and molecular spectra of molecules performing large amplitude motions are often very complex, the study of large amplitude motions will continue to be of great interest to molecular physics, chemistry or astrophysics in the future.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission
Research funding: None declared.
Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. Hougen, JT, Bunker, PR, Johns, JWC. The vibration-rotation problem in triatomic molecules allowing for a large-amplitude bending vibration. *J Mol Spectrosc* 1970;34:136. https://doi.org/10.1016/0022-2852(70)90080-9.
2. Hougen, JT. A rotational Hamiltonian for the ground vibrational state of hydrazine. *J Mol Spectrosc* 1981;89:296. https://doi.org/10.1016/0022-2852(81)90025-4.
3. Wilson, EBJr, Howard, JB. The vibration rotation energy levels of polyatomic molecules I. Mathematical theory of semirigid asymmetrical top molecules. *J Chem Phys* 1936;4:260. https://doi.org/10.1063/1.1749833.
4. Darling, BT, Dennison, DM. The water vapor molecule. *Phys Rev* 1940;57:128. https://doi.org/10.1103/physrev.57.128.
5. Nielsen, HH. The vibration-rotation energies of molecules. *Rev Mod Phys* 1951;23:90. https://doi.org/10.1103/revmodphys.23.90.
6. Dennison, DM, Uhlenbeck, GE. The two-minima problem and the ammonia molecule. *Phys Rev* 1932;41:313. https://doi.org/10.1103/physrev.41.313.
7. Kemp, JD, Pitzer, KS. Hindered rotation of the methyl groups in ethane. *J Chem Phys* 1936;4:749. https://doi.org/10.1063/1.1749784.
8. Koehler, JS, Dennison, DM. Hindered rotation in methyl alcohol. *Phys Rev* 1940;57:1006. https://doi.org/10.1103/physrev.57.1006.
9. Wilson, EBJr, Decius, JC, Cross, PC. *Molecular vibrations*. New York: Mc Graw-Hill Book Co.; 1955.
10. Thorson, WR, Nakagawa, I. Dynamics of the quasi-linear molecule. *J Chem Phys* 1960;33:994. https://doi.org/10.1063/1.1731399.
11. Freed, KF, Lombardi, JR. Considerations on the rotation-vibration of triatomic molecules. *J Chem Phys* 1966;45:591. https://doi.org/10.1063/1.1727613.
12. Kirtman, B. Interactions between ordinary vibrations and hindered internal rotation. I. Rotational energies. *J Chem Phys* 1962;37:2516. https://doi.org/10.1063/1.1733049.
13. Kirtman, B. Interactions between ordinary vibrations and hindered internal rotation. II. Theory of internal rotation fine structure in some perpendicular bands of ethane-type molecules. *J Chem Phys* 1964;41:775. https://doi.org/10.1063/1.1725960.
14. Meyer, R, Günthard, HH. General internal motion of molecules, classical and quantum-mechanical Hamiltonian. *J Chem Phys* 1968;49:1510. https://doi.org/10.1063/1.1670272.

15. Quade, CR. The interaction of a large amplitude internal motion with other vibrations in molecules. The effective Hamiltonian for the large amplitude motion. *J Chem Phys* 1976;65:700. https://doi.org/10.1063/1.433084.
16. Louck, JD, Galbraith, HW. Eckart vectors, Eckart frames, and polyatomic molecules. *Rev Mod Phys* 1976;48:69. https://doi.org/10.1103/revmodphys.48.69.
17. Hougen, JT. A group-theoretical treatment of electronic, vibrational, torsional, and rotational motions in the dimethylacetylene molecule. *Can J Phys* 1964;42:1920. https://doi.org/10.1139/p64-182.
18. Hougen, JT. Vibrational motions in dimethyl acetylene. *Can J Phys* 1965;43:935. https://doi.org/10.1139/p65-089.
19. Bunker, PR, Hougen, JT. Normal modes, coriolis coupling, and centrifugal distortion in molecules with nearly free internal rotation: CH3CCCH3 and CH3CCSiH3. *Can J Phys* 1967;45:3867. https://doi.org/10.1139/p67-324.
20. Bunker, PR. Dimethylacetylene: an analysis of the theory required to interpret its vibrational spectrum. *J Chem Phys* 1967;47:718. https://doi.org/10.1063/1.1711945.
21. Bunker, PR. Erratum: dimethylacetylene: an analysis of the theory required to interpret its vibrational spectrum. *J Chem Phys* 1968;48:2832. https://doi.org/10.1063/1.1669535.
22. Papoušek, D, Stone, JMR, Špirko, V. Vibration-inversion-rotation spectra of ammonia. A vibration-inversion-rotation Hamiltonian for NH3. *J Mol Spectrosc* 1973;48:17–37, https://doi.org/10.1016/0022-2852(73)90132-X.
23. Redding, RW. Rotational levels of double-minimum potentials: the linear BAB molecule with unequal bond lengths. *J Mol Spectrosc* 1971;38:396. https://doi.org/10.1016/0022-2852(71)90122-6.
24. Sarka, K. The vibration - rotation Hamiltonian of the tetratomic quasilinear molecule. *J Mol Spectrosc* 1971;38:545. https://doi.org/10.1016/0022-2852(71)90081-6.
25. Moule, DC, Ramachandra Rao, CVS. Vibration-rotation-inversion Hamiltonian of formaldehyde. *J Mol Spectrosc* 1973;45:120. https://doi.org/10.1016/0022-2852(73)90181-1.
26. Stone, JRS. The large amplitude motion-rotation Hamiltonian for tetratomic molecules, application to HCNO and HNCO. *J Mol Spectrosc* 1975;54:1–9. https://doi.org/10.1016/0022-2852(75)90002-8.
27. Danielis, V, Papoušek, D, Špirko, V, Hokák, M. Vibration-inversion-rotation spectra of ammonia. *J Mol Spectrosc* 1975;54:339. https://doi.org/10.1016/0022-2852(75)90165-4.
28. Špirko, V, Stone, JMR, Papoušek, D. Vibration-inversion-rotation spectra of ammonia: centrifugal distortion, coriolis interactions, and force field in 14NH3, 15NH3, 14ND3, and 14NT3. *J Mol Spectrosc* 1976;60:159–78.
29. Hardwick, JL, Brand, JCD. Anharmonic potential constants and the large amplitude bending vibration in nitrogen dioxide. *Can J Phys* 1976;54:80. https://doi.org/10.1139/p76-010.
30. Brand, JCD, Ramachandra Rao, CVS. The vibration-rotation problem in triatomic molecules allowing for two large-amplitude stretching vibrations. *J Mol Spectrosc* 1976;61:360. https://doi.org/10.1016/0022-2852(76)90327-1.
31. Weber, WH, Ford, GW. Carbon suboxide as a quasilinear molecule with a large amplitude bending mode. *J Mol Spectrosc* 1976;63:445. https://doi.org/10.1016/0022-2852(76)90306-4.
32. Istomin, VA, Stepanov, NF, Zhilinskii, BI. Vibration-rotation problem for triatomic molecules with two large-amplitude coordinates. *J Mol Spectrosc* 1977;67:265. https://doi.org/10.1016/0022-2852(77)90042-x.
33. Bunker, PR, Stone, JMR. The bending-rotation Hamiltonian for the triatomic molecule and application to HCN and H2O. *J Mol Spectrosc* 1972;41:310. https://doi.org/10.1016/0022-2852(72)90209-3.

34. Bunker, PR, Landsberg, BM. The rigid bender and semirigid bender models for the rotation-vibration Hamiltonian. *J Mol Spectrosc* 1977;67:374. https://doi.org/10.1016/0022-2852(77)90048-0.
35. Hoy, AR, Bunker, PR. A precise solution of the rotation bending Schrödinger equation for a triatomic molecule with application to the water molecule. *J Mol Spectrosc* 1979;74:1. https://doi.org/10.1016/0022-2852(79)90019-5.
36. Jensen, P, Bunker, PR, Hoy, AR. The equilibrium geometry, potential function, and rotation-vibration energies of CH2 in the X̃3B1 ground state. *J Chem Phys* 1982;77:5370. https://doi.org/10.1063/1.443785.
37. Jensen, P, Bunker, PR. The application of the nonrigid bender Hamiltonian to a quasilinear molecule. *J Mol Spectrosc* 1983;99:348. https://doi.org/10.1016/0022-2852(83)90319-3.
38. Bunker, PR. Quasilinear and quasiplanar molecules. *Annu Rev Phys Chem* 1983;34:59. https://doi.org/10.1146/annurev.pc.34.100183.000423.
39. Jensen, P, Bunker, PR. The nonrigid bender Hamiltonian using an alternative perturbation technique. *J Mol Spectrosc* 1986;118:18. https://doi.org/10.1016/0022-2852(86)90220-1.
40. Jensen, P. The nonrigid bender Hamiltonian for calculating the rotation-vibration energy levels of a triatomic molecule. *Comput Phys Rep* 1983;1:1.
41. Špirko, V. The inversional dependence of hyperfine quadrupole coupling in 14NH3. *Mol Phys* 1979;38:1761. ibid. 43 (1981) 723, https://doi.org/10.1080/00268977900102841.
42. Špirko, V. Vibrational anharmonicity and the inversion potential function of NH3. *J Mol Spectrosc* 1983;101:30. https://doi.org/10.1016/0022-2852(83)90004-8; Špirko V, Kraemer WP, ibid. 133 (1989) 331, Anharmonic potential function and effective geometries for the NH3 molecule, https://doi.org/10.1016/0022-2852(89)90196-3.
43. Kręglewski, M. Vibration-inversion-internal rotation-rotation Hamiltonian for methylamine. *J Mol Spectrosc* 1978;72:1. https://doi.org/10.1016/0022-2852(78)90038-3; ibid. 133 (1989) 10, The geometry and inversion-internal rotation potential function of methylamine.
44. Kręglewski, M. The effect of the large amplitude vibration on the rotational structure of the 3500 Å transition of formaldehyde: part I. *J Mol Struct* 1979;55:135, https://doi.org/10.1016/0022-2860(79)80193-3.
45. Kręglewski, M. Correlation diagrams for quasi-symmetric top molecules with a single rotor. *J Mol Spectrosc* 1984;105:8, https://doi.org/10.1016/0022-2852(84)90100-0.
46. Kręglewski, M, Jensen, P. Determination of the skeletal bending potential function for SiH3NCO from the microwave spectrum. *J Mol Spectrosc* 1984;103:312, https://doi.org/10.1016/0022-2852(84)90058-4.
47. Kręglewski, M. Methyl isothiocyanate as a quasi-symmetric top. A reinterpretation of the microwave spectrum. *Chem Phys Lett* 1984;112:275. https://doi.org/10.1016/0009-2614(84)80549-7.
48. Wierzbicki, A, Koput, J, Kręglewski, M. The large-amplitude motions in quasi-symmetric top molecules with internal C3v rotors. *J Mol Spectrosc* 1983;99:102. https://doi.org/10.1016/0022-2852(83)90296-5.
49. Kręglewski, M. Wpływa drgań o dużej amplitudzie na widma oscylacyjno-rotacyjne i strukturę cząsteczek typu bąka quasi-symetrycznego. *Wiad Chem* 1986;40:803.
50. Koput, J. Characteristic patterns in microwave spectra of quasi-symmetric top molecules of the WH3XYZ type. *J Mol Spectrosc* 1986;118:448. https://doi.org/10.1016/0022-2852(86)90181-5.
51. Quade, CR. Contributions of the interaction of internal rotation with other vibrations to the effective potential energy for internal rotation in molecules with symmetric internal rotors. *J Chem Phys* 1980;73:2107. https://doi.org/10.1063/1.440405.

52. Guan, Y, Quade, CR. Curvilinear coordinate formulation for vibration-rotation-large amplitude internal motion interactions. I. The general theory. *J Chem Phys* 1986;84:5624. https://doi.org/10.1063/1.449923.
53. Szalay, V. On the separation of the large- and small-amplitude internal motions of a molecule. *J Mol Spectrosc* 1985;110:172. https://doi.org/10.1016/0022-2852(85)90221-8.
54. Pyka, J, Foltynowicz, I, Makarewicz, J. A numerical approach to the internal large amplitude motion Hamiltonian of a polyatomic molecule. *J Mol Spectrosc* 1990;143:137. https://doi.org/10.1016/0022-2852(90)90266-s.
55. Makarewicz, J, Łodyga, W. Self-consistent internal axes for a rotating-vibrating triatomic molecule. *Mol Phys* 1988;64:899. https://doi.org/10.1080/00268978800100613.
56. Makarewicz, J. Hartree theory for rovibrational states of molecules. *Mol Phys* 1987;61:547. https://doi.org/10.1080/00268978700101311.
57. Makarewicz, J. Self-consistent approach to the bending-rotation interactions in the H2O molecule. *J Mol Spectrosc* 1988;130:316. https://doi.org/10.1016/0022-2852(88)90080-x.
58. Tennyson, J. The calculation of the vibration-rotation energies of triatomic molecules using scattering coordinates. *Comput Phys Rep* 1986;4:1.
59. Manning, MF. Energy levels of a symmetrical double minima problem with applications to the NH3 and ND3 molecules. *J Chem Phys* 1935;3:136. https://doi.org/10.1063/1.1749619.
60. Cleeton, CE, Williams, NH. Electromagnetic waves of 1.1 cm wave-length and the absorption spectrum of ammonia. *Phys Rev* 1934;45:234. https://doi.org/10.1103/physrev.45.234.
61. Sheng, H-y, Barker, EF, Dennison, DM. Further resolution of two parallel bands of ammonia and the interaction between vibration and rotation. *Phys Rev* 1941;60:786. https://doi.org/10.1103/physrev.60.786.
62. Swalen, JD, Ibers, JA. Potential function for the inversion of ammonia. *J Chem Phys* 1962;36:1914. https://doi.org/10.1063/1.1701290.
63. Coon, JB, Naugle, NW, McKenzie, RD. The investigation of double-minimum potentials in molecules. *J Mol Spectrosc* 1966;20:107. https://doi.org/10.1016/0022-2852(66)90046-4.
64. Kręglewski, M. The story of the ammonia molecule: ten years of investigation of molecular inversion. *J Mol Struct* 1983;100:179, https://doi.org/10.1016/0022-2860(83)90091-1.
65. Kręglewski, M. The geometry and inversion-internal rotation potential function of methylamine. *J Mol Spectrosc* 1989;133:10. https://doi.org/10.1016/0022-2852(89)90239-7.
66. Ohashi, N, Murase, N, Yamanouchi, K, Sugie, M, Takeo, H, Matsumura, C, et al.. Analysis of the inversion splittings in methylhydrazine. *J Mol Spectrosc* 1989;138:497. https://doi.org/10.1016/0022-2852(89)90015-5.
67. Łodyga, W, Kręglewski, M. Two-dimensional potential function for hydrazine determined from rotation-vibration spectra. *Chem Phys Lett* 1993;210:303, https://doi.org/10.1016/0009-2614(93)89137-7.
68. Gulaczyk, I, Kręglewski, M. Multi-dimensional proton tunneling in 2-methylmalonaldehyde. *J Mol Struct* 2020;1220:128733. https://doi.org/10.1016/j.molstruc.2020.128733.
69. Kozłowska, ED. *Wiad Chem* 1983;37:765.
70. Groner, P, Sullivan, JF, During, JR. *Vibrational spectra and structure*. Amsterdam: Elsevier; 1978, vol. 8. chapter 6.
71. Lin, CC, Swalen, JD. Internal rotation and microwave spectroscopy. *Rev Mod Phys* 1959;31:841. https://doi.org/10.1103/revmodphys.31.841.
72. Tsuboi, M, Tamagake, K, Hirakawa, AY, Yamaguchi, J, Nakagawa, H, Manocha, AS, et al.. Internal rotation in ethylamine: a treatment as a two-top problem. *J Chem Phys* 1975;63:5177. https://doi.org/10.1063/1.431300.

73. Flaud, JM, Camy-Peyret, C, Johns, JWC, Carli, B. The far infrared spectrum of H2O2. First observation of the staggering of the levels and determination of the cis barrier. *J Chem Phys* 1989;91:1504. https://doi.org/10.1063/1.457110.
74. Meyer, R, Caminati, W, Hollenstein, H. Torsional motions in methyl glycolate. *J Mol Spectrosc* 1989;137:87. https://doi.org/10.1016/0022-2852(89)90271-3.
75. Meyer, R. Flexible models for intramolecular motion, a versatile treatment and its application to glyoxal. *J Mol Spectrosc* 1979;76:266. https://doi.org/10.1016/0022-2852(79)90230-3.
76. Brown, C, Hogg, DR, McKeen, DC. Infrared spectra of some 2-nitrobenzenesulphenate esters, 2-NO2C6H4S.O.R. *Spectrochim Acta* 1970;26A:26A. https://doi.org/10.1016/0584-8539(70)80247-1.
77. Jensen, P. HCNO as a semirigid bender: the degenerate v4 state. *J Mol Spectrosc* 1983;101:422. https://doi.org/10.1016/0022-2852(83)90146-7.
78. Jensen, P. C3O2 as a semirigid bender: the degenerate v5 state. *J Mol Spectrosc* 1984;104:59. https://doi.org/10.1016/0022-2852(84)90245-5.
79. Mills, IM. Born-Oppenheimer failure in the separation of low-frequency molecular vibrations. *J Phys Chem* 1984;88:532. https://doi.org/10.1021/j150647a043.
80. Egawa, T, Fukuyama, T, Yamamoto, S, Takabayashi, F, Kambara, H, Ueda, T, et al.. Molecular structure and puckering potential function of cyclobutane studied by gas electron diffraction and infrared spectroscopy. *J Chem Phys* 1987;86:6018. https://doi.org/10.1063/1.452489.
81. Egawa, T, Yamamoto, S, Ueda, T, Kuchitsu, K. Two-dimensional analysis of the ring-puckering and methylene-rocking modes of cyclobutane. *J Mol Spectrosc* 1987;126:231. https://doi.org/10.1016/0022-2852(87)90094-4.
82. Tecklenberg, JL. Kinetic energy functions for the ring-bending and ring-twisting vibrations of asymmetrically substituted six-membered rings. *J Mol Spectrosc* 1989;137:65. https://doi.org/10.1016/0022-2852(89)90269-5.
83. Laane, J. Determination of vibrational potential energy surfaces from Raman and infrared spectra. *Pure Appl Chem* 1987;59:1307. https://doi.org/10.1351/pac198759101307.
84. Robiette, AG, Borgers, TR, Strauss, HL. New interpretation of the far infrared spectrum and ring-puckering potential of azetidine. *Mol Phys* 1981;42:1519. https://doi.org/10.1080/00268978100101121.
85. Kilpatrick, JE, Pitzer, KS, Spitzer, R. The thermodynamics and molecular structure of cyclopentane1. *J Am Chem Soc* 1947;69:2483. https://doi.org/10.1021/ja01202a069.
86. Berry, RS. Correlation of rates of intramolecular tunneling processes, with application to some group V compounds. *J Chem Phys* 1960;32:933. https://doi.org/10.1063/1.1730820.
87. Berry, RS. Time-dependent measurements and molecular structure: ozone. *Rev Mod Phys* 1960;32:447. https://doi.org/10.1103/revmodphys.32.447.
88. Herzberg, G. *Molecular spectra and molecular structure: electronic spectra and electronic structure of polyatomic molecules*. Malabar, Fl, USA: Krieger Publishing Company; 1991a, vol. 3.
89. Herzberg, G. *Molecular spectra and molecular structure: infrared and raman of polyatomic molecules*. Malabar, Fl, USA: Krieger Publishing Company; 1991b, vol. 2.
90. Bunker, PR, Jensen, P. *Molecular symmetry and spectroscopy*. Ottawa: NRC Research Press; 1998.
91. Longuet-Higgins, HC. The symmetry groups of non-rigid molecules. *Mol Phys* 1963;6:445–60. https://doi.org/10.1080/00268976300100501.
92. Hougen, JT. Vibronic and rotational energy levels of a linear triatomic molecule in a3Π electronic state. *J Chem Phys* 1962;36:1874. https://doi.org/10.1063/1.1701283.
93. Wilson, EB, Lin, CC, Lide, DR. Calculation of energy levels for internal torsion and over-all rotation. I. CH3BF2 type molecules. *J Chem Phys* 1955;23:136–42. https://doi.org/10.1063/1.1740512.

94. Howard, JB. The normal vibrations and the vibrational spectrum of C2H6. *J Chem Phys* 1937;5:442–50. https://doi.org/10.1063/1.1750052.
95. Wilson, EB. Nuclear spin and symmetry effects in the heat capacity of ethane gas. *J Chem Phys* 1938;6:740–5. https://doi.org/10.1063/1.1750160.
96. Kasuya, T, Kojima, T. Internal motions of hydrazine. *J Phys Soc Jpn* 1963;18:364–8. https://doi.org/10.1143/jpsj.18.364.
97. Watson, JKG. The correlation of the symmetry classifications of states of nonrigid molecules. *Can J Phys* 1965;43:1996–2007. https://doi.org/10.1139/p65-193.
98. Swalen, JD, Costain, CC. Internal rotation in molecules with two internal rotors: microwave spectrum of acetone. *J Chem Phys* 1959;31:1562–74. https://doi.org/10.1063/1.1730653.
99. Myers, RJ, Wilson, EB. Application of symmetry principles to the rotation-internal torsion levels of molecules with two equivalent methyl groups. *J Chem Phys* 1960;33:186–91. https://doi.org/10.1063/1.1731075.
100. Groner, P. Molecular symmetry: why permutation-inversion (PI) groups don't render the point groups obsolete. *J Mol Spectrosc* 2018;343:34–43. https://doi.org/10.1016/j.jms.2017.07.011.
101. Watson, JKG. *Vibrational spectra and structure*. During, JR, editor. Amsterdam: Elsevier; 1977, vol. 6, chapter 1.
102. Groner, P. Large-amplitude motion tunneling parameters in effective rotational Hamiltonians from rotation-internal rotation theory. *J Mol Spectrosc* 1992;156:164. https://doi.org/10.1016/0022-2852(92)90101-s.
103. Dalton, BJ. Nonrigid molecule effects on the rovibronic energy levels and spectra of phosphorous pentafluoride. *J Chem Phys* 1971;54:4745. https://doi.org/10.1063/1.1674750.
104. Hougen, JT. Summary of group theoretical results for microwave and infrared studies of H2O2. *Can J Phys* 1984;62:1392. https://doi.org/10.1139/p84-186.
105. Ohashi, N, Hougen, JT. The torsional-wagging tunneling problem and the torsional-wagging-rotational problem in methylamine. *J Mol Spectrosc* 1987;121:474. https://doi.org/10.1016/0022-2852(87)90064-6.
106. Oda, M, Ohashi, N, Hougen, JT. A group-theoretical formalism for the large-amplitude vibration-rotation problem in methylamine-d1. *J Mol Spectrosc* 1990;142:57. https://doi.org/10.1016/0022-2852(90)90292-x.
107. Kleiner, I, Hougen, JT, Suenram, RD, Lovas, FJ, Godefroid, M. The ground torsional state of acetaldehyde. *J Mol Spectrosc* 1991;148:38. https://doi.org/10.1016/0022-2852(91)90032-6.
108. Hougen, JT, Ohashi, N. Group theoretical treatment of the planar internal rotation problem in (HF)2. *J Mol Spectrosc* 1985;109:134. https://doi.org/10.1016/0022-2852(85)90056-6.
109. Hougen, JT. A generalized internal axis method for high barrier tunneling problems, as applied to the water dimer. *J Mol Spectrosc* 1985;114:395. https://doi.org/10.1016/0022-2852(85)90234-6.
110. Coudert, LH, Hougen, JT. Tunneling splittings in the water dimer: further development of the theory. *J Mol Spectrosc* 1988;130:86. ibid., 139 (1990) 259. https://doi.org/10.1016/0022-2852(88)90286-x.
111. Coudert, LH, Hougen, JT. Tunneling splittings in (XY3)2-type dimers. *J Mol Spectrosc* 1991;149:73. https://doi.org/10.1016/0022-2852(91)90144-y.
112. Hougen, JT, Ohashi, N. Tunneling splitting patterns in two forms of the methanol-water dimer. *J Mol Spectrosc* 1993;159:363. https://doi.org/10.1006/jmsp.1993.1135.
113. Hollas, JM. *Modern spectroscopy*. England: John Willey & Sons Ltd; 2004.
114. Hougen, JT. Strategies for advanced applications of permutation-inversion groups to the microwave spectra of molecules with large amplitude motions. *J Mol Spectrosc* 2009;256:170–85. https://doi.org/10.1016/j.jms.2009.04.011.

115. Kasuya, T. Microwave studies of internal motions of hydrazine molecule. *Sci Pap Inst Phys Chem Res* 1962;56:1–39.
116. Papoušek, D, Sarka, K, Špirko, V, Jordanov, B. *Collect Czech Chem Commun* 1971;36:890–905.
117. Merer, AJ, Watson, JKG. Symmetry considerations for internal rotation in ethylene-like molecules. *J Mol Spectrosc* 1973;47:499–514. https://doi.org/10.1016/0022-2852(73)90097-0.
118. Łodyga, W, Kręglewski, M, Makarewicz, J. The Inversion–Torsion potential function for hydrazine. *J Mol Spectrosc* 1997;183:374–87. https://doi.org/10.1006/jmsp.1997.7271.
119. Tsunekawa, S, Kojima, T, Hougen, JT. Analysis of the microwave spectrum of hydrazine. *J Mol Spectrosc* 1982;95:133–52. https://doi.org/10.1016/0022-2852(82)90243-0.
120. Łodyga, W, Makarewicz, J. Torsion-wagging tunneling and vibrational states in hydrazine determined from its ab initio potential energy surface. *J Chem Phys* 2012;136:174301.
121. Ohashi, N, Lafferty, WJ, Olson, WB. Fourier transform spectrum of the torsional band of hydrazine. *J Mol Spectrosc* 1986;117:119–33. https://doi.org/10.1016/0022-2852(86)90096-2.
122. Ohashi, N, Olson, WB. Fourier transform spectrum of the second torsional band of hydrazine. *J Mol Spectrosc* 1991;145:383–91. https://doi.org/10.1016/0022-2852(91)90125-t.
123. Ohashi, N, Henry, L, Chazelas, J, Valentin, A. Further analysis of the fourier transform spectrum of the antisymmetric amino-wagging band of hydrazine. *J Mol Spectrosc* 1993;161:560–74. https://doi.org/10.1006/jmsp.1993.1258.
124. Ohashi, N, Masue, M. Explicit expressions for torsional-wagging-rotational Hamiltonian matrix elements in hydrazine-like molecules. *J Mol Spectrosc* 1991;150:238–54. https://doi.org/10.1016/0022-2852(91)90207-q.
125. Gulaczyk, I, Kręglewski, M, Valentin, A. The N–N stretching band of hydrazine. *J Mol Spectrosc* 2003;220:132–6. https://doi.org/10.1016/s0022-2852(03)00106-1.
126. Giguère, PA, Liu, ID. On the infrared spectrum of hydrazine. *J Chem Phys* 1952;20:136–40. https://doi.org/10.1063/1.1700155.
127. Tsuboi, M, Hamada, Y, Henry, L, Chazelas, J, Valentin, A. Amino wagging and inversion in hydrazine: K′ = 1 levels in the first excited state of the antisymmetric wagging vibration. *J Mol Spectrosc* 1984;108:328–42. https://doi.org/10.1016/0022-2852(84)90189-9.
128. Gulaczyk, I, Kręglewski, M. The symmetric amino-wagging band of hydrazine: assignment and analysis. *J Mol Spectrosc* 2008;249:73–7. https://doi.org/10.1016/j.jms.2008.02.012.
129. Gulaczyk, I, Pyka, J, Kręglewski, M. The effective Hamiltonian for the coupling between inversion–torsion states of hydrazine. *J Mol Spectrosc* 2007;241:75–89. https://doi.org/10.1016/j.jms.2006.11.001.
130. Suggs, HJ, Luskus, LJ, Kilian, HJ, Mokry, JW. *Exhaust gas composition of the F-16 emergency power unit*. Technical report USAF. SAM-TR-79-2, 2019.
131. Fourikis, N, Takagi, K, Morimoto, M. Detection of interstellar methylamine by its 2_{02}->1_{10} A_{a^{-}} state transition. *APJ (Acta Pathol Jpn)* 1974;191:L139. https://doi.org/10.1086/181570.
132. Kaifu, N, Morimoto, M, Nagane, K, Akabane, K, Iguchi, T, Takagi, K. Detection of interstellar methylamine. *APJ (Acta Pathol Jpn)* 1974;191:L135. https://doi.org/10.1086/181569.
133. Muller, S, Beelen, A, Guélin, M, Aalto, S, Black, JH, Combes, F, et al.. Molecules atz = 0.89. *A&A* 2011;535:A103. https://doi.org/10.1051/0004-6361/201117096.
134. Glavin, DP, Dworkin, JP, Sandford, SA. Detection of cometary amines in samples returned by Stardust. *Met Planet Sci* 2008;43:399. https://doi.org/10.1111/j.1945-5100.2008.tb00629.x.
135. Birge, RR, Hubbard, LM. Molecular dynamics of cis-trans isomerization in rhodopsin. *J Am Chem Soc* 1980;102:2195. https://doi.org/10.1021/ja00527a008.
136. Balogh-Nair, V, Nakanishi, K. The stereochemistry of vision. In: Tamm, Ch., editor. *New comprehensive biochemistry*. Amsterdam Mathies: Elsevier; 1982, vol 3 Stereochemistry.

137. Brito Cruz, RA, Cruz, CH, Pollard, WT, Shank, CV. Direct observation of the femtosecond excited-state cis-trans isomerization in bacteriorhodopsin. *Science* 1988;240:777. https://doi.org/10.1126/science.3363359.
138. Boxer, SG, Goldstein, RA, Lockhart, DJ, Middendorf, TR, Takiff, L. Excited states, electron-transfer reactions, and intermediates in bacterial photosynthetic reaction centers. *J Phys Chem* 1989;93:8280 and references therein. https://doi.org/10.1021/j100363a004.
139. Nakamura, S, Irie, M. Thermally irreversible photochromic systems. A theoretical study. *J Org Chem* 1988;53:6136. M. Irie. J Appl Phys 28 (1989). https://doi.org/10.1021/jo00261a035.
140. Uchida, K, Nakayama, Y, Irie, M. Thermally irreversible photochromic systems. Reversible photocyclization of 1,2-Bis(benzo[b]thiophen-3-yl)ethene derivatives. *Bcsj* 1990;63:1311. https://doi.org/10.1246/bcsj.63.1311.
141. Muszkat, KA, Fischer, E. Structure, spectra, photochemistry, and thermal reactions of the 4a,4b-dihydrophenanthrenes. *J Chem Soc B* 1967:662. https://doi.org/10.1039/j29670000662.
142. Gegiou, D, Muszkat, KA, Fischer, E. Temperature dependence of photoisomerization. V. Effect of substituents on the photoisomerization of stilbenes and azobenzenes. *J Am Chem Soc* 1968;90:3907. https://doi.org/10.1021/ja01017a002.
143. Frederick, JH, Fujiwara, Y, Penn, JH, Yoshihara, K, Petek, H. Models for stilbene photoisomerization: experimental and theoretical studies of the excited-state dynamics of 1,2-diphenylcycloalkenes. *J Phys Chem* 1991;95:2845. https://doi.org/10.1021/j100160a040.
144. Bâldea, I. Floppy molecules as candidates for achieving optoelectronic molecular devices without skeletal rearrangement or bond breaking. *Phys Chem Chem Phys* 2017;19:30842–51. https://doi.org/10.1039/c7cp06428d.
145. Frisbie, CD. Designing a robust single-molecule switch. *Science* 2016;352:1394–5. https://doi.org/10.1126/science.aag0827.
146. Kumar, AS, Ye, T, Takami, T, Yu, B-C, Flatt, AK, Tour, JM, et al.. Reversible photo-switching of single azobenzene molecules in controlled nanoscale environments. *Nano Lett* 2008;8:1644–8. https://doi.org/10.1021/nl080323+.
147. Mativetsky, JM, Pace, G, Elbing, M, Rampi, MA, Mayor, M, Samorì, P. Azobenzenes as light-controlled molecular electronic switches in nanoscale Metal–Molecule–Metal junctions. *J Am Chem Soc* 2008;130:9192–3. https://doi.org/10.1021/ja8018093.
148. Yasuda, S, Nakamura, T, Matsumoto, M, Shigekawa, H. Phase switching of a single isomeric molecule and associated characteristic rectification. *J Am Chem Soc* 2003;125:16430–3. https://doi.org/10.1021/ja038233o.
149. Zhang, X, Wen, Y, Li, Y, Li, G, Du, S, Guo, H, et al.. Molecularly controlled modulation of conductance on azobenzene monolayer-modified silicon surfaces. *J Phys Chem C* 2008;112:8288–93. https://doi.org/10.1021/jp711808p.
150. Ferri, V, Elbing, M, Pace, G, Dickey, MD, Zharnikov, M, Samorì, P, et al.. Light-powered electrical switch based on cargo-lifting azobenzene monolayers. *Angew Chem Int Ed* 2008;47:3407–9. https://doi.org/10.1002/anie.200705339.
151. Bâldea, I. A quantum chemical study from a molecular transport perspective: ionization and electron attachment energies for species often used to fabricate single-molecule junctions. *Faraday Discuss* 2014;174:37–56.

Donata Pluskota-Karwatka and Marcin Hoffmann

4 Computational studies on statins photoactivity

Abstract: Statins are popular drugs widely prescribed to control hypercholesterolaemia and to prevent cardiovascular diseases. Synthetic statins constitute a group of pharmaceuticals which are very sensitive to exposure to light in both UVA and UVB ranges. Light, by causing drugs degradation, can essentially change their pharmaceutical properties leading even to the loss of therapeutic activity and/or to the formation of deleterious photoproducts. Drugs which exhibit photochemical reactivity may elicit undesired adverse effects. A detailed understanding of mechanisms involved in molecular basis of these effects origin is very important for evaluating the photobiological risk associated with therapy in which drugs prone to exposure to light are involved. In this work we critically discussed finding regarding the mechanisms of synthetic statins phototransformation. We showed inconsistency of some previously reported facts and revised earlier presented studies. We also completed the lack of information on pitavastatin photobehaviour. This all together resulted in proposal of new schemes for the statins photodecomposition. We reviewed data derived from both experimental and computational methods. Studies of photochemical problems by the use of theoretical methods enable getting insight into areas of some fascinating events that experimental techniques can touch only indirectly. Besides effect of light, phenomenon of statins' sensitivity to pH and resulting implications were discussed. Statins undergo pH-dependent interconversion between their pharmacologically active hydroxy acid and inactive lactone forms, and it was shown that for both forms, drugs' interactions should be considered. Knowledge of the statins interconversion mechanisms is important for understanding how differences in the structures of their molecules can affect the drugs' activity.

Keywords: computational chemistry, HMG-CoA reductase inhibitors, lactonization, photodegradation, statins

4.1 Introduction

Photostability of drugs became an important topic in the field of pharmaceutical research. The number of drugs being photochemically instable is steadily increasing, and photoinstability remains one of the major pathways of drugs decomposition. Photosensitivity of drugs can have severe medicinal consequences, may result

This article has previously been published in the journal Physical Sciences Reviews. Please cite as: Pluskota-Karwatka, D., Hoffmann, M. Computational studies on statins photoactivity *Physical Sciences Reviews* [Online] 2021, 6. DOI: 10.1515/psr-2020-0033

https://doi.org/10.1515/9783110678215-004

in potency loss, in altered efficacy as well as in adverse biological effects. This is because drugs decomposition also involves generation of reactive oxygen species or energy transfer. These processes can in turn contribute to formation of diverse photoproducts which are able to interact with biomolecules and thus constitute a source of various side effects. Studies of the mechanisms of photochemical reactions and identification of photodegradation products are essential for establishing conditions of safe and efficient treatment. They may also provide information on the mode of stabilization of the active ingredients in a drug product.

3-Hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors known as statins belong to the group of drugs which are susceptible to light exposure. Lovastatin, pravastatin and simvastatin are derived from fungal metabolites [1], while fluvastatin (FLV), atorvastatin (ATV), pitavastatin (PTV) and rosuvastatin (RSV) (Figure 4.1.1) are fully synthetic compounds [2].

Molecules of synthetic statins consist of two structural components; a dihydroxyheptanoic or heptenoic acid unit and a central ring system with lipophilic substituents [3] (Figure 4.1.1). The dihydroxy acid component is essential for the drugs' activity; only the linear forms of statins are biologically active; however, *in vivo* δ-lactone forms of drugs are enzymatically hydrolysed to the corresponding β-hydroxy acids [4].

Statins act by blocking biosynthesis of cholesterol. This leads to the increase in the number of high affinity low-density lipoprotein (LDL) receptors and thus LDL and cholesterol blood levels decrease [2[, [3], [4]. Due to the structural similarity to 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA), (Figure 4.1), statins competitively inhibit HMG-CoA reductase which catalyses the conversion of HMG-CoA to mevalonate, the rate-limiting step in *de novo* cholesterol synthesis [5]. Although mechanism of action is common for all statins, differences in the compounds' structures may contribute to differences in potency of HMG-CoA reductase inhibition [4]. Statins are very sensitive to light, possess the ability to absorb UV radiations in both UVA and UVB ranges and exhibit high photochemical reactivity [6], [7], [8],[9], [10]. Upon exposure to light, a variety of drugs can elicit undesired effects such as phototoxicity, photoallergy or even photocarcinogenicity [11], [12], [13]. To evaluate the photobiological risk of a therapy with photosensitive drug, it is necessary to elucidate the mechanisms involved. Such mechanisms are often very complex; however, photophysical and photochemical studies are important for understanding the key early events resulting in potential drug phototoxicity.

A very intriguing statins' feature is also their pH sensitivity *in vivo*. 3,5-Diol moieties of the statins' heptenoic and heptanoic acid side chains undergo reversible pH-dependent lactonization [14], [15]. At the physiological and higher pH, lactone form is unstable and hydrolyses yielding the β-hydroxy acid form. Under acidic conditions, intramolecular esterification leads to the formation of lactone. *In vivo* both forms coexist in equilibrium, and it was shown that for both forms, drugs' interactions should be considered [16], [17]. Studies on the statins interconversion mechanisms may help in

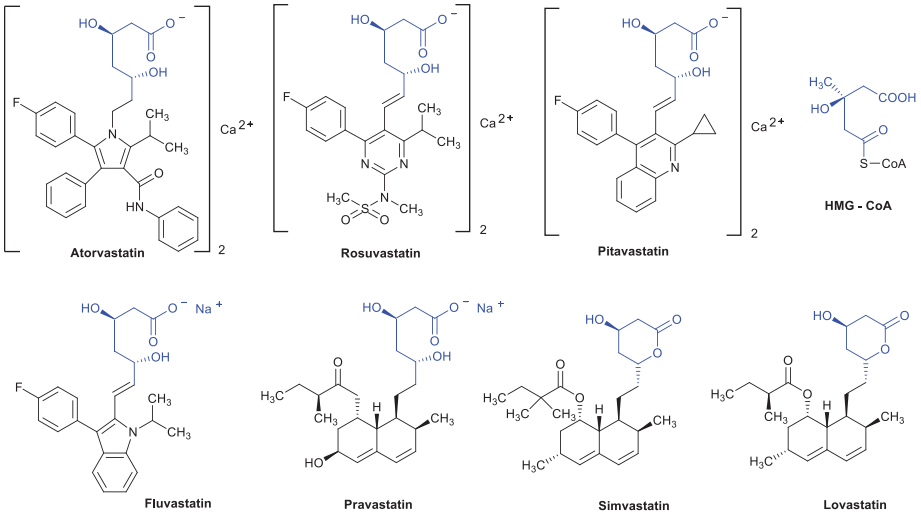


Figure 4.1.1: Structures of statins and HMG-CoA. Fluvastatin, rosuvastatin, pitavastatin, atorvastatin and pravastatin are presented in their linear β-hydroxy acid forms, while simvastatin and lovastatin in their cyclic β-hydroxy-δ-lactone forms.

understanding how differences in the structures of the molecules affect the drugs' activity.

The primary focus of this review is to critically summarize the recent findings regarding the statins photodegradation and lactonization. The work mainly deals with mechanisms responsible for statins' transformations and includes studies performed with use of computational chemistry methods which often enable to explore such areas of photochemical processes that experiment can touch only indirectly.

4.2 Photochemistry of rosuvastatin and pitavastatin

One serious consequence of the presence in the synthetic statins' structures of conjugated aromatic ring systems is the compounds' susceptibility to light-induced decomposition. In the rosuvastatin molecule, the pyrimidine ring bonded to fluorophenyl moiety and conjugated with the heptenoic acid side chain constitutes a centre of the lipophilic unit. In pitavastatin, the pyrimidine ring is replaced by a pyridine one. Molecules of both compounds contain similar π-electron systems which are involved in photodegradation processes. Upon exposure to light, the statins undergo photolysis that leads to the formation of analogous sets of photoproducts [6], [9], [11], [18], (Figure 4.2.1, 4.2.2).

The photochemical behaviour of rosuvastatin in water under solar and UV irradiation was studied by Astarita et al. [18]. The statin was found to decompose

completely after 4 days of exposure to sunlight and after 8 h of irradiation with a 500 W UV lamp. Regardless of the irradiation source formation of the same photoproducts was confirmed by TLC and NMR analyses. Irradiation of RSV under argon atmosphere resulted in the identification of RSV1 and RSV2. RSV1 (diastereomeric mixture) was determined as rosuvastatin main and primary photoproduct. Its photostability was evaluated by irradiation with solar light. After 3 days most of the starting material was transformed into RSV2 and RSV3 indicating that these compounds are the secondary photoproducts of rosuvastatin. On the basis of these findings, it was proposed that photochemical electrocyclization is the main light-induced reaction leading to the formation of RSV1 through undetected intermediate [18], (RSVI1, Scheme 4.2.1). This suggestion was based on the fact that this kind of cyclization is a characteristic reaction for *o*-vinylbiphenyl compounds [19], [20] to which rosuvastatin could be considered to belong. Support for the intermediate formation can be derived from a transient absorption spectrum of RSV obtained in laser flash photolysis experiments [11]. The spectrum showed a long-lived individuum that was assigned to the intermediate of RSV cyclization. The intermediate associated with RSV cyclization must be considered as a mixture of two diastereomeric 8a,9-dihydrophenanthrenes. Each of them contains a novel stereogenic centre. Due to loss of aromaticity, these intermediates are unstable and undergo a thermal intramolecular [1], [5] sigmatropic hydrogen shift to yield two diastereomers of RSV1. Absolute configurations (Figure 4.2.3) of the diastereomeric photoproducts in the forms of their lactones were determined by combination of calculation method with data obtained from spectroscopic and spectrometric studies [6].

RSV2 and RSV3 were proposed to arise from benzylic radical (RSVI2, Scheme 4.2.1), generated by photo-induced cleavage of the bond between the heptenoic acid side chain and the carbon atom of the RSV1 newly formed ring [18]. RSV2 can be afforded by abstraction of a hydrogen radical from a side-chain radical fragment or other surrounding molecule, while RSV3 can be formed by the loss of hydrogen radical. RSV3 was also thought to be produced by photooxidation of RSV2; however, this hypothesis was not verified by experimental studies. Alternatively, both secondary photoproducts of RSV were assumed to arise from the benzylic radical disproportionation [18] (Scheme 4.2.1).

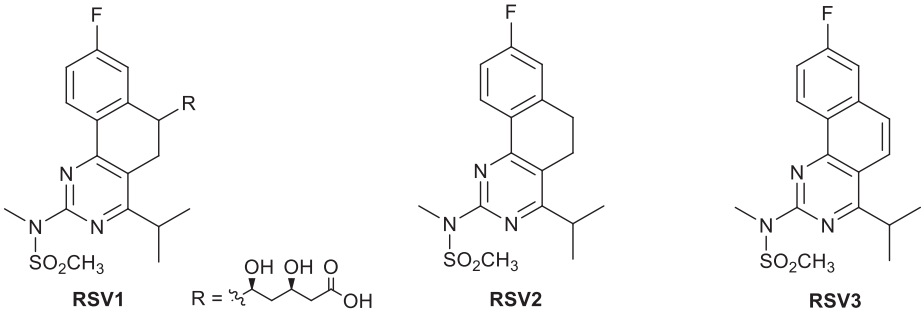


Figure 4.2.1: Structures of rosuvastatin photoproducts.

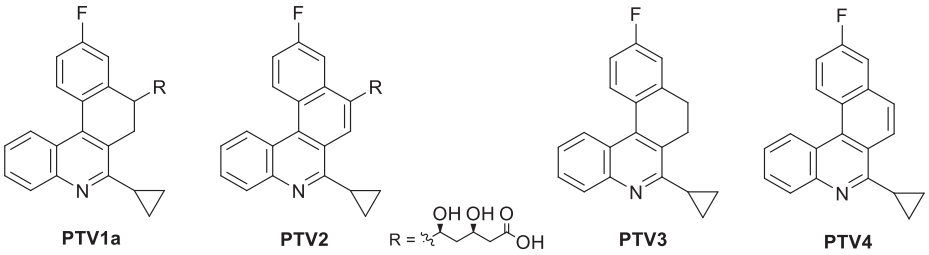


Figure 4.2.2: Proposed structures of pitavastatin photoproducts.

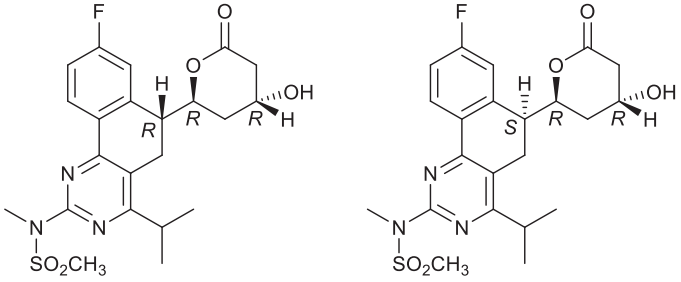
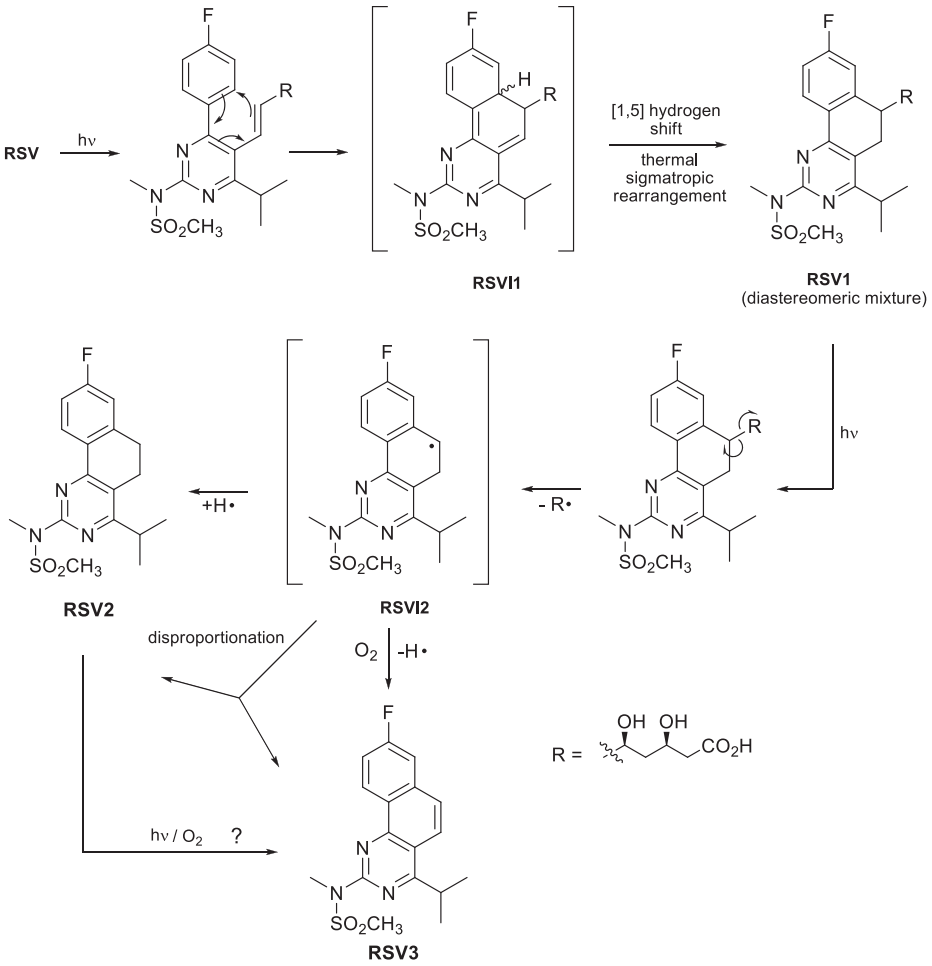


Figure 4.2.3: Absolute configuration of two diastereomers of RSV1 in the form of lactones.

Interesting results were provided by investigations aimed at photobehaviour of RSV1 [11]. It was found that deactivation pathways of singlet excited state of the compound may occur *via* fluorescence, phosphorescence or intersystem crossing. Moreover, laser flash photolysis studies revealed a triplet–triplet energy transfer from the RSV1 triplet excited state to thymidine, resulting in the formation of the nucleoside cyclobutane dimers. Investigations of RSV1-mediated oxidation character performed with the using of tryptophan indicated that an electron transfer process giving rise to the tryptophanyl radical as well as singlet oxygen-mediated oxidation takes place [11].

Findings regarding photophysical characterization of RSV and pathways suggested for its photochemical transformation as well as photoproducts formation are summarized in Scheme 4.2.2.

Phototransformation of pitavastatin was examined by Grobelny et al. [9]. The studies resulted in identification of four photoproducts, whose structures were suggested on the basis of HRMS spectra obtained for these compounds [9] (Figure 4.2). Due to similarities in structures of the respective RSV and PTV photoproducts, it seems that analogous mechanisms are responsible for both statins photodegradation. However, the lack in the scientific literature of comprehensive studies devoted to the statin photobehaviour made impossible to verify this thesis. For this reason, recently in our laboratory, studies aimed at clarifying photophysical and photochemical properties of



Scheme 4.2.1: Mechanisms proposed for RSV photoproducts formation.

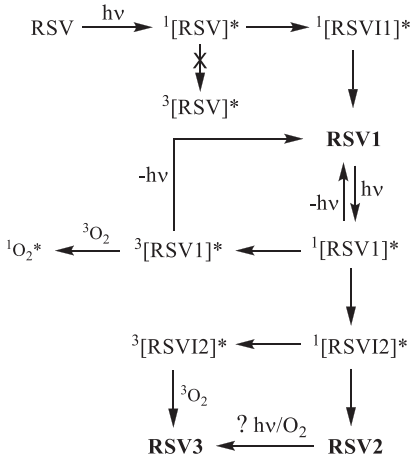
pitavastatin were undertaken [21]. Our attention was focused on elucidating mechanisms responsible for pitavastatin primary photoproducts formation. In our investigations both experimental and computational methods were used.

On the basis of data obtained from transient absorption spectroscopy, it was found that PTV in its excited singlet state ($^1PTV^*$) undergoes an intersystem crossing yielding the excited triplet state $^3PTV^*$ which was, however, generated with a low efficiency due to the fact that the statin exhibits fluorescent properties. This finding was confirmed by nanosecond laser flash photolysis (LFP).

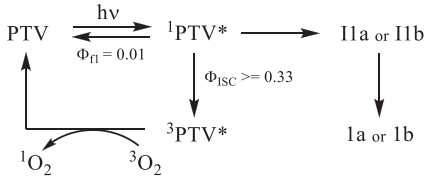
LEP revealed also that long-lived intermediate (most likely I1a or I1b) is formed directly from PTV being in its excited singlet state. This intermediate was supposed to be a precursor of PTV primary photoproduct 1a (or PTV1b, see Scheme 4.2.3). On

the basis of these results, it can be concluded that upon excitation PTV yields both singlet and triplet states, which undergo different, parallel reactions.

Interestingly, PTV was shown to generate the triplet state. This finding confirms that PTV being in its excited singlet state undergoes not only conversion to the intermediates I1a/I1b, but also an intersystem crossing to the excited triplet state (Scheme 4.2.3). This observation is in the perfect agreement with the results obtained from studies performed using nanosecond LFP and femtosecond transient absorption spectroscopy techniques. Determination of quantum yield of $^1O_2^*$ generation by PTV was challenging because formation of the primary photoproduct was observed immediately after exposure of the statin solution to the light. However, the initial value of the yield was estimated to be 0.33 ± 0.03. Moreover, it was found that singlet oxygen is generated not only by PTV but likely also by a photoproduct 1a (or 1b) formed in the photoreaction [21].



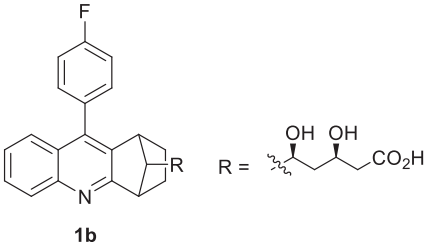
Scheme 4.2.2: Photochemical transformation of rosuvastatin and its primary photoproduct.



Scheme 4.2.3: Photochemical transformation of pitavastatin [21].

The structure of PTV primary photoproduct (PTV1a, Figure 4.2.2) was previously suggested by Grobelny et al. We assumed the structure, but also proposed an alternative one (1b, Figure 4.2.4). This proposition was made basing on the fact that in the literature

there are examples of different organic systems containing cyclopropane ring which are thermally and photochemically instable [22].

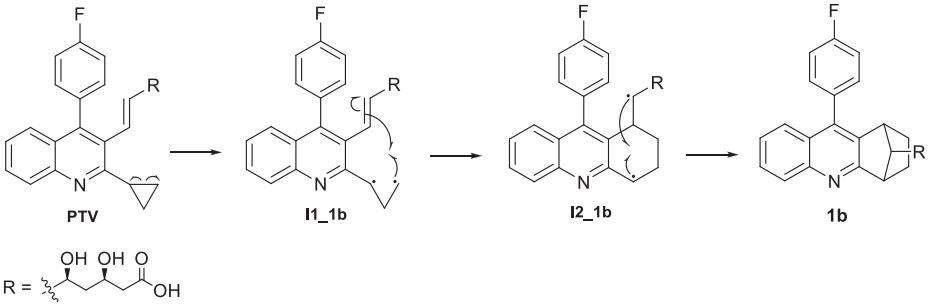


1b

Figure 4.2.4: An alternative structure proposed for pitavastatin primary photoproduct.

In silico studies revealed that the structure 1b is characterized by about 15 kcal/mol lower energy comparing to the 1a. Formation of 1b was considered by mechanism *via* diradical species (Scheme 4.2.4) in triplet and/or singlet state

Mechanisms of both compounds, 1a and 1b formation were examined by the use of quantum-chemical methods. The details can be found in the reference [21].



Scheme 4.2.4: Mechanism of 1b formation [21].

4.3 Photochemistry of fluvastatin

Photobehaviour of fluvastatin is determined by the presence in its molecule of π-electron system derived from indol ring bonded to fluorophenyl moiety and conjugated with the acid side chain. Photostability of FLV was examined in water as well as in methanol solutions, upon exposure to solar and UV irradiation generated by a 500 W high-pressure Mercury lamp [7] and two HPW 125 Philips lamps [12], and the statin was found to exhibit high photochemical reactivity expressed by formation of a range of photoproducts (Figure 4.3.1).

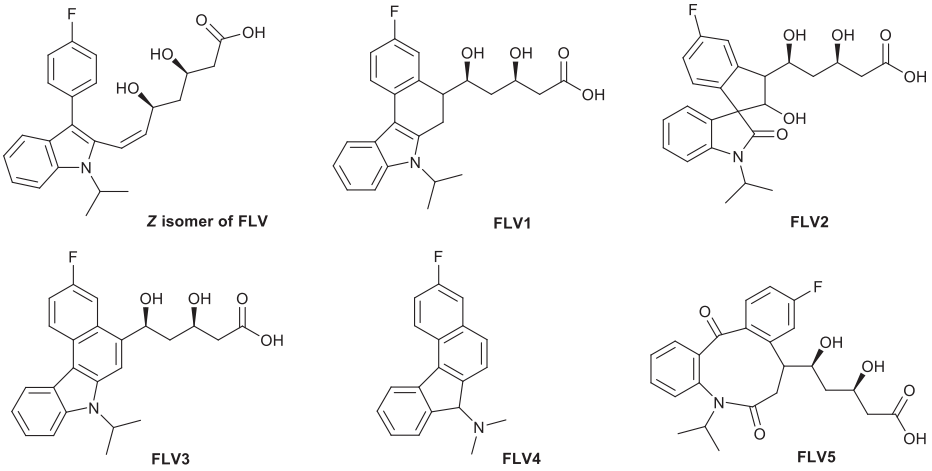
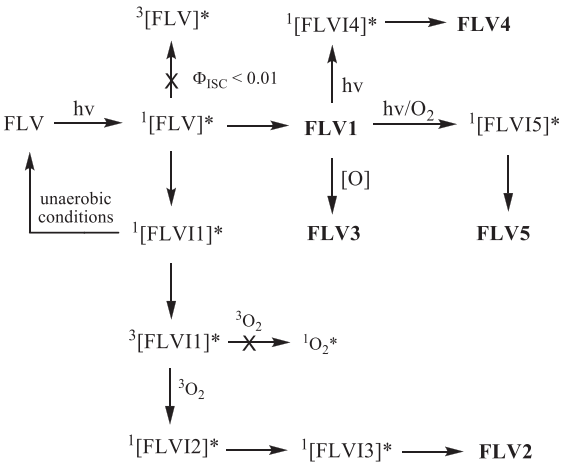


Figure 4.3.1: Structures of fluvastatin photoproducts.

However, reports on the FLV photostability are inconsistent. Investigations of Mielcarek et al. resulted in the identification of *Z*-isomer of fluvastatin and FLV1 (Figure 4.3.1), as two photoproducts detected after 2 min of exposure to light at λ = 365 nm [23, 24]. It should be noted that formation of these compounds was determined only on the basis of HPLC-MS studies. In the work of Cermola et al., the FLV1 was also found to be fluvastatin degradation product. However, the study refers to results received after 1 h of the statin exposure to light at λ > 300 nm and under these conditions formation of additional to FLV1 photoproducts was reported [7] (Figure 4.3.1, FLV2-FLV5). Moreover, the authors did not propose fluvastatin *Z*-isomer to be involved in the drug photodecomposition and found that FLV1 was the only photoproduct formed under argon atmosphere. In their study photostability of FLV1, FLV3 and FLV4 was also evaluated by irradiating of these compounds for 4 h with UV light (λ > 300 nm). On the basis of NMR analysis, the authors indicated that FLV1 was converted to FLV3, FLV4 and FLV5, while FLV3 and FLV4 were stable under the applied conditions. Prolonged irradiation of FLV3 and FLV4 resulted in the slow decomposition of the former compound, while FLV4 remained unchanged. One can reflect on the fact why FLV2 was not subjected to stability studies. The photoproducts' structures presented by Cermola et al. were determined by spectroscopic methods. Fluvastatin photodegradation was investigated also by Viola et al. with respect to the drug phototoxicity. In their study formation of almost the same set of photoproducts as reported by Cermola et al. was observed except FLV3 which was not identified [12].

Recently, the detailed kinetic studies on fluvastatin photodegradation were undertaken by our research group. We focused our attention on fluvastatin primary photochemistry and applied both experimental and computational methods to get insight into the possible mechanisms responsible for the drug photoproducts formation [25].

Upon a few seconds of exposure of air-saturated aqueous solution of FLV to UV light (355 nm, OPSL CW laser, 10 mW reaching the sample), a significant decrease in absorption was observed suggesting the statin fast photodegradation. After 60 s of irradiation with light at λ = 355 nm and 10 mW power, FLV1 and two diastereomers of FLV2 (FLV2a, FLV2b) were identified as FLV primary photoproducts. Using transient absorption spectroscopy technique, a transient individuum assigned to the intermediate state ^{1}FLVI1* (Scheme 4.3.1) was observed. It was also found that generation of singlet oxygen is not involved in the drug photodecomposition indicating that the excited triplet state of fluvastatin is not populated efficiently. On the basis of these findings, we proposed the scheme for FLV photochemical transformation (Scheme 4.3.1). This scheme shows that in the excited S_1* state fluvastatin can afford FLV1 but can also yield an intermediate ^{1}FLVI1* which under anaerobic conditions can be converted back to FLV but may also undergo intersystem crossing forming ^{3}FLVI1*. This intermediate exhibits a lifetime sufficiently long to undergo reaction with oxygen yielding FLVI2, then FLVI3 and finally two diastereomers of FLV2.



Scheme 4.3.1: Fluvastatin photochemical transformation.

Based on the literature data referring to photochemical behaviour of *o*-vinylbiphenyl compounds, Cermola et al. suggested pathways expected for fluvastatin photoproducts formation. FLV1 was assumed to be formed by photochemical electrocyclization followed by a thermal sigmatropic [1], [5] hydrogen shift, the mechanism analogous to that presented in scheme 1 for formation of RSV1. We investigated this mechanism theoretically using the density functional theory (DFT) methods and our studies support this mechanism. We found and analysed two reaction pathways for cyclization leading to FLV1 formation in which two fluvastatin conformers differing in the orientation of the acid side chain with respect to the fluorobenzene ring are involved [25]. The pathway in which fluvastatin conformer with the heptenoic

acid chain situated in front of the fluorobenzene ring is involved was proposed to be the favourable reaction route because this conformer was characterized by ca. 7 kcal/mol lower relative energy compared to the second possible one having the acid chain placed behind the fluorobenzene ring [25]. To get a deeper insight into the mechanism leading to the formation of FLV1, we performed further calculations in order to find avoided crossings or conical intersections between ground and excited states. The obtained results suggest that indeed after FLV is excited, it relaxes its geometry and may then through CI/AC reach the geometry of the ground state TS1 structure and further intermediate A.I1 (Figure 4.3.2).

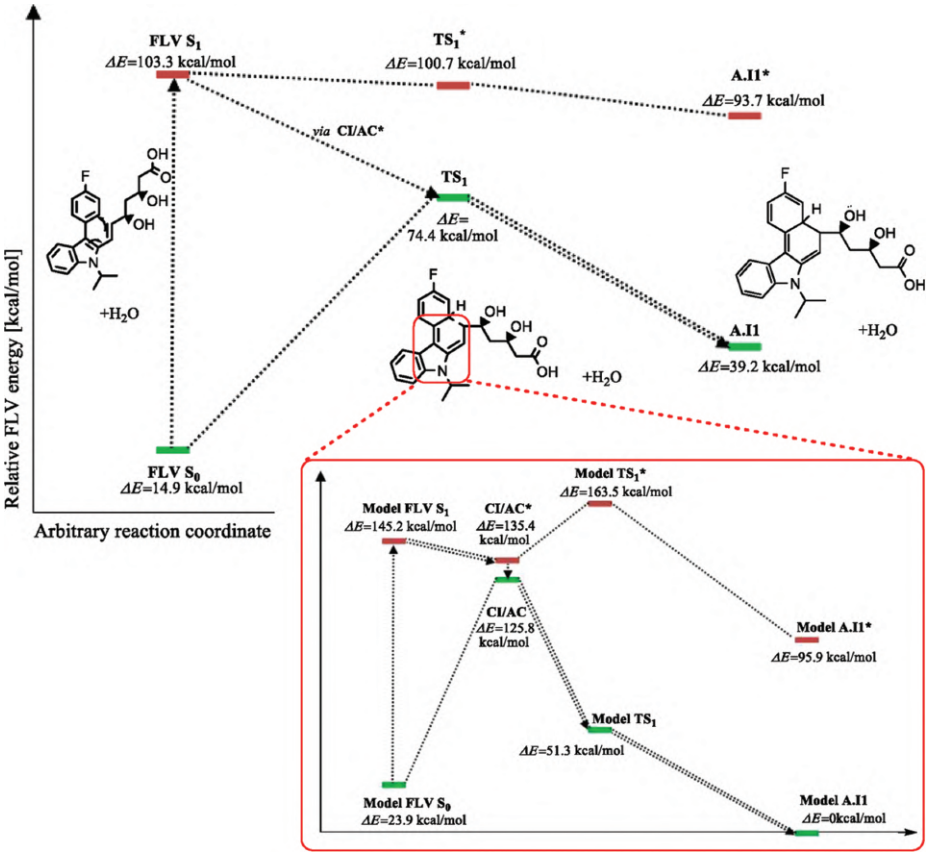
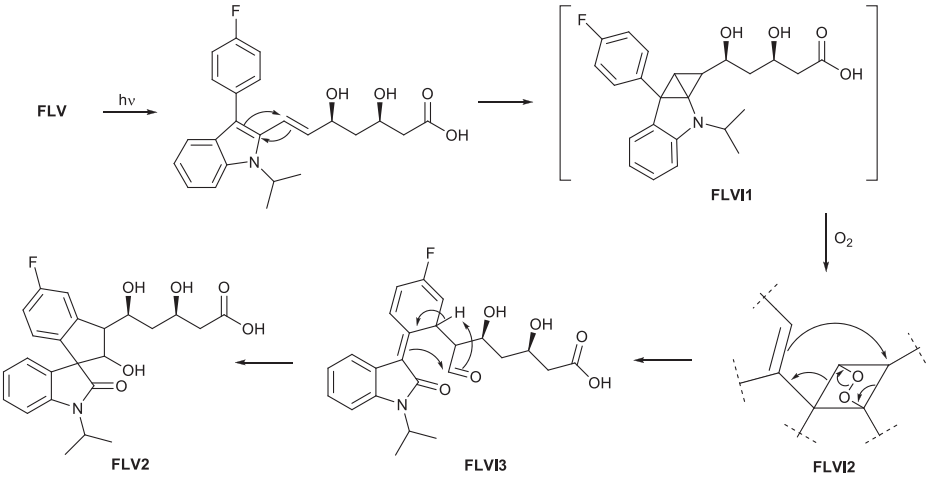


Figure 4.3.2: Energy profile of mechanism for A.I1 formation – an intermediate leading to FP1 formation (pathway A). The red rectangle presents the energy profile of mechanism for A.I1 model fragment formation [25].

It was proposed that diastereomeric FLV2a and FLV2b are formed by oxygen addition to fluvastatin followed by rearrangement of the resulted diepoxidic intermediate [7]. It was also suggested that singlet oxygen is involved in the fluvastatin photooxygenation,

but it was not excluded that other oxygenating species instead of singlet oxygen are responsible for the diepoxide generation. Contrary to these suggestions, we found that the lifetime of the excited singlet ^{1}FLV* was too short for the reaction with oxygen (or any other efficient bimolecular reaction); therefore, we proposed entirely new mechanism for arising of FLV2 (Scheme 4.3.2).



Scheme 4.3.2: Mechanisms proposed for FLV2 formation [26].

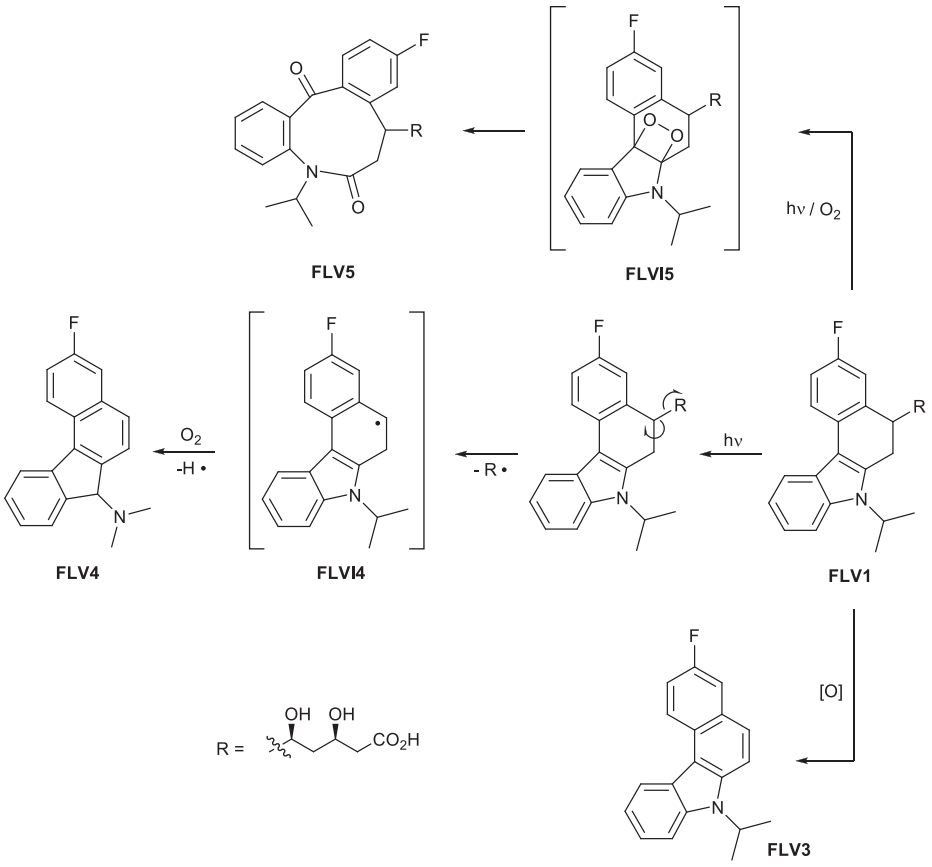
This mechanism assumes valence isomerization of fluvastatin and formation of the intermediate FLVI1 which is involved in photooxygenation leading to FLV2a and FLV2b. This mechanism is in good agreement with experimental data which indicate that the transient absorption does not decay to the initial value confirming the existence of a relatively long-lived species (FLVI1) that is able to undergo a bimolecular reaction. DFT calculations showed that from two possible routes for reaction with oxygen, addition up to the flat intermediate FLVI1* ring was energetically more favourable [25].

Formation of FLV3, FLV4 and FLV5 was proposed as depicted in Scheme 4.3.3.

4.4 Photochemistry of atorvastatin

Atorvastatin (ATV) molecule contains pyrrole ring having two phenyl substituents, and π-electron system of these moieties is the major factor that determines photoreactivity of the statin.

It was reported that sunlight irradiation of atorvastatin aqueous solution leads to the formation of four photoproducts (ATV1–ATV4, Figure 4.4.1), those structures were elucidated by NMR and MS techniques [8].



Scheme 4.3.3: Mechanisms proposed for FLV3, FLV4 and FLV5 formation.

The obtained results indicate that photoproducts ATV1 and ATV2 arose from oxidation of the pyrrole moiety followed by an alkyl or an aryl shift with a lactam ring formation. ATV3 is a lactone form of ATV2 yielded by lactonization of the γ-hydroxyacid chain, while ATV4 is the ATV2 derivative afforded by photochemical electrocyclization followed by oxidation under aerobic conditions of an initially formed intermediate. The same photoproducts were shown to be formed upon ATV exposure to UV irradiation generated by a 500 W high-pressure Mercury lamp [8]. Oxygen involvement in the ATV phototransformation was confirmed by the fact that the drug left unaltered after irradiation carried out under argon atmosphere. Results of ATV photodegradation studies performed in the presence of radical inhibitor, singlet oxygen quencher and singlet oxygen sensitizer suggested that the drug may undergo singlet oxygen–mediated photooxygenation [8]. On the basis of these findings and literature data, Cermola et al. proposed a plausible mechanistic explanation of ATV1 and ATV2 formation [8] (Scheme 4.4). As it is shown in the scheme, energy transfer

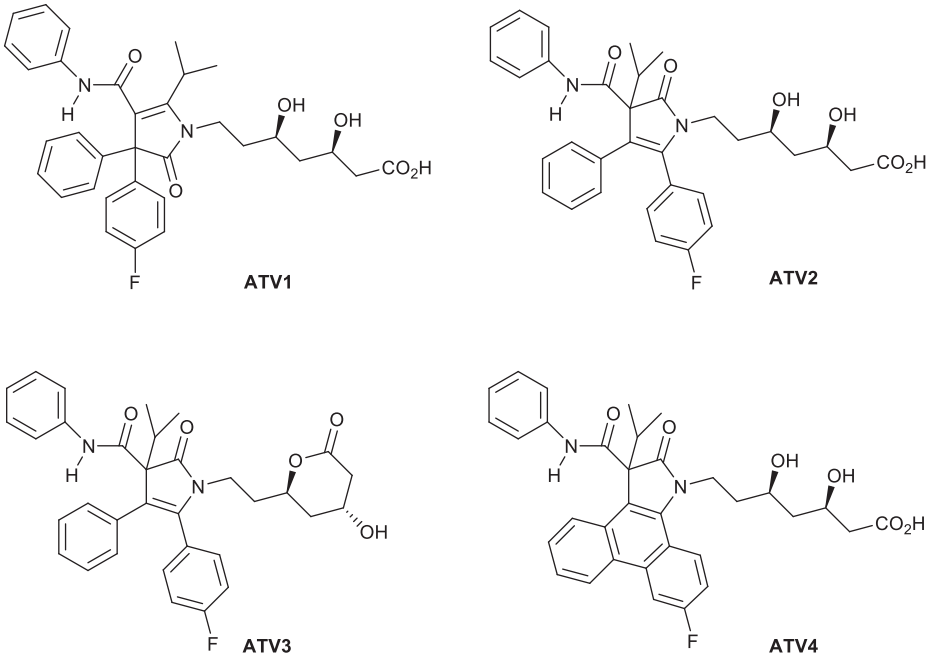
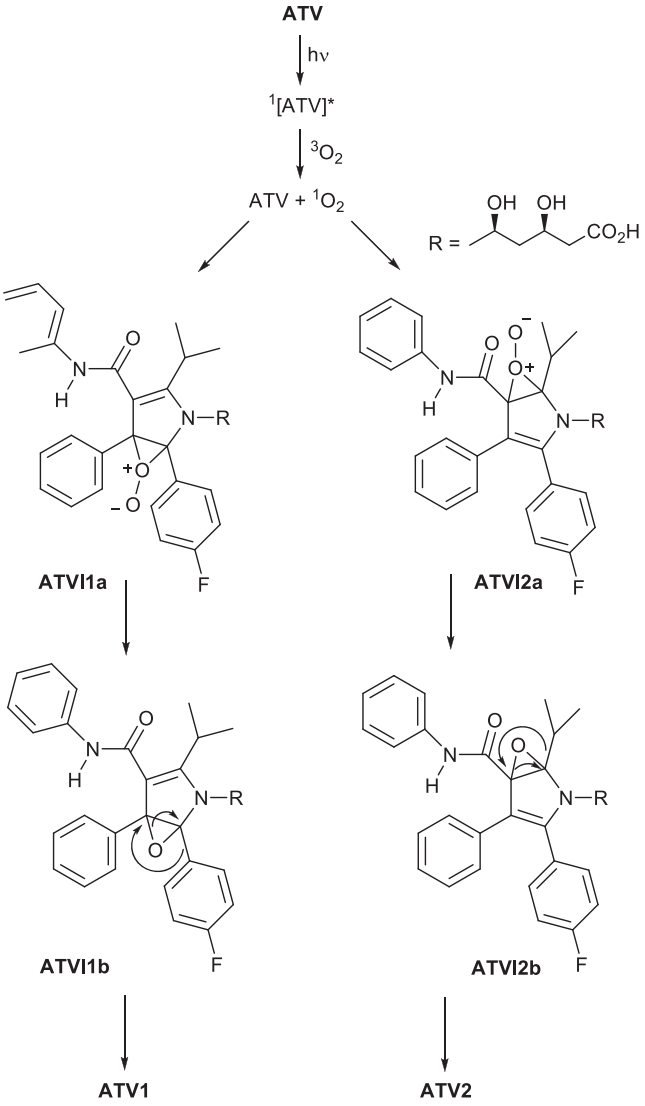


Figure 4.4.1: Structures of atorvastatin photoproducts [26].

from the excited ATV molecule to oxygen gives rise to singlet oxygen which may undergo addition to the ATV pyrrole moiety yielding intermediates ATVI1a and ATVI2a which evolve to oxiranes ATVI1b and ATVI2b. ATV1 and ATV2 are formed after migration of aryl and alkyl group respectively. It should be stressed that the oxiranes were not detected in experimental studies, and this fact was explained by probable rearrangement of ATVI1b and ATVI2b to the corresponding carbonyl compounds [8]. Support for involvement of the oxiranes in ATV oxidation was provided by using dimethyldioxirane, a known epoxidizing agent for various unsaturated compounds, in reaction with ATV. Indeed after 4 h, ATV1 and ATV2 were detected in the reaction mixture [8].

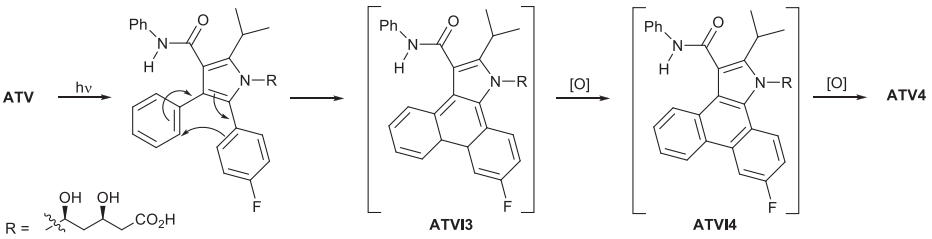
Findings reported by Cermola et al. were evaluated in the later study which dealt with ATV photophysical properties [13]. To elucidate mechanism of ATV photolysis, steady-state and time-resolved spectroscopic experiments were performed to get insight into generation and behaviour of ATV-related excites states [13]. Transient absorption spectrum obtained from LFP studies exhibited two maxima at λ = 360 and 580 nm, and monoexponential decay (τ = 41 μs), on which oxygen had no significant influence. Based on these observations it was concluded that the detected long-lived individuum does not correspond to the triplet–triplet transition of ATV, and the drug direct involvement in singlet oxygen generation was ruled out. Taking into account these results and considering atorvastatin photochemistry, the transient spectrum



Scheme 4.4.1: Mechanistic explanation of ATV1 and ATV2 formation [26].

was assigned to *trans*-dihydrophenanthrene, the intermediate (ATVI3, Scheme 4.4.2) of the stilbene-like photocyclization.

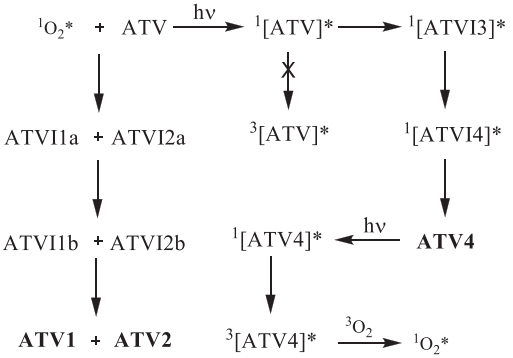
This assignment is in agreement with the literature data [27], [28] which show that absorption spectra of related species exhibit two maxima in the region 300–360 and 440–640 nm, and with the earlier reported ATV photoreactivity resulted in isolation and structural characterization of ATV4 [8]. Photochemical formation and behaviour of ATV4 were investigated in detail, and ATVI4 (Scheme 4.4.2) was proposed to



Scheme 4.4.2: ATV photocyclization leading to the formation of intermediate ATVI3.

be a real (first-stage) singlet oxygen sensitizer which after self-sensitization would be oxygenated to ATV4 [13]. This hypothesis was confirmed by photolysis of ATV in deaerated methanol in the presence of iodine as oxidant. Under these conditions formation of ATVI4 was detected by HPLC-MS analysis. However, due to the intermediate nonpersistence caused by its instability under air, ATVI4 seems to be a very poor photosensitizer, while ATV4 appears to be a good candidate for such activity. Indeed the ability of ATV4 to sensitize singlet oxygen was confirmed by spin trapping experiments, through conversion of TEMP (2,2,6,6-tetramethylpiperidine) to the stable free-radical TEMPO ((2,2,6,6-tetramethylpiperidin-1-yl)oxyl). Thus, it was demonstrated that ATV4 behaves as an efficient singlet oxygen photosensitizer giving rise to oxidation of biomolecules (for example, tryptophan) and also atorvastatin [13]. Oxidation of ATV leads to the formation of ATV1 and ATV2 and is in agreement with the nature of the steady-state photolysis products, resulting from photooxidation of the pyrrole moiety [8].

On the basis of all these results, scheme illustrating ATV photochemical transformations can be proposed (Scheme 4.4.3).



Scheme 4.4.3: Atorvastatin photochemical transformation.

Photochemistry of ATV was also evaluated with respect to influence of the phenolic hydroxyl group of *ortho* hydroxy atorvastatin metabolite [29] (ATV-OH, Figure 4.4.2). Based on similarity between the transient absorption spectra obtained for ATV and ATV-OH, and on the finding that, like for ATV, transient absorption of ATV-OH was not quenched by oxygen, it was concluded that the same kind of intermediates are formed upon direct excitation of ATV and ATV-OH [29]. It was also demonstrated that singlet oxygen reactivity towards ATV-OH is almost the same like for ATV indicating that the presence of the ATV-OH phenolic group does not affect the reaction. This fact was explained by the results of theoretical calculations which showed that introduction of the hydroxyl group to ATV structure did not change its electron density [29].

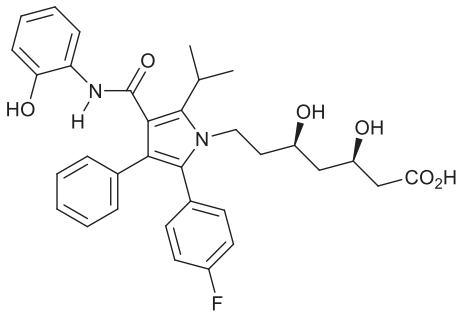


Figure 4.4.2: Atorvastatin *ortho* hydroxy metabolite.

4.5 Effect of pH

Statins undergo pH-dependent interconversion between their pharmacologically active hydroxy acid and inactive lactone forms, both forms are found in human plasma after the drugs administration [14], [15], [30], and in case of many statins the lactone form is as abundant as the hydroxy acid form [16], [17]. Moreover, most of the metabolites of statins present in human plasma were hypothesized to be generated by interconversion of lactone metabolites which are formed by CYP3A4 [17] (cytochrome P450). Crystal structures of the catalytic portion of human HMG-CoA reductase bound to six different statins clearly indicate that linear forms of statins are predisposed for binding to the enzyme, as the terminal carboxylate group forms salt bridges with Lys692 and Lys735, while the δ hydroxy group serves as a charge-assisted hydrogen bond donor to Glu559 and as a hydrogen bond acceptor from Lys691 [4], [31]. Although only the linear forms reduce cholesterol levels, drug interactions should be considered for both acid and lactone forms of statins as these forms exhibit different behaviour in terms of inhibitory effects on CYPs metabolic activities and MDR1 (multidrug resistance protein 1) transporting activity [16], [17].

Indeed atorvastatin and rosuvastatin in their acid forms have no or minimal effects on CYP2C9 or CYP3A4/5 activities, while their lactone forms were elucidated to inhibit CYP2C9 or CYP3A4/5 activities, and reported drug interactions of these statins were explained by the action of their lactone forms generated in the body, not by that of the acid forms present in the medicines [16]. The lactone forms of statins are more lipophilic than the corresponding acid forms, and the lipophilicity is thought to be one of the important factors having impact on the inhibitory effects [16]. This is not surprising as the affinity for CYPs is defined by lipophilicity [16].

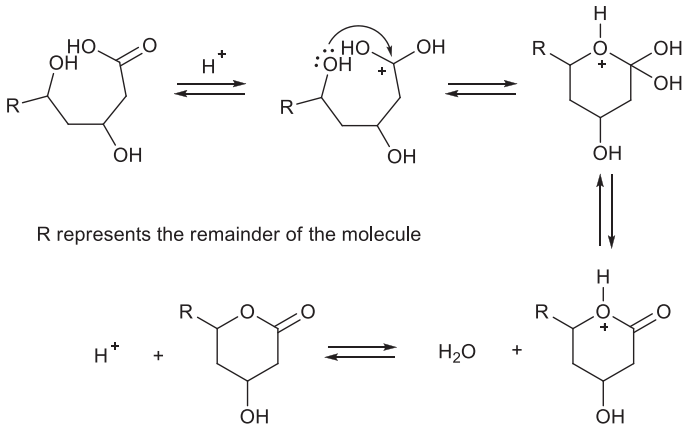
The kinetics of the interconversion and the equilibrium between these two forms were studied experimentally as a function of pH, buffer concentration and temperature at a fixed ionic strength of 0.5 M for the case of atorvastatin [32]. The dependence of the observed rate constants for lactone formation and hydrolysis, k_1 and k_2, respectively, and the associated equilibrium constant, K_{eq}, on pH is presented in Table 4.5.1.

Table 4.5.1: The pH dependencies of the rate constants for lactone formations (k_1) and hydrolysis (k_2) and the associated equilibrium constant (K_{eq}) at 80 °C and μ = 0.5 M [32].

Reaction medium	pH	k_1	k_2	K_{eq}
HCl	1.30	13.6[a]	20.2[a]	0.673[a]
HCl	1.64	5.62	7.67	0.733
HCl	1.98	2.57	3.36	0.766
Formate buffer	3.00	2.28×10^{-1}	3.42×10^{-1}	0.667
Formate buffer	3.50	8.80×10^{-2}	1.51×10^{-1}	0.583
Formate buffer	4.00	2.98×10^{-2}	5.13×10^{-2}	0.581
Acetate buffer	4.12	2.66×10^{-2}	5.64×10^{-2}	0.472
Acetate buffer	4.60	1.39×10^{-2}	5.59×10^{-2}	0.248
Acetate buffer	5.10	1.15×10^{-2}	1.02×10^{-1}	0.113
Phosphate buffer	5.73	1.51×10^{-2}	5.16×10^{-1}	0.029
Phosphate buffer	6.32	–	1.41	–
Phosphate buffer	6.86	–	5.20	–
Borate buffer	8.48	–	123[a,b]	–
Borate buffer	8.70	–	214[a,b]	–
Borate buffer	9.0	–	372[a,b]	–

[a] Extrapolated from the corresponding Arrhenius relationship.
[b] Buffer catalysis was observed, and these values represent the buffer-independent rate constants.

It was showed that the acid-catalysed reaction is reversible (Scheme 4.5.1), while the base-catalysed can be treated as an irreversible process [32]. At pH < 6, an equilibrium favouring the hydroxy acid form was established, whereas at pH > 6 the equilibrium was not detectable and greatly favoured the hydroxy acid form [32]. The acid-catalysed lactone hydrolysis of several HMG-CoA reductase inhibitors including lovastatin and simvastatin was studied earlier at pH 2, and reversibility of the process was demonstrated [33].



Scheme 4.5.1: Proposed mechanism for the acid-catalysed lactonization of the hydroxy acid form and acid-catalysed hydrolysis of the lactone form of atorvastatin [32].

The lactonization-hydrolysis mechanism of atorvastatin was later investigated theoretically using the DFT method and the obtained results fully supported the previously reported findings [31]. Four reaction pathways connecting both forms of the statin were found and analysed. Interconversions under acidic conditions as one-step processes were determined as unfavourable under physiological conditions due to high (35 kcal mol^{-1}) activation energy barriers [31]. Other reaction pathways with no intermediates were characterized by significantly lower activation energy barrier. Thus reaction leading from the hydroxy acid form of atorvastatin to the corresponding lactone form went through energy barrier of about 19 kcal mol^{-1}. The reverse reaction was endoergic and had activation energy barrier of 23 kcal mol^{-1} [31]. The overall equilibrium in the presence of carboxylic acid was slightly shifted towards the lactonization (exoergic reaction), which is in agreement with the experimental data. Under basic conditions the hydroxy acid form was found to be much more stable. The activation energy barrier calculated for the hydrolysis reaction was less than 10 kcal mol^{-1}, while for the lactone formation it nearly amounted to 28 kcal mol^{-1} [31]. Shifting of the lactonization-hydrolysis equilibrium towards the atorvastatin carboxylate salt is in line with the experimental findings [32, 33].

The mechanism of the hydroxy acid–lactone interconversion of fluvastatin under both acidic and basic conditions was also subjected to theoretical studies [34]. Similar to the case of atorvastatin, one-step, direct interconversion between these two forms observed in an acidic environment occurred to be unfavourable due to the high activation barrier (>40 kcal mol^{-1}) [34]. Other obtained pathways presented more probable solutions. The activation barriers under acidic conditions were slightly higher than for atorvastatin (22 leading towards lactone and 28 kcal mol^{-1} for the reverse reaction) [31]. Under basic conditions the activation energy barrier for the hydrolysis was significantly lower (9 kcal mol^{-1}) than for the reverse reaction (28 kcal mol^{-1})

resulting in shifting the equilibrium towards the hydroxy acid form [34]. Results obtained from the DFT-based investigation of the hydroxy acid–lactone interconversion of fluvastatin also fully supported findings derived from experimental measurements on the pH-dependent character of the reaction.

Results of the theoretical studies performed for atorvastatin and fluvastatin suggest that in basic conditions the lactone form of the former is less stable by about 18 kcal mol^{-1}. Moreover, activation energy barriers for hydrolysis were found in vacuo to be 6 and 10 kcal mol^{-1} for FLV and ATV, respectively [31, 34]. In the mildly acidic conditions the energy span of the lactonization reaction was slightly smaller for ATV (about 19 kcal mol^{-1}) compared to that for FLV (22 kcal mol^{-1}). This is most likely due to the more flexible dihydroxy acid side chain of ATV. The lack of a double bond between carbon atoms in this chain causes the molecule of ATV to be more flexible than the molecule of FLV. Thanks to its flexibility and lower activation barriers for lactonization and hydrolysis reactions, the ATV molecule can adopt more chemical structures and conformers more easily than FLV [31].

4.6 Summary and conclusions

Statins, since their discovery in the 1970s by Japanese microbiologist Akira Endo, have become popular drugs widely prescribed to control hypercholesterolaemia and to prevent cardiovascular diseases [35], [36], [37]. This is due to their high effectiveness and relatively few side effects. These compounds, however, especially synthetic statins whose molecular structures are based on conjugated bonds systems, are very prone to exposure to light in both UVA and UVB ranges. Among the synthetic statins, rosuvastatin, atorvastatin and fluvastatin exhibit phototoxicity which is thought to be mediated by their photoproducts formed according to common mechanism that assumes photochemical electrocyclization followed by sigmatropic hydrogen shift [7], [11], [13]. Statins of natural origin, although not being chromophores of UVA, strongly potentiate the UVA-induced damage towards cultured human keratinocytes [38].

In this work, on the basis of the data derived from both experimental and computational methods, we critically discussed finding regarding the mechanisms of synthetic statins phototransformation. We showed inconsistency of some previously reported facts and revised earlier presented studies. We also completed the lack of information on pitavastatin photobehaviour. This all together resulted in proposal of new schemes for the statins photodecomposition.

We also discussed the phenomenon of statins' sensitivity to pH and resulting implications. Apart from being used to treat lipid disorders, statins were shown to exhibit activities promising for the fight against other diseases [39, [40], [41], [41], [44]. Together with development of new therapeutic targets, interactions with other drugs deserve particular attentions. In this context influence of pH on statins' structures and properties should be carefully considered in order to understand the impact of

structural differences on the drugs' activity and to predict the possible drug interactions which are important determinants of safety for patients.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. Endo, A. A historical perspective on the discovery of statins. *Proc Jpn Acad Ser B Phys Biol Sci* 2010;86:484–93.
2. Schachter, M. Chemical, pharmacokinetic and pharmacodynamic properties of statins: an update. *Fundam Clin Pharmacol* 2005;19:117–25. https://doi.org/10.1111/j.1472-8206.2004.00299.x.
3. Istvan, E. Statin inhibition of HMG-CoA reductase: a 3-dimensional view. *Atherosclerosis Suppl* 2003;4:3–8. https://doi.org/10.1016/S1567-5688(03)00003-5.
4. Istvan, ES, Deisenhofer, J. Structural mechanism for statin inhibition of HMG-CoA reductase. *Science* 2001;292:1160–4. https://doi.org/10.1126/science.1059344.
5. Bellosta, S, Paoletti, R, Corsini, A. *History and development of HMG-CoA reductase inhibitors*, Schmitz, G and Torzewski, M, editors. Switzerland: Birkhauser Verlag; 2002, n.d.
6. Litvić, M, Šmic, K, Vinković, V, Filipan-Litvić, M. A study of photodegradation of drug rosuvastatin calcium in solid state and solution under UV and visible light irradiation: the influence of certain dyes as efficient stabilizers. *J Photochem Photobiol Chem* 2013;252:84–92. https://doi.org/10.1016/j.jphotochem.2012.11.008.
7. Cermola, F, DellaGreca, M, Iesce, MR, Montanaro, S, Previtera, L, Temussi, F, et al.. Irradiation of fluvastatin in water. *J Photochem Photobiol Chem* 2007;189:264–71. https://doi.org/10.1016/j.jphotochem.2007.02.011.
8. Cermola, F, DellaGreca, M, Iesce, MR, Montanaro, S, Previtera, L, Temussi, F. Photochemical behavior of the drug atorvastatin in water. *Tetrahedron* 2006;62:7390–5. https://doi.org/10.1016/j.tet.2006.05.027.
9. Grobelny, P, Viola, G, Vedaldi, D, Dall'Acqua, F, Gliszczyńska-Świgło, A, Mielcarek, J. Photostability of pitavastatin—a novel HMG-CoA reductase inhibitor. *J Pharmaceut Biomed Anal* 2009;50:597–601. https://doi.org/10.1016/j.jpba.2008.10.004.
10. Sobotta, L, Kachlicki, P, Marczak, L, Kryjewski, M, Mielcarek, J. Photochemical activity of glenvastatin, a HMG-CoA reductase inhibitor. *J Photochem Photobiol Chem* 2011;224:1–7. https://doi.org/10.1016/j.jphotochem.2011.08.008.
11. Nardi, G, Lhiaubet-Vallet, V, Leandro-Garcia, P, Miranda, MA. Potential phototoxicity of rosuvastatin mediated by its dihydrophenanthrene-like photoproduct. *Chem Res Toxicol* 2011;24:1779–85. https://doi.org/10.1021/tx200341f.
12. Viola, G, Grobelny, P, Linardi, MA, Salvador, A, Basso, G, Mielcarek, J, et al.. The phototoxicity of fluvastatin, an HMG-CoA reductase inhibitor, is mediated by the formation of a benzocarbazole-like photoproduct. *Toxicol Sci Off J Soc Toxicol* 2010;118:236–50. https://doi.org/10.1093/toxsci/kfq228.

13. Montanaro, S, Lhiaubet-Vallet, V, Iesce, M, Previtera, L, Miranda, MA. A mechanistic study on the phototoxicity of atorvastatin: singlet oxygen generation by a phenanthrene-like photoproduct. *Chem Res Toxicol* 2009;22:173–8. https://doi.org/10.1021/tx800294z.
14. Jemal, M, Ouyang, Z, Powell, ML. Direct-injection LC-MS-MS method for high-throughput simultaneous quantitation of simvastatin and simvastatin acid in human plasma. *J Pharmaceut Biomed Anal* 2000;23:323–40.
15. Zhao, JJ, Xie, IH, Yang, AY, Roadcap, BA, Rogers, JD. Quantitation of simvastatin and its beta-hydroxy acid in human plasma by liquid-liquid cartridge extraction and liquid chromatography/tandem mass spectrometry. *J Mass Spectrom* 2000;35:1133–43.
16. Sakaeda, T, Fujino, H, Komoto, C, Kakumoto, M, Jin, J, Iwaki, K, et al.. Effects of acid and lactone forms of eight HMG-CoA reductase inhibitors on CYP-mediated metabolism and MDR1-mediated transport. *Pharm Res* 2006;23:506–12. https://doi.org/10.1007/s11095-005-9371-5.
17. Fujino, H, Saito, T, Tsunenari, Y, Kojima, J, Sakaeda, T. Metabolic properties of the acid and lactone forms of HMG-CoA reductase inhibitors. *Xenobiotica Fate Foreign Compd Boil Syst* 2004;34:961–71. https://doi.org/10.1080/00498250400015319.
18. Astarita, A, DellaGreca, M, Iesce, MR, Montanaro, S, Previtera, L, Temussi, F. Polycyclic compounds by sunlight exposure of the drug rosuvastatin in water. *J Photochem Photobiol Chem* 2007;187:263–8. https://doi.org/10.1016/j.jphotochem.2006.10.007.
19. Sajimon, MC, Lewis, FD. Photocyclization of a conformationally constrained 2-vinylbiphenyl. *Photochem Photobiol Sci* 2005;4:789–91. https://doi.org/10.1039/b511270b.
20. Lewis, FD, Sajimon, MC, Zuo, X, Rubin, M, Gevorgyan, V. Competitive 1,2- and 1,5-hydrogen shifts following 2-vinylbiphenyl photocyclization. *J Org Chem* 2005;70:10447–52. https://doi.org/10.1021/jo051730y.
21. Jarmużek, D, Pedzinski, T, Hoffmann, M, Siodła, T, Pluskota-Karwatka, D. Phototransformations of pitavastatin - the inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *J Photochem Photobiol Chem* 2020;389:112243. https://doi.org/10.1016/j.jphotochem.2019.112243.
22. Liu, Y, Wang, Q-L, Chen, Z, Zhou, C-S, Xiong, B-Q, Zhang, P-L, et al.. Oxidative radical ring-opening/cyclization of cyclopropane derivatives. *Beilstein J Org Chem* 2019;15:256–78. https://doi.org/10.3762/bjoc.15.23.
23. Mielcarek, J, Kula, M, Zych, R, Grobelny, P. Kinetic studies on fluvastatin photodegradation in solutions. *React Kinet Catal Lett* 2005;86:119–26. https://doi.org/10.1007/s11144-005-0302-6.
24. Mielcarek, J, Grobelny, P, Osmałek, T. Identification of photoproducts of fluvastatin in solutions. *J Planar Chromatogr* 2009;22:137–40. https://doi.org/10.1556/JPC.22.2009.2.11.
25. Jarmużek, D, Pedzinski, T, Hoffmann, M, Siodła, T, Salus, K, Pluskota-Karwatka, D. Experimental and theoretical studies on fluvastatin primary photoproduct formation. *Phys Chem Chem Phys* 2017;19:21946–54. https://doi.org/10.1039/C7CP01094J.
26. Pluskota-Karwatka, D, Hoffmann, M. Transformations of statins: effect of light and pH. *Curr Org Chem* 2018;22:1926–39. https://doi.org/10.2174/1385272822666180913112356.
27. Lednev, IK, Mathivanan, N, Johnston, LJ. Photochemistry of stilbene adsorbed on silica gel and NaX zeolite. A diffuse reflectance laser flash photolysis study. *J Phys Chem* 1994;98:11444–51. https://doi.org/10.1021/j100095a030.
28. Barik, R, Bhattacharyya, K, Das, PK, George, MV. Photochemical transformations of 1-imidazolyl-1,2-dibenzoylalkenes. Steady-state and laser flash photolysis investigations. *J Org Chem* 1986;51:3420–8. https://doi.org/10.1021/jo00368a003.
29. Alarcón, E, González-Béjar, M, Gorelsky, S, Ebensperger, R, Lopez-Alarcón, C, Netto-Ferreira, JC, et al.. Photophysical characterization of atorvastatin (Lipitor®) ortho-hydroxy metabolite: role of hydroxyl group on the drug photochemistry. *Photochem Photobiol Sci* 2010;9:1378–84. https://doi.org/10.1039/C0PP00102C.

30. Ando, H, Tsuruoka, S, Yanagihara, H, Sugimoto, K, Miyata, M, Yamazoe, Y, et al.. Effects of grapefruit juice on the pharmacokinetics of pitavastatin and atorvastatin. *Br J Clin Pharmacol* 2005;60:494–7. https://doi.org/10.1111/j.1365-2125.2005.02462.x.
31. Hoffmann, M, Nowosielski, M. DFT study on hydroxy acid–lactone interconversion of statins: the case of atorvastatin. *Org Biomol Chem* 2008;6:3527–31. https://doi.org/10.1039/B803342K.
32. Kearney, AS, Crawford, LF, Mehta, SC, Radebaugh, GW. The interconversion kinetics, equilibrium, and solubilities of the lactone and hydroxyacid forms of the HMG-CoA reductase inhibitor, CI-981. *Pharm Res* 1993;10:1461–5.
33. Kaufman, MJ. Rate and equilibrium constants for acid-catalyzed lactone hydrolysis of HMG-CoA reductase inhibitors. *Int J Pharm* 1990;66:97–106. https://doi.org/10.1016/0378-5173(90)90389-L.
34. Grabarkiewicz, T, Grobelny, P, Hoffmann, M, Mielcarek, J. DFT study on hydroxy acid–lactone interconversion of statins: the case of fluvastatin. *Org Biomol Chem* 2006;4:4299–306. https://doi.org/10.1039/B612999B.
35. Nissen, SE, Tuzcu, EM, Schoenhagen, P, Crowe, T, Sasiela, WJ, Tsai, J, et al.. Statin therapy, LDL cholesterol, C-reactive protein, and coronary artery disease. *N Engl J Med* 2005;352:29–38. https://doi.org/10.1056/NEJMoa042000.
36. Mills, EJ, Rachlis, B, Wu, P, Devereaux, PJ, Arora, P, Perri, D. Primary prevention of cardiovascular mortality and events with statin treatments: a network meta-analysis involving more than 65,000 patients. *J Am Coll Cardiol* 2008;52:1769–81. https://doi.org/10.1016/j.jacc.2008.08.039.
37. Hiro, T, Kimura, T, Morimoto, T, Miyauchi, K, Nakagawa, Y, Yamagishi, M, et al.. Effect of intensive statin therapy on regression of coronary atherosclerosis in patients with acute coronary syndrome: a multicenter randomized trial evaluated by volumetric intravascular ultrasound using pitavastatin versus atorvastatin (JAPAN-ACS [Japan assessment of pitavastatin and atorvastatin in acute coronary syndrome] study). *J Am Coll Cardiol* 2009;54:293–302. https://doi.org/10.1016/j.jacc.2009.04.033.
38. Quiec, D, Mazière, C, Auclair, M, Santus, R, Gardette, J, Redziniak, G, et al.. Lovastatin enhances the photocytotoxicity of UVA radiation towards cultured N.C.T.C. 2544 human keratinocytes: prevention by cholesterol supplementation and by a cathepsin inhibitor. *Biochem J* 1995;310:305–9. https://doi.org/10.1042/bj3100305.
39. Sparks, DL, Kryscio, RJ, Sabbagh, MN, Connor, DJ, Sparks, LM, Liebsack, C. Reduced risk of incident AD with elective statin use in a clinical trial cohort. *Curr Alzheimer Res* 2008;5:416–21.
40. Eckert, GP, Wood, WG, Müller, WE. Statins: drugs for Alzheimer's disease? *J Neural Transm* 1996;112:1057–71. https://doi.org/10.1007/s00702-004-0273-1.
41. Ajith, TA, Riji, T, Anu, V. In vitro anti-oxidant and DNA protective effects of the novel 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitor rosuvastatin. *Clin Exp Pharmacol Physiol* 2008;35:625–9. https://doi.org/10.1111/j.1440-1681.2007.04853.x.
42. Demierre, M-F, Higgins, PDR, Gruber, SB, Hawk, E, Lippman, SM. Statins and cancer prevention. *Nat Rev Cancer* 2005;5:930–42. https://doi.org/10.1038/nrc1751.
43. Chan, KKW, Oza, AM, Siu, LL. The statins as anticancer agents. *Clin Cancer Res Off J Am Assoc Cancer Res* 2003;9:10–9.
44. Staedler, D, Chapuis-Bernasconi, C, Dehmlow, H, Fischer, H, Juillerat-Jeanneret, L, Aebi, JD. Cytotoxic effects of combination of oxidosqualene cyclase inhibitors with atorvastatin in human cancer cells. *J Med Chem* 2012;55:4990–5002. https://doi.org/10.1021/jm300256z.

Maciej Staszak

5 Artificial intelligence in the modeling of chemical reactions kinetics

Abstract: The work presents a selection of recent papers in the field of modeling chemical kinetics by the use of artificial intelligence methods. Due to the fact that kinetics of the chemical reaction is the key element of industrial reactor design and analysis, the work is focused on the presentation of the quality of modeling, the assembly of neural network systems and methods of training required to achieve acceptable results. The work covers a wide range of classes of chemical processes and modeling approaches presented by several authors. Because of the fact that the methods of neural networks training require huge amounts of data, many approaches proposed are intrinsically based on classical kinetics modeling like Monte Carlo methods, quantum *ab initio* models or classical Arrhenius-like approaches using mass balance rate equations. The work does not fully exhaust the area of artificial intelligence because of its very broad scope and very fast evolution, which has been greatly accelerated recently. However, it is a contribution to describing the current state of science in this field.

Keywords: artificial neural network, chemical kinetics, machine learning

5.1 Concise and brief description of the artificial intelligence methods

Machine learning (ML) is a subset of artificial intelligence (AI), which provides systems with the ability to automatically learn and improve the expected response in terms of modeling, based on experimental, training data [1]. Deep learning is a subset of machine learning that uses neural networks to analyze various factors in a similar way to the human learning process (Figure 5.1). The key idea of ML is that it is possible to create algorithms that execute the learning process using training data and create predictions and generalizations based on it. The AI system is at the most basic level built of neural networks, which in turn are built of neurons, connected by weights and typically formed into layers.

Machine learning systems are trained on properly prepared sets of samples, measurements, experiments which are formed as sets of data [2]. The selection of specific characteristics of the material presented to the learning process is an

This article has previously been published in the journal Physical Sciences Reviews. Please cite as: Staszak, M. Artificial intelligence in the modeling of chemical reactions kinetics, *Physical Sciences Reviews* [Online] 2021, 6. DOI: 10.1515/psr-2020-0079

https://doi.org/10.1515/9783110678215-005

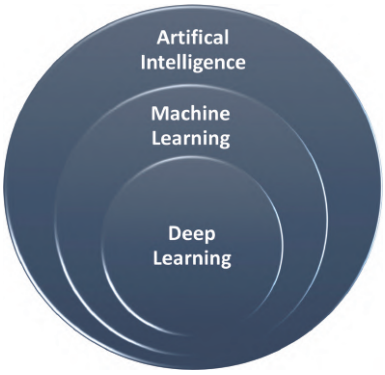


Figure 5.1: The framework of artificial intelligence approaches.

important, preliminary element of learning. The main goal is to present the data properly and to extract the most influential parameters due to the course of the process analyzed. It is an important element in the case of supervised learning, when there are teaching data and data set separately for validation of the taught network. During the learning process, the neural network learns, i.e., it conducts convergence to the right solution. The validation set is then used to take the appropriate test of the quality of the learning achieved. In the case of unsupervised learning, networks are taught using unmarked input data. The algorithm of unsupervised learning is designed to see patterns and relationships on its own.

It is possible to realize the learning process using different algorithms. Depending on the algorithm, the accuracy or speed of convergence is different. It is possible to combine different algorithms in order to achieve better results, it is team learning. There are a number of approaches based on how neural networks are taught. Machine learning algorithms are usually divided into four groups (Figure 5.2): supervised learning [3], unsupervised learning [4], semi-supervised learning [5] and reinforcement learning [6].

The process of supervised learning consists of presenting the data of the training agent, which contains the input information and the required output response. The network that has passed the learning process is then tested using a validation kit that checks whether the learned attribute or property is correct. If training outcomes are unsatisfactory, the learning process is repeated; either using the networks modified learning settings or using additional data. The learning process continues until the model reaches the desired accuracy level of the learners' data. This type of learning is commonly used for classification and regression. During unsupervised learning, only data is presented without additional information about the requirements for the expected solution or learning result. This is to enable the neural network to search for patterns independently. Unsupervised learning is often used to divide data into groups according to similarity. Unsupervised learning is used for in-depth data analysis.

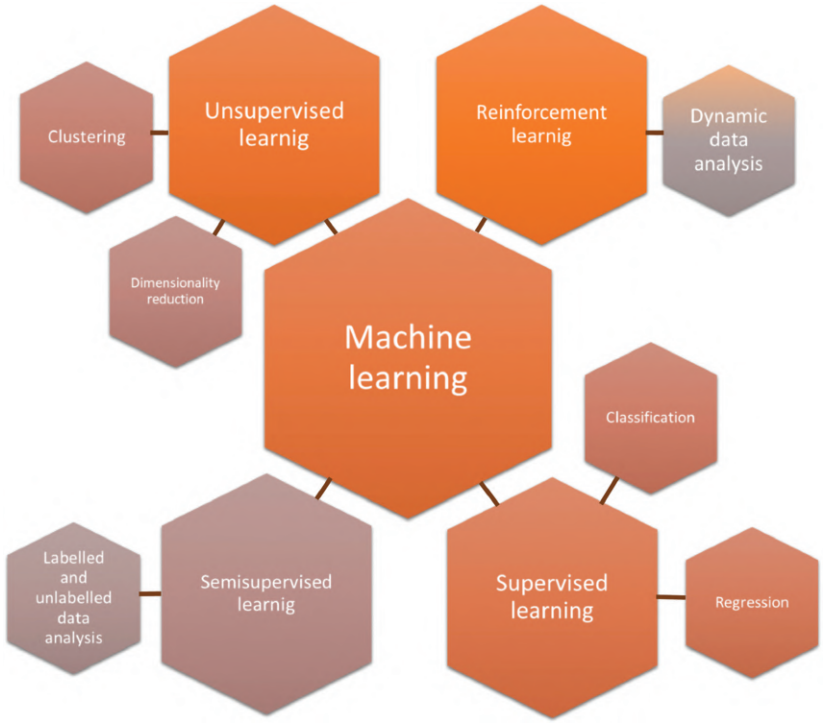


Figure 5.2: Types of machine learning methods.

Many times the neural network has the ability to recognize patterns that are unnoticeable to people because of the inability to process such large amounts of numerical data. The operator does not know a priori what he is trying to find, but surely there are some patterns and the system can detect them. Semisupervised algorithms implement the learning process of neural networks, so that the input data is a mixture of marked and unmarked samples. The neural network is designed to find hidden patterns that will allow you to organize the data and make predictions on your own with some requirements for the expected result. Reinforcement learning on the other hand is an ML area that involves taking appropriate action to maximize rewards in a given situation. It is used to find the best possible behavior or to determine the path to take in a specific situation. Reinforcement learning differs from supervised learning in that in supervised learning, the learners' data are marked accordingly, i.e., they have a matching answer key, while in reinforcement learning no answer key data is used, but the procedure of reinforcement learning decides what to do to perform a given task. In the absence of a set of training data, it is obliged to learn from its experience. Reinforcement learning is the process of making decisions in a sequential way. This means that the state at the output of the neural network depends on the state of the current input, and the next input depends on the output of the previous step.

5.2 Kinetics of chemical reactions in industrial applications

One of the central elements of the chemical reaction nature is how fast it runs. The research on the kinetics of chemical processes has been conducted for a long time and in classical approaches addresses the problem in the form of appropriate mathematical equations. Kinetics, in other words, the rate of chemical reaction, is typically described by well-known models: power law, Langmuir–Hinshelwood [7], Eley–Rideal [8], Mars–van Krevelen [9] or Michaelis–Menton [10] to name just the most popular (see Table 5.1). In some cases researchers may also use own approaches specified and suited precisely to select chemical reaction system, many times it is a closed form or some analytic expression for black-box model which renders the nature of chemical reaction kinetics without explaining its mechanism.

Table 5.1: Selection of most common kinetic models used for chemical reactions.

Name	Model	Description
Power law	$r = -\frac{dC}{dt} = kC^n$	General kinetics typical for homogenous reactions; k – rate constant, n – exponential coefficient, sometimes stoichiometric one.
Langmuir–Hinshelwood	$r = -\frac{dC_A}{dt} = kC_S^2 \frac{K_1K_2C_AC_B}{(1+K_1C_A+K_2C_B)^2}$	General kinetics for superficial reactions, two reactants A and B adsorb and react at a surface; C_S – is the concentration of all sites, K_i – adsorption constant, k – rate constant.
Eley–Rideal	$r = -\frac{dC_A}{dt} = kC_SC_B \frac{K_1C_A}{K_1C_A+1}$	General kinetics for superficial reactions, one reactant A adsorbs and reacts with other in the fluid phase; C_S – is the concentration of all sites, K_i – adsorption constant, k – rate constant.
Mars–van Krevelen	$r = -\frac{dC}{dt} = \frac{k_{red}C^{n_1} \cdot k_{ox}C^{n_2}}{k_{red}C^{n_1}+k_{ox}C^{n_2}}$	Oxidation, hydrodesulfurization, NOx removal; k_{red} – reduction rate constant, k_{ox} – oxidation rate constant, n_i – parameters
Michaelis–Menton	$r = -\frac{dC}{dt} = \frac{r_{max}C}{K_M+C}$	Enzymatic reactions; r_{max} – maximum feasible rate, K_M – Michaelis constant.

Anyway, the mathematical formulation of kinetic model is required to calculate the time evolution of reactants concentrations. From a design point of view, kinetics is of fundamental importance when estimating the volume and/or shape of the chemical reactor. Also the knowledge of kinetics is particularly important when designing reactor control systems.

In addition to mass law effect presented, rate of chemical reaction can change under pressure. The rate coefficients and the products of many reactions running in the gaseous phase and at high temperatures change when an inert gas is added to the reaction mixture. The effects of such processes are called fall-offs. These phenomena are caused by exothermic or endothermic reactions occurring faster than the heat transfer process, which means that the reacting molecules do not have Boltzmann energy distribution. Increased pressure increases the rate of heat transfer between the reacting molecules and the rest of the system, which causes the fall-off effect to decrease. The values of rate constants k are typically estimated by researchers in the temperature-dependent form of Arrhenius equation with frequency and beta factor, and activation energies provided.

From the industrial point of view, the kinetics of the chemical reaction is a key data necessary for design, analysis and optimization tasks. In classical design methods, the knowledge of reaction kinetics allows to determine the required volume of reaction volumes, either in zero-dimensional approach as for a tank reactor or in one-dimensional approach in case of column or tubular apparatus. In the increasingly used finite element methods such as computational fluid dynamics (CFD), the knowledge of the chemical reaction kinetics allows to design the appropriate reactor geometry and optimize it in terms of performance or operating time under given conditions. It follows that the ability to calculate source expressions of balance equations is crucial and the calculation time plays a very important role here. In many cases, taking into account the course of a chemical reaction whose rate, and thus time constant, is significantly different from the rates of other concurrent processes, makes the problem poorly or ill conditioned from a numerical point of view. Such a negative effect of solving stiff numerical problems, causes that the time required to get a solution to the design or optimization task increases significantly, even when using computers with very efficient architecture. In order to remove such problems, researchers look for solutions and possibilities in the field of machine learning.

5.3 Reasons for artificial intelligence models use in chemical kinetics

Taking into account the fact that classical kinetic models are sufficient, it would seem that there is no greater need to model the paths and rates of a chemical reaction using other methods. Artificial neural network methods do not explain the arising

phenomena, but only present them in the form of black-box responses. However, classical models, even those containing physicochemical interpretable constants of rate or equilibrium constants, provide opportunities to perform wide generalization. When deep machine learning techniques are used, additional possibilities of creating more general models, capable of predicting the paths of reactions for which they were not originally trained, open up. The ability of an artificial neural network to learn a single kinetics for a selected reaction would not contribute much to design or research applications. On the other hand, the ability to create generalizations, to find distant relationships, and to bind apparently distant processes and phenomena makes it possible to create a single neural network, using deep learning, that would not only be able to give an answer about the reactions it learned in the course of training but also about those it did not know originally. Such broadly conceived deep machine learning has only become available in recent years, mainly due to the enormous advances in computer hardware. Deep learning requires a huge amount of computational effort, which until a few years ago was not widely available. An additional benefit of using artificial neural networks is that it is a kind of data storage, understood as knowledge of the kinetics of various chemical reactions compressed in the form of weights on the dendrites of neurons of the artificial network. Existing machine learning platforms bring artificial intelligence models to common use, not only in industrial or academic applications but especially in common homebrew applications. MLNET framework [11] from Microsoft, TensorFlow [12] from Google or Keras [13] from Open-ended Neuro-Electronic Intelligent Robot Operating System (ONEIROS) project to name only a few, allow not only to create deep learning systems on their own but also to exchange already trained network models between users. Collections of pretrained artificial neural networks operate using Open Neural Network Exchange (ONNX) format [14], which constitutes so called open AI ecosystem, are widely understood (in the sense of compatibility) and shared between various neural networks frameworks and platforms. In this way, the application of neural networks and deep learning algorithms in the field of chemical reaction kinetics brings enormous cognitive potential. The teaching of neural networks, especially the algorithms of reinforced learning, seems to be an excellent direction for further research in this area. A certain limitation in this regard may be the demand for a large amount of data for the neural network to train, which in principle, should come only from the experiment. However, experimental investigation in this area is a research field that has already been perfectly mastered in a wide class of chemical reactions.

5.4 Selection of recent papers on artificial intelligence methods in prediction of kinetics of various chemical processes

5.4.1 Neural network training with Arrhenius kinetics for equilibrium reactions

The method presented in [15] reflects the calculation of the source term in the form realized by means of an artificial neural network, which allows to interpret the equations as a series of connected neural layers. The structure of neural network reflects the process by means of process variables, namely the concentrations and temperatures. The neural network output on the other hand is defined as the resultant source terms set (Figure 5.3).

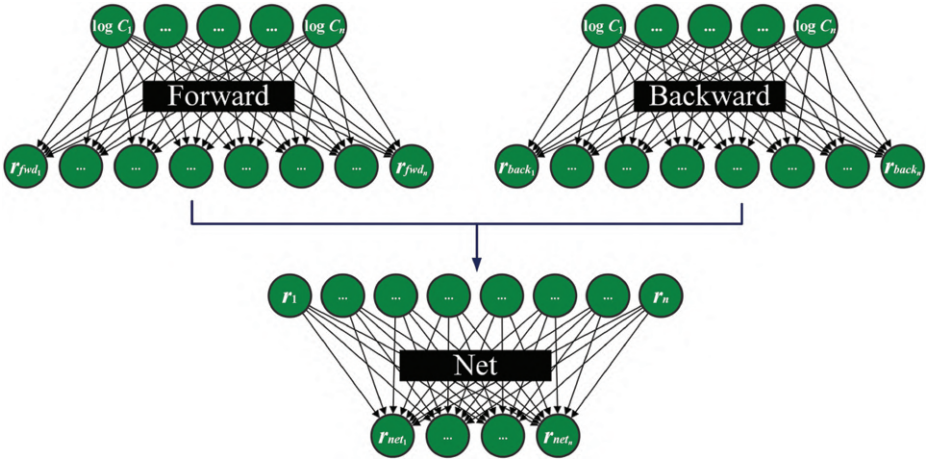


Figure 5.3: Neural network which resembles the reaction system; forward, backward and net reactions system terms.

The interpretation realized by means of an artificial neural network allows to replace parts of the classically formulated kinetic model by a trained, approximate approach by means of an artificial neural network. Authors use additional, open source software Cantera [16] to obtain training data for neural network. The main goal is to use an approximate approach using a neural network in order to reduce computational costs (or increase speed) compared to accurate kinetic models. In this work, the neural network proxy model has been trained as a substitute for the calculation of speed and balance constants. The use of trained artificial neural networks (ANNs) in this way eliminates the input dependence on the concentration vector of NS dimensional species, which greatly simplifies the training process and allows to maintain certain physical characteristics during the source calculation (i.e., the Arrhenius form for the forward course constants). Thanks to zero-dimensional simulations of automatic ignition and detonation in a one-dimensional channel model based on several mechanisms of different

complexity, the results finally showed the feasibility of this approach for complex non-linear approaches. The authors concluded that the acceleration achieved by approximate ANNs of speed and equilibrium constants depended both on the approximate ANN architecture and on the complexity of the chemical mechanism. For simple reaction mechanisms, the advantage of the neural network over the kinetic accurate model was ultimately less noticeable because the upper limit of acceleration is achieved by the neural network, and in this context is limited by the number of components and the amount of running chemical reactions. This is clearly due to the simplicity of the original kinetic model, which can be calculated very efficiently. However, for complex reaction mechanisms, the realistic acceleration achieved by the neural network in relation to the exact form of the kinetic model is significant. In general, the analysis has shown that the computational advantage provided by approximate models realized by artificial neural networks can be significant. On the other hand, learned neural networks used to replace the individual algorithms of the original kinetic models, should not be interpreted as just a proper acceleration technique.

5.4.2 Catalytic cracking

Predicting and understanding the fluid catalytic cracking (FCC) process in a real industrial environment is still difficult because of its huge complexity that is influenced by many extremely non-linear and interrelated factors. In the Yang et al. work [17] a hybrid predictive framework for FCC was developed, integrating a data-driven deep neural network with a physical kinetic model, powered by an appropriate amount of high quality data from the modem FCC automated process. The results show that the hybrid model exhibits the best predictions for all assessment criteria, such as average absolute percentage error, Pearson's coefficient and standard deviation. Authors assure that a data-driven hybrid, deep learning process with a mechanical kinetics model creates a better approach to quickly predict and optimize complex response processes such as FCC.

A standard neural network was proposed as a representative of data-only models. In this case the neural network consisted of four layers, which authors classify as deep neural networks (DNN). Compared to a simple neural network, the essence of DNN provides higher level of abstraction and the ability to learn more beneficial features through a greater number of hidden layers and massive training data to ultimately improve the accuracy of classification or prediction. DNN presented contained inputs for six elements, i.e., weight percent of components in the crude oil (saturates, aromatic, resins and asphalt), weight of carbon residues and microactivity index. The output layer was a vector built up by weight of five industrial products: diesel, petrol, LPG, dry gas and coke (Figure 5.4).

In turn, the physicochemical model was developed on the basis of the eight lumped kinetics schemes. In this model, raw materials were lumped into saturated

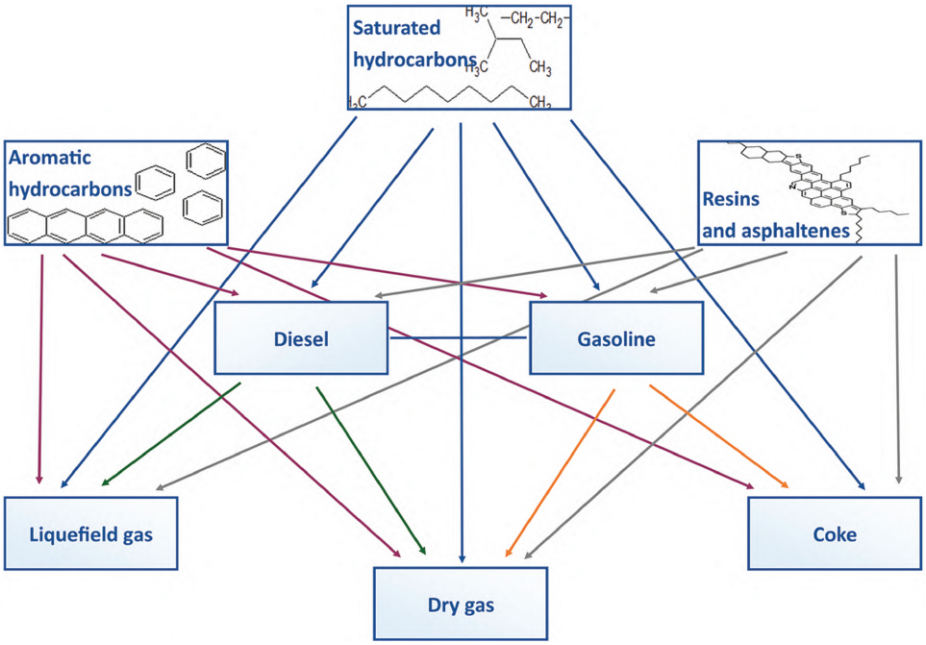


Figure 5.4: DNN model for catalytic cracking.

hydrocarbon (HS), aromatic hydrocarbon (HA), resin and phosphate (HR) solids, and the products were lumped into lumps on a gas pipeline, diesel, liquefied petroleum gas, dry gas and coke, from which eight lumped reaction network was formed. The following assumptions were applied for the complex FCC reaction system:

- All reactions are irreversible first order.
- Plug flow reactor condition has been assumed.
- The same catalyst activity applied for all reactions are.
- Deactivation of the catalyst during the reaction is only related to the catalyst residence time.

Raw data was gathered from the industrial FCC unit at a petrochemical plant in China. In particular, the operational variables and material properties data analyzed in this study come from the Distributed Control System and Laboratory Information Management System installed at the FCC plant. The collected data are arranged in a key-value pair format, where a time stamp and a corresponding value were recorded as key and value. The initial database was created on the basis of the such a layout, which is then prepared based on the modeling needs.

In summary, the hybrid model proposed by authors, effectively embeds the physically relevant lumped kinetics model in the data-driven deep neural network model, allowing for better data correlation based on a clear reflection of the

physical process rules. It is suggested that such approach applies not only to FCC, but can be widely used in other complex processes and chemical reaction engineering.

5.4.3 Photochemical reactions

Time-resolved spectroscopes play an essential role in explaining the basic mechanisms of light-induced processes, in particular the study of relaxation models for electronically excited particles. Determining such models on the basis of experimentally obtained data in the temporal and spectral domain often presents a numerical challenge and is often not free from ambiguity. Kollenz et al. [18] demonstrate the analysis of time domain laser spectroscopy data using a deep learning neural network to get the right relaxation kinetic model. The presented Deep Kinetic Spectroscopy Network (DeepSKAN) can predict kinetic models consisting of up to five excitation states, resulting in 103 possible different classes of kinetic models (Figure 5.5).
By indicating the probabilities for each path of the top-k models normalized by total probabilities, relaxation paths can be determined for a given data set. Authors as-

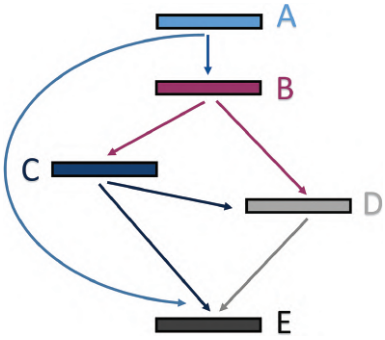


Figure 5.5: Model of reaction paths and activation states.

sure that architecture and training by the use of DeepSKAN is resistant to experimental noise and typical preanalysis errors such as time-zero corrections. The application of DeepSKAN to experimental data has been demonstrated for three different photoinduced processes: transient absorption of retinal isomerization, transient infrared relaxation spectroscopy of photoactivated DRONPA and transient absorption of dynamics in lycopene.

The key problem in the application of spectroscopy is to propose the kinetic model, explaining how the molecular system develops after light excitation. These models include transitions between electron states in transient absorption

spectroscopy and time resolved fluorescence spectroscopy or transitions between vibration states in the time domain of infrared spectroscopy.

The DeepSKAN was developed in Python (v. 3.6) based on the open source libraries Keras (v. 2.2.4) and TensorFlow (v. 1.13.1) for deep learning and the NVIDIA driver CUDA (v. 10.1) for GPU acceleration and optimization. Authors provided versions of the free software language, frameworks and drivers used, although it is typical that newer will also suit for the tasks presented.

The authors proposed a network architecture similar to that which is typically used in image processing. In this case two image axes encode physical spatial dimensions of the image. As a result, neural convolution neural networks (CNNs) use symmetric filters analyzing both axes equally, by using square convolution kernels. In contrast, time resolved data image representations define an affine space where axes with two different time-encoded dimensions (probe delay) and spectral information (e.g., probe wavelength, wavenumber, etc.) are used. Therefore, the information obtained by convolution along each axis is different; convolution along a time axis can reveal velocity constants, while convolution along a spectral axis can reveal individual spectrum contributions. In time-resolved spectroscopy, the spectra of a given state and changes in population in that state are independent.

Authors proposed a way of managing in the case of small, experimental data set which could be too small to obtain high quality results. In such case, the authors propose to use a set of data on simulated transient state absorption signals so that the neural network can be trained successfully. For each training case, the kinetic model must be preselected by randomized choice. A system with five states (one base state and four excited states) results in a total of 20 possible reaction paths, giving 220 different possible kinetic models. The number of models and the complexity of the predictive task has been considerably reduced by using the following assumptions:

- The kinetic model is acyclical – no loops in the transition paths.
- There is only one initial populated state.
- Each state can split into up to two other states.

This resulted in reduction to 103 physically feasible kinetic models that the network could select from the data presented. The first two assumptions are physically motivated, because no molecular system is able to excite from a base state without a pump pulse (typically generated by a high-powered laser). The third assumption was applied for practical reasons, because in most situations it is technically impossible to differentiate between a branch up to two states from a branch up to three or more states. Branching to three or more states is also very seldom detected for molecular systems.

The authors have demonstrated the use of a deep learning toolkit to address complex spectroscopic problems. The presented modeling method guarantees simplification

of the initial analysis of complex spectroscopic data with the potential to improve objectivity by reducing dependence on user input.

5.4.4 Laminar and turbulent combustion modeling

Applicability of the artificial neural network was tested for flame computation by following cases:

- Laminar flame with syngas mixture was used with the fuel which includes CO_2 and H_2O as diluents in a $CO–H_2$ fuel mixture.
- Methane-air laminar flames with five reacting species (CH_4, O_2, CO_2, H_2O, N_2).
- Syngas-air flames taking into account flame–turbulence interaction.

Artificial neural network for kinetic modeling of reactive flow simulation by the large eddy simulation CFD model (LES) was introduced. The emphasis was placed on replacing stiff ordinary differential equations (ODE) with ANN solutions for chemical kinetics calculations. The type of reverse propagation of ANN code was developed and tested in laminar and turbulent methane/air precombustion and synthesis/air gas processes. ANN is trained using independent preflame calculations and then used as described in LES model for CFD flow simulation. The results indicate that the accuracy of ANN predictions for turbulent flow calculations is highly dependent on the initial training data set. If the ANN is well trained, it can successfully predict chemical status using less memory than a conventional tabular method and in a more computationally efficient way than stiff ODE solvers. Direct numerical simulation (DNS) of flows for large Reynolds number values is not yet easily feasible because the mesh resolution dictated by the solution's methodology still generates huge computational demand. In the field of reactive flow calculations, a great effort has been put into capturing time-dependent processes by LES calculations. In order to capture changes in the species composition of state spaces due to chemical reactions, a stiff nonlinear ODE system should be resolved like backward differentiation Gear's predictor-corrector method or an implicit Runge–Kutta method of fifth order.

Authors use LES approach with a mixing and combustion model as a frame for the flow calculations. For this purpose, Favre-filtered Navier Stokes equations for continuity, momentum, total energy and preservation of the components were fully compressible, dynamic and multicomponent.

$$\frac{\partial \bar{\rho}}{\partial t} + \frac{\partial}{\partial x_i}\left[\bar{\rho}\tilde{u}_i\right] = 0 \tag{5.1}$$

momentum:

$$\frac{\partial}{\partial t}\left(\bar{\rho}\tilde{u}_i\right) + \frac{\partial}{\partial x_j}\left[\bar{\rho}\tilde{u}_i\tilde{u}_j + \bar{p}\delta_{ij} - \bar{\tau}_{ij} + \tau_{ij}^{sgs}\right] = 0 \tag{5.2}$$

total energy:

$$\frac{\partial}{\partial t}\left(\bar{\rho}\tilde{E}\right)+\frac{\partial}{\partial x_j}\left[\bar{\rho}\tilde{u}_j\tilde{E}+\tilde{u}_j\bar{p}+\bar{q}_j-\tilde{u}_i\bar{\tau}_{ij}+H_j^{sgs}\right]=0 \tag{5.3}$$

components conservation:

$$\frac{\partial}{\partial t}\left(\rho\tilde{Y}_k\right)+\frac{\partial}{\partial x_j}\left[\rho\tilde{u}_j\tilde{Y}_k-\bar{\rho}\bar{D}_k\frac{\partial\tilde{Y}_k}{\partial x_j}+\phi_{i,k}^{sgs}+\theta_{i,k}^{sgs}\right]=\bar{R}_k \tag{5.4}$$

Here ~ represents the averaging Favre operator and is scaled with the use of density, where overbar means spatial filtration with a top-hat filter. Subgrid (*sgs*) shear stress $\left(\tau_{ij}^{sgs}\right)$, subnet viscosity $\left(H_i^{sgs}\right)$, subnet diffusion flux $\left(\theta_{jm}^{sgs}\right)$ appear as a result of the filtering operation and are also modeled. The rate at which the k component is formed/consumed is taken into account by the variable R_k.

The classical formulation for chemical reactions analyzed by the authors was defined by mass balance of reactive components expressed by kinetic equations:

$$q_i=k_{fwd,i}\prod_k c_k^{\nu'_{k,i}}-k_{bck,i}\prod_k c_k^{\nu''_{k,i}} \tag{5.5}$$

and

$$R_k=\sum_{i=1}^{N_{react}}\nu_{k,i}q_i \tag{5.6}$$

Typical Arrhenius type equation is used for temperature dependence modeling. Authors derive kinetic data by using PREMIX module of CHEMKIN [19].

In this particular study, a multilayer type of artificial neural network perceptron (MLP-ANN) was used to replace chemical kinetics calculations. For data training, a back propagation algorithm is used which is essentially based on a gradient descent procedure. For an MLP-ANN back-propagation training procedure, the algorithm generally consists of two parts:

- forward propagation of the input data,
- back-propagation of the error.

Training data was submitted into the network into randomly selected neurons of the input layer to avoid unfavorable memorization of a particular input pattern and to achieve generalization about the nature of the process, not the network setup. The authors selected a specific ANN architecture adjusted to each case separately, which was done on a basis of a compromise between number of hidden layers, number of neurons on each layer and the ANN error value achieved.

Authors highlight difference between the use of rigid ODE solvers ANN where the latter is very time-efficient and the memory effective. It is important to mention that

the article focuses on determining the reliability of ANN forecasts, which is highly dependent on the training procedure. For premixed flames, the ANN can be trained based on a laminar premixed flame solution but such approach is not very effective for turbulence calculations as turbulence disturbs the composition state space locally. Authors provided also tests approach that gave better results as shown in the case of turbulent, premixed methane and syngas-air flames. As long as the composition state space is covered by a sufficient number of chemical states which can be used as training data set, the ANN can be successfully used to replace stiff ODE solvers.

5.4.5 Enzymatic reaction kinetics

The study [20] investigated the usefulness of ANN to assess the rate of enzymatic reactions. The study was conducted on the basis of a model reaction, the enzymatic hydrolysis of maltose, catalyzed by the enzyme amyloglucosidase. The influence of substrate (maltose) and product (glucose) concentration on the rate of enzymatic reaction was studied. The reaction (5.7) taken into account reads:

$$\text{Maltose} + H_2O \xrightarrow{Amyloglucosidase} 2\text{Glucose} \tag{5.7}$$

Enzymatic hydrolysis of maltose was performed in 5 ml of reaction solution (phosphate buffer, 100 mM, pH 5.5). The reaction solution was heated to 37 °C using a water continuous stirred batch reactor, the reaction was started by adding amyloglucosidase enzyme to the solution. Data on the dynamics of the concentrations C were obtained by measuring the amount of glucose produced during the enzymatic reaction. Transient experimental data were adjusted to Boltzmann's function and then transformed into a differential form (5.8) characterizing the kinetics of the reaction.

$$-\frac{dC}{dt} = \frac{(A_1 - A_2)e^{(t-t_0)/\Delta x}}{\Delta x\left(1 + e^{(t-t_0)/\Delta x}\right)^2} \tag{5.8}$$

The selection of the Boltzmann's function and its derivative was justified by its popularity in the field of biochemistry to render the kinetics of accuracy required. The A_1, A_2, t_0 and Δx were the coefficients to be fitted based on experimental transient evolution of concentrations during enzymatic reaction. In this way the model described by the classical speed equation was obtained. The model built in this way served as a training set for teaching the neural network. Data obtained from seven time courses were used to train ANN, and another set of data obtained from eight time courses was used to test the trained network.

The artificial neural network was built with The Neural Network Toolbox of MATLAB software. The feedforward neural network was formed, which consists of three neurons in the input layer, four neurons in the hidden layer and one neuron

in the output layer. Neural network inputs are maltose concentration, glucose concentration and initial glucose concentration, and system output is the rate of reaction. The network was trained until the square mean value between the training set values and the outputs was less than $1{\cdot}10^{-4}$. Authors report that the determination coefficient (R^2) showed good correlation between the estimated and experimental data sets for both the training (0.992) and the validation data (0.965). In the further part of the study the estimated ANN data were used in the numerical solution of the batch reactor modeling equation to obtain time evolution of concentrations. The authors present the use of ANN for estimating the kinetics of the reaction, pointing out some relevant advantages. First of all, ANN estimates the reaction rate without using any kinetic model equation. Secondly, estimating the reaction rate without the use of a kinetic model eliminates the risk resulting from the wrong choice of the kinetic model. Thirdly, the use of ANN requires smaller amount of experimental data than conventional kinetic model with similar accuracy. This is due to the general feature of neural networks, i.e., the ability to generalize. Fourthly, when using ANN, there is no need to derive kinetic model. As a result, the classical kinetic model does not impose any limitations on the actual experimental analysis of the kinetics of the reaction system. On the other hand ANN does not provide any explanation on physicochemical nature of the process.

5.4.6 Deep learning based on quantum modeling of chemical reactions paths

Data on the activation energy of chemical reactions are difficult to access due to the complexity of the implementation of experimental studies, especially in the case of reactions that are characterized by high rates. Prediction of activation energy is useful for computer-aided generation of reaction mechanisms and planning of organic synthesis. The authors [21] propose a deep learning model for prediction of activation energy of selected reactions and geometry of product molecules. The neural network model is trained on a new, diverse set of quantum chemistry reaction data for the gaseous phase. The authors have developed a deep learning model to predict activation energies for a wide range of reaction types, which does not depend on any additional input data and requires only graphical representation of the reagents. Such a model represents a deep learning neural network built to estimate kinetics by automatically generating the reaction mechanism and allows for the quantitative ranking of candidates for possible reactions. Data based on large-scale quantum chemistry calculations were used. The authors rightly state that high performance experiments are also beginning to become a valuable source of new data.

The authors use the directed message passing neural network (D-MPNN) method to predict molecular properties. The D-MPNN maintains two representations for information placed on the bond between the atoms v and w: one of the atoms v to ($mtvw$) and one of the atoms w to ($mtwv$). Consequently, instead of aggregating information

from adjacent atoms, D-MPNN aggregates information from adjacent bonds. The information about each bond is updated based on all other, that is, not belonging to a given bonded atom, incoming information about bonds. Because of this structure, with the information located on the bonds and the distinction between the two directions of the bond messages, the D-MPNN has more control over the information flow through the molecule and therefore can build more informative molecular representations [22].

The D-MPNN constructs the learned representation of the molecule by transferring information between the elements of the chart using information related to atomic bonds. Binding information contains characteristics of the atom, such as atomic number, and the initial properties of the bond, using a neural network. The neural network iteratively updates the geometry of the molecules by using information from adjacent bonds. Since the mapping of atoms between substrate and product is known, the difference between all atoms of products and substrates is calculated to obtain so-called fingerprints of atomic differences. In the next step each fingerprint is used by the same neural network to create a reaction path encoding. The final stage of the process is the reading phase, in which the network learns the linear combination of elements in the encoding of the reaction to obtain the estimated activation energy.

The model is trained using a dataset for the gaseous phase of elementary atomic reactions obtained from Density Functional Theory (DFT) calculations with appropriately adopted functional databases. These data include a set of organic chemistry reactions with carbon, hydrogen, nitrogen and oxygen atoms.

A promising application of this model is to enable the discovery of new species of reactions in large set of possible chemical mechanisms. Reaction candidates can be generated from each molecule in given mechanism by systematically changing the bonds to calculate potential products. The deep learning model can then evaluate which candidates have the lowest barriers and justify further evaluation. This process of discovering the reaction would be done without initial template structures, while the software for generating conventional reaction mechanisms relies on such templates to limit acceptable chemistry.

In order to assess whether the model's chemical reaction acquisition is chemically justified, molecule encoding has been embedded in two-dimensional space using the stochastic neighbor t-SNE method. This method is a nonlinear and unsupervised technique used primarily for the exploration and visualization of multidimensional data [23]. The activation energies gradients analyzed by the authors indicate that the model has learned to organize reactions that were not presented to the model during training, in a correct way that correlates with chemical reactivity. Moreover, different regions identified by neural network in the chemical space representing the reactions correspond to different types of reactions. Since the types of reactions are defined by the types of bonds, the reactions within each type are associated with many different functionalities and chemical properties; nevertheless,

the model learns to group reactions of the same type into individual, separate subsets. The authors additionally tested the situation if the model behaves correctly when different functional groups are present in a given compound. The effects of replacing hydrogen atoms with substituent containing different functional groups were analyzed and model predictions were verified using DFT quantum calculations. This analysis indicated that the neural network model correctly identifies the fact that the replacement of an electronegative group near the reaction center has a strong impact on activation energy, while the replacement of any functional group far from the reaction center of a given reaction does not significantly affect activation energy. The authors conclude that activation energies for various gaseous phase organic chemistry reactions can be accurately predicted using the deep learning method.

5.4.7 Coupling microscale kinetics and macroscale chemistry using AI Random Forest algorithms, catalysis on $RuO_2(110)$ surface

The coupling of reactor models with the heterogeneous nature of the chemical reaction requires simultaneous consideration of phenomena from atomic level to reactor level. This includes orders of magnitude of differences both spatially and temporally, making direct coupling impractical from a calculation point of view. In addition, however, the strong nonlinearity and stiffness of the resulting equation systems require specific numerical procedures related to the reactive source members in the macroscale governing equations. One of the classical approaches to modeling kinetics at the microscale for subsequent use at the macroscale is the Monte Carlo method. Unfortunately, Monte Carlo methods, whose basic concept is to use randomness to solve problems that can be deterministic in principle, are very demanding due to the calculation load.

Machine learning techniques (ML) can overcome these limitations because they are specifically designed to work with large dimensional data sets. The authors [24] have proposed collaborative learning methods and ANNs for effective tabulation and regression of response rate and turnover frequency. In particular, collaborative learning methods, such as Random Forest (RF), have been recognized as an effective solution to overcome the limitations of mean-field or Monte Carlo kinetic models.

In order to approximate the computationally intensive kinetic models of the first principles in the catalytic way, the authors suggest using a design procedure to generate a training set for machine learning algorithms using an RF ensemble learning method. The proposed methodology was used to array mean field and kinetic models of Monte Carlo kMC in order to combine them with reactor simulations in CFD technique. The trained Machine Learning algorithms with the results of the procedure are then incorporated into both macroscopic reactor models and computational fluid dynamics simulations. First, the plug flow reactor model is used to make a direct comparison with the full classical kinetic model solution. Secondly the CFD simulation of

the complex three-dimensional geometry is calculated using a tabular kMC model for the oxidation of CO on ruthenium oxide.

The authors verified the proposed approach using the process of CO oxidation on the surface of $RuO_2(110)$ catalyst. In this model a grid representation of the active surface was used, taking into account different types of sites, i.e., bridge and cus (coordinatively unsaturated sites). The simulations used a grid consisting of 20 × 20 surface sites (200 bridge and 200 cus) and periodic boundary conditions proposed in [25]. The kMC model takes into account all elementary processes that can occur on the grid: dissociative adsorption of oxygen, associative desorption of oxygen, unimolecular adsorption and desorption of carbon oxide together with surface reactions according to the Langmuir–Hinshelwood CO + O model. Readsorption of CO_2 together with surface diffusion of CO and O are not taken into account due to high partial pressure of the components under consideration. For a given set of operating conditions (i.e., CO, O_2 partial pressures and temperature), the result of the simulation with the classic kMC is a steady state catalytic turnover frequency and the corresponding surface coverage distribution. For data from the classical kMC model, turnover frequency and corresponding ranges are tabulated using ExtraTrees. Such a model can be included in simulation calculations combining the chemistry of the reaction catalyzed on the surface with the reactor size scale.

Machine learning methods were developed in order to carry out effective and accurate discovery learning (process where the neural network determines rules by examples presented). The RF algorithm is a machine learning method used mainly for classification or regression problems. As a tabulation/interpolation method, RF presents the ability to effectively handle large data sets with a large number of input dimensions (i.e., independent variables), called descriptors. Here, simulation data from complex kinetic systems, e.g., the classical kMC model, which depend on several descriptors, such as partial pressure and temperature, are tabulated. RF model shows good tabular and regression parameters even for relatively small training datasets.

The authors have included the kMC kinetic model with the first principles into the CFD simulation with the tabular turnover frequency ExtraTrees, enabling coupling between the exact chemistry description at microscale and process size scale in the reactor. The authors conclude that the simulation of CO oxidation on a complex three-dimensional open-cell foam geometry shows promising possibilities of artificial intelligence methods in predicting the concentrations of components in the gaseous phase and on the surface during the course of complex surface-catalyzed chemical reactions.

5.4.8 Hydrogen oxidation

In the paper [26] authors use a deep learning of the ANN feedforward type to model the H_2 oxidation reaction. The machine learning algorithm was adapted to this

case, in such a way which enabled to act as a universal approximation, improving in areas of high nonl inearity, noise and background signal bias. This approach allows for flexibility in both network architecture and loss functions, as well as for scalability on modern architectures. The primary goal of implementing a deep neural network is to approximate functions f:$x \rightarrow y$ based on input x and output y. After selecting function f, the network weights are adjusted by a process called stochastic gradient descent in which the model:

- iteratively performs predictions on randomly selected sets of training examples,
- determines the quality of the prediction using the L loss function,
- updates parameter weights in a process called back propagation.

This process is repeated until the learning error is at an acceptable minimum. The trained model can then be used to generate predictions for data the network has not seen. In case of investigated training data, the authors use in particular the training autoregression function f:$(T_t, p_t, Y_t) \rightarrow (T_t+\Delta t, p_t+\Delta t, Y_t+\Delta t)$ for a specific period of time Δt for hydrogen oxidation reaction.

The authors proposed to assembly network architecture of 12 jointly trained subnetworks, one for each variable, namely temperature, pressure and each of the 10 reagents. Each subnetwork consists of two densely connected hidden layers (every input is connected to every output) with 30 neurons per layer. During the training, authors also apply the uniform regulation of L1 (lasso – absolute penalty value) and L2 (ridge – square penalty value) in each node to prevent overturning and encourage more generalized function. The term "penalty" or tuning parameter denoted by the authors by θ, but in many other sources by λ [27], was introduced to avoid undesirable limitation of the prediction of the neural network only to the datasets for which it was trained. This is particularly important for the integration of transient series of values: if the network is not immune to small errors in the input data, errors will spread, and time evolution will deviate from proper solution. Both L1 and L2 normalization encourage small parameter weights, but L1 normalization has the additional quality of reducing insignificant parameters to zero.

The authors use the Cantera open-source library [16] to generate a synthetic, combustion of H_2/O_2 mixtures data using the Westbrook H_2-air reaction mechanism [28]. The size of the data set and parameter space were different for each experiment. The network is trained for the initial and final process conditions: temperature T, pressure p and concentration of the components $Y = (Y_0, \ldots Y_k)$. Neural network is trained more efficiently and effectively when data is normalized and scaled to a small range, so authors used logarithmic normalization for their input states.

The neural network has been implemented using the TensorFlow [12] machine learning library. In order to improve the neural network predictive ability of free radicals, which have very low concentrations compared to major species, subnetworks of species have been trained independently over 100 epochs (a single use of all the teaching cases contained in the training set). The full network was trained

for 1000 epochs in a collection of 400 examples each. The training sets were reshaped and selected before each epoch to facilitate training and avoid local minima. The data set was randomly divided into training, test and validation sets containing 80, 10 and 10% of data respectively.

The authors validated their model in the H_2 oxidation reaction, testing its accuracy and effectiveness in stiff regimes. Authors state that the neural network model was built using equally spaced values across the parameter space, and some regions of the teaching space are more problematic to model than others. Consequently the prediction accuracy was lower close the boundary of the training set which was particularly visible at low temperatures.

5.5 Summary

Although quite a few of the cases presented above claim to use deep learning techniques it seems to be a bit of exaggeration, at least at the neural network cognitive level. Deep neural network refers more to the cognitive capabilities of the artificial intelligence than to the way it is assembled. Although it is obvious that deep learning demands that the network construction is made of higher number of layers and neurons. Each case presented although was successful to show the potential of trained neural network prediction in typical engineering and design calculations. The real deep learning algorithms allows to make huge generalizations, which should be understood in regards of chemical kinetics, as recognition of hidden patterns, reaction paths, finding relations difficult to catch by traditional models and finally ability to predict kinetics for chemical reactions that the network was never trained for. In the case of reinforced training it extends the possibilities to teach the artificial intelligence system also to recognize and predict paths for even different classes of chemical processes which were not presented in the course of training process. Taking into consideration that some successful neural network applications were already done to solve the Schrödinger equation [29] to create mapping between a confining electrostatic potential and the ground-state energy, kinetic energy, and first excited state of a bound electron, such model-based reinforcement training could be implemented by incorporating selected quantum models. As a typical example, the chemical reaction path determinable based on well-known *ab initio* transition state theory calculations [30] could be interpreted and considered as training set for artificial neural network modeling.

Future perspective for modeling the chemical reactions paths and rates by artificial intelligence systems is promising. Artificial neural network might be understood in simple picture as a huge set of regression variables, which in fact are weights on the neurons of the neural layers, to be adjusted during the training process. Thus the ability of neural network to generate comprehensive and generalized responses to given problem allows to create predictions even for the chemical and

physical systems it was not trained. Such behavior resembles human brain functions in its fundamental cognitive roles. The process of training the neural network is burdened with huge calculation work and demand for big training datasets. However the use of already trained neural model is fast. Therefore, the main obstacle in this field is the requirement to use huge training datasets and training time.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. Lecun, Y, Bengio, Y, Hinton, G. Deep learning. *Nature* 2015;521:436–44. Nature Publishing Group. https://doi.org/10.1038/nature14539.
2. Srivastava, N, Hinton, G, Krizhevsky, A, Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;15:1929–58.
3. Palmer, J, Chakravarty, A. Supervised machine learning. In: *An introduction to high content screening*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2015:231–45 pp.
4. Oja, E. Finding clusters and components by unsupervised learning. *Lect Notes Comput Sci* (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2004;3138:1–15. https://doi.org/10.1007/978-3-540-27868-9_1.
5. Feng, Y, Xie, M, Wang, L. Semi-supervised learning method of constructive neural networks. *Proceedings of 2015 IEEE advanced information technology, electronic and automation control conference, IAEAC 2015*. Chongqing, China: IEEE; 2016:1020–3 p. https://doi.org/10.1109/IAEAC.2015.7428711.
6. Wang, Q, Zhan, Z. Reinforcement learning model, algorithms and its application. *Proceedings 2011 international conference on mechatronic science. Electric engineering and computer, MEC 2011*. Jilin, China: IEEE; 2011:1143–6 p. https://doi.org/10.1109/MEC.2011.6025669.
7. Irvine, WM. Langmuir–Hinshelwood mechanism. In *Encyclopedia of astrobiology*. Berlin Heidelberg: Springer; 2011: 905 p.
8. Irvine, WM. Eley–Rideal mechanism. In *Encyclopedia of astrobiology*. Berlin Heidelberg: Springer; 2011:485 p.
9. Mora-Briseño, P, Jiménez-García, G, Castillo-Araiza, CO, González-Rodríguez, H, Huirache-Acuña, R, Maya-Yescas, R. Mars van Krevelen mechanism for the selective partial oxidation of ethane. *Int J Chem React Eng* 2019;17. https://doi.org/10.1515/ijcre-2018-0085.
10. Ross, JRH. The kinetics and mechanisms of catalytic reactions. *Contemporary catalysis*. Amsterdam, Netherlands: Elsevier; 2019:161–86 p.
11. Microsoft. *Learn ML.NET|Free tutorials, courses, videos, and more|.NET*. Available from: https://dotnet.microsoft.com/learn/ml-dotnet [Accessed 30 Aug 2020].
12. TensorFlow. Available from: https://www.tensorflow.org/ [Accessed 30 Aug 2020].
13. Keras. *Keras: the Python deep learning API*. Available from: https://keras.io/ [Accessed 30 Aug 2020].
14. ONNX. *ONNX|Home*. Available from: https://onnx.ai/ [Accessed 30 Aug 2020].

15. Barwey, S, Raman, V. *A neural network inspired formulation of chemical kinetics [Online]*; 2020. Available from: http://arxiv.org/abs/2008.08483 [Accessed 02 Sep 2020].
16. Goodwin, DG, Speth, RL, Moffat, HK, Weber, BW. Cantera: an object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. *Zenodo*; 2018. https://doi.org/10.5281/ZENODO.1174508.
17. Yang, F, Dai, C, Tang, J, Xuan, J, Cao, J. A hybrid deep learning and mechanistic kinetics model for the prediction of fluid catalytic cracking performance. *Chem Eng Res Des* 2020;155:202–10. https://doi.org/10.1016/j.cherd.2020.01.013.
18. Kollenz, P, Herten, DP, Buckup, T. Unravelling the kinetic model of photochemical reactions via deep learning. *J Phys Chem B* 2020;124:6358–68. https://doi.org/10.1021/acs.jpcb.0c04299.
19. Ansys. *Chemkin-Pro: chemistry effects predicting simulation software|Ansys*. Available from: https://www.ansys.com/products/fluids/ansys-chemkin-pro [Accessed 29 Aug 2020].
20. Baş, D, Dudak, FC, Boyaci, IH. Modeling and optimization III: reaction rate estimation using artificial neural network (ANN) without a kinetic model. *J Food Eng* 2007;79:622–8. https://doi.org/10.1016/j.jfoodeng.2006.02.021.
21. Grambow, CA, Pattanaik, L, Green, WH. Deep learning of activation energies. *J Phys Chem Lett* 2020;11:2992–7. https://doi.org/10.1021/acs.jpclett.0c00500.
22. Swanson, K. *Message passing neural networks for molecular property prediction*. Massachusetts, USA: Massachusetts Institute of Technology; 2019.
23. Van Der Maaten, L, Hinton, G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;9: 2579–605.
24. Bracconi, M, Maestri, M. Training set design for machine learning techniques applied to the approximation of computationally intensive first-principles kinetic models. *Chem Eng J* 2020;400:125469. https://doi.org/10.1016/j.cej.2020.125469.
25. Reuter, K, Scheffler, M. Composition, structure, and stability of (formula presented) as a function of oxygen pressure. *Phys Rev B Condens Matter* 2002;65:1–11. https://doi.org/10.1103/PhysRevB.65.035406.
26. Sharma, AJ, Johnson, RF, Kessler, DA, Moses, A. Deep learning for scalable chemical kinetics. *AIAA scitech 2020 forum*. Orlando, USA: American Institute of Aeronautics and Astronautics (AIAA); 2020. https://doi.org/10.2514/6.2020-0181.
27. Melkumova, LE, Shatskikh, SY. Comparing ridge and LASSO estimators for data analysis. *Procedia Eng* 2017;201:746–55. https://doi.org/10.1016/j.proeng.2017.09.615.
28. Marcuś, M, Conaire, M, Curran, HJ, Simmie, JM, Pitz, WJ, Westbrook, CK. A comprehensive modeling study of hydrogen oxidation; 2015.
29. Mills, K, Spanner, M, Tamblyn, I. Deep learning and the Schrödinger equation. *Phys Rev A* 2017;96:042113. https://doi.org/10.1103/PhysRevA.96.042113.
30. Truong, TN, Truhlar, DG. Ab initio transition state theory calculations of the reaction rate for OH + CH4 → H2O+CH3. *J Chem Phys* 1990;93:1761–9. https://doi.org/10.1063/1.459103.

Magdalena Olkiewicz, Bartosz Tylkowski, Josep M. Montornés, Ricard Garcia-Valls and Iwona Gulaczyk

6 Modelling of enzyme kinetics: cellulose enzymatic hydrolysis case

Abstract: Enzymes as industrial biocatalysts offer numerous advantages over traditional chemical processes resulting on improvements in process economy and environmental sustainability. Because enzymes are extensively used in different industrial areas, the enzyme kinetics is an important factor for industry as it is able to estimate the extent of substrate conversion under known conditions and evaluate reactor performance. Furthermore, kinetic modelling is useful in the analysis, prediction, and optimization of an enzymatic process. Thus, kinetic modelling is a powerful tool for biochemical reaction engineering. In addition to the aforementioned, in the industrial technology, modelling together with simulation play a key role because they help to understand how a system behaves under specific conditions, and thus they allow saving on costs and lead times. Enzymatic conversion of renewable cellulosic biomass into biofuels is at the heart of advanced bioethanol production. In the production of bioethanol from cellulosic biomass, enzymatic hydrolysis of cellulose to fermentable sugars accounts for a large portion (~30%) of the total production costs. Therefore, a thorough understanding of enzymatic hydrolysis is necessary to create a robust model which helps designing optimal conditions and economical system. Nevertheless, it is a challenging task because cellulose is a highly complex substrate and its enzymatic hydrolysis is heterogeneous in nature, and thus the whole process of cellulose conversion to glucose involves more steps than classical enzyme kinetics. This chapter describes the bases of enzyme kinetic modelling, focussing on Michaelis-Menten kinetics, and presents the models classification based on the fundamental approach and methodology used. Furthermore, the modelling of cellulose enzymatic hydrolysis is described, also reviewing some model examples developed for cellulose hydrolysis over the years. Finally, the application of enzyme kinetics modelling in food, pharmaceutical and bioethanol industry is presented.

Keywords: cellulose hydrolysis, enzymatic hydrolysis, enzymes, kinetic modelling, Michaelis-Menten kinetics

This article has previously been published in the journal Physical Sciences Reviews. Please cite as: Olkiewicz, M., Tylkowski, B., Montornés, J. M., Garcia-Valls, R., Gulaczyk, I. Modelling of enzyme kinetics: cellulose enzymatic hydrolysis case *Physical Sciences Reviews* [Online] 2021, 6. DOI: 10.1515/psr-2020-0039

https://doi.org/10.1515/9783110678215-006

6.1 Introduction

Kinetic modelling is a powerful tool for biochemical reaction engineering. The useful model provides guidance in the design process or enables a better understanding of the phenomena and interactions. With appropriate utilization of the models, the reaction can be improved ahead of testing and go into the test laboratory with higher confidence to achieve successful results. Once the model is built, it is used to design, optimize, operate and control the reactors, and further to scale up the whole industrial process. In addition, modelling and simulation together play a key role in the industrial technology because they help understanding how a system behaves under specific conditions, and thus they allow saving on costs and lead times [1], [2].

Enzymes, as industrial biocatalysts, offer numerous advantages over traditional chemical processes resulting in improvements in process economy and environmental sustainability. Thus, enzyme catalysis has been implemented in several industries, such as food and pharmaceutical, with recent trends for biofuel production [3]. The constantly growing demand for food, pharmaceutical and biofuel products requires the development of tailor-made, low-cost products manufactured by sustainable processes and for this reason, modelling plays an important role in these sectors of industry. One of the key elements of enzyme kinetics modelling is the grouping of model compound reaction pathways and kinetics into a prediction of a real system. However, in practice, this procedure might not be so simple [4]. For example, in the case of cellulose, the most abundant and sustainable biopolymer that enzymes can degrade into soluble sugars, the reaction mechanism and the morphology of cellulosic biomass are very complex, resulting in difficulties in the reaction modelling [5].

Enzymatic conversion of renewable cellulosic biomass into biofuels is at the heart of advanced bioethanol production. In the production of bioethanol from cellulosic biomass, enzymatic hydrolysis of cellulose to fermentable sugars accounts for a large portion (~30%) of the total production costs. Therefore, a thorough understanding of enzymatic hydrolysis is necessary to create a robust model which helps to design optimal conditions and economical systems [6]. Nevertheless, it is a challenging task because cellulose is a highly complex substrate and its enzymatic hydrolysis is heterogeneous in nature, and thus the whole process of cellulose conversion to glucose involves more steps than classical enzyme kinetics [7], [8]. In addition, the cellulose hydrolysis mechanism is greatly affected by several factors, such as enzyme characteristics (adsorption, inhibition, synergism, activity, composition), substrate characteristics (degree of polymerization, crystallinity, accessible surface area, hemicellulose and lignin content) and operating conditions (pH, temperature and agitation) [5], [7], [8], [9], [10]. Including all of them into a kinetic model of enzymatic cellulose hydrolysis is one of the most challenging subjects in engineering for this process. In this scenario, mechanistic models are very useful because they take account of the mechanisms through which changes occur in the reaction system. Usually, it is desirable that a mechanistic model incorporates key

information about the comprehension of the process while, on the other hand, not being too complex in an attempt to capture overall phenomena [8].

This chapter describes the bases of enzyme kinetic modelling, focusing on Michaelis-Menten (M-M) kinetics, and presents the models classification based on the fundamental approach and methodology used. Furthermore, the modelling of cellulose enzymatic hydrolysis is described, also reviewing some model examples developed for cellulose hydrolysis over the years. Finally, the application of enzyme kinetics modelling in food, pharmaceutical and bioethanol industry is presented.

6.2 Modelling of enzyme kinetics

Because enzymes are extensively used in different industrial areas, the enzyme kinetics is an important factor for industry as it is able to estimate the extent of substrate conversion under known conditions and evaluate reactor performance. In addition, mathematical modelling is useful in the analysis, prediction and optimisation of an enzymatic process. Since the most famous M-M kinetics model was developed, the biochemistry has advanced tremendously, and then enzymes kinetics modelling has become an important factor for developing any enzymatic industrial processes. Since then, different modelling approaches have been developed for enzyme kinetics. This section describes the principles and importance of M-M kinetics in modelling of enzymatic reactions, methodologies developed over the years and models classification based on the fundamental approach and methodology used.

6.2.1 The Michaelis-Menten kinetic

Enzymes are highly specific catalysts for biochemical reactions that show a high selectivity for a single reactant, or substrate. Enzyme reactions are modelled as a multistep mechanism (Figure 6.1), where first, the enzyme (E) and substrate (S) bind to each other to form the enzyme-substrate complex (ES). Then, the substrate is transformed into the product(s), giving enzyme-product complex (EP). Finally, the product(s) dissociates from the enzyme, releasing the enzyme to react again [11].

This enzymatic reaction is schematically described within the following multistep mechanism (Equation (6.1)):

$$E + S \underset{k_r}{\overset{k_f}{\rightleftarrows}} ES \underset{k_{catr}}{\overset{k_{cat}}{\rightleftarrows}} EP \underset{k_{asoc}}{\overset{k_{dis}}{\rightleftarrows}} E + P \tag{6.1}$$

where k_f, k_r, k_{cat}, k_{catr}, k_{dis}, k_{asoc} are forward rate constant, reverse rate constant, catalytic rate constant, catalytic revers rate constant, dissociation constant and association constant, respectively.

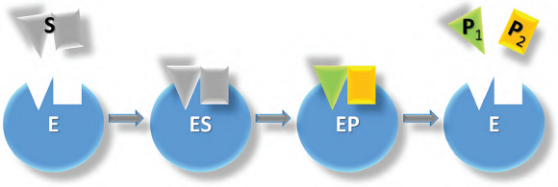


Figure 6.1: Model of enzymatic reaction. S–substrate, E–enzyme, ES–enzyme-substrate complex, EP–enzyme-product complex, P_1, P_2–products.

However, because the rate of catalysis is measured by observing formation of the product, most of the models of biochemical systems use the simplified form of the reaction, where the catalytic step and product dissociation is condensed into a single, irreversible step (Equation (6.2)) [11], [12]:

$$E + S \underset{k_r}{\overset{k_f}{\rightleftarrows}} ES \overset{k_{cat}}{\rightarrow} E + P \tag{6.2}$$

The simplified form of the reaction, known as M-M enzyme-catalysed model, describes reversible binding of E and S into the ES complex, and irreversible conversion of the complex into the product (P) and the free E. The model was proposed in 1913 by Leonor Michaelis and Maud Leonora Menten, who published their work about the reaction kinetics of the enzyme invertase [13]. They investigated a quantitative theory of enzyme kinetics and proposed a mathematical model for it. In their original analysis, Michaelis and Menten based their equation on the assumption that the substrate, the concentration of which greatly exceeds the enzyme, is in an instantaneous equilibrium with the complex (equilibrium approximation), thus there is no change in the substrate concentration (dS/dt = 0). This simplification results in the following equation describing the rate of enzymatic reactions (Equation (6.3)):

$$\upsilon = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_d + [S]} \tag{6.3}$$

where $K_d = k_r/k_f$ is the dissociation constant of the complex ES, $V_{max} = k_{cat}[E]_0$ is the maximum reaction velocity and $[E]_0$ is the initial enzyme concentration.

Their work was further developed by G.E. Briggs and J.B.S. Haldane, who formulated a slightly different assumption (quasi-steady-state approximation) in 1925 [14], whereby the concentration of the complex is set to be constant over small time scales. They assumed that after initiation of the reaction, the ES complex is in equilibrium ($dES/dt = 0$). This results on the final form of the M-M equation (Equation (6.4)):

$$\upsilon = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_M + [S]} \tag{6.4}$$

where $K_M = k_r + k_{cat}/k_f$ is the Michaelis constant, roughly the dissociation constant of the complex.

Therefore, the final form of M-M kinetics (Equation (6.4)) adopts the Briggs-Haldane quasi-steady-state assumption. The M-M model is one of the simplest and best-known approaches to enzyme kinetics. The equation has been used to predict the rate of product formation in enzymatic reactions for more than a century and is still widely considered today a starting point in modelling enzymatic activity. Although, other models of the same type and often based on the same assumptions also exist, they are frequently faced consolidated scepticism [15]. Furthermore, it must be pointed out that the assumption of irreversibility in the M-M model (Equation (6.2)) is only true when the concentration of substrate greatly exceeds the enzyme or the energy released in the reaction is very large. In situations where neither of these two conditions hold, the M-M equation breaks down, and more complex modelling approaches taking the reversibility of the reaction into account must be done [12], [13], [14], [15]. In addition, besides reversibility, in some cases, it is necessary to include the effects of enzyme inhibition in the model. There are several types of inhibition: competitive (binding to the same site as the substrate), noncompetitive or allosteric (binding to a different site) and uncompetitive (binding only to the ES complex). Including the inhibition effects to M-M model, the following equations (Equations (6.5)–(6.7)) are obtained [11]:

- Competitive inhibition

$$\upsilon = \frac{d[P]}{dt} = \frac{V_{max}[S]}{[S] + K_M\left(1 + \frac{I}{K_I}\right)} \tag{6.5}$$

where $K_I = k_{ri}/k_{fi}$ is enzyme-inhibitor dissociation constant, k_{fi} and k_{ri} are forward inhibitor rate constant and reverse inhibitor rate constant, respectively.

- Allosteric (noncompetitive) inhibition

$$\upsilon = \frac{d[P]}{dt} = \frac{V_{max}[S]}{([S] + K_M)\left(1 + \frac{I}{K_I}\right)} \tag{6.6}$$

- Uncompetitive inhibitions

$$\upsilon = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_M + [S]\left(1 + \frac{I}{K_I}\right)} \tag{6.7}$$

6.2.2 Modelling over the years

As described by Galanakis et al. [4], enzyme kinetics modelling arises from the late 60s, when the theory of subsite mapping was developed to describe proteases

activity. The subsite theory postulates that the active site of an enzyme consists of several subsites which are responsible for binding the substrate, and one catalytic site which is responsible for hydrolysing the bond. The theory was developed to explain the strong dependence of Michaelis constant and maximal reaction rate on the chain length of the polymers [16]. Thereafter, subsite theory was also used to describe the enzymatic hydrolysis of polysaccharides and, in this case, amino acid chain was replaced with a glucose unit [16], [17], [18]. As the subside theory explains the dependence of Michaelis constant on the polymer length, the models based on subside mapping usually include the M-M equation.

At the same time, other models of polysaccharides enzymatic hydrolysis based on M-M equation which additionally included inhibition [19], [20], [21] and reversibility [20], [22]. Parameters of kinetics were developed. The reversibility and inhibition parameters were not described adequately via subsite mapping [4]. The presented models, developed on M-M equation or subsite mapping and M-M equation, are both deterministic. The deterministic models predict the average behaviour of the system and offered satisfactory solutions on time, but they ignore random variation.

The dramatic increase of kinetics equations for more complex biological systems (with different input variables) that needs multireaction kinetics model and the complexity of data processing in practice, forced the scientists to try different approaches. In particular, stochastic models such as Monte Carlo (MC) were developed for this purpose because they can capture different sources of unexplained variation in the behaviour of the system [23], [24].

More recently, empirical models based on artificial neural network were also applied for polysaccharides enzymatic hydrolysis [25], [26]. The multivariable neural networks method lets model large data in a complex system involving many parameters and performs correlations without requiring mechanistic assumptions of how the output depends on the inputs.

6.2.3 Classification of models

Galanakis et al. classified enzyme kinetics models into three major groups (deterministic, empirical and stochastic) in accordance with the implemented methodology [4], described in the abovementioned section. On the other hand, one of the methods of distinguishing between types of models is to consider the level of understanding on which the model is based. In this case, as Figure 6.2 shows, models are divided into mechanistic (they take account of the mechanisms through which changes occur) and empirical (no account is taken of the mechanism by which changes to the system occur). In addition, depending on the biological question addressed, mechanistic models can be further subdivided into deterministic (when only the average behaviour of the system is of interest) and stochastic (when unexplained variation in the behaviour of the system matters too) [27].

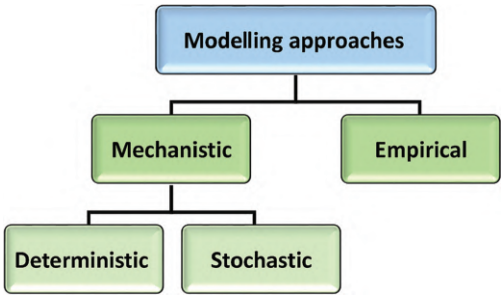


Figure 6.2: Classification of enzyme kinetics models.

As presented in Figure 6.2, there are two types of modelling approaches, empirical and mechanistic modelling. Empirical models describe relations between the variables in an experimental data set using mathematical correlations without requiring detailed knowledge of an underlying mechanism of the system. Because empirical models do not incorporate any knowledge or hypotheses about the inner structural connectivity of the system, they are also known as extrinsic. Thus, an empirical model usually can only accurately predict conditions represented by the data set that was used to build the model. Nevertheless, the empirical models are easy to develop and are useful in enzyme characterization and substrate preparation. By contrast, mechanistic models incorporate an underlying mechanism of the system by which changes in the system are thought to occur, giving initial conditions and future system behaviour can be predicted. Mechanistic models are a mathematical formulation of the internal operation of a system in terms of its constituent parts and mechanisms. As these models address the underlying dynamics of the process, they can be extensively used in every stage. Mechanistic models are developed from the reaction mechanisms, mass transfer considerations and other physical parameters that affect the extent of enzymatic hydrolysis. Thus, these models are quite useful in describing the reaction mechanism of enzymatic hydrolysis of cellulosic biomass [27], [28].

Mechanistic models differ in their complexity based on the intended use of the model and the type of outcome they predict. Thereby, mechanistic models can be further subdivided into either deterministic or stochastic. In deterministic models, results are fully determined by the parameter values chosen and the initial conditions. Thus, a given parameter set will always yield the same model output every time the model is run. By contrast, stochastic models yield different results each time when initialized with the same parameters and initial conditions. This is because deterministic models ignore random variation, and always predict the same outcome from a given starting point. On the other hand, the stochastic models involve inherent randomness and may predict the distribution of possible outcomes. Therefore, deterministic models describe the average behaviour of a system, while the average of multiple stochastic realizations tends to approximate the deterministic solution [27].

6.3 Modelling of cellulose enzymatic hydrolysis

Cellulose is the most abundant and sustainable biopolymer that can be degraded into soluble sugars. In general, cellulose is the major component of lignocellulosic biomass. Enzymatic conversion of renewable cellulosic biomass into biofuels is at the heart of advanced biofuels production. However, the complex heterogeneous reaction mechanism involved in cellulose enzymatic hydrolysis and the complicated morphology of biomass make the modelling of enzymatic hydrolysis challenging. Therefore, this section describes in detail the mechanism of cellulose hydrolysis and review some examples of the models developed for cellulose hydrolysis over the years.

6.3.1 Cellulose hydrolysis

As presented in Figure 6.3, cellulose is a linear, long-chain polymer consisting of D-glucose molecules linked by β-1,4 bonds, with cellobiose as the smallest repetitive unit. In the cellulose structure, the chains are in layers, held together by van der Waals forces with intramolecular and intermolecular hydrogen bonds being present. Cellulose present in lignocellulose is composed of crystalline and amorphous structures, each showing different water solubility and digestibility on enzymatic attack. However, the cellulose structure strongly depends on the biomass origin and therefore can widely vary in chain length and degree of interaction between the chains [9], [29]. This physical arrangement and structural properties of cellulose that affect surface availability and accessibility of cellulose glycosidic bonds are thus a critical consideration for predicting hydrolysis kinetics [29].

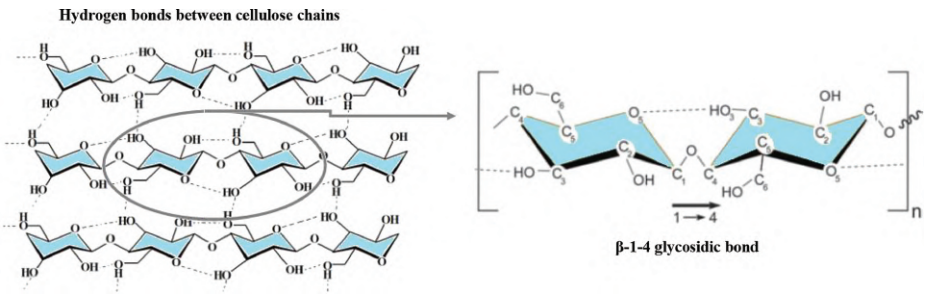


Figure 6.3: Molecular structure of a cellulose.

The enzyme capable to hydrolyse the complex cellulose structure into simple glucose monomer is called cellulase. Cellulase refers to a group of three main enzymes that hydrolyse β-1,4 glycosidic bonds in cellulose, which are endoglucanase (endo-β-1,4-glucanase), exoglucanase (exo-β-1,4-glucanase) and β-glucosidase. Figure 6.4 shows the model representation of the different enzymes action modes during cellulose hydrolysis. As presented in Figure 6.4, endoglucanase internally cleaves cellulose chains with random actions on β-1,4 glycosidic bonds that leads to binary breakages without any position preference. The endoglucanase therefore increases the availability of cellulose chain ends, which is a substrate for exoglucanase. Although endoglucanase can act on cellulose of different degrees of crystallinity, primarily acts on amorphous regions of native cellulose. On the other hand, exoglucanase processively cleaves only glycosidic bonds in the chain ends of cellulose to release cellobiose (a disaccharide). Dependently on the exoglucanase type, CBH-I and CBH-II, they cut off cellulosic polymers from reducing and nonreducing chain ends, respectively. Exoglucanase acts directly towards crystalline cellulose, displacing individual cellobiose chains from the surface of crystalline regions, making it accessible for β-glucosidase action. Then, β-glucosidase hydrolyses cellobiose into glucose monomers [9], [29], [30]. All three enzyme classes must be present in the system to produce glucose efficiently. Therefore, the cellulose hydrolysis involves the synergistic action of endoglucanase, exoglucanase, and β-glucosidases, and sometimes, if present, also exoglucosidase. Exoglucosidase forms glucose via a non-β-glucosidase pathway that acts at the end of the polysaccharide chain removing successive glucose units (Figure 6.4). It must be added that some fungal species can produce exoglucosidase (exo-β-1,4-glucosidase); however, it is not commonly

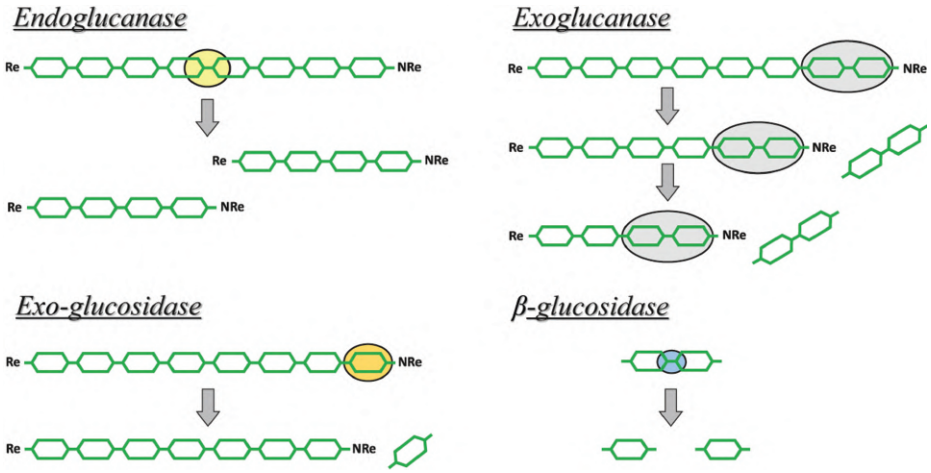


Figure 6.4: Model representation of different enzymes action modes during cellulose hydrolysis. Re, reducing end; NRe, nonreducing end.

present in cellulase preparations. Although the presence of exoglucosidase is not usual, if it is in the system, it must be taken into consideration for modelling hydrolysis kinetics [9].

The enzymatic hydrolysis of cellulosic is therefore a multistep, heterogeneous process characterised by an insoluble reactant (cellulose) and a soluble catalysts (enzymes). Figure 6.5 shows the reaction scheme, where the reaction actually occurs in two phases: solid and liquid. In the solid phase (heterogeneous reaction), insoluble cellulose is initially broken down at the solid-liquid interface (with enzyme adsorption/desorption) via the synergistic actions of the enzymes (endoglucanase/exoglucanase). Endoglucanase attacks the internal glucosidic bonds randomly, and exoglucanase cleaves cellobiose and releases it into the solution. This initial degradation in solid phase is then accompanied by liquid-phase hydrolysis (homogenous reaction) of soluble cellobiose into glucose by the action of β-glucosidase [30], [31].

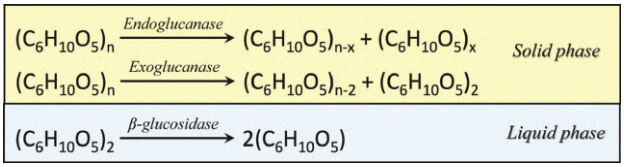


Figure 6.5: Reaction scheme of cellulose hydrolysis occurring in two phases.

Because the cellulose hydrolysis consists of various reactions in different phases (Figure 6.5), the model of cellulose hydrolysis should take into account the multireaction system in which different reactions are modelled; heterogeneous reactions for cellulose breakdown to cellobiose and homogeneous reaction for cellobiose hydrolysis to glucose. Moreover, because cellulose is a highly complex substrate and its enzymatic hydrolysis is heterogeneous in nature, the whole process of cellulose conversion to glucose involves more steps than classical enzyme kinetics. The major steps, described by Bansal et al. [7], are as follows:

1. Adsorption of cellulases onto cellulose via the binding domain.
2. Location of a bond susceptible to hydrolysis on the cellulose surface (chain end for exoglucanase, cleavable bond for endoglucanase).
3. Formation of ES complex (binding of a molecule of cellulose in the active side of the catalytic domain).
4. Hydrolysis of the β-glucosidic bond and simultaneous forward sliding of the enzyme along the cellulose chain.
5. Desorption of cellulases from the cellulose and their transfer to liquid phase or repetition of step 4 or steps 2/3 if only the catalytic domain detaches from the chain.
6. Hydrolysis of cellobiose to glucose by β-glucosidase (classical homogenous enzyme kinetic).

The accordance with the abovementioned description of hydrolysis steps, adsorption of enzymes on the cellulose plays an important role in cellulose hydrolysis, as it is the initiation step. Thus, the kinetic model of cellulose hydrolysis should include the adsorption step. Nevertheless, the M-M–based models, developed initially for homogenous reaction (such as the action of β-glucosidase on soluble cellulose), have been used by many authors to describe the whole hydrolysis process, neglecting the adsorption step, and the models have in general worked satisfactorily [7], [9], [10]. However, some models attempting to predict reducing sugar production over extended batch reactions have encountered difficulties and discrepancies, mainly due to changing reaction dynamics caused by changing substrate structure and loss of enzyme activity [9]. It is therefore important that a kinetic model incorporates vital information about the whole cellulose hydrolysis steps. Because cellulose hydrolysis is a heterogeneous reaction, the classic M-M kinetics is inadequate to explain the action of cellulases (i.e., endoglucanase and exoglucanase) on insoluble cellulose. The overall rate depends on the amount of adsorbed cellulases. In general, adsorption model based on Langmuir adsorption isotherm has been extensively used in modelling of cellulose hydrolysis [7], [31], [32]. Considering the homogenous and heterogeneous reactions, the cellulose hydrolysis should be modelled by both M-M and Langmuir kinetics, respectively. In addition, there are other factors related to both, enzyme characteristics (adsorption, inhibition, synergism, activity, composition) and substrate characteristics (degree of polymerization, crystallinity, accessible surface area, hemicellulose and lignin content), which affect the hydrolysis of cellulose [5], [7], [9], [10]. Including all of them into a kinetic model of enzymatic cellulose hydrolysis is one of the most challenging subjects in engineering for this process. Over the past decades, many models of enzymatic hydrolysis of cellulose have been developed and efforts to enhance understanding of this reaction have been made. These works are found in four main reviews about modelling of enzymatic cellulose hydrolysis, which were published over past years [7], [29], [32], [33]. In the next section, we review some examples of the models developed for cellulose hydrolysis over the years. On the other hand, in Section 6.4, kinetic models developed for different lignocellulosic biomass (the real biomass containing cellulose) are presented.

6.3.2 Model examples used in cellulose hydrolysis

As stated in Section 6.2.3 there are two principal approaches to model enzymatic hydrolysis, empirical and mechanistic. Over the years, various empirical models have been developed to simulate the enzymatic hydrolysis of cellulose [7], [32], [33]. These empirical models have been generally used to correlate hydrolysis either with the structural properties of the substrate or with time. They can help in quantifying the effects of various substrates and enzyme properties on hydrolysis. They

can be simple to construct but require large sets of experimental data. Nevertheless, empirical models lack a theoretical foundation and cannot be applied beyond the range of the data under which they were developed. Therefore, they do not provide any insight into the mechanistic details of the process [6], [7]. Because the mechanism of cellulose hydrolysis is very complex and it is still not well known, more recent works have emphasized on developing detailed mechanistic models that are solved numerically or stochastically rather than empirically. Thus, this section focuses only on the mechanistic models, which include the factors that ought to be taken into consideration in establishing a comprehensive mechanistic kinetic model. Table 6.1 provides a list of some mechanistic models published in the literature over the years, along with their methodology and factors affected the rate of hydrolysis.

As reviewed previously [7], [32], different models have been developed for cellulose hydrolysis since 1975. They were usually based on one of the two methodologies: M-M or Langmuir adsorption (Ads). In 1984, Holtzapple et al. [34] developed a generalized mechanistic model for enzymatic cellulose hydrolysis termed the Holtzapple–Caram–Humphrey-1 (HCH-1) model. The HCH-1 model is a modified M-M model which additionally includes an adsorption step, noncompetitive inhibition by glucose and a parameter that describes the fraction of insoluble substrate available to bind with enzyme. The inclusion of these parameters makes the model more appropriate for heterogeneous cellulose hydrolysis than the M-M model. This was confirmed by good agreement with experimental data, which appeared better than older models based on M-M or Langmuir adsorption [34]. Furthermore, in an additional study, the model was extended including the noncompetitive inhibition by cellobiose [35]. Recently, the HCH-1 model has been used to describe the hydrolysis of lime-pretreated corn stover with the purpose to determine the inhibition pattern, which in the initial model was assumed to be noncompetitive [5]. The study identified a noncompetitive inhibition pattern by glucose for the corn stover–cellulase reaction system, thereby validating the assumptions of the HCH-1 model. This result confirmed that the HCH-1 model is able to predict well not only the hydrolysis of pure cellulose substrate [34], but also lignocellulosic substrate [5]. More recently, the HCH-1 was modified to extend its application to integrated enzymatic hydrolysis for long-term process (i.e., 10 days). The modified model accounts for stochastic MC method for sensitivity analysis. Excellent fits to the data were obtained for the model that uses the mix of enzymes without the need to model each enzyme component individually. Moreover, the sensitivity analyses not only determine the parameters which have greater influence on results, but also can provide direction for further modification of the HCH-1 model to apply it to different lignocellulose biomass [6].

The first stochastic model for hydrolysis of insoluble polysaccharides was developed by Fenske et al. in 1999 [24]. This model used MC simulations for an enzyme mix of endo and exoglucanases, evaluating the synergistic action. Despite the results proved that autosynergism is occurred, it was a theoretical study and the results were not validated experimentally. Anyway, this model did not capture the

Table 6.1: Mechanistic models of cellulose enzymatic hydrolysis.

Substrate feature	Enzyme	Methodology	Inhibition	No of presented equations	Reference
Solka Floc, ASA	*T. viride* + β-glycosidase	Ads, modified M-M	Noncompetitive by glucose	15	[34]
Solka Floc, ASA	*T. viride* + β-glycosidase	Ads, modified M-M	Noncompetitive by glucose and cellobiose	4	[35]
Textile cotton waste, pretreated cellulose pulp	*T. viride* (endo + exo), *A. niger* (β-glycosidase)	M-M	Evaluation of inhibition type by cellobiose and glucose	7	[21]
Cellodextrins DP	β-glycosidase	M-M	Competitive by glucose	31	[36]
Cellulose[a], ASA	Cellulase[a], (end + exo), S	Ads, MC	Type of inhibition was not specified.		[24]
α-cellulose, ASA + (A, C)	*T. reesei* + β-glycosidase	Ads, M-M	Competitive by glucose and cellobiose	20	[9]
Avicel, ASA	*T. reesei* (Cel7A) –exoglucanase	M-M	Evaluation of inhibition type by cellobiose	3	[37]
α-cellulose, A	*Spezyme CP* (Genencor), (endo + exo + β-glycosidase)	Ads	Noncompetitive by cellobiose and glucose (endo/exo), Competitive by glucose (β-glycosidase)	5	[28]
Lime-pretreated corn stover, ASA	*T. reesei* + β-glycosidase	Same model as Holtzapple et al. [34]	Noncompetitive by glucose	7	[5]

(continued)

Table 6.1 (continued)

Substrate feature	Enzyme	Methodology	Inhibition	No of presented equations	Reference
Cellulose[a], DP, ASA	Multiple enzymes, S	Ads, M-M	Competitive by glucose and cellobiose	7	[38]
Cellulose[a], ASA + (A, C), DP	Cellulase[a] (exo + endo + β-glycosidase), S	MC	Type of inhibition was not specified.	12	[39]
Cellulose[a], A, ASA, DP	Cellulase[a] (exo + endo + β-glycosidase), S	Ads, M-M, MC	Noncompetitive by cellobiose and glucose	42	[30]
α-cellulose (C8002)	*Novozymes Cellic* (CTec2) (endo + exo + β-glycosidase)	Modified model of Holtzapple et al. [34], MC	Noncompetitive by glucose	12	[6]

[a]Theoretical study. A, amorphous structure; Ads, Langmuir adsorption; ASA, accessible surface area; C, crystalline structure; DP, degree of polymerisation; MC, Monte Carlo; M-M, Michaelis-Menten; S, synergism.

actual structural properties of cellulose and multi enzyme dynamics. Latterly, more comprehensive stochastic model of cellulose hydrolysis based on MC technique was developed by Kumar and Murthy [39]. The model captured the most important structural properties of cellulose (crystallinity, degree of polymerization, accessibility) and enzyme (mode of actions, synergism, inhibition), and showed qualitative and quantitative agreement with experimental data. In addition, the model was effective in capturing the dynamic behaviour of cellulose hydrolysis during action of individual as well as multiple enzymes. The major advantage of this model is that it can be applied for different substrates, different enzyme characteristics and different reaction conditions after performing independent characterization experiments [39]. Niu et al. [30] also proposed a detailed mechanistic model based on stochastic approach. However, in this case the model was developed for amorphous cellulose and noncompetitive inhibition, so the cellulose structure and inhibition pattern were fixed.

As presented in Table 6.1, in most cases, inhibition is either competitive [9], [36], [38] or noncompetitive [5], [6], [30], [34], [35]. It can also be a combination of both [28]. Beltrame et al. [21] investigated the inhibition pattern for endo and exoglucanase on the hydrolysis of textile cotton and cellulose pulp. Uncompetitive and competitive inhibition on endoglucanase and exoglucanase, respectively, appeared to occur for both substrates. More recently, Bezerra and Dias [37] tested eight different M-M models of inhibition by cellobiose to establish the mode of inhibition on exoglucanase (*Trichoderma reesei* (Cel7A)) during cellulose hydrolysis. It was found that cellulose hydrolysis follows a model that takes into account competitive inhibition on exoglucanase by cellobiose, agreeing with Beltrame's result. Nevertheless, it was stated that the discrepancy between different types of inhibition reported is a result of enzyme concentration and enzyme/substrate concentration ratio, source of cellulase enzyme complex, and/or the hydrolysis time over which the experiments were conducted [5], [9].

6.4 Industrial applications

Kinetic modelling is a powerful tool for biochemical reaction engineering. At the industrial level, the usual purpose of modelling is to improve process efficiency and/or to provide a quantitative basis for process design, control and optimisation. In addition, modelling together with simulation play a key role in the industrial technology because they help to understand how a system behaves in specific conditions, and thus they allow reducing costs and lead times. In other words, if a company wants to introduce a new technology which was tested successfully in a laboratory, it is more useful and cheaper to develop a simulation model of the system and evaluate the costs and benefits of such new solution first than to buy and implement the new solution directly in a plant. This allows saving on money and

time in case the new technology does not work as expected once scaled up. Therefore, modelling and simulation techniques contribute together to simplification and economic favouring of the design, realization, tests and running a live operation of control systems [1], [2].

Enzymes as industrial biocatalysts offer numerous advantages over traditional chemical processes, therefore they have been implemented in foods, pharmaceutical and biofuel industry, among others. The constantly growing food and pharmaceutical industry and increasing demand for their products require the development of tailor-made, low-cost products manufactured by sustainable processes and for this reason, modelling plays an important role in these sectors of industry. Enzyme kinetics modelling has found a broad range of applications in the enzymatic hydrolysis of milk proteins as it is an important process in the food industry that improves the functional properties and reduces allergenicity. Empirical model based on artificial neural networks was successfully used to predict the degree of hydrolysis of skimmed goat milk proteins with subtilisin and trypsin as a function of the operating conditions, namely the reaction temperature, the enzyme-substrate ratio and the time of hydrolysis. Significant correlation between degree of hydrolysis predicted using the models and degree of hydrolysis obtained under experimental conditions was achieved [40]. On the other hand, the hydrolysis of milk protein has been also predicted by stochastic model. In the study by Gao et al. [41], stochastic model based on MC technique was performed for enzymatic hydrolysis of lactose with β-galactosidase for the purpose of real-time applications and the model predicted the reaction kinetics with good accuracy [41]. Empirical model has been also successfully applied in the production of pharmaceuticals such as L-carnitine. Canovas et al. [42] proposed a model which describes all the activities of different enzymes involved in the biotransformation of crotonobetaine into L-carnitine, and was successfully implemented in batch and continuous membrane reactor [42]. Furthermore, the authors optimized the model for a continuous reactor and experimentally verified the predicted optimal parameter profiles. The optimization gave 90% increase in the L-carnitine production rate resulting also in a complete coincidence between the theoretical and the experimental findings [43]. Another pharmaceutical product of a great importance which production was empirically modelled is lactic acid. Lactic acid is also widely used in food and biodegradable plastics manufacturing which is nowadays even a bigger market for lactic acid than pharmaceutical and food industry. Bioconversion of paper mill sludge to lactic acid was modelled using simultaneous saccharification and fermentation (SSF) process which combined the enzymatic hydrolysis of paper mill sludge into glucose and the fermentation of glucose into lactic acid in one reactor. The proposed model of one-step process fitted satisfactorily with the experimental data, thus the SSF is a good choice for bioconversion of waste cellulosic materials to lactic acid [44].

Treatment of byproducts produced in food, agriculture or paper industry is another important area of kinetics modelling applicability as real substrates (i.e., byproducts)

do not show the same characteristic as pure ones and usually they need to be pretreated before use. For this reason, predicting optimal conditions can improve conversion efficiency. Many examples exist in the literature, that is, optimisation of model for enzymatic hydrolysis of cellulose from differently pretreated dairy manures [45], from apple pomace [46], from sugarcane bagasse and straw [8], [10], [47], from corn stover [48], [49], and optimization of model for SSF of rice straw [50] or paper mill sludge [44], [51].

Recently, the enzyme kinetic modelling has been widely used in the bioethanol industry. Bioethanol is one of the most promising alternatives to petroleum fuel. However, to be more competitive, all stages of its production process must be simple, inexpensive, efficient and easy to control. In the production of bioethanol from cellulosic biomass, enzymatic hydrolysis of cellulose to fermentable sugars accounts for a large portion (~30%) of the total production costs; therefore, a thorough understanding of enzymatic hydrolysis is necessary to help design optimal conditions and economical systems [6]. Although different models were applied for cellulose hydrolysis, as described in Section 6.3.2, from an industrial point of view, it is also necessary to develop a kinetic model being able to describe the process as realistically as possible. In this context, modelling should be focused on the enzymatic hydrolysis of lignocellulosic biomass, where cellulose is combined with hemicellulose and lignin, rather than on the model substrate (i.e., cellulose), as real substrates do not show the same characteristics as pure cellulose [52]. However, modelling the enzymatic hydrolysis of lignocellulosic materials is probably one of the most challenging subjects in bioreactor engineering due to the complexity of the reaction process, influenced by a number of hydrolysis conditions. Although empirical models for prediction of enzymatic hydrolysis of lignocellulosic biomass can also be found in the literature [46], [53], the mechanistic models have been studied preferably. During the last two decades, many efforts have been made to understand the reaction mechanism and develop a suitable model for lignocellulosic biomass hydrolysis to support advanced bioethanol production [5], [8], [10], [45], [47], [48], [49], [52], [54]. The first model that included also competitive xylose inhibition besides glucose and cellobiose inhibition was proposed by Kadam et al. [48]. The direct conversion of cellulose to glucose by exo-β-glucosidase and the temperature effect as well as the reactivity substrate were also incorporated in this model. It was one of the most complete mechanistic models for enzymatic hydrolysis of lignocellulosic biomass which presented a multireaction kinetics model for batch enzymatic hydrolysis of pretreated corn stover. This model has been experimentally verified [48], [49], evaluated in applications of simulation and optimization process [49], [52], [55], extensively analysed statistically [54], and adapted, with some improvements, to model the enzymatic hydrolysis of pretreated sugarcane straw [10]. Angarita et al improved the model taking into account the conversion of hemicellulose to xylose and lignin content. The model was able to predict with reasonable accuracy the concentration of glucose formed in a batch enzymatic hydrolysis of sugarcane straw at high-solids concentration [10]. More recently, also

based on Kadam's model, a comprehensive stochastic model was developed to quantify the overall impact of uncertain model parameters associated with the acid pre-treatment and enzymatic hydrolysis. Because bioethanol can be produced from different feedstock, the influence of variability in feedstock composition, among others, was studied. The stochastic simulations showed evidently the expected influence of variability in feedstock composition and kinetic parameters on the product yield, which could also affect the optimal batch time. The stochastic model also demonstrated that pre-treatment and hydrolysis of lignocellulosic biomass are affected by several uncertainties, which must be systematically considered for a robust process design [52]. All publications reviewed previously are about the modelling of lignocellulosic biomass hydrolysis to fermentable sugars because this step accounts for a large portion of the total bioethanol production costs. The next step in bioethanol production is a fermentation. Besides, the two process steps (i.e., hydrolysis and fermentation) can also be performed simultaneously and is called SSF. For industrial application, the one-step process is preferable as it reduces operating cost because of the lower number of reactors needed. It has been estimated that the capital investment can be reduced by more than 20 % with SSF compared with separate hydrolysis and fermentation processes [56], [57]. Therefore, several kinetic models describing SSF [50], [56], [58] or cofermentation (SSCF) [51], [55], [57] have been proposed for different reactor configurations: batch [50], [51], [56], [57], [58], fed-batch [55], [57] and continuous [55], [56], [58]. The important issue was to demonstrate that the process under SSF is beneficial particularly when high product ethanol concentration is desired. The operating time of SSF process was reduced in comparison with the one needed for the two-step process [50].

6.5 Conclusions

Enzymes as industrial biocatalysts offer numerous advantages over traditional chemical processes resulting on improvements in process economy and environmental sustainability. Because enzymes are extensively used in different industrial areas, the enzyme kinetics is an important factor for industry as it is able to estimate the extent of substrate conversion under known conditions and evaluate reactor performance. In addition, mathematical modelling is useful in the analysis, prediction and optimisation of an enzymatic process.

In enzymatic process modelling, there are two main approaches empirical and mechanistic. Empirical models describe relations between the variables in an experimental data set using mathematical correlations without requiring detailed knowledge of an underlying mechanism of the system. Empirical models are replaced by mechanistic models when more knowledge about the process mechanisms, through which changes in the system occur, is needed. Among mechanistic models, deterministic models are easier and faster to apply and readily lend themselves to computer

applications. However, in some cases, for more complex biological systems with different input variables stochastic models are more accurate.

Cellulase hydrolysis is one of the most complex biological reaction. The complex heterogeneous reaction mechanism involved in cellulose enzymatic hydrolysis and the complicated morphology of biomass make enzymatic hydrolysis difficult to model. Classical homogenous enzyme catalysis is modelled by M-M kinetics and heterogeneous catalysis by Langmuir kinetics. Considering the heterogeneous and homogenous reactions that occur during the entire process, the cellulose hydrolysis should be modelled by both, Langmuir and M-M kinetics. In addition, other factors related to enzyme characteristics (adsorption, inhibition, synergism, activity, composition) and substrate characteristics (degree of polymerization, crystallinity, accessible surface area, hemicellulose and lignin content), which affect the hydrolysis of cellulose, should be included in the model. In addition, to scale up to the industrial level, it is important to model the deviation of the process from ideal conditions. Thus, stochastic modelling is recommended because it involves inherent randomness, thus may predict the distribution of possible outcomes.

Acknowledgements: The chapter was prepared as a dissemination part of EURECAT-ACCIÓ project (Regional Agency for the Business Competitiveness of the Generalitat de Catalunya) which is gratefully acknowledged.
Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. Bruzzone, AG, Madeo, F, Dallorto, M, Poggi, D, Ferrando, A. Different modeling and simulation approaches applied to industrial process plants. *Simulation Series* 2013;45:39–46. ISSN: 07359276. ISBN: 978-162748039-0.
2. Rümpler, C, Chechare, R, Zacharias, A. Arc modeling in industrial applications. *Plasma Phys Technol* 2019;6:200–7. https://doi.org/10.14311/ppt.2019.2.200.
3. Chapman, J, Ismail, AE, Dinu, CZ. Industrial applications of enzymes: recent advances, techniques, and outlooks. *Catalysis* 2018;8:1–26. https://doi.org/10.3390/catal8060238.
4. Galanakis, CM, Patsioura, A, Gekas, V. Enzyme kinetics modeling as a tool to optimize food industry: a pragmatic approach based on amylolytic enzymes. *Crit Rev Food Sci Nutr* 2015;55:1758–70. https://doi.org/10.1080/10408398.2012.725112.
5. O'Dwyer, JP, Zhu, L, Granda, CB, Holtzapple, MT. Enzymatic hydrolysis of lime-pretreated corn stover and investigation of the HCH-1 model: inhibition pattern, degree of inhibition, validity of simplified HCH-1 model. *Bioresour Technol* 2007;98:2969–77. https://doi.org/10.1016/j.biortech.2006.10.014.
6. Liang, C, Gu, C, Raftery, J, Karim, MN, Holtzapple, M. Development of modified HCH-1 kinetic model for long-term enzymatic cellulose hydrolysis and comparison with literature models. *Biotechnol Biofuels* 2019;12:1–13. https://doi.org/10.1186/s13068-019-1371-5.
7. Bansal, P, Hall, M, Realff, MJ, Lee, JH, Bommarius, AS. Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnol Adv* 2009;27:833–48. https://doi.org/10.1016/j.biotechadv.2009.06.005.
8. Pratto, B, de Souza, RBA, Sousa, RJr, da Cruz, AJG. Enzymatic hydrolysis of pretreated sugarcane straw: kinetic study and semi-mechanistic modeling. *Appl Biochem Biotechnol* 2016;178:1430–44. https://doi.org/10.1007/s12010-015-1957-8.
9. Gan, Q, Allen, SJ, Taylor, G. Kinetic dynamics in heterogeneous enzymatic hydrolysis of cellulose: an overview, an experimental study and mathematical modelling. *Process Biochem* 2003;38:1003–18. https://doi.org/10.1016/s0032-9592(02)00220-0.
10. Angarita, JD, de Souza, RBA, da Cruz, AJG, Biscaia, ECJr., Secchi, AR. Kinetic modeling for enzymatic hydrolysis of pretreated sugarcane straw. *Biochem Eng J* 2015;104:10–9. https://doi.org/10.1016/j.bej.2015.05.021.
11. Blackwell, KT. Enzyme kinetics, modeling of. In: Jaeger, D, Jung, R, editors. *Encyclopedia of Computational Neuroscience*, 1st ed. vol 2. New York, USA: Springer Press; 2015:1110–13 pp. book-chapter.
12. Wong, MKL, Krycer, JR, Burchfield, JG, James, DE, Kuncic, Z. A generalised enzyme kinetic model for predicting the behaviour of complex biochemical systems. *FEBS Open Bio* 2015;5:226–39. https://doi.org/10.1016/j.fob.2015.03.002.
13. Michaelis, L, Menten, M. Die kinetik der invertinwirkung. *Biochem Z* 1913;49:333–69.
14. Briggs, GE, Haldane, JBA. Note on the kinetics of enzyme action. *Biochem J* 1925;19:338–9. https://doi.org/10.1042/bj0190338.
15. Tummler, K, Lubitz, T, Schelker, M, Klipp, E. New types of experimental data shape the use of enzyme kinetics for dynamic network modelling. *FEBS J* 2014;281:549–71. https://doi.org/10.1111/febs.12525.
16. Hiromi, K. Interpretation of dependency of rate parameters on the degree of polymerization of substrate in enzyme-catalyzed reactions. Evaluation of subsite affinities of exo-enzyme. *Biochem Biophys Res Commun* 1970;40:1–6. https://doi.org/10.1016/0006-291x(70)91037-5.

17. Torgerson, EM, Brewer, LC, Thoma, JA. Subsite mapping of enzymes. Use of subsite map to simulate complete time course of hydrolysis of a polymeric substrate. *Arch Biochem Biophys* 1979;196:13–22. https://doi.org/10.1016/0003-9861(79)90546-0.
18. Thoma, JA, Crook, C. Subsite mapping of enzymes. Double inhibition studies. *Eur J Biochem* 1982;122:613–8. https://doi.org/10.1111/j.1432-1033.1982.tb06483.x.
19. Dwivedi, CP, Ghose, TK. A model on hydrolysis of bagasse of cellulose by enzyme from Trichoderma reesei QM9414. *J Ferment Technol* 1979;57:15–24.
20. Beschkov, V, Marc, A, Engasser, JM. A kinetic model for the hydrolysis and synthesis of maltose, isomaltose and maltotriose by glucoamylase. *Biotechnol Bioeng* 1984;26:22–6. https://doi.org/10.1002/bit.260260106.
21. Beltrame, PL, Carniti, P, Focher, B, Marzetti, A, Sarto, V. Enzymatic hydrolysis of cellulosic materials: a kinetic study. *Biotechnol Bioeng* 1984;26:1233–8. https://doi.org/10.1002/bit.260261013.
22. Cabral, JMS, Novais, JM. Cardoso JP Modeling of immobilized glucoamylase reactors. *Ann N Y Acad Sci* 1983;413:535–41. https://doi.org/10.1111/j.1749-6632.1983.tb47933.x.
23. McDermott, JB, Libanati, C, LaMarca, C, Klein, MT. Quantitative use of model compound information: Monte Carlo simulation of the reactions of complex macromolecules. *Ind Eng Chem Res* 1990;29:22–9. https://doi.org/10.1021/ie00097a004.
24. Fenske, JJ, Penner, MH, Bolte, JP. A simple individual-based model of insoluble polysaccharide hydrolysis: the potential for autosynergism with dual-activity glycosidases. *J Theor Biol* 1999;199:113–8. https://doi.org/10.1006/jtbi.1999.0938.
25. Bryjak, J, Murlikiewicz, K, Zbicinski, I, Stawczyk, J. Application of artificial neural networks to modeling of starch hydrolysis by glucoamylase. *Bioproc Biosyst Eng* 2000;23:351–7. https://doi.org/10.1007/s004499900170.
26. O'Dwyer, J, Zhu, L, Holtzapple, MT. Developing a fundamental understanding of biomass structural features responsible for enzymatic digestibility. In: AIChE Annual Meeting Conference Proceedings no. 32c; 2004:713–5 pp.
27. Vlazaki, M, Huber, J, Restif, O. Integrating mathematical models with experimental data to investigate the within-host dynamics of bacterial infections. *Pathog Dis* 2019;77:1–13. https://doi.org/10.1093/femspd/ftaa001.
28. Peri, S, Karra, S, Lee, YY, Karim, MN. Modeling intrinsic kinetics of enzymatic cellulose hydrolysis. *Biotechnol Prog* 2007;23:626–37. https://doi.org/10.1021/bp060322s.
29. Jeoh, T, Cardona, MJ, Karuna, N, Mudinoor, AR, Nill, J. Mechanistic kinetic models of enzymatic cellulose hydrolysis–a review. *Biotechnol Bioeng* 2017;114:1369–85. https://doi.org/10.1002/bit.26277.
30. Niu, H, Shah, N, Kontoravdi, C. Modelling of amorphous cellulose depolymerisation by cellulases, parametric studies and optimisation. *Biochem Eng J* 2016;105:455–72. https://doi.org/10.1016/j.bej.2015.10.017.
31. Hosseini, SA, Shah, N. Modelling enzymatic hydrolysis of cellulose part I: population balance modelling of hydrolysis by endoglucanase. *Biomass Bioenergy* 2011;35:3841–8. https://doi.org/10.1016/j.biombioe.2011.04.026.
32. Zhang, YHP, Lynd, LR. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulose systems. *Biotechnol Bioeng* 2004;88:797–824. https://doi.org/10.1002/bit.20282.
33. Sousa, RJr., Carvalho, ML, Giordano, RLC, Giordano, RC. Recent trends in the modeling of cellulose hydrolysis. *Braz J Chem Eng* 2011;28:545–64. https://doi.org/10.1590/s0104-66322011000400001.
34. Holtzapple, MT, Caram, HS, Humphrey, AE. The HCH-1 model of enzymatic cellulose hydrolysis. *Biotechnol Bioeng* 1984a;26:775–80. https://doi.org/10.1002/bit.260260723.

35. Holtzapple, MT, Caram, HS, Humphrey, AE. Determining the inhibition constants in the HCH-1 model of cellulose hydrolysis. *Biotechnol Bioeng* 1984b;26:735–57. https://doi.org/10.1002/bit.260260719.
36. Nassar, R, Chou, ST. Stochastic analysis of stepwise cellulose degradation. *Chem Eng Sci* 1991;46:1651–7. https://doi.org/10.1016/0009-2509(91)87012-2.
37. Bezerra, RMF, Dias, AA. Discrimination among eight modified Michaelis–Menten kinetics models of cellulose hydrolysis with a large range of substrate/enzyme ratios. *Appl Biochem Biotechnol* 2004;112:173–84. https://doi.org/10.1385/abab:112:3:173.
38. Levine, SE, Fox, JM, Blanch, HW, Clark, DS. A mechanistic model of the enzymatic hydrolysis of cellulose. *Biotechnol Bioeng* 2010;107:37–51. https://doi.org/10.1002/bit.22789.
39. Kumar, D, Murthy, GS. Stochastic molecular model of enzymatic hydrolysis of cellulose for ethanol production. *Biotechnol Biofuels* 2013;6:1–20. https://doi.org/10.1186/1754-6834-6-63.
40. Espejo-Carpio, FJ, Pérez-Gálvez, R, Guadix, A, Guadix, EM. Artificial neuronal networks (ANN) to model the hydrolysis of goat milk protein by subtilisin and trypsin. *J Dairy Res* 2018;85:339–46. https://doi.org/10.1017/s002202991800064x.
41. Gao, L, Guo, Q, Lin, H, Pan, D, Huang, X, Lin, J, Lin, J. Modeling of lactose enzymatic hydrolysis using Monte Carlo method. *Electron J Biotechnol* 2019;40:78–83. https://doi.org/10.1016/j.ejbt.2019.04.010.
42. Canovas, M, Maiquez, JR, Obón, JM, Iborra, JL. Modeling of the biotransformation of crotonobetaine into L-(–)-carnitine by Escherichia coli strains. *Biotechnol Bioeng* 2002;77:764–75. https://doi.org/10.1002/bit.10157.
43. Alvarez-Vasquez, F, Canovas, M, Iborra, JL, Torres, NV. Modeling, optimization and experimental assessment of continuous L-(–)-carnitine production by Escherichia coli cultures. *Biotechnol Bioeng* 2002;80:794–805. https://doi.org/10.1002/bit.10436.
44. Lin, J-Q, Sang-Mok, L, Yoon-Mo, K. Modeling and simulation of simultaneous saccharification and fermentation of paper mill sludge to lactic acid. *Microb Biotechnol* 2005;15:40–7.
45. Liao, W, Liu, Y, Wen, Z, Frear, C, Chen, S. Kinetic modeling of enzymatic hydrolysis of cellulose in differently pretreated fibers from dairy manure. *Biotechnol Bioeng* 2008;101:441–51. https://doi.org/10.1002/bit.21921.
46. Gama, R, Van Dyk, JS, Burton, MH, Pletschke, BI. Using an artificial neural network to predict the optimal conditions for enzymatic hydrolysis of apple pomace. *3 Biotech* 2017;7:1–10. https://doi.org/10.1007/s13205-017-0754-1.
47. de Godoy, CM, Machado, DL, da Costa, AC. Batch and fed-batch enzymatic hydrolysis of pretreated sugarcane bagasse–assays and modelling. *Fuel* 2019;253:392–9. https://doi.org/10.1016/j.fuel.2019.05.038.
48. Kadam, KL, Rydholm, EC, McMillan, JD. Development and validation of a kinetic model for enzymatic saccharification of lignocellulosic biomass. *Biotechnol Prog* 2004;20:698–705. https://doi.org/10.1021/bp034316x.
49. Hodge, DB, Karim, MN, Schell, DJ, McMillan, JD. Model-based fed-batch for high-solids enzymatic cellulose hydrolysis. *Appl Biochem Biotechnol* 2009;152:88–107. https://doi.org/10.1007/s12010-008-8217-0.
50. Ko, J, Su, WJ, Chien, IL, Chang, DM, Chou, SH, Zhan, RY. Dynamic modeling and analyses of simultaneous saccharification and fermentation process to produce bio-ethanol from rice straw. *Bioproc Biosyst Eng* 2010;33:195–205. https://doi.org/10.1007/s00449-009-0313-1.
51. Morales-Rodriguez, R, Gernaey, KV, Meyer, AS, Sin, G. A mathematical model for simultaneous saccharification and co-fermentation (SSCF) of C6 and C5 sugars. *Chin J Chem Eng* 2011a;19:185–91. https://doi.org/10.1016/s1004-9541(11)60152-3.

52. Verma, SK, Fenila, F, Shastri, Y. Sensitivity analysis and stochastic modelling of lignocellulosic feedstock pretreatment and hydrolysis. *Comput Chem Eng* 2017;106:23–39. https://doi.org/10.1016/j.compchemeng.2017.05.015.
53. Holtzapple, MT, Caram, HS, Humphrey, AE. A Comparison of two empirical models for the enzymatic hydrolysis of pretreated poplar wood. *Biotechnol Bioeng* 1984c;26:936–41. https://doi.org/10.1002/bit.260260818.
54. Sin, G, Meyer, AS, Gernaey, KV. Assessing reliability of cellulose hydrolysis models to support biofuel process design—identifiability and uncertainty analysis. *Comput Chem Eng* 2010;34:1385–92. https://doi.org/10.1016/j.compchemeng.2010.02.012.
55. Morales-Rodriguez, R, Meyer, AS, Gernaey, KV, Sin, G. Dynamic model-based evaluation of process configurations for integrated operation of hydrolysis and co-fermentation for bioethanol production from lignocellulose. *Bioresour Technol* 2011b;102:1174–84. https://doi.org/10.1016/j.biortech.2010.09.045.
56. López-Pérez, PA, Puebla, H, Velázquez Sánchez, HI, Aguilar-López, R. Comparison tools for parametric identification of kinetic model for ethanol production using evolutionary optimization approach. *Int J Chem React Eng* 2016;14:1201–9. https://doi.org/10.1515/ijcre-2016-0045.
57. Wang, R, Unrean, P, Franzén, CJ. Model-based optimization and scale-up of multi-feed simultaneous saccharification and co-fermentation of steam pre-treated lignocellulose enables high gravity ethanol production. *Biotechnol Biofuels* 2016;9:2–13. https://doi.org/10.1186/s13068-016-0500-7.
58. Shao, X, Lynd, L, Wyman, C, Bakker, A. Kinetic modeling of cellulosic biomass to ethanol via simultaneous saccharification and fermentation: Part I. Accommodation of intermittent feeding and analysis of staged reactors. *Biotechnol Bioeng* 2009;102:59–65. https://doi.org/10.1002/bit.22048.

Francis M. Gaitho, Genene T. Mola and Giuseppe Pellicane

7 Computational approach to the study of morphological properties of polymer/fullerene blends in photovoltaics

Abstract: Organic solar cells have the ability to transform solar energy efficiently and have a promising energy balance. Producing these cells is economical and makes use of methods of printing using inks built on solvents that are well-matched with a variety of cheap materials like flexible plastic or paper. The primary materials used to manufacture organic solar cells include carbon-based semiconductors, which are good light absorbers and efficient charge generators. In this article, we review previous research of interest based on morphology of polymer blends used in bulk heterojunction (BHJ) solar cells and introduce their basic principles. We further review computational models used in the analysis of surface behavior of polymer blends in BHJ as well as the trends in the field of polymer surface science as applied to BHJ photovoltaics. We also give in brief, the opportunities and challenges in the area of polymer blends on BHJ organic solar cells.

Keywords: polymer blends, morphology, computational, bulk heterojunction, photovoltaics

7.1 Introduction

Sun is primarily the source of non-polluting energy, but we have not managed to utilize it fully. Most solar devices in use today are not economically viable because the cost of manufacturing the inorganic semiconductor solar devices is very high. As a result, scientists got the motivation to develop organic photovoltaic (OPV) structures as well as devices because they cost less and are flexible and easy to obtain. The easiness of fabricating OPVs forms the foundation of future competitive engineering methods. One of the successful techniques of manufacturing OPVs is the solution-processed BHJ photovoltaic cells (PCs). BHJ is a state-of-the-art device, which blends conjugated (electrical conducting) polymers closely with soluble fullerene byproducts. Over the last decade, scientists have succeeded to refine the Power Conversion Efficiency (PCE) of the polymer BHJ PCs from 4 % to above 10 % [1–4]. Shortly after the discovery of BHJ PCs, the quantity of the scientific journals

This article has previously been published in the journal Physical Sciences Reviews. Please cite as: Gaitho, F. M., Mola, G. T., Pellicane, G. Computational approach to the study of morphological properties of polymer/fullerene blends in photovoltaics *Physical Sciences Reviews* [Online] 2018, 3. DOI: 10.1515/psr-2017-0102

https://doi.org/10.1515/9783110678215-007

in this field continued to increase with a significant percentage of the publications dealing with the morphological structure of the BHJ PCs. Figure 7.1 shows the trend of the publications in the areas of knowledge of BHJ PCs in terms of their efficiency, morphological properties, blend-surface segregation and computational models, over the last decade. The trends clearly indicate progressive research and a promising future for these devices.

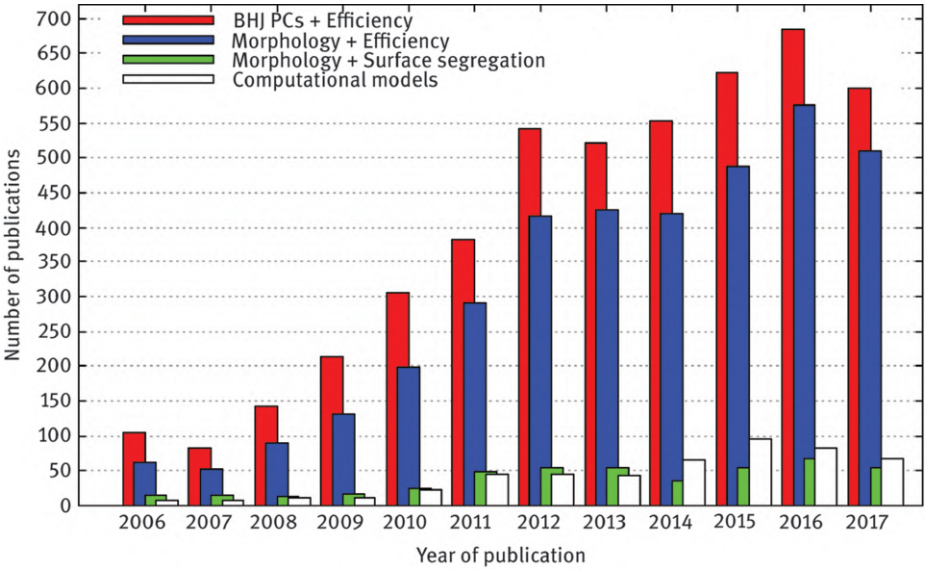


Figure 7.1: Trend of the publications in the area of knowledge of efficiency, morphology and surface segregation of blends in BHJ PCs from web of science (searched by the key words: "Bulk heterojunction photovoltaic cell + efficiency" (Red), "Morphology + Efficiency" (Blue), "Morphology + surface segregation" (Green) and "Computational models" (White).

The basic working principle of the BHJ PC is presented in Figure 7.2. The contact area between the donor and acceptor organics is large, a configuration that enables the excited hole–electron pair (exciton) to reach the donor–acceptor boundary and separate before recombination takes place. The donor material absorbs the solar radiations to produce excitons, which then travel to the donor–acceptor interface, where they separate because of the difference in the energy levels of the organics. Electrons in the higher energy states move through the acceptor and move out through the cathode. Holes move from end to end of the donor material to the anode. Control of morphology of donor/acceptor blend is important in order to allow large charge generating interface, to suppress loss of excitons and to ensure good percolation pathways for transportation of holes and electrons to the anode and cathode respectively [5].

The general efficiency of this kind of PC is much lower compared to that of inorganic PCs [6]. Properties of materials making up the active layer of the polymer

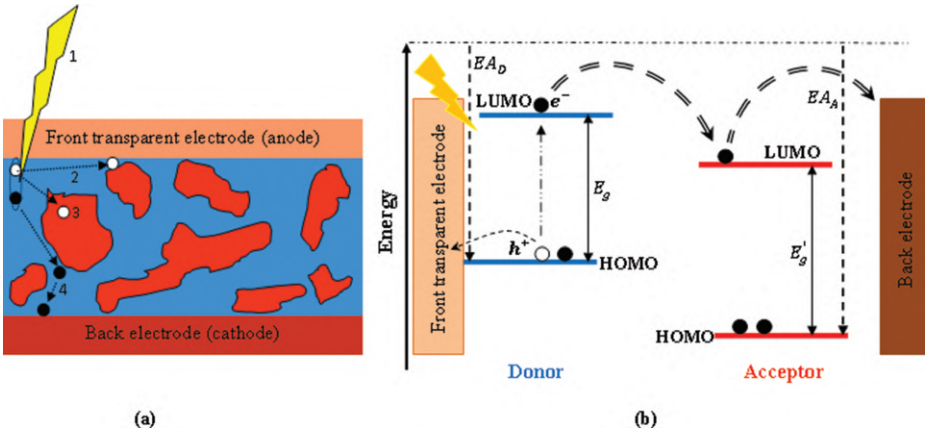


Figure 7.2: Schematic representation of working mechanism of bulk heterojunction photovoltaic cell **(a)**: 1- Incident photon radiation creates an exciton, 2- Hole diffusion toward anode, 3- Hole trapped in an isolated acceptor molecule, 4- Free electron moves toward cathode. Red/blue region represents acceptor/donor blend **(b)**: Donor/acceptor system showing energy levels, HOMO-highest occupied molecular orbital, LUMO-lowest unoccupied molecular orbital, EA_D-electron affinity in donor molecule, EA_A- electron affinity in acceptor molecule.

cells are the primarily the efficiency limiting factors. An ideal polymer solar cell has high charge carrier mobility and a wide-ranging absorption in the range of the solar band that enhances absorption of radiations from the sun and also achievable in terms of production, despite low efficiency levels [7]. However, there are certain experimental [5, 8, 9] and computational [10–12] methods suitable to improve the efficiency of the device such as morphology optimization of its active layer. This process involves optimization of the horizontal phase-separated morphology, a perpendicular scattering of the constituents making the composite film as well as the vertical phase gap. Deibel and Dyakonov [13] suggested that good control of the surface segregation is very crucial for developing an efficient BHJ. Therefore, a good choice of the solvent and the strengthening of the solution treated polymer help to produce an inner structure with better-quality detachment of the immobile electron–hole pair and the resulting charge transference [14, 15]. For improved charge collection in the photoactive region of BHJ device, thermal annealing must be done effectively. Studies by Jo et al. [1] and Kumar et al. [15] revealed that different time-dependent annealing treatments achieve different nanoscale BHJ morphologies. The different time-dependent annealing treatments produce various levels of device performance and efficiencies as presented in Figure 7.3.

The curves also show that solar cells fabricated without annealing treatments performed poorly. The annealing process achieves a better-quality blend morphology, which further results in increased induced photon-to-current conversion efficiency (IPCE) of the solar cells. Morphological properties of polymer solar cells are

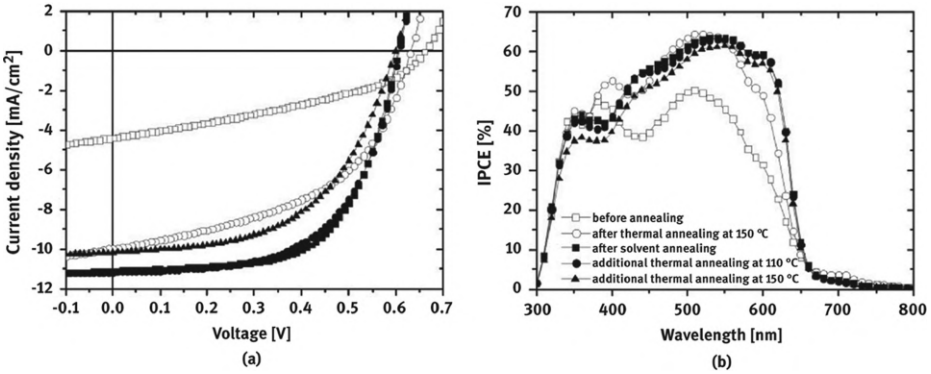


Figure 7.3: (a) Performance of BHJ organic PCs having P3HT: PCBM energetic layers with different hardening processes that depend on time. (b) Incident Photon-to-Current Conversion Efficiency (IPCE) bands of P3HT: PCMB BHJ PCs with different hardening treatments [1].

therefore imperative to their functioning. Several studies have also investigated the contribution of film thickness to segregation, for instance, a study by Verploegen [16], proved that after annealing thin films of polymer:fullerene in the ratio of 1:3 at 160°C, they displayed dewetting [17–19]. Segregation is a condition where the donors (polymers) are in excess at the region near the substrate, and the acceptors (fullerenes) are in excess near the top surface of the device. Such distribution of components is critical in avoidance and control of charge carrier seepage. However, from the studies reviewed herein, the results relating to the surface separation contradict. Nevertheless, most of the studies tend to agree with the annealing results in increased concentration of the acceptors near the cathode. Such results are imperative in the process of improving the efficiency of BHJ organic PCs.

In the past few years, computational modeling has attracted attention in accelerating the exploration for proficient conjugated photovoltaic materials [20]. By prediction and extrapolation, it provides information about physical properties of the BHJ PC and factors that affect the performance. Indeed, it is possible to improve absorption of light, chain conformations, electron configurations, operation, as well as properties essential in the design of these devices, through computational modeling. In the next section of this review, we explore computational models on surface segregation and morphology of the polymer blends in BHJ PCs. The blends used comprise of mixtures of *linear* chain polymers and fullerene derivatives whose chain architecture is basically *cyclic*.

7.2 Computational models for polymer blends in BHJ PCs

7.2.1 Course-grained model

For a period of time now, the study of structural, morphological and dynamic properties of cyclic and linear polymers have shown an increasing trend owing to the nature of their chain architecture. Lack of chain ends in cyclic polymers, for instance, results in enormous topological limitations; this has generated great interest in the investigation of properties such as local density, radius of gyration, diffusion and percolation of the polymer chains. Experimental studies have been conducted on linear and branched polymers in dilute solutions [21–25]. Computational work has also been done on the behavior of the polymers at the surface and in the bulk of cyclic/linear blends [26–29]. A study by Cates and Deutsch [30] indicated that cyclic polymers approximate Gaussian in their short chain range and the linear polymers display conformational statistics in their long chain regimes. It was found that as the chain length changes the radius of gyration varies as in the expression below;

$$\frac{1}{3} \leq R_g^2 \sim N^{2v} \leq \frac{1}{2} \tag{7.1}$$

where R_g^2 is the square radius of gyration, N is the total number of atoms and v is a constant. A Gaussian chain possesses an exponent equal to 1. Thus, it turned out that size scaling of a ring polymer generally depends on the simulation model [31], which has activated the quest of many researchers to find a better understanding of polymer scaling properties.

Polymer organic PCs stand out as the cheap and large-scale-manufacture technology for the future. However, the current efficiency of organic PCs is too low for economic viability. There is great need to increase the current efficiency from about 11 % to levels comparable with other technologies. Computational models can play a significant role in making these devices achieve the desired effectiveness. There are several computational models for conducting polymer blends. The coarse-grained (C-G) models hosted by molecular dynamic simulations of small systems have been used to study the morphology of the photoactive layer of BHJ PCs [32–34]. Do et al. [33] and Huang et al. [34] developed C-G models of P3HT/PCBM (poly(3-hexylthiophene)/[6,6]-phenyl C_{61}-butyric acid methyl ester) blend in which P3HT monomers were represented by three centers of mass:- of the thiophene ring, that of the first three side-chain methyl groups and that of the carbons of the last three side-chain methyl groups. In this model, the center of mass of PCBM represented a single bead. The main aim of this C-G model was to examine the structural formation and dynamics of the polymer/fullerene blend used in BHJ at a microstructure scale. The structures with C-G sites are represented in Figure 7.4.

Starting from a perfect polymer atomistic model, the C-G model is configured at temperatures close to 600 K in order to accommodate fluid state and also avoid

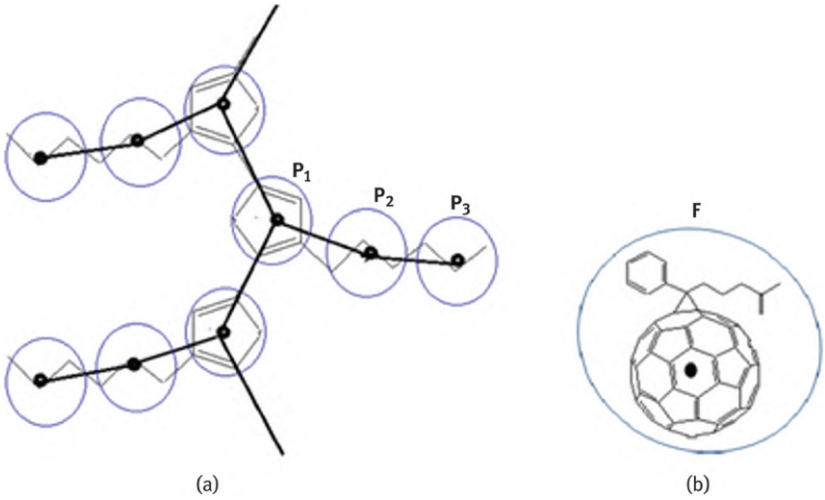


Figure 7.4: Chemical structures with C-G sites (a) Polymer; poly(3-hexylthiophene)-(P3HT), (b) Fullerene; [6,6]-phenyl C_{61}-butyric acid methyl ester (PCBM)[26].

solidification of the fullerene molecule. By Boltzmann inversion iterative procedure, the positions of centers of mass (sites) are set to give a general atomistic system having non-bonded polymer-fullerene particles, bond angles and two-faced polymer distribution [26]. Varying weight ratios of 12-mers P3HT were conveniently chosen for the atomistic system at normal pressure. Polymer–polymer C–G interactions were optimized in simulations of P3HT:P3HT at T = 550 K. Then Polymer–fullerene and fullerene–fullerene C–G interactions were optimized as P3HT-PCBM and PCBM-PCBM, respectively, in simulations of 1.85:1 w/w P3HT: PCBM at the same temperature.

7.2.2 The bead-spring model

The other important model is the bead-spring model [35, 36]. Very recently, researchers [37, 38] have extensively used this model to study the behavior of linear-cyclic polymer blends at the polymer/air interface using molecular dynamics (MD) simulations. Linear and cyclic polymers were modeled as monomers each of mass m, and sharing common boundaries. The connections form open or closed loops which interact via springs while centers of mass of the monomers act as the interaction beads, see Figure 7.5.

The dynamics of bead-and-spring in this model is achieved by summing the external forces at the center of mass of the monomer using the expression;

$$F = F_f + F_E + F_B \tag{7.2}$$

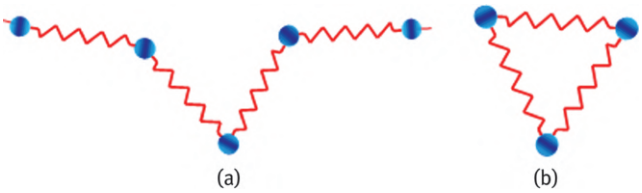


Figure 7.5: Sketch of configurations of the bead-and-spring nature of (a) linear and (b) cyclic polymer chains.

where, $F_f = -\left[\frac{m}{\beth}\right]V$, $F_B = \sqrt{\frac{k_B T m}{\Delta t \beth}}$, and F_E, are the Stokes' drag, the Brownian force and the effective spring force respectively. The factor $\beth = 0.5$ represents the damping factor, V the velocity of the bead, while T the temperature of solvent molecules colliding with the beads during the Brownian motion. The time-step is represented by Δt, while k_B is the Boltzmann constant. A combination of two potentials; the finitely extensible nonlinear elastic (FENE) [35] and Weeks–Chandler–Andersen (WCA) [39] were used in this model. Interactions between the beads were approximated by Lennard–Jones (LJ) potential at a limit expanse of $r_c = 2.5\ \sigma$. All simulations were done using LAMMPS package [40]. For the time step, they used:

$$\Delta t = 0.005\tau; \text{ with } \tau = \sigma\sqrt{\frac{m}{\epsilon}} \text{ and temperature } T = \epsilon/k_B$$

(ϵ and σ are the energy and distance parameters for LJ potential)

A2/τ damping factor of the Langevin thermostat was used to maintain constant temperature and pressure in all the simulations. The total potential was obtained by combining FENE and LJ potentials;

$$E(r) = \begin{cases} -\frac{1}{2}Kr_0^2 \ln\left[1 - \left(\frac{r}{r_0}\right)^2\right] + 4 \in \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 + \frac{1}{4}\right], r \le r_c \\ -\frac{1}{2}Kr_0^2 \ln\left[1 - \left(\frac{r}{r_0}\right)^2\right], r > r_c \end{cases} \tag{7.3}$$

where, σ and $\in$ are LJ parameters related by $K = 30 \in /\sigma^2$ and $r_0 = 3/2\sigma$ gives the maximum bond size while r is the inter-bead separation. Figure 7.6 shows that the LJ and FENE potentials restrict the lower and upper limits of bond lengths, respectively, between the adjacent monomers.

In their analysis on adsorption properties of cyclic-linear polymer blends at the interface in a low cyclic chain length regime, Megnidio et al. [38] further observed that as the distance from the interface increases, the local density of either species $\rho_i = \frac{N_{i,slice}}{V}$, tends to the value of density in the bulk, $\rho_i^0 = \frac{N_i^0}{V}$. Here,

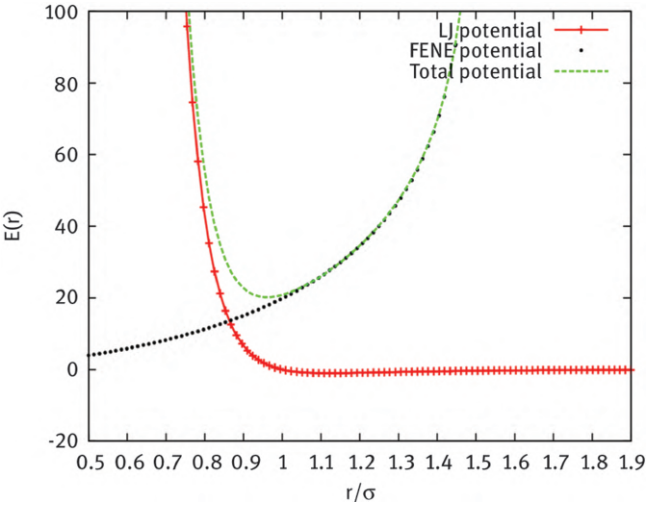


Figure 7.6: FENE and LJ potentials [38].

$i \equiv c(\text{cyclic}), l(\text{linear})$, $N_{i,\text{slice}}$ is the number of particles of polymer species i in a slice of thickness σ perpendicular to the interface and V_{slice} is the volume of this slice. A similar observation was made with the local composition of linear polymers that migrate to the interface [37];

$$c = \frac{N_{l,\text{slice}}}{N_{c,\text{slice}} + N_{l,\text{slice}}} \tag{7.4}$$

which, also corresponds to the value in the bulk $c_0 = N_l^0 / N_c^0 + N_l^0$ as presented in Figure 7.7.

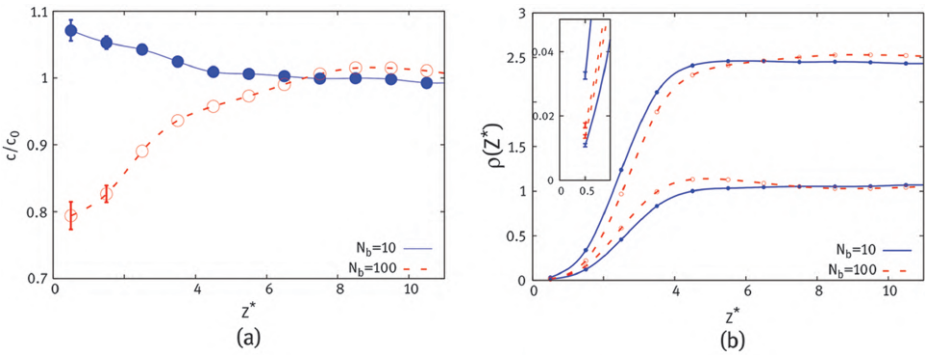


Figure 7.7: (a) Linear polymer local composition as a function of distance from interface. (b) Particle density of linear polymers as a function of distance from interface (inset: region very close to interface [38].

In BHJ OPV applications, the active layer is a blend of two components as discussed earlier in this section, namely; a conjugated hole-conveying polymer and a solvable fullerene in form of linear and cyclic chains respectively. Since the two components have different surface energies as well as the energies of integration, nanoscale phase separation occurs after casting the solutions to form the film. A study conducted by Kong et al. [41], concluded that surface segregation occasionally occurs during the first evaporation stage of the host solvent. Further in their study, Pellicane et al. [37] and Megnidio et al. [38] observed that the total energy per cyclic monomer is greater than that of linear for the entire distance from the surface of the blend as in Figure 7.8. Due to their low flexibility the cyclic chains pack much less than the linear chains and show less number of pair interactions among the beads in their shorter chain length regime. This explains why linear polymers achieve lower interfacial free energy in addition to minimizing their surface energy more than cyclic [42] as well as optimizing their entropy by exposing the chain ends [37, 38]. Enrichment of linear polymers at the interface also arises from entropy – and enthalpy – resulting processes. However, the total energy per bead for both linear and cyclic polymers is nearly the same.

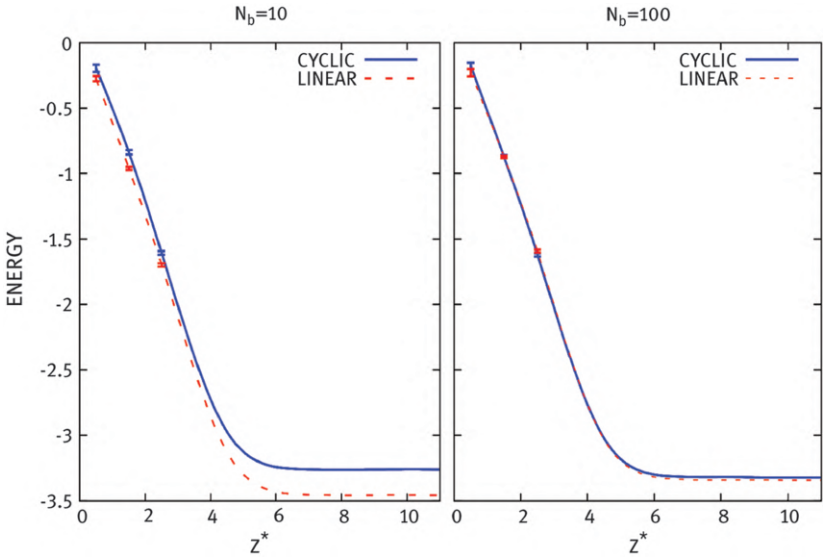


Figure 7.8: Energy per monomer of linear and cyclic chains as a function of distance from interface [37].

7.2.3 Process-device model

As discussed previously, the processes that affect the effectiveness of the polymer-fullerene solar cells include generation rate of excitons, their diffusion rate and the dissociation of the excitons at the donor/acceptor crossing point. Ray et al. [43] noted that most of the developed models fail to capture the mechanism of the primary BHJ solar cells. Instead, the study used a model that relates the efficiency of the cell to the useful parameters of the organic material. Taking the exciton generation rate to be G_{ex} and the diffusion rate to be D_{ex}, the simple model becomes; $L_{ex} = \sqrt{D_{ex}\tau_{ex}}$, within the range of 10 to 20 nm. With this model, we only realize an acceptable level of efficiencies when the donor–acceptor polymer blends get phase-segregated through the annealing process. Therefore, any photo-excited excitons get to the donor/acceptor boundary within a length that is approximately equal to L_{ex}, despite their point of location [43].

In this section, we explore the model of the random morphology and the regular structure with its role in refining the effectiveness of the BHJ polymer–fullerene photovoltaic solar cells. We also echo that 3D simulations of polymer phase separation predict optimal film thickness and maximum annealing time.

Ray et al. [43] provided a clear description of the spinodal segment separation as a product of the annealing interval by using Cahn–Hilliard (C–H) equation, i. e.

$$\frac{\partial \varnothing}{\partial t} = M_o\left(\nabla^2 \frac{\partial \mathrm{f}}{\partial \varnothing} + 2k\nabla^4 \varnothing\right) \tag{7.5}$$

The contest flanked by the entropy of conjoining polymers having chain lengths N_A and N_B as well as the energy of interaction of the pairs of the material[1] represents the primary cause of the spinodal phase segregation. Here, $\varnothing$ is the polymer–fullerene volume fraction. The quantitative expression of free energy density function containing these factors is illustrated in eq. (7.6);

$$f = \frac{\mathrm{k}_B T}{v^3}\left[\frac{\varnothing \ln(\varnothing)}{N_A} + \frac{(1-\varnothing)\ \ln\ (1-\varnothing)}{N_B} + \chi\varnothing(1-\varnothing)\right] \tag{7.6}$$

The parameter M_o in eq. (7.5) is the diffusion parameter at absolute temperature T and volume v, which depends on the solvent, Boltzmann factor k_B is also taken into consideration in this expression. Together with the gradient energy coefficient k, M_o affect the moving part of the entire process. This moving part determines the rate at which the mixture reaches the lowest level of the free energy curve. Ray et al. [43] also noted that the active layer morphology of BHJ has an average domain width $W(t_a)$ determined by the Lifshitz-Slyozov law [44], given as:

1 Flory Parameter x represents the interaction energy between the material pairs.

$$W(t_a) \sim \left[D_{eff} t_a\right]^n \rightarrow n \sim \frac{1}{3} \tag{7.7}$$

The polymer–fullerene morphology in BHJ organic PCs at any given time t_a allows only a small part of the photo-excited electron/hole excitons to disperse to the charge extrication zone before recombining. The solution of the three-dimensional exciton diffusion law given below gives the relation of exciton flux (*Jex*) to the morphology of the film. Considering the case where t tends to infinity $(t \rightarrow \infty)$ and at a photon absorption coefficient α, the solution of the equation gives the stable situation of total exciton (n_{ex}) fluctuation for each unit space of the interfacial area on the morphologies resulting from different annealing instants.

$$\frac{\partial n_{ex}}{\partial t} = D_{ex} \nabla^2 n_{ex} - \frac{n_{ex}}{\tau_{ex}} + G_0 e^{-\alpha z} \tag{7.8}$$

One of the solutions to eq. (7.8) is the carrier diffusion term eq. (7.9) which represents the hole–electron (h–e) transport in the blend.

$$D_{h,e} \nabla^2 n_{h,e} - \frac{n_{h,e}}{\tau_{h,e}} + G_{h-e} = 0 \tag{7.9}$$

For a case that every exciton is entering the interfacial regions gets detached, into a hole–electron (h–e) pair, and the charge carriers reach the electrodes, the limit of the maximum current varies depending on the thickness of the film. Therefore, the computational models could be extrapolated to present practical methods that could be used to predict morphology and improve efficiency of the current organic devices.

7.3 Importance of morphological properties in the development of BHJ PCs

BHJ is currently one of the most efficient ways to construct polymer solar cells. It involves mixing of the polymers having excess electrons and fullerides with a small percentage of electrons. This method of construction allows scientist to study the composition of polymeric active film that would result to the highest *PCE*. Currently, the *PCE* achieved in organic PCs is close to 11% [5, 45, 46]. A study by Liang et al. [7, 47] explored the performance of polymer PCs and found that open-circuit voltage (V_{oc}), short-circuit current (J_{sc}) and the fill factor (*FF*) are primary parameters that determine the working of the polymer PCs. The study related the three parameters with the following equation:

$$PCE = \frac{(V_{oc} \times J_{sc} \times FF)}{(I_p \times M)} \tag{7.10}$$

The parameter $\boldsymbol{I_p}$ represents the power density and $\boldsymbol{M}$ accounts for the mismatch factor. An ideal structure of BHJ achieves the highest value of *PCE*. However, the structure of the typical BHJ organic solar cell should have a form that allows a broad absorption. It should also have a high hole mobility, energy levels that match to the fulleride and be compatible with the fulleride [48–52]. It is however hard to have a blueprint of a polymer that meets all the above qualities. Since the already developed polymer solar cells have issues related to the nature of the materials, their structure and morphological properties remain the most important factors to check in order to achieve better performing BHJ organic PCs. Some BHJ device structures and respective optimized parameters as well as PCEs obtained in various studies over the last decade are summarized in Table 7.1.

7.4 Past, current and future trends in the application of polymer interfaces in BHJ PCs

The discovery of electrical conduction on conjugated polymer by Alan Heeger, Alan, MacDiarmid and Hideki Shirawaka who won a noble price for chemistry in 2000, marked the beginning of use of organic polymers in the fabrication of photonic and optoelectronic devices [53, 54]. Since then, intensive research efforts were carried out to be able to produce efficient and thin film flexible organic molecule-based photovoltaic devices. Until the discovery of BHJ structural design in the mid-1990s, the PCE of ordinary organic PCs remained low (<1%) because of low charge mobility and exciton dissociation efficiency due to the short exciton life time in most organic semiconductors [55]. The BHJ technology is defined by ultrafast charge transfer process between the polymer and fullerene molecules (Figure 7.9) resulting in efficient exciton dissociation process which makes possible the fabrication of more efficient, low-cost and stable PC devices [56].

7.4.1 Conjugated polymers used in BHJ PCs

Investigations have shown that in the presence of certain dopants, some conjugated polymers, notably polyacetylene, poly(sulfur nitride) and Poly(isothianaphthalene) [66], could be made highly conducting. In their study of the Al/polyacetylene/graphite PC, Weinberger et al. [67] reported low V_{oc} of only 0.3 V and quantum efficiency (*QE*) of 0.3%. In the early 1990s, Scott et al. [68] had observed that poly(p-phenylene vinylene) (PPV) in the ITO/PPV/Al LEDs and PCs had a V_{oc} value of 1 V and *PCE* of 0.1% in white light. Investigations with different polythiophenes as the active material in the cell reported equally low V_{oc}, QE and low efficiencies owing to relaxation of delocalized excitons in the energy gap, which in turn resulted in large spectral shift and limited attainable voltage and *PCE* [69]. The quest to have better *PCE* saw

Table 7.1: Summary of reported optimized parameter and PCE of BHJ PCs (chemical structures and names of polymers can be obtained from respective references).

Year	Device structure	J_{sc} (mA/cm^2)	V_{oc} (V)	FF (%)	PCE (%)	Ref.
2007	ITO/PEDOT:PSS/PCPDTBT:$PC_{70}BM$/Al	16.2	0.62	55	5.5	[57]
2008	ITO/PEDOT:PSS/PCPDTBT/$PC_{71}BM$/Br/Al	15.7	0.61	53	5.1	[58]
2009	ITO/PEDOT:PSS/PBDTTT-CF:$PC_{70}BM$/Ca/Al	15.2	0.79	67	7.7	[59]
2010	ITO/PEDOT:PSS/PTB7:$PC_{70}BM$/Ca/Al	14.6	0.74	69	7.0	[7]
2011	ITO/PEDOT:PSS/PBDTTT-C-T:$PC_{71}BM$/Ca/Al	17.5	0.74	59	7.6	[60]
2012	ITO/PEDOT:PSS/PBDTTPD:PCMB/BCP/Al	10.9	0.93	70	7.1	[61]
2013	ITO/PEDOT:PSS /PDTP-DFBT:$PC_{71}BM$/MoO_3/Ag	17.8	0.68	65	7.9	[80]
2014	ITO/PEDOT:PSS/ PBDT-TS1:$PC_{71}BM$/Mg/Al	17.9	0.80	70	9.9	[62]
2015	ITO/ZnO/PNTz4T:$PC_{71}BM$ /MoO_x//Ag	19.4	0.71	73	10.1	[63]
2015	ITO/ZnO/GEN-2/PEDOT:PSS/ZnO/PEO/PTB7-Th/MoO_x/ Ag (polymer tandem cell)	15.5	0.85	77	10.3	[81]
2016	ITO/ZnO$(Li_2CO)_3$/PTB7:$PC_{71}BM$/PEDOT:PSS/MoO_3/Ag	18.3	0.73	73	10.1	[64]
2017	ITO/ZnO/PffBT4T-2OD:$PC_{71}BM$/MoO_x/Al	17.7	0.75	70	9.4	[65]

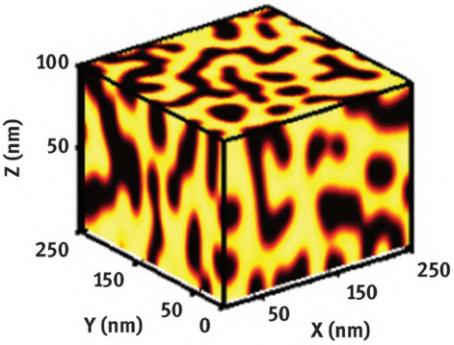


Figure 7.9: Simulated phase segregated morphology of polymer-fullerene blend – 3D view. Polymer – bright regions and fullerene – bark regions[43].

further investigations in other conjugated polymers (see Figure 7.10) as candidates in PCs. The bi-continuous architecture of electron acceptor and electron donor materials in BHJ PCs provides largely dispersed acceptor/donor interfaces, thereby allowing excitons to dissociate into free charge carriers and increasing the PCE to above 2% [70]. In their computational study on pentacene-based organic molecules and derivatives blended with fullerene acceptors in a BHJ PC model, Pramanik et al. [71] performed calculations on density functional theory/time dependent-density functional theory (DFT/TDDFT) and obtained optoelectronic properties which revealed that this model offered type II band alignment. They also estimated the relative positions of HOMO and LUMO for pentacene diacid as −5.16 eV and −3.08 eV, respectively. By introducing different functional groups and combining with a good acceptor like fullerene one may expect a type II band alignment, which is one of the most important criteria in determining the performance of PCs.

7.4.2 Inverted structure BHJ PCs

Currently, more attention lies on the solution processed inverted BHJ PCs, (Figure 7.11 shows comparison between conventional and inverted PCs). The inverted structure has emerged a better candidate to remedy the bottleneck of low air stability arising from use of metal anodes of low work function [72]. It is reported that one section of the active layer has a high concentration on the transparent conducting oxide (cathode) side and the other has a high concentration on the metal oxide (anode) side. This is a favorable condition for the process of charge transport in photovoltaic devices [73, 74]. The inverted structure is compatible with roll-to-roll device fabrication and exhibits an excellent environmental stability and its performance has reached more than 11

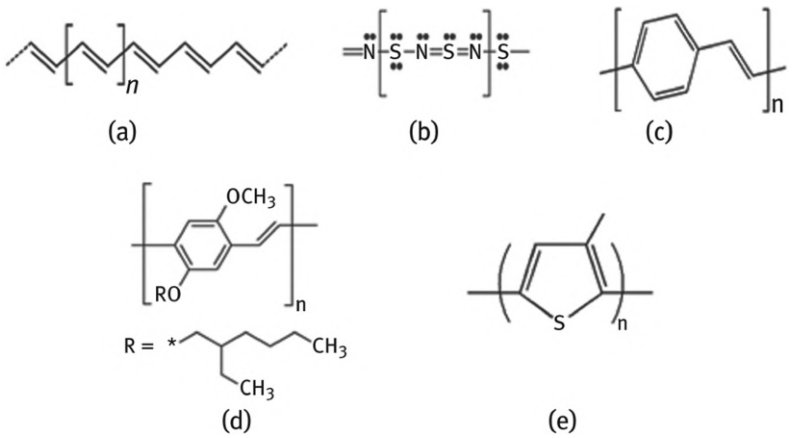


Figure 7.10: Molecular structures of some conjugated polymers used in BHJ PCs; (a) trans-polyacetylene (**PA**), (b) Resonance poly(sulfur nitride) (**SN_X**), (c) Poly(p-phenylene vinylene) (**PPV**), (d) Poly(2-methoxy-5-(20-ethylhexyloxy)-1,4-phenylvinylene) (**MEH–PPV**), (e) Poly(3-alkylthiophene) (**P3AT**).

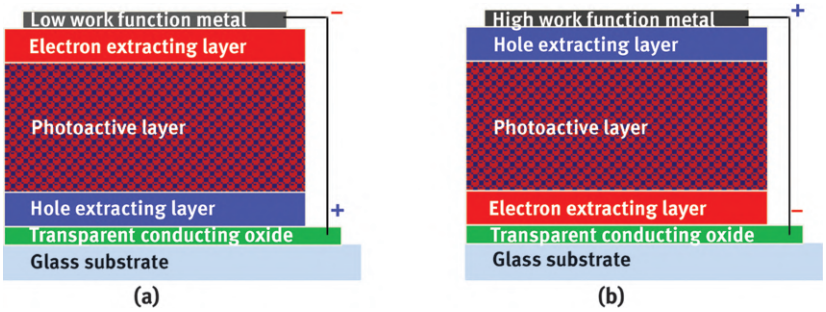


Figure 7.11: A general comparison of BHJ PC designs; (a) Conventional, (b) Inverted.

% PCE through careful tailoring of their energy levels, energy gaps and mixing derivatives of polymers and fullerenes [73–75].

7.4.3 Challenges and opportunities

The main challenges in the realization of thin film organic solar cells today is stability of the polymer in an ambient environment and the low efficiency of the devices. The other challenges associated with organic BHJ devices include, but are not limited to, low external *QE* due to insufficient absorption by the photoactive layer, degradation and decrease in performance with time arising from temperature fluctuations. Presence of impurities in different photoactive polymer combinations (see Table 7.1) also affects

interfacial charge carrier generation rate and diffusion length of excitons [70, 76, 77]. Present day research has found out that BHJ PCs suffer charge-carrier mobility imbalance resulting in space charge build-up since mobility of holes is always a step lower than that of electrons. Consequently, this results in low FF and PCE [4]. In order to overcome these difficulties, the photoactive layer should be made thin enough to prevent recombination of charge carriers. This may however be unfavorable for the absorption process in the device. Prediction from different simulation results has shown that, to obtain a PC with improved FF, external QE and PCE, balanced carrier mobility must be achieved. Appropriate compositions in the photoactive layer must accurately be computed to obtain a good balance point.

Based on PCE figures of Table 7.1, the future appears bright for computational and experimental scientists whose research is inclined toward film morphology of dynamic layers of BHJ PCs. The emerging trend in a number of studies is the use of transparent metal oxides (MO_x) as superior category of interface materials having exceptional optical lucidity, good electrical conductivity and bendable work utility [4, 13, 78, 79]. Recent research has also embraced hybrid tandem PC by stacking an organic cell on top of an inorganic cell. The hybrid device can absorb higher energy photons without energy loss caused by thermalization and hence operate at higher voltages [80, 81]. The future of these devices lies in achieving a level of efficiency equal or even higher than that of stand-alone inorganic photovoltaic devices [75]. It is expected that this technology will provide cheap or entirely new devices having mechanical stability, impact resistance and excellent optical transparency.

7.5 Conclusion

This article has reviewed available literature on polymer/fullerene BHJ PCs. Considerable progress has been made in understanding the properties of polymer/fullerene BHJ PCs. Most studies have revealed the importance of the role played by polymeric materials blending the photoactive layer, their interaction during phase separation and their relation to performance of the devices. The wide range of conjugated polymers being synthesized with fullerene makes the study of morphological properties difficult to accomplish. Computational models have been developed to help elucidate the choice of best material combination during synthesis, since most experimental processes are largely based on trial-and-error approach. With such models, it is has become easy to predict the morphology of BHJ devices; besides the intrinsic utility of these models, the charge transport process in some organic material is complex and in most cases may only be modeled numerically [43, 82].

There are numerous opportunities for refining the productivity of the BHJ carbon-based photovoltaic devices. Adjustment of the morphological properties of the BHJ organic devices is currently a viable method of improving the present low-

energy levels of these devices. Proper adjustment and controlling of segregation in BHJ organic PCs hold a great opportunity to improve the effectiveness of the devices. The current level of efficiency is only about 11 %, which is not economically viable for industrial production. Current computational models predict that it is possible to attain higher levels of efficiency that are comparable to those of the silicon-based photovoltaic devices. However, the knowledge available about the effect of morphological conditions of the devices on their effectiveness is contradicting. Processing of BHJ PCs, characterization and designing of the instrument, optical modeling, electrical modeling and accurate categorization of PCE and external QE have become a challenge, nevertheless some new technology such as, inverted BHJ, tandem, tandem-hybrid and ternary PCs have emerged to pave way to better cell performance.

Notes

Flory Parameter x represents the interaction energy between the material pairs.

Acknowledgement: The authors are thankful to the Physics staff in UKZN for fruitful discussions and comments.

References

[1] Jo J, Kim SS, Na SI, Yu BK, Kim DY. Time-dependent morphology evolution by annealing processes on polymer:fullerene blend solar cells. Adv Funct Mater. 2009;19:866–74.
[2] Wang M, Zhu L, Zhou M, Jiang C, Li Q. High efficiency organic bulk-heterojunction solar cells applying a new system of co-additives. Mater Lett. 2016;166:227–30.
[3] Nelson J. Polymer:fulerene bulk heterojunction solar cells. Mater Today. 2011;14:462–70.
[4] Brady MA, Su GM, Chabinyc ML. Recent progress in the morphology of bulk heterojunction photovoltaics. Soft Matt. 2011;7:11065–77.
[5] Scharber MC, Sariciftci NS. Efficiency of bulk-heterojunction organic solar cells. Prog Poly Sci. 2013;38:1929–40.
[6] Sharma GD, Mikroyannidis JA, Roy SSS,MS, Thomas KRJ. Efficient bulk heterojunction photovoltaic devices based on diketopyrrolopyrrole containing small molecule as donor and modified PCBM derivatives as electron acceptors. Organ Electr. 2012;13:1763–64.
[7] Liang Y, Xu Z, Xia J, Tsai ST, Wu Y, Li G, et al. For the bright future-bulk heterojunction polymer solar cells with power conversion efficiency of 7.4%. Adv Mater. 2010;22:135–38.
[8] Yuan Y, Reece T, Sharma J,P, Poddar S, Ducharme S, Gruverman A, et al. Efficiency enhancement in organic solar cells with ferroelectric polymers. Nat Mater. 2011;10:296–302.
[9] Liu C, Zhan X, Li Z, He Y, Li J, Shen L, et al. Efficiency improvement of inverted organic solar cells via introducing a series of polyfluorene dots in electron transport layer. J Phys Chem C. 2015;29:16462–67.

[10] Pastore M, Mosconi E, Angelis FD, Grätzel M. A computational investigation of organic dyes for dye-sensitized solar cells–benchmark, strategies, and open issues. J Phys Chem C. 2010;15:7205–12.
[11] Watkins PK, Walker AB, Verschoor GLB. Dynamical Monte Carlo Modelling of organic solar cells–the dependence of internal quantum efficiency on morphology. Nano Lett. 2005;9:1814–18.
[12] Arı M, Kanat Z, Dinçer H. Design, computational screening and synthesis of novel non-peripherally tetra hexylthio-substituted phthalocyanines as bulk heterojunction solar cell materials. Sol Ener Mater Sol Cell. 2016;134:1–8.
[13] Deibel C, Dyakonov V. Polymer–fullerene bulk heterojunction solar cells. Rep Prog Phys. 2010;73:1–39.
[14] Zhou H, Zhang Y, Seifter J, Collins SD, Luo C, Bazan GC, et al. High-efficiency polymer solar cells enhanced by solvent treatment. Adv Mater. 2013;25:1646–52.
[15] Kumar CV, Cabau L, Viterisi A, Biswas S, Sharma GD, Palomares E. Solvent annealing control of bulk heterojunction organic solar cells with 6.6% efficiency based on a benzodithiophene donor core and dicyano acceptor units. J Phys Chem C. 2015;119:20871–79.
[16] Verploegen E, Mondal R, Bettinger CJ, Sok S, Toney MF, Bao Z. Effects of thermal annealing upon the morphology of polymer–fullerene blends. Adv Funct Mater. 2010;20:3519–29.
[17] Krishnaswamy RK, Yang Q. Influence of phase segregation on the mechanical properties of binary blends of polyethylenes that differ considerably in molecular weight. Poly. 2007;48:5348–54.
[18] Bernardo G, Bucknall DG. Recent progress in the understanding and manipulation of morphology in polymer: fullerene photovoltaic cells. Optoelectr Adv Mater Dev. 2013;9:208–27.
[19] Xavier P, Rao P, Bose S. Nanoparticle induced miscibility in LCST polymer blends: critically assessing the enthalpic and entropic effects. Phys Chem Chem Phys. 2016;18:47–64.
[20] Raba A, Cordan AS, Leroy Y. Two-dimensional simulation of organic bulk heterojunction solar cell: influence of the morphology. J Nanosci Nanotech. 2013;13:5164–69.
[21] Ragnetti M, Geiser D, Hocker H, Oberthur RC. Small angle neutron scattering (SANS) of cyclic and linear polystyrene in toluene. Macromol Chem. 1985;186:1701–09.
[22] Roovers J. Dilute-solution properties of ring polystyrenes. J Polym Sci Polym Phys Ed. 1985;23:1117–26.
[23] Hadziioannou G, Cotts PM, Tenbrinke G, Han CC, Lutzs P, Strazielle C, et al. Thermodynamic and hydrodynamic properties of dilute solutions of cyclic and linear polystyrenes. Macromol. 1987;20:493–97.
[24] Geyler S, Pakula T. Conformation of ring polymers in computer simulated melts. Macromol Chem Rapid Comm. 1988;9:617–23.
[25] Iyer BVS, Lele AK, Shanbhag S. What is the size of a ring polymer in a ring-linear blend? Macromol. 2007;40:5995–6000.
[26] M¨Uller M, Wittmer JP, Cates ME. Topological effects in ring polymers II: influence of persistence length. Phys Rev E. 2000;61:4078–89.
[27] Brown S, Lenczycki T, Szamel G. Influence of topological constraints on the statics and dynamics of ring polymers. Phys Rev E. 2001;63:1–4.
[28] Brown S, Szamel G. Computer simulation study of the structure and dynamics of ring polymers. J Chem Phys. 1998;109:6184–92.
[29] M¨Uller M, Wittmer JP, Cates ME. Topological effects in ring polymers: a computer simulation study. Phys Rev E. 1996;53:5063–74.
[30] Cates ME, Deutsch JM. Conjectures on the statistics of ring polymers. J Phys. 1986;47:2121–28.
[31] Halverson JD, Grest GS, Grosberg AY, Kremer K. Rheology of ring polymer melts: from linear contaminants to ring-linear blends. Phys Rev Lett. 2012;108:1–5.

[32] Lee CK, Pao CW, Chen. CW. Multiscale molecular simulations of the nanoscale morphologies of P3HT: PCBM blends for bulk heterojunction organic photovoltaic cells. Ener Environ Sci. 2011;4:4124–32.
[33] Huang DM, Faller R, Do K, Moule AJ. Coarse-grained computer simulations of polymer/fullerene bulk heterojunctions for organic photovoltaic applications. J Chem Theory Comput. 2010;6:526–37.
[34] Huang DM, Moule AJ, Faller R. Characterization of polymer–fullerene mixtures for organic photovoltaics by systematically coarse-grained molecular simulations. Fluid Phas Equil. 2011;302:2–25.
[35] Kremer K, Grest GS. Dynamics of entangled linear polymer melts: a molecular-dynamics simulation. J Chem Phys. 1990;92:5057–64.
[36] Grest GS, Kremer K. Molecular dynamics simulation for polymers in the presence of a heat bath. Phys Rev. 1986;33:3628–31.
[37] Pellicane G, Tchoukouegno MM, Mola TG, Tsige M. Surface segregation driven by polymer topology. Phys Rev E. 2016;93:1–5.
[38] Megnidio-Tchoukouegno M, Gaitho FM, Mola GT, Tsige M, Pellicane G. Unravelling the surface composition of symmetric linear-cyclic polymer blends. Fluid Phas Equil. 2017;441:33–42.
[39] Weeks JD, Chandler J, Andersen HC. Role of repulsive forces in determining the equilibrium structure of simple liquids. J Chem Phys. 1971;54:5237–47.
[40] Plimpton SJ. Fast parallel algorithms for short-range molecular dynamics. J Comput Phys. 1995;117:1–19.
[41] Kong J, Hwang IW, Lee K. Top-down approach for nanophase reconstruction in bulk heterojunction solar cells. Adv Mater. 2014;26:1–8.
[42] Li TSI, Walker GC. Interfacial free energy governs single polystyrene chain collapse in water and aqueous solution. J Am Chem Soc. 2010;132:6530–40.
[43] Ray B, Nair PR, Garcia RE. Modeling and optimization of polymer based Bulk Heterojunction (BH) solar cell. Birck NCN Publicat. 2009;(Paper 521):1–5.
[44] Lifshitz M, Slyozov VV. The kinetics of precipitation from supersaturated solid solutions. J Phys Chem Solid. 1961;19:35–50.
[45] Ji Y, Du C, Xu X, Hou T, Li Y. Characterising the morphology and efficiency of polymer solar cell by experiments and simulations. Molec Simul. 2016;42:836–45.
[46] Masahiko S, Itaru O, Yasuhito S, Kazuo T, Takashi O, Satoru I, et al. Highly efficient and stable solar cells based on thiazolothiazole and naphthobisthiadiazole copolymers. Sci Rep. 2015;5:1–5.
[47] Liang Y, Feng D, Wu Y, Tsai ST, Li G, Ray C, et al. Highly efficient solar cell polymers developed via fine-tuning of structural and electronic properties. J Am Chem Soc. 2009;131:7792–99.
[48] Walker B, Tamayo AB, Dang XD, Zalar P, Seo JH, Garcia A, et al. Nanoscale phase separation and high photovoltaic efficiency in solution-processed, small-molecule bulk heterojunction solar cell. Adv Funct Mater. 2009;19:3063–69.
[49] Nam YM, Huh J, Jo WH. Optimization of thickness and morphology of active layer for high performance of bulk-heterojunction organic solar cells. Sol Ener Mater Sol Cell. 2010;94:1118–24.
[50] Allen JE, Black CT. Improved power conversion efficiency in bulk heterojunction organic solar cells with radial electron contacts. Acs Nano. 2011;5:7986–91.
[51] Zhou S, Sun JX. Mobility dependent effciencies of organic bulk-heterojunction solar cells with recommbination via tail. Int J Moder Phys B. 2013;27:28–32.

[52] Tchoukouegno MM, Pellicane G, Tsige M, Mola GT. Nano-scale morphology dependent performance of thin film organic solar cells. J Mater Sci: Mater Electr. 2016;28:214–21.
[53] Facchetti A. π-Conjugated polymers for organic electronics and photovoltaic cell applications. Chem Mater. 2011;23:733–58.
[54] Shirakawa H, McDiarmid A, Heeger A. Focus article: twenty-five years of conducting polymers. Chem Comm. 2003;1:1–4.
[55] Yu G, Gao J, Hummelen JC, Wudl F, Heeger AJ. Polymer photovoltaic cells: enhanced efficiencies via a network of internal donor-acceptor heterojunctions. Sci. 1995;270:1789–91.
[56] Mola GT, Newayemedhin A. Correlation between LUMO offset of donor/acceptor molecules to an open circuit voltage in bulk heterojunction solar cell. Phys B: Cond Matt. 2014;445:56–59.
[57] Peet J. Efficiency enhancement in low-bandgap polymer solar cells by processing with alkane dithiols. Nat Mater. 2007;6:497–500.
[58] Lee JK, Ma WL, Brabec CJ, Yuen J, Moon SJ, Kim JY, et al. Processing additives for improved efficiency from bulk heterojunction solar cells. J Am Chem Soc. 2008;139:3619–23.
[59] Chan H. Polymer solar cells with enhanced open-circuit voltage and efficiency. Nat Photon. 2009;3:649–53.
[60] Hou L. Replacing alkoxy groups with alkylthienyl groups–a feasible approach to improve the properties of photovoltaic polymers. Angew Chem Int Ed Engl. 2011;50:9697–6702.
[61] Aich BR, Lu J, Beaupré S, Leclerc M, Tao Y. Control of the active layer nanomorphology by using coadditives towards high-performance bulk heterojunction solar cells. Org Electron. 2012;13:1736–41.
[62] Zhang S, Ye L, Zhao W, Yang B, Wang Q, Hou J. Realizing over 10% efficiency in polymer solar cell by device optimization. Sci China Chem. 2014;58:248–56.
[63] Varun V, Kazuaki K, Takeshi K, Tomoyuki K, Itaru O, Kazuo T, et al. Efficient inverted polymer solar cells employing favourable molecular orientation. Nat Photon. 2015;9:403–08.
[64] Sungho N, Gyoelim B, Park S, Lee BR, Cha MJ, Lim DC, et al. Highly efficient inverted bulk-heterojunction solar cells with a gradiently-doped ZnO layer. Ener Environ Sci. 2016;9:240–46.
[65] Yiwei Z, Nicholas SW, Tao W, David LG. Fabricating high performance conventional and inverted polymer solar cells by spray coating in air. Vacuum. 2017;139:154–58.
[66] Swager TM. 50th Anniversary perspective: conducting/semiconducting conjugated polymers. A personal perspective on the past and the future. Macromol. 2017;50:4867–86.
[67] Weinberger BR, Akhtar M, Gau C. Polyacetylene photovoltaic devices. Synth Met. 1982;4:187–97.
[68] Scott JC, Karg S, Carter SA. Bipolar charge and current distributions in organic light-emitting diodes. J Appl Phys. 1993;82:1454–60.
[69] Holger S, Krebs FC. A brief history of the development of organic and polymeric photovoltaics. Sol Ener Mater Sol Cell. 2004;83:125–46.
[70] Zhang T, Birgersson E, Luther J. Closed-form expressions correlating exciton transport and interfacial charge carrier generation with the donor/acceptor morphology in organic bulk heterojunction solar cells. Phys B-Cond Matt. 2015;456:267–74.
[71] Pramanik A, Sarkar B, Pal S, Sarkar P. Pentacene–fullerene bulk-heterojunction solar cell–a computational study. Phys Lett A. 2015;379:1036–42.
[72] Chen D, Zhang C, Wang Z, Zhang J, Feng Q, Xu S, et al. Performance comparison of conventional and inverted organic bulk heterojunction solar cells From optical and electrical aspects. IEEE Trans Electr Dev. 2013;60:451–57.
[73] Oseni SO, Mola GT. Properties of functional layers in inverted thin film organic solar cells. Sol Ener Mater Sol Cell. 2017;160:241–56.
[74] Oseni SO, Mola GT. The effect of uni- and binary solvent additives in PTB7:PC61BM based solar cells. Sol Ener Mater Sol Cell. 2017;150:66–72.

[75] Yi Z, Ni W, Zhang Q, Li M, Kan B, Wan X, et al. Effect of thermal annealing on active layer morphology and performance for small molecule bulk heterojunction organic solar cells. J Mater Chem C. 2014;2:7247–55.
[76] Wong YQ, Wong HY, Tan CS, Meng HF. Performance optimization of organic solar cells. IEEE Photo J. 2014;6:1–27.
[77] Arbouch I, Karzazi Y, Hammouti BB. Organic photovoltaic cells: operating principles, recent developments and current challenges – review. Phys Chem News. 2014;72:73–84.
[78] Li G, Chu CW, Shrotriya V, Huang J, Yanga Y. Efficient inverted polymer solar cells. Appl Phys Lett. 2006;88:1–5.
[79] Karagiannidis PG, Kalfagiannis N, Georgiou D, Laskarakis A, Hastas NA, Pitsalidis C, et al. Effects of buffer layer properties and annealing process on bulk heterojunction morphology and organic solar cell performance. J Mater Chem. 2012;22:14624–32.
[80] You JY, Dou L, Yoshimura K, Kato T, Ohya K, Moriarty T, et al. A polymer tandem solar cell with 10.6% power conversion efficiency. Nat Comm. 2013;4:1–10.
[81] Zhou H, Zhang Y, Mai CK, Collins SD, Bazan GC, Nguyen TQ, et al. Polymer homo-tandem solar cells with best efficiency of 11.3%. Adv Mater. 2015;27:1767–73.
[82] MacKenzie RCI, Kirchartz T, Dibb GFA, Nelson J. Modeling nongeminate recombination in P3HT:PCBM solar cells. J Phys Chem C. 2011;115:9806–13.

Irene Tsibranska, Serafim Vlaev, Daniela Dzhonova,
Bartosz Tylkowski, Stela Panyovska and
Nadezhda Dermendzhieva

8 Modeling and assessment of the transfer effectiveness in integrated bioreactor with membrane separation

Abstract: Integrating a reaction process with membrane separation allows for effective product removal, favorable shifting of the reaction equilibrium, overcoming eventual inhibitory or toxic effects of the products and has the advantage of being energy and space saving. It has found a range of applications in innovative biotechnologies, generating value-added products (exopolysaccharides, antioxidants, carboxylic acids) with high potential for separation/ concentration of thermosensitive bioactive compounds, preserving their biological activity and reducing the amount of solvents and the energy for solvent recovery. Evaluating the effectiveness of such integrated systems is based on fluid dynamics and mass transfer knowledge of flowing matter close to the membrane surface – *shear deformation* rates *and shear stress at the membrane interface, mass transfer* coefficients. A Computational Fluid Dynamics (CFD)-based approach for assessing the effectiveness of integrated stirred tank bioreactor with submerged membrane module is compiled. It is related to the hydrodynamic optimization of the selected reactor configuration in two-phase flow, as well as to the concentration profiles and analysis of the reactor conditions in terms of reaction kinetics and mass transfer.

Keywords: computational fluid dynamics, effectiveness, integration of processes, mass transfer, membrane bioreactor

8.1 Introduction

The advantage of the integrated bioreactor with membrane separation is its ability to separate products with sufficient difference in molecular mass from complex systems with varying concentrations and composition. The research interest in the field includes innovative membrane bioreactor (MBR) designs, less energy consumption, effective membrane fouling control (rotating, baffled, vibrating etc.) and

This article has previously been published in the journal Physical Sciences Reviews. Please cite as: Tsibranska, I., Vlaev, S., Dzhonova, D., Tylkowski, B., Panyovska; S., Dermendzhieva, N. Modeling and assessment of the transfer effectiveness in integrated bioreactor with membrane separation *Physical Sciences Reviews* [Online] 2020, 5. DOI: 10.1515/psr-2020-0063

https://doi.org/10.1515/9783110678215-008

beneficial economic estimates [1–7]. In side-stream MBRs, the quality of mixing in the reactor vessel and the fouling conditions in the membrane section (the shear stress field to its surface) are clearly distinguished, whereas a submerged membrane bioreactor (sMBR) is a hybrid vessel with flow characteristics affected both by the turbulence promoter (aeration and/or mixer) and the membrane module.

The integrated membrane reactors compete successfully with conventional continuous reactors with suspended biomass and agitation [8], the sMBR being of special interest [3, 9, 10]; the latter includes creating conditions for high product quality at reduced energy consumption and reduced amounts of the solvents used [11], as well as beneficial economic estimates such as lower operating costs of the integrated operation [12, 13]. Innovative MBR designs attempt the use of submerged membrane module sMBRs to achieve better membrane performance, less energy consumption, effective and green/sustainable fouling control. The importance of stirred sMBRs is increasing in parallel with the integrated processing of value-added molecules where retention and concentration of biocatalysts promote continuous hybrid operation. Examples of such operation in presence of components with high fouling potential are reported [3, 14].

The CFD's potential for application in a sMBR is most often realized in hydrodynamic studies, especially concerning membrane fouling and aeration [15]. Recent publications report CFD combined hydrodynamic and mass transfer simulation of membrane modules with different configuration and stirred sMBRs [16, 17].

8.1.1 Scope of application

Membrane modules immersed in the bioreactor, or side-stream ones connected in recycle [18–20] have been used in different biotechnological processes for separation of thermally unstable products, including in two-phase systems [21]. They have proven their efficiency in biological treatment of waste flows with high content and variety of organic contaminants, as well as an effective barrier against many active pharmaceutical ingredients, pesticides, alkylphenols from production of nonionic surfactants and other endocrine disrupters and resistant to degradation organic pollutants [22–27]. The design of this type of integrated reactors is the subject of further analysis and classification [28, 29] for biofuels production [12, 30–32] for separating the products of the trans-esterification (mono-, di- and triglycerides, free glycerol and unreacted excess methanol); separation of carboxylic acids from industrial fermentation [33–37] including continuous ones [38–40]; separation of amino acids and peptides – e.g., of L-glutamic acid from the fermentation broth by a sequence of membranes [41]; *p*-nonylphenol removal under nitrifying conditions [42] in a sMBR using tubular membrane module.

Attractive applications of the sMBR are found in microbial operation [43, 44]. The sMBR configuration was studied and discussed on the example of production

and recovery of proteins, surfactants, polysaccharides and metabolites [45]; fructo-oligosaccharides [46]; emulsion-stabilizing exopolysaccharides by production oriented Antarctic yeast process [47, 48] etc. Integration of the bioreactor with membrane separation in submerged configuration is developing in new directions such as submerged osmotic MBRs, membrane electrobioreactors, photocatalytic MBRs, reverse osmosis (RO) or forward osmosis (FO) membrane bioreactors [9, 49, 50]. Some examples, illustrating the diversity of applications, are summarized in Table 8.1

8.1.2 Research trends

The current trends concerning basic research in the area of integrated bioreactors with membrane separation can be summarized as follows:

The engineering part of the reactor design including membrane separation, its optimal geometry and conditions for the integrated process [68] have been the subject of considerable scientific interest. The comparative study itself of the hydrodynamics in membrane bioreactors of both configurations – inside submerged versus outside submerged membranes (i.e., side-stream ones connected in recycle) – is an object of CFD modeling [69]. The latter is also important in scale-up methodologies for MBR [70].

Recent developments in membrane separation technology and increased implementation of various membrane techniques for water purification and recovery of value-added products bring attention to the *immersed modular systems in stirred tanks* [69]. Combined knowledge on the performance of the integrated systems, based on the fluid flow in stirred tank reactors and the separation dynamics of immersed membranes, is needed. At condition of negligible effect of permeation on the hydrodynamics in the vicinity of the membrane surface, the impeller-induced cross-flow was examined [71]. In the CFD simulations, the rate-of-deformation tensor was targeted, as determined by the local gradients of the component velocities near the membrane interface [71]. CFD was applied to demonstrate the effect of reactor configurations (the relative position between membrane and impeller, effect of impeller design) on the fluid flow pattern in submerged membrane reactors [72]. Continuous perfusion-based bioprocessing in disposable vessels has extended the importance of membranes onto a research related to single-use bioreactors [47, 72, 73].

Fouling reduction has been searched by different means for turbulence intensification; stirred vessels with turbulent impellers and aeration have been proposed [74–77]. The need of fouling control is an important condition for proper production using membranes in submerged configuration [78]. The essential means for fouling mitigation are related to hydrodynamics and fouling problems have been analyzed in view of the shear level at the membrane surface. The *relation between flux and shear stress* created by stirring are usually qualitatively defined [70, 71], or limits of influence are considered, where

Table 8.1: Application area of sMBR.

Application area	Membrane process	Membrane module	sMBR design and operation details	Reactor volume
Water treatment				
Oily WW [51]	MF	PVDF hollow fiber	Pendulum type oscillation	30 L
Landfill leachate [52]	UF	PVDF hollow fiber	Periodic feeding and intermittent aeration	20 L
Fruit-juice industry WW [53]	UF	PES flat plate	Aerated tank	35 L
Yogurt production WW [54]	MF	Flat-plate	No data	Industrial (600 m^3/d)
Paper mill WW [55]	MF	Hollow fiber	Continuously fed, aerated tank	10 L
Coal gasification WW [56]	MF	Flat sheet (chlorinated polyvinyl chloride)	Long-term bio-augmented (CH_3OH, granular activated carbon, inorganic carbon added)	7 L
Phenol-contaminated saline water [57]	UF	Four tubular membranes	Extractive MBR (triphasic, aeration)	No data
Potable WW reuse [58]	FO	Cellulose triacetate with embedded polyester mesh	Osmotic MBR	Pilot scale
High strength WW treatment for water reclamation [59]	MF	Polyethylene hollow fiber	Sponge-sMBR, aeration	7 L
Secondary WW effluents for water reclamation[60]	UF	Flat sheet (PS)	Gravity-driven sMBR	Lab
Heavy metals removal from synthetic industrial WW [61]	FO	Flat-sheet, dense non-porous cellulose triacetate membrane	Osmotic membrane bioreactor, aerated	4.5 L
Production of soluble microbial products [62]-	MF	Flat sheet (chlorinated polyethylene)	Anaerobic, baffled, recycle biogas used for mixing, fouling control;	3 L

Digestion of molassa treating of bio-refinery sidestreams [63]	MF	Flat plate chlorinated polyethylene	Anaerobic, magnetically induced membrane vibration, with biogas recirculation	10 L
Sodium gluconate production [64]	MF	Tubular (ceramic)	Fermentor + membrane module coupled in recycle mode	50 L
Continuous hydrolysis of lignocellulosic biomass, [2], [65]	UF	Flat sheet placed at the bottom of the reactor (PES)	Stirred tank reactor (pitched blade impeller) with bottom placed membrane	1.5 L
Phenol hydrogenation to cyclohexanone [3]	MF	Tubular ceramic (a-Al_2O_3)	Stirred tank bioreactor with L-shaped membrane tube	2 L
Syngas fermentation [66], [67]	MF	Hollow fiber (polysulfone)	Stirred (2 Rushton turbines) sMBR with full cell retention	1 L

FO, forward osmosis; MBR, membrane bioreactor; MF, microfiltration; PES, polyethersulfone; PVDF, polyvinylidene fluoride; sMBR, submerged membrane bioreactor; UF, ultrafiltration; WW, wastewater.

membrane fouling could be effectively controlled by shear stress [79]. Using combined impeller and membrane designs allows application in highly *viscous and complex fluids* where both interface and transport problems occur [80]. Complex hydrodynamics in sMBR may lead to restricted fluid mobility and cause difficulties in both the fermentation reaction and the product recovery. Depending on the input variables, unfavorable hydrodynamic conditions could be generated in the integrated processing vessels that have to be revealed in advance to avoid problems of inefficient product recovery [81]. In a stirred sMBR, shear control on fouling should be balanced with shear effect on cell viability, so that turbulence-promoted membrane should assure reduced fouling and in the same time avoid the risk for cell physiological damage [47].

Mass transfer investigations are focused on the following:

a) *Sherwood number expressions*, accounting for a concentration dependent viscosity and the effect of the transmembrane pressure drop; prediction of the steady state permeate flux, as well as the transient flux decline profile [82]. Exploring the effect of fouling-layer thickness on mass transfer numerically is an important step in the search of reliable *Sh* number correlations accounting for the fouling-layer thickness and including the *Re* and *Sc* numbers [83].
b) *Concentration polarization* (CP) phenomena leading to reduced permeate flux, which is a major problem in membrane separation and remains in the focus of the mass transfer studies and modeling. Because of the accumulation of retained solutes, reversibly and immediately occurring, the driving force for filtration decreases and the permeating solvent meats an increasing transport resistance. The solute distribution within the CP layer and the effect of the increased concentration at the membrane interface has been solved both for steady- and unsteady-state conditions [84]; in dead-end (stirred and unstirred) [85] and cross-flow filtration modes [84]; alone or coupled with the membrane layer [86] or gel layer growth on the membrane surface [87]. Fluid dynamic modeling was applied to analyze the mass transfer characteristics in nanofiltration including CP [88]. Examples with reverse osmosis simulations and experimental validation show that the used CFD models properly predict both permeate flux and the evolution of the polarization layer [84]. The effects of variable permeate flux, boundary layer thickness and rejection coefficients on the concentration profile close to the membrane surface are discussed in the case of nanofiltration of polyphenols from natural extracts [89].
c) *Gas presence*, which is an effective tool to solve problems related to CP and fouling in the integrated process [20]. In a typical sMBR, enhanced shear stress on the deposited cake layer is achieved through upward blowing of air bubbles. A significant variation of the permeate flux with the rate of air blowing is observed and explained by the created shearing stress at the membrane surface. At higher aeration rates, increased back transport of foulants occurs thus resulting in reduced CP and rate of cake layer formation [90]. The use of a gas/

liquid two-phase flow leads to enhanced mass transfer [91], the sMBR being of special interest [76]. Local Sherwood numbers and shear stress distributions at the membrane surface around bubbles at different gas velocities were approached by CFD simulations [92]. Optimum operation conditions for submerged in aerated bioreactor membrane modules are searched [20] in direction of positioning of the membrane modules; gas/liquid ratio; bubble size; transmembrane pressure; obtaining sufficient wall shear stress to create friction on the membrane surface. In the case of stirred sMBR, the effect of gas flow on shear at various tip velocities of the stirrer is studied together with the sparging intensity and rheology influence [47]. The gas phase was found to reduce wall shear, but to increase shear uniformity.

d) *Mass transfer* coefficients

The mass transfer and the shear stress distribution at the membrane surface are two main factors directly affecting the effective membrane operation. Models, coupling hydrodynamic and mass transfer calculations, are found, such as the ones proposed for NF in a slit-type channel [93] and applied to NF and RO in a spacer-filled channel [94]; the models of a polarization layer, during RO in a slit-type channel [84] and in a roto-dynamic filtration system [95]. Fluid to membrane mass transfer coefficients are obtained by combined hydrodynamic and mass transfer modeling, based on shear stress distribution along the membrane surface [96] and the relation between shear stress and mass transfer coefficient [97, 98]. The latter was applied in CFD study of submerged hollow fiber [98] and tubular [17] membrane module. The relation of mass transfer coefficient to the velocity field in the viscous sublayer near the wall was used, as proposed by Reiss and Hanratty [99] and widely used in membrane studies by the electrochemical technique [100] for measuring shear rate at the membrane surface.

This study presents an overview of the flow and mass transfer characteristics in a stirred sMBR with in-line tubular membrane module and the development of evaluation method for assessing the integrated process effectiveness based on CFD modeling and simulation. The impeller is considered as prospective turbulence promoter in a process involving production and recovery of value-added biomass from non-Newtonian viscous biofluids. The effectiveness of the integrated process is defined as the ratio of the product spatial-average mass transfer coefficient to the value determined at maximum product proliferation rate in the film at the feed side of the membrane.

8.2 Research results

To be able to discriminate between effective processing, a detailed study of the sMBR flow behavior and species transport is required. Both issues come down to evaluation

of mass transfer rate across the membrane interface. The research analysis has been exemplified by a dynamic membrane filtration system of a stirred sMBR illustrated in Figure 8.1. Details of component size have been reported elsewhere [17, 47].

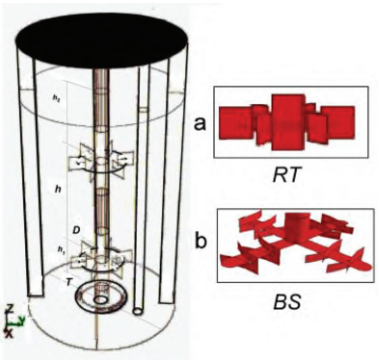


Figure 8.1: An example of bioreactor setup schematic including the tubular membrane module and alternative impeller designs: **a**. configuration 2RT with dual six flat-blade impeller (RT); **b**. configuration 2 MV equipped with dual backswept (BS) impeller.

8.2.1 Flow behavior

Velocity and shear stress distribution inside the example bioreactor and at the example membrane surface were studied depending on the *impeller geometry and mixing intensity* [70, 101]. The characteristics of the impeller and stirring conditions are given in Table 8.2. Simulation and computer visualization of the flow behavior using a Reynolds-averaged Navier-Stokes model and CFD methodology were used at different mixing conditions, thus allowing enhanced access of the retentate fluid to the membrane surface, as well as possible low membrane fouling potential related to microfiltration practice.

Vis-à-vis the conventional six flat-blade impeller (Figure 8.2, case 2RT) imposes predominantly radial circulation across the membrane interface, whereas the curved blades (Figure 8.2, case 2 MV) produce a multivortex mixed radial and axial flow pattern of the backswept type accompanied by different flow velocity distribution along the membrane-fluid interface. In case of gas presence, change of the flow structure occurs due to the tendency of the air to propagate around the impellers, which hinders the mixing at higher rpm. The same sparger position leads to central and peripheral gas hold-up zones for RT impeller and predominantly central gas hold-up zones around the BS impellers. Consequently, the different structure of gas dispersion influences the shear force [101], the focus being improved velocity gradient and enhanced shear profile close to the membrane [81].

Table 8.2: Details on the stirred tank submerged membrane bioreactor used in this study.

Type of impeller, Figure 8.1	Geometry	Liquid viscosity	Gas presence	Stirring velocity Re number	Spec. power input P/V_r
Radial, six flat-blade	T = 0.4 m D/T = 1/3 a = 25 mm	Power-law fluid n = 0.34–0.78, K = 0.02–0.55 Pa s^n	No	200–750 rpm Re = (0.16–7)·10^4	0.1–4 W/dm^3
Radial, six flat-blade, dual impeller,	T = 0.16 m D/T = 0.412 h/D = 1.67 h_1/D = 0.76 h_2/D = 1.06 a = 20 mm	Power-law fluid 0.34 < n < 0.78 K = 0.02–0.1 Pa s^n	Nonaerated Aerated 4.5–9 dm^3/min	400–750 rpm Re = (0.1–2)·10^4	0.1–5 W/dm^3
Curved-4 blade backswept, dual impeller	T = 0.16 m D/T = 0.425 h/D = 1.62 h_1/D = 0.84 h_2/D = 0.84 a = 20 mm	Power-law fluid n = 0.78, K = 0.02–0.1 Pa s^n	Nonaerated Aerated 4.5 dm^3/min	600–1100 rpm Re = (0.13–2)·10^4	1.7–8.6 W/dm^3 0.1–4.5 W/dm^3

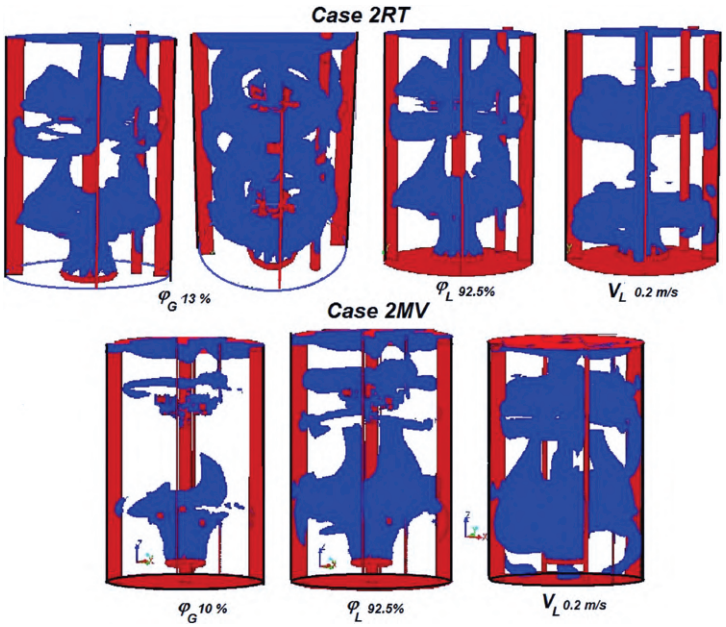


Figure 8.2: Impeller view and *iso*-surfaces of gas hold-up 10%, liquid hold-up 90 % and liquid velocity 0.2 m/s in a stirred reactor compared at close values of specific input power: **a**. case 2RT – radial flat-blade (RT) impeller 400 rpm, 1.5 W/dm^3; **b**. Case 2 MV – curved-blade (BS) impeller 750 rpm, 1.35 W/dm^3.

Figure 8.3 illustrates the effect of stirring velocity and gas presence on the average bulk and membrane surface shear rate. The values of the latter at 400 rpm are used as comparison base. As expected, stirring velocity has an important effect, stronger at the membrane surface, where more than 3-fold increase is observed in the range of the studied stirring intensities (400 to 750 rpm). At the same time, the average bulk shear rate is increased by a factor of 1.7. The gas phase is found to reduce both bulk and wall shear, but to increase shear uniformity [47]. The negative effect on the membrane mean near-wall shear rate is proved by the comparison, shown in Figure 8.3, as well as by calculations at two different gas flow rates [47]. The explanation is found in the fact that increasing the gas volume fraction decreases the contact of the liquid at the membrane-liquid interface, as well as in flow structural changes. As can be seen from Figure 8.3, the reduction is in the order of 0.73–0.75 times for the bulk shear rate vs. 0.80–0.93 for the wall shear rate; a closer analysis of the shear distribution along the tubular membrane surface showed improved longitudinal shear uniformity [47].

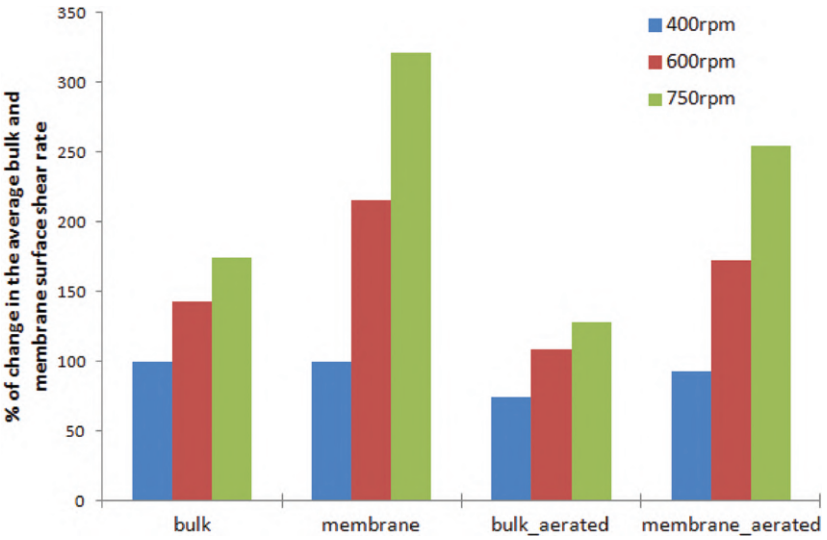


Figure 8.3: Average bulk and membrane surface shear rate.

8.2.2 Species transport

Depending on the specific hydrodynamic resistance considered, the separate steps of mass transfer across the membrane surface – external (this study), internal pore diffusional [102] or in-channel one [97] are to be considered. Referring to the sMBR studied recently [17], the membrane boundary layer size δ has been identified before evaluation of specific mass transfer rate, k_m. δ is an important variable constituting

the mass transfer driving force, namely, the concentration gradient at the membrane surface. Concentration profiles at the retention side of the tubular membrane of Figure 8.1 were elaborated by CFD; a plot summary is illustrated in Figure 8.4. Results calculated at position facing the impeller discharge and dimensionless axial distance Z = $2z/T$ = 0.625 are presented. Moderate gas flow rate of 1 vvm are supposed, stirring intensities 400, 600 and 750 rpm and a power-law fluid (consistency coefficient K = 0.02 Pa s^n to 0.1 Pa s^n, flow index n = 0.78) are included. The analysis of the solute mass fraction profiles in the boundary layer radially to the membrane surface also includes the interimpeller and subsurface sMBR area (at high mixing rate of 750 rpm).

The liquid velocity and contact at the membrane are the major variables affecting the concentration boundary layer δ_c and the respective membrane mass transfer performance at the retention side. The first one is illustrated by the concentration profiles at different stirring intensities. The liquid contact at the membrane is affected by the gas present and was discussed previously [17] for gas flow rates 1 to 2 vvm. It was shown that within the range of the aeration rates used in practice, the impact of gas velocity (approx. 0.5–1 cm/s) is low compared with the impeller-induced liquid velocity (0.5–2 m/s). Furthermore, at these stirring and gas flow velocities, the regime is closer to uniform gas distribution and δ_c is affected mainly by liquid velocity and only partially by gas velocity. Deviations from the aforementioned due to gas bypass and larger local gas hold-up variations may be observed at high stirring intensities (750 rpm) and high gas flow rates in the subsurface and interimpeller sections (Figure 8.4 and [17]).

The results [17] showed that the product transport rate was maximal in case of mixing intensity of 750 rpm, gas flow rate of 1 vvm and water-like fluid that exhibited minor boundary layer thickness of 20 µm.

8.2.3 Mass transfer

Membrane mass transfer analyses rely on the relationship between membrane shear stress and mass transfer rate, as derived by Reiss and Hanratty [99, 103] based on their proposed electrochemical diffusion experiments; the latter lead to a formulation binding up shear rate $\dot{\gamma}$ and liquid film mass transfer coefficient, k_m in the recurrent relationship:

$$k_m = 0.862\left(\frac{\dot{\gamma} D_{eff}^2}{d_e}\right)^{\frac{1}{3}}$$

In a separate study, a range of k_m values at a variety of conditions were determined for the bioreactor. Combined with the contours of linear velocity at four horizontal cross sections, the spatial transfer characteristic of the tubular membrane based on

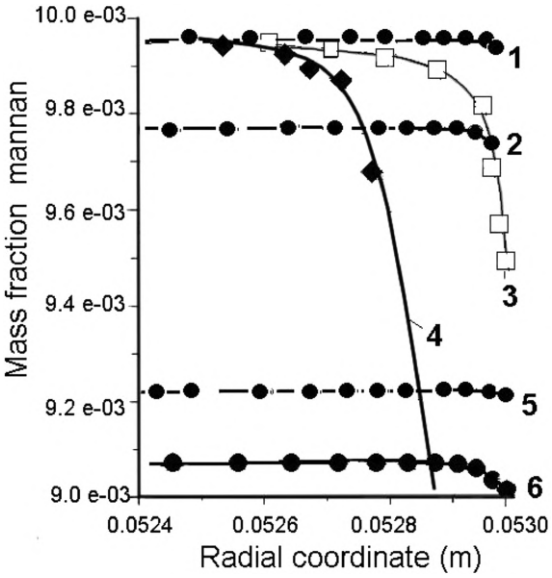


Figure 8.4: An overview of the mass transfer driving force at various conditions: results at dimensionless axial distance Z = 2*z*/*T* = 0.625, impeller position and water-like fluid: 1 – 750 rpm, 2 – 400 rpm, 5 – 600 rpm, 6 – viscous fluid, 400 rpm, 3 – 750 rpm, subsurface position, 4 – 750 rpm, interimpeller position.

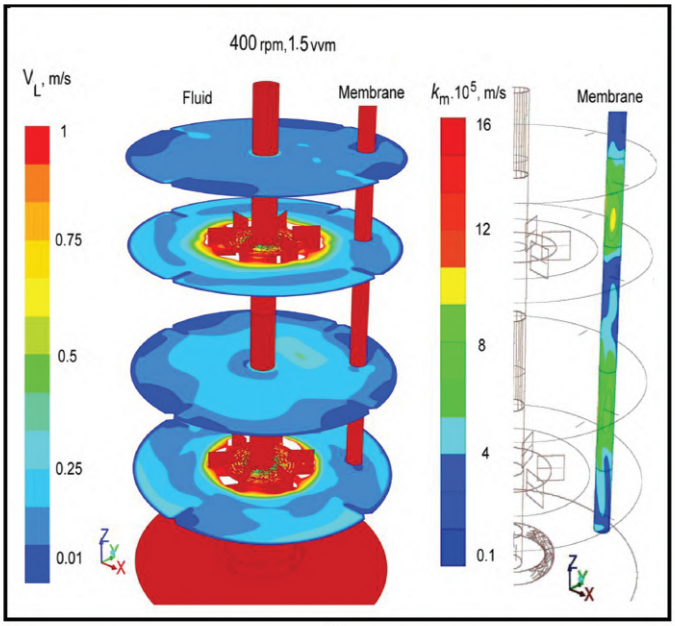
retention-side area-averaged parameter k_m is visualized by CFD in Figure 8.5. The case of high mixing intensity of 750 rpm with minor boundary layer thickness of 20 μm (depicted in Figure 8.4) was found to produce the maximum specific mass transfer rate of $k_{\mathrm{m}} \sim 16 \cdot 10^{-5}$ m/s.

8.3 Discussion

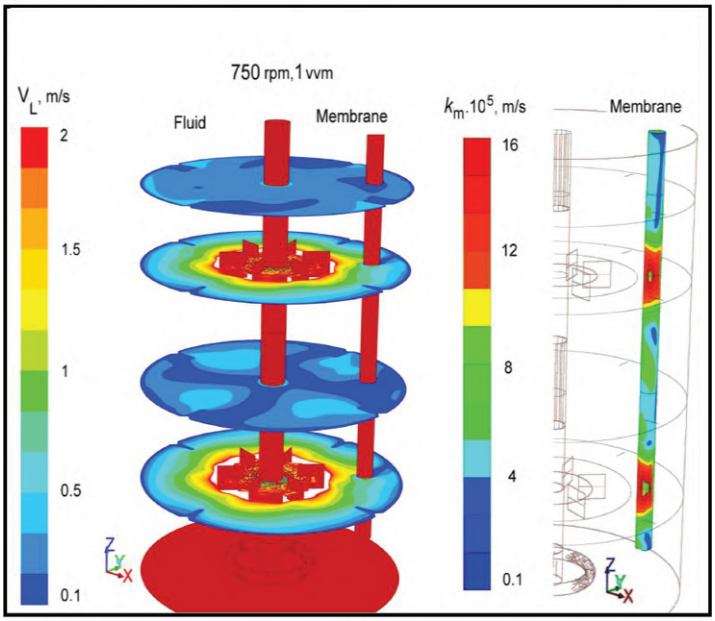
The results obtained were interpreted in the context of membrane effectiveness.

8.3.1 Inference on immersed membrane effectiveness

There are numerous studies and publications devoted to mass transfer, yet unique parameter to represent MBR mass transfer effectiveness and thus to serve for comparison of various MBR units has been hardly pointed out or highlighted. Membrane mass transfer has been studied by various parameters, most often the flux being the target value [104, 105]. The solute flux is related to diffusivity and the mass transfer driving force – the concentration differences and the thickness of the transfer limiting stagnant layers across the membrane interface – the one at the feed side of the



a.



b.

Figure 8.5: Contours of k_m spatial z-distribution along membrane surface showing membrane mass transfer relative to impeller positions and relative to the vessel bulk fluid velocity contours at **a.** 400 rpm, 1.5 vvm; **b.** 750 rpm and 1 vvm.

membrane, the one through the membrane pores and the one at the permeate side. Thus, the concentration differences and the thickness of the relevant layers have been analyzed. Most simply, the relationship $J = k_m \ln(\Delta w_m/\Delta w_b)$ has been derived and put forward the concentration differences $[(w_m - w_p)/(w_b - w_p)]$ and the mass transfer coefficient, k_m (m/s), as important variables [97]. In advance, both the nondimensional concentration boundary layer [106] and k_m [107–110] have been preferred parameters, but there were also others. Studies deal also with permeability coefficient [111], permeation velocity or boundary layer thickness [106], permeate flux [112], pore diffusion transfer coefficient [102], even with diffusion limiting electrical current intensity [113]. Studying the energy efficiency, the permeate processing cost of membrane designs was assessed by flux and specific Sherwood number [104, 114].

Mass transfer effectiveness has been discussed rarely. It has been represented as flux enhancement parameter ϕ formulated as a ratio of flux with impact on flux without impact [112]; permeate flux enhancement factor [115]; mass transfer enhancement parameter, for example, by a function of diffusion-limited current enhancement in flat-sheet membranes [113]. In other studies, [116] permeate flux enhancement at the feed side of the membrane has been measured and mass transfer enhancement factor E has been formulated as Sherwood number ratio [117]. Referring to FO and being limited to specification for water, the authors have pointed out that a membrane-characteristic parameter has to be selected and they refer to flux ratio [105].

Referring to sMBR hydrodynamics and assuming that CFD has the modeling potential to determine shear rate $\dot{\gamma}$ and k_m in various most complicated cases using also the relationships between flux J and k_m and the recurrent relationship between k_m and $\dot{\gamma}$ [97, 100, 118], it looks most pertinent to characterize membrane effectiveness by k_m. But there is also another argument to support this inference. Referring to the work of Reiss [101, 103] quoted in previous studies [98, 113, 118], shear rate appears to be the most appropriate flow parameter that could be directly related to the local rate of mass transfer. Consequently, using k_m helps to outline the hydrodynamic perspective of the MBR unit.

Because effectiveness is defined as the degree of achievement of the target value, one should formalize it to occur as extreme specific mass transfer rate of the membrane module considered. Table 8.3 contains k_m values obtained in various sMBR studies related to mass transfer; the overview allows assessing the differences in the species transport potentials of the various MBR units.

Deduced from Table 8.3, shear rates and corresponding k_m values are extreme in dynamic cross-flow filtration (DCF) units that perform shear-enhanced filtration. Assuming the effectiveness E_{ff} to be defined as the degree of achievement of the target value, it should be formulated in relative units, that is, as a ratio of k_m values. The proper way should be to take the ratio of a local value of $k_{m,i}$ corresponding to the local velocity gradient, and an extreme value $k_{m,\,max}$, determined as the maximum desirable target value attainable at the most favorable performance conditions.

Table 8.3: Overview of specific mass transfer rate data reported in recent MBR studies.

Ref	Separation system studied	Reference conditions	k_m, m/s	Remark
[109, 112]	Gas-sparged UF/vertical tubular module	Turbulent liquid film Re > 1000	$6\cdot\cdot10^{-7}$ – $2\cdot10^{-5}$	Calculated: Correlation for turbulent falling film flow
[97]	Gas-sparged UF/vertical tubular module	Physical modeling of tubular air-lift MBR; Re∼400–1800 (U_L∼ 0.2 m/s; U_G∼0·01 m/s) (Sh∼30, $D = 7\cdot10^{-10}$ m^2/s)	$\sim1.10^{-6}$ 5.10^{-6} Sh∼30	Measured and calculated
[108]	Stirred cell UF/dead-end bottom membrane	Dextran 50 kDa Dextran 1000 Da	$\sim 1\cdot10^{-5}$ $\sim 4\cdot10^{-5}$	Calculated: Reiss eqn.
[108]	Cross-flow system UF	Dextran 50 kDa Dextran 1000 Da	$1.6\cdot10^{-6}$ $5.9\cdot10^{-6}$	Calculated:film model- based relationships
[107]	Stirred cell, flat bottom membrane	Re∼6500–39000 (0.3 m.s up to 0.9 m/s)	$3\cdot10^{-5}$–$1\cdot10^{-4}$	Reported measurement and CFD Sh∼50–100
[119]	Spacer-filled channels of spiral wound membrane module	Re∼70–200	$8\cdot10^{-5}$–$2.5\cdot10^{-4}$	Calculated from Sh number, reported CFD and measurements
[116]	Unobstructed channel at the feed side of a membrane module	Electro-osmotic flow with permeation, Re∼200–2000	$\sim 3\cdot10^{-5}$	CFD
[120]	Spacer-filled membrane modules	Laminar flow, Re∼75–100	$2\cdot10^{-5}$–$1\cdot10^{-4}$	CFD
[110]	Spacer-filled membrane modules	Laminar flow, Re∼75–200	$2\cdot10^{-5}$–$1.3\cdot10^{-4}$	CFD
[17]	Dual impeller sMBR/vertical tubular module	Re∼$1\cdot10^3$–$27\cdot10^3$	$2\cdot10^{-5}$–$1.6\cdot10^{-4}$	CFD
[121]	Spiral wound module; spacer-filled channels	Re∼50–170	$2\cdot10^{-5}$–$8\cdot10^{-5}$	CFD
[122, 123]	Rotating disc system, 2500 rpm Vibrating system VSEP 60.75 Hz	$\dot{\gamma}$∼$1.5\cdot10^4$–$4.6\cdot10^5$ 1/s $\dot{\gamma}$∼$1.12\cdot10^4$–$1.5.10^4$ 1/s	$3.4\cdot10^{-4}$–$5\cdot10^{-4}$ $1.5\cdot10^{-4}$–$3.5\cdot10^4$	Calculated using Reiss equation

sMBR, submerged membrane bioreactor; UF, ultrafiltration.

This way a value of E_{ff} could be assigned to the various zones of the membrane module, and what is more, if based on a widely accepted standard, one could compare the effectiveness E_{ff} of different membrane vessels.

As being inferred from Table 8.3, shear rates and corresponding k_m values are extreme in DCF systems, for example, rotating disc membrane units that perform shear-enhanced filtration. In so far that the highest mass transfer coefficients are characteristic for DCF systems [122] and in view of their high reported shear rates of the order of 10^5 s^{-1}, one could expect the highest k_m values to be of the order of $5{\cdot}10^{-4}$ m/s. Such k_m values should be characteristic for 100% effectiveness of membrane reactors. However, as pointed out by the author [122], the result is obtained at the cost of enhanced energy consumption, for example, specific energy consumption up to 30 kWh/m^3 for RDM (rotating disc membranes) compared with less than 10 kWh/m^3 in the case of the example sMBR of this study.

In this study, we have assumed k_m to be the parameter that characterizes the mass transfer effectiveness by taking the ratio of the product local or spatial-average mass transfer coefficient to the value determined at maximum product proliferation rate in the film at the feed side of the membrane, $E_{ff} = k_{m,i}/k_{m,\,max}$. As seen by the concentration profiles and the k_m values determined in Figure 8.4 and Table 8.3, an almost negligible concentration boundary layer thickness and highest mass transfer rate correspond to the most rigorous mixing hydrodynamics, that is, impeller speed of 750 rpm and position facing the impeller discharge, angle of attack 0°. At such conditions, the thickness of the boundary layer is minimal (ca. $2{\cdot}10^{-5}$ m), k_m is at the maximum of $1.6{\cdot}10^{-4}$ m/s and effectiveness is marked as 100%. After the spatial distribution of local k_m, the effectiveness is spread in parallel to k_m, as illustrated along four exploration lines in Figure 8.6. Comparing these data of agitated sMBR with sMBR promoted by bubbling or airlift effectiveness at such conditions, it is found to be higher than the one of airlift or other units that exhibit lower k_m values of the order of 10^{-5} m/s (cf. Table 8.3). As an extension of Figure 8.5, the membrane mass transfer effectiveness of the particular stirred sMBR in water-like biofluid (e.g., feed apparent viscosity 6.7 mPa s) at 400 rpm and 1 vvm is shown to vary vertically and angularly in Figure 8.7. Figure 8.8 contains a summary of overall effectiveness E_{ff} – values at various conditions. The results can be used to determine permeate fluxes and compare the energy efficiency of various DCF systems.

8.4 Conclusion

Based on an analysis of previous and present studies of fluid dynamics and mass transfer, a unified approach to formulate membrane effectiveness to mass transfer is proposed. It is formulated as fraction of a maximum achievable specific mass transfer rate assumed to be the target value for the particular system. The study suggests

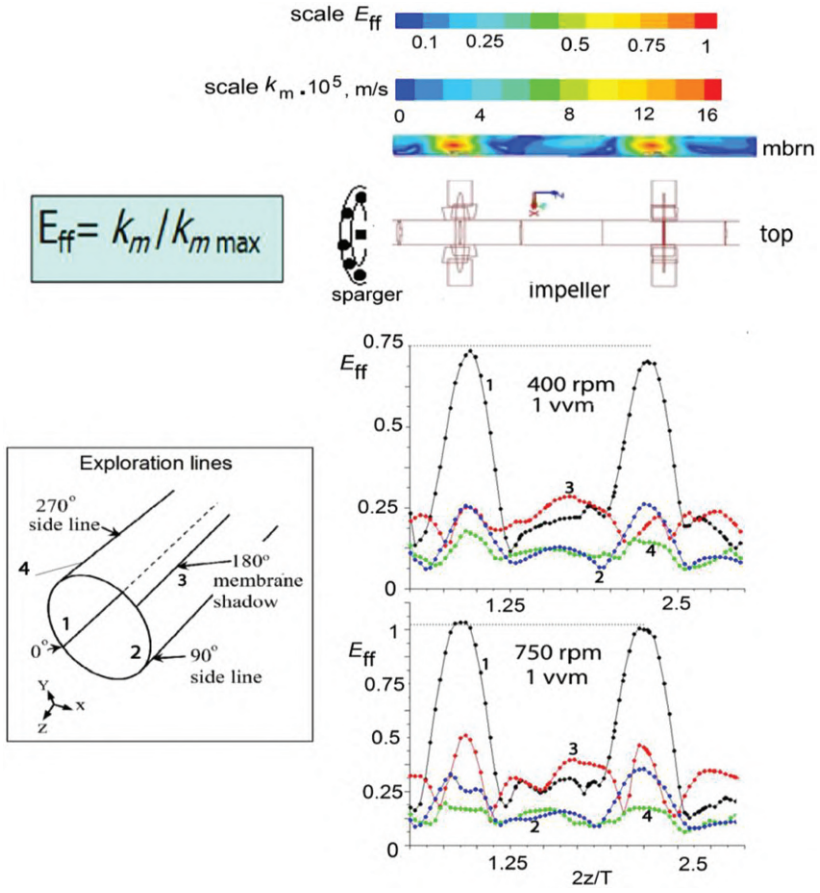


Figure 8.6: Formulation of mass transfer effectiveness E_{ff} as partial maximum specific mass transfer rate, k_m. Example axial variation of effectiveness E_{ff} along the four exploration lines shifted at 90 degrees from one another: line 1 at 0° facing the impeller discharge, line 2 at 90°, line 3 at 180°, and 4 at 270°.

the implementation of a three-step procedure for evaluation of the mass transfer potential, including (1) CFD evaluation of membrane area spatial shear rate distribution, (2) CFD analysis of species transport to uncover the component mass transport resistances allowing for marking down the maximum achievable specific mass transfer rate, and (3) determination and mapping of membrane mass transfer effectiveness E_{ff} parameter ratio. Once obtained, the CFD-generated E_{ff} parameter allows comparison of various membrane systems or extension of the studies to determine membrane systems energy efficiency. The case of study is exemplified by agitated sMBRs with immersed tubular membrane module and the concept is illustrated by contour plots and X–Y plots allowing comparison of different DCF systems.

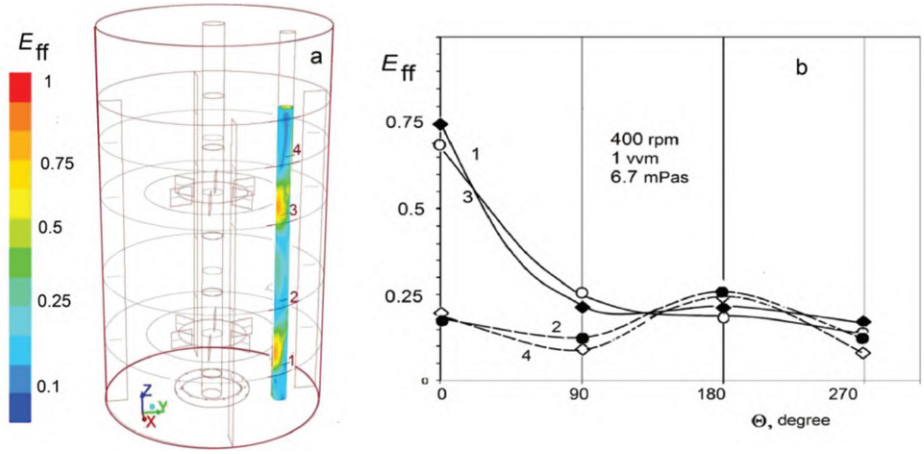


Figure 8.7: Vertical and angular variation of membrane mass transfer effectiveness in water-like (e.g., of apparent viscosity 6.7 mPa s) feed at impeller speed 400 rpm, gas flow rate 1 vvm shown as **a**. contours facing the impeller; **b**. *X*–*Y* plots corresponding to four levels and four angular positions.

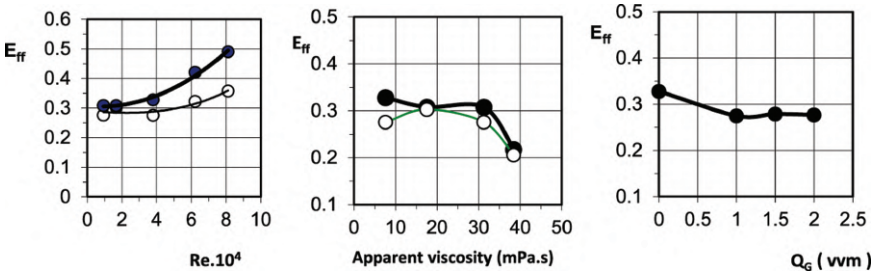


Figure 8.8: Overall membrane-average effectiveness at various Reynolds number, apparent viscosity and aeration rate: ● no gas, ○ gassed.

Acknowledgments: The authors gratefully acknowledge the financial support of the National Science Fund of the Republic of Bulgaria according to Contract DN 07/11/15.12.2016.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. Robles, Á, Ruano, MV, Charfi, A, Lesage, G, Heran, M, Harmand, J, et al. A review on anaerobic membrane bioreactors (AnMBRs) focused on modelling and control aspects. *Bioresour Technol* 2018;270:612–26. https://doi.org/10.1016/j.biortech.2018.09.049.
2. Mohammadmahdi, M, Stickel, J, Wickramasinghe, SR. Investigation of a submerged membrane reactor for continuous biomass hydrolysis. *Food Bioprod Process* 2015;96: 189–97. https://doi.org/10.1016/j.fbp.2015.07.001.
3. Jiang, H, Qu, Z, Li, Y, Huang, J, Chen, R, Xing, W. One-step semi-continuous cyclohexanone production via hydrogenation of phenol in a submerged ceramic membrane reactor. *Chem Eng J* 2016;284:724–32. https://doi.org/10.1016/j.cej.2015.09.037.
4. Wei, P, Zhang, K, Gao, W, Kong, L, Field, R. CFD modeling of hydrodynamic characteristics of slug bubble flow in a flat sheet membrane bioreactor. *J Membr Sci* 2013;445:15–24. https://doi.org/10.1016/j.memsci.2013.05.036.
5. Rezakazemi, M, Maghami, M, Mohammadi, T. High loaded synthetic hazardous wastewater treatment using lab-scale submerged ceramic membrane bioreactor. *Period Polytech – Chem Eng* 2018;62:299–304. https://doi.org/10.3311/PPch.11459.
6. Krzeminski, P, Leverette, L, Malamis, S, Katsou, E. Membrane bioreactors–a review on recent developments in energy reduction, fouling control, novel configurations, LCA and market prospects. *J Membr Sci* 2017;527:207–27. https://doi.org/10.1016/j.memsci.2016.12.010.
7. Nguyenhuynh, T, Nithyanandam, R, Chong, CH, Krishnaiah, D. Configuration modification of a submerged membrane reactor for enzymatic hydrolysis of cellulose. *Biocatal Agric Biotechnol* 2017;12:50–8. https://doi.org/10.1016/j.bcab.2017.08.013.
8. Bakonyi, P, Nemestóthy, N, Simon, V, Bélafi-Bakó, K. Fermentative hydrogen production in anaerobic membrane bioreactors: a review. *Bioresour Technol* 2014;156:357–63. https://doi.org/10.1016/j.biortech.2014.01.079.
9. Galinha, CF, Sanches, S, Crespo, JG. Chapter 6-Membrane bioreactors. In: *Fundamental modelling of membrane systems*. Amsterdam, Netherlands: Elsevier; 2018. https://doi.org/10.1016/B978-0-12-813483-2.00006-X.
10. Dosta, J, Nieto, JM, Vila, J, Grifoll, M, Mata-Álvarez, J. Phenol removal from hypersaline wastewaters in a Membrane Biological Reactor (MBR): operation and microbiological characterisation. *Bioresour Technol* 2011;102:4013–20. https://doi.org/10.1016/j.biortech.2010.11.123.
11. Martínez, MB, Van der Bruggen, B, Negrin, ZR, Alconero, PL. Separation of a high-value pharmaceutical compound from waste ethanol by nanofiltration. *J Ind Eng Chem* 2012;18:1635–41. https://doi.org/10.1016/j.jiec.2012.02.024.
12. Wei, P, Cheng, L-H, Zhang, L, Xu, X-H, Chen, H-L, Gao, C-J. A review of membrane technology for bioethanol production. *Renew Sustain Energy Rev* 2014;30:388–400. https://doi.org/10.1016/j.rser.2013.10.017.
13. Murthy, GS, Sridhar, S, Shyam, S, Shankaraiah, B, Ramakrishna, M. Concentration of xylose reaction liquor bynanofiltration for the production of xylitol sugar alcohol. *Separ Purif Technol* 2005;44:221–8. https://doi.org/10.1016/j.seppur.2005.01.009.
14. Lyagin, E, Drews, A, Bhattacharya, S, Kraume, M. Membrane reactor systems for parallel continuous screening and characterisation of biocatalysts. *Chem Eng Trans* 2012;27:319–24. https://doi.org/10.3303/CET1227054.
15. Tsibranska, I, Vlaev, S, Tylkowski, B. The problem of fouling in submerged membrane bioreactors – model validation and experimental evidence. *Phys Sci Rev* 2018;3:1. https://doi.org/10.1515/psr-2017-0143.

16. Haddadi, B, Jordan, C, Miltner, M, Harasek, M. Membrane modeling using CFD: combined evaluation of mass transfer and geometrical influences in 1D and 3D. *J Membr Sci* 2018;563:199–209. https://doi.org/10.1016/j.memsci.2018.05.040.
17. Vlaev, SD, Dzhonova-Atanasova, D, Tsibranska, I. CFD evaluation of mass transfer distribution heterogeneity along the membrane-liquid interface in stirred submerged membrane bioreactors. *Chem Eng Process* 2020;147:107738. https://doi.org/10.1016/j.cep.2019.107738.
18. Firdaous, L, Dhulster, P, Amiot, J, Gaudreau, Al, Lecouturier, D, Kapel, R, et al. Concentration and selective separation of bioactive peptides from an alfalfa white protein hydrolysate by electrodialysis with ultrafiltration membranes. *J Membr Sci* 2009;329:60–7. https://doi.org/10.1016/j.memsci.2008.12.012.
19. Valadez-Blanco, R, Ferreira, FRC, Jorge, RF, Livingston, AG. A membrane bioreactor for biotransformations of hydrophobic molecules using organic solvent nanofiltration (OSN) membranes. *J Membr Sci* 2008;317:50–64. https://doi.org/10.1016/j.memsci.2007.04.032.
20. Wibisono, Y, Cornelissen, ER, Kemperman, AJB, vander Meer, WGJ, Nijmeijer, K. Two-phase flow in membrane processes: a technology with a future. *J Membr Sci* 2014;453:566–602. https://doi.org/10.1016/j.memsci.2013.10.072.
21. Priac, A, Morin-Crini, N, Druart, C, Gavoille, S, Bradu, C, Lagarrigue, C, et al. Alkylphenol and alkylphenol polyethoxylates in water and wastewater: a review of options for their elimination. *Arabian J Chem* 2017;10:S3749–73. https://doi.org/10.1016/j.arabjc.2014.05.011.
22. Peeva, L, da Silva Burgal, J, Valtcheva, I, Livingston, AG. Continuous purification of active pharmaceutical ingredients using multistage organic solvent nanofiltration membrane cascade. *Chem Eng Sci* 2014;116:183–94. https://doi.org/10.1016/j.ces.2014.04.022.
23. Tambosi, J, de Sena, R, Favier, M, Gebhardt, W, José, H, Schröder, H, et al. Removal of pharmaceutical compounds in membrane bioreactors (MBR) applying submerged membranes. *Desalination* 2010;261:148–56. https://doi.org/10.1016/j.desal.2010.05.014.
24. Yangali-Quintanilla, V, Maeng, SK, Fujioka, T, Kennedy, M, Amy, G. Proposing nanofiltration as acceptable barrier for organic contaminants in water reuse. *J Membr Sci* 2010;362:334–34. https://doi.org/10.1016/j.memsci.2010.06.058.
25. Nguyen, LN, Hai, FI, Kang, J, Price, WE, Nghiem, LD. Removal of emerging trace organic contaminants by MBR-based hybrid treatment processes. *Int Biodeterior Biodegrad* 2013;85:474–82. https://doi.org/10.1016/j.ibiod.2013.03.014.
26. Wu, B, Kitade, T, Chong, TH, Uemura, T, Fane, AG. Impact of membrane bioreactor operating conditions on fouling behavior of reverse osmosis membranes in MBR–RO processes. *Desalination* 2013;311:37–45. https://doi.org/10.1016/j.desal.2012.11.020.
27. Mutamim, NSA, Noor, ZZ, Hassan, MAA, Yuniarto, A, Olsson, G. Membrane bioreactor: applications and limitations in treating high strength industrial wastewater. *Chem Eng J* 2013;225:109–19. https://doi.org/10.1016/j.cej.2013.02.131.
28. Mutamim, NSA, Noor, ZZ, Hassan, MAA, Olsson, G. Application of membrane bioreactor technology in treating high strength industrial wastewater: a performance review. *Desalination* 2012;305:1–11. https://doi.org/10.1016/j.desal.2012.07.033.
29. Teixeira, CB, Junior Madeira, JV, Macedo, GA. Biocatalysis combined with physical technologies for development of a green biodiesel process. *Renew Sustain Energy Rev* 2014;33:333–43. https://doi.org/10.1016/j.rser.2014.01.072.
30. Abels, C, Carstensen, F, Wessling, M. Membrane processes in biorefinery applications. *J Membr Sci* 2013;444:285–317. https://doi.org/10.1016/j.memsci.2013.05.030.
31. Othman, R, Mohammad, AW, Ismail, M, Salimon, J. Application of polymeric solvent resistant nanofiltration membranes for biodiesel production. *J Membr Sci* 2010;348:287–97. https://doi.org/10.1016/j.memsci.2009.11.012.

32. López-Garzón, CS, Straathof, AJ. Recovery of carboxylic acids produced by fermentation. *Biotechnol Adv* 2014;32:873–904. https://doi.org/10.1016/j.biotechadv.2014.04.002.
33. Lubsungneon, J, Srisuno, S, Rodtong, S, Boontawan, A. Nanofiltration coupled with vapor permeation-assisted esterification as an effective purification step for fermentation-derived succinic acid. *J Membr Sci* 2014;459:132–42. https://doi.org/10.1016/j.memsci.2014.02.006.
34. Ghaffar, T, Irshad, M, Anwar, Z, Aqil, T, Zulifqar, Z, Tariq, A, et al. Recent trends in lactic acid biotechnology: a brief review on production to purification. *J Radiat Res Appl Sci* 2014;7:222–9. https://doi.org/10.1016/j.jrras.2014.03.002.
35. Sikder, J, Chakraborty, S, Pal, P, Drioli, E, Bhattacharje, C. Purification of lactic acid from microfiltrate fermentation broth by cross-flow nanofiltration. *Biochem Eng J* 2012;69:130–7. https://doi.org/10.1016/j.bej.2012.09.003.
36. Sikder, J, Roy, M, Dey, P, Pal, P. Techno-economic analysis of a membrane-integrated bioreactor system for production of lactic acid from sugarcane juice. *Biochem Eng J* 2012;63:81–7. https://doi.org/10.1016/j.bej.2011.11.004.
37. Dey, P, Pal, P. Direct production of l (+) lactic acid in a continuous and fully membrane-integrated hybrid reactor system under non-neutralizing conditions. *J Membr Sci* 2012;389:355–62. https://doi.org/10.1016/j.memsci.2011.10.051.
38. Pal, P, Dey, P. Process intensification in lactic acid production by three stage membrane integrated hybrid reactor system. *Chem Eng Process* 2013;64:1–9. https://doi.org/10.1016/j.cep.2012.12.006.
39. Dey, P, Pal, P. Modelling and simulation of continuous l (+) lactic acid production from sugarcane juice in membrane integrated hybrid-reactor system. *Biochem Eng J* 2013;79:15–24. https://doi.org/10.1016/j.bej.2013.06.014.
40. Kumar, R, Vikramachakravarthi, D, Pal, P. Production and purification of glutamic acid: a critical reviewtowards process intensification. *Chem Eng Process* 2014;81:59–71. https://doi.org/10.1016/j.cep.2014.04.012.
41. Buitrón, G, Torres-Bojorges, AX, Cea-Barcia, G. Removal of p-nonylphenol isomers using nitrifying sludge in a membrane sequencing batch reactor. *Chem Eng J* 2015;281:860–8. https://doi.org/10.1016/j.cej.2015.07.018.
42. Carstensen, F, Apel, A, Wessling, M. In-situ productrecovery: submerged versus external loop membranes. *Areview. J Membr Sci.* 2012;394–395:1–36. https://doi.org/10.1016/j.memsci.2011.11.029.
43. Drioli, E, Brunetti, A, Di Profio, G, Barbieri, G. Process intensification strategies and membrane engineering. *Green Chem* 2012;14:1561–72. https://doi.org/10.1039/c2gc16668b.
44. Dhariwal, A. *Significance of submerged ceramic membrane systems for production-oriented bioprocesses* [PhD Thesis]. Saarbrücken: der Universität des Saarlandes; 2007.
45. Sanchez, O, Guio, F, Garcia, D, Silva, E, Caicedo, L. Fructooligosaccharides production by Aspergillus sp. N74 in amechanically agitated airlift reactor. *Food Bioprod Process* 2008;86:109–15. https://doi.org/10.1016/j.fbp.2008.02.003.
46. Vlaev, SD, Rusinova-Videva, S, Pavlova, K, Kuncheva, M, Panchev, I, Dobreva, S. Submerged culture process forbiomass and exopolisaccharide production by Antarctic yeast:some engineering considerations. *Appl Microbiol Biotechnol* 2013;97:5303–13. https://doi.org/10.1007/s00253-013-4864-3.
47. Vlaev, SD, Tsibranska, I, Dzhonova-Atanasova, D. Hydrodynamic characterization of dual-impeller submerged membrane bioreactor relevant to single-use bioreactor options. *Chem Eng Res Des* 2018;132:930–41. https://doi.org/10.1016/j.cherd.2018.02.004.
48. Andrić, P, Meyer, AS, Jensen, PA, Dam-Johansen, K. Reactor design for minimizing product inhibition during enzymatic lignocellulose hydrolysis II. Quantification of inhibition and

suitability of membrane reactors. *Biotechnol Adv* 2010;28:407–25. https://doi.org/10.1016/j.biotechadv.2010.02.005.
49. Hasan, SW, Elektorowicz, M, Oleszkiewicz, JA. Start-up period investigation of pilot-scale submerged membrane electro-bioreactor (SMEBR) treating raw municipal wastewater. *Chemosphere* 2014;97:71–7. https://doi.org/10.1016/j.chemosphere.2013.11.009.
50. Giwa, A, Dindi, A, Kujawa, J. Membrane bioreactors and electrochemical processes for treatment of wastewaters containing heavy metal ions, organics, micropollutants and dyes: recent developments. *J Hazard Mater* 2019;370:172–95. https://doi.org/10.1016/j.jhazmat.2018.06.025.
51. Qin, L, Fan, Z, Xu, L, Zhang, G, Wang, G, Wu, D, et al. A submerged membrane bioreactor with pendulum type oscillation (PTO) for oily wastewater treatment: membrane permeability and fouling control. *Bioresour Technol* 2015;183:33–41. https://doi.org/10.1016/j.biortech.2015.02.018.
52. Remmas, N, Melidas, P, Zerva, I, Kristoffersen, JB, Nikolai, S, Tsiamis, G, et al. Dominance of candidate saccharibacteria in a membrane bioreactor treating medium age landfill leachate: effects of organic load on microbial communities, hydrolytic potential and extracellular polymeric substances. *Bioresour Technol* 2017;238:48–56. https://doi.org/10.1016/j.biortech.2017.04.019.
53. Demirkol, GT, Dizge, N, Acar, TO, Salmanli, OM, Tufekci, N. Influence of nanoparticles on filterability of fruit-juice industry wastewater using submerged membrane bioreactor. *Water Sci Technol* 2017;76:705–11. https://doi.org/10.2166/wst.2017.255.
54. Pellegrin, ML, Aguinaldo, J, Arabi, S, Sadler, ME, Greiner, AD, Wong, J, et al. Membrane processes. *Water Environ Res* 2018;90:1457–536. https://doi.org/10.2175/106143018x15289915807272.
55. Erkan, HS, Engin, GO. The investigation of paper mill industry wastewater treatment and activated sludge properties in a submerged membrane bioreactor. *Water Sci Technol* 2017;76:1715–25. https://doi.org/10.2166/wst.2017.351.
56. Jia, S, Han, H, Zhuang, H, Hou, B. Bio-augmented submerged membrane bioreactor as an effective application for treatment of biologically pretreated coal gasification wastewater. *J Chem Technol Biotechnol* 2015;91:1532–9. https://doi.org/10.1002/jctb.4751.
57. Yeo, BJL, Goh, S, Livingston, AG, Fane, AG. Controlling biofilm development on the extractive membrane bioreactor. *Separ Sci Technol* 2017;52:113–21. https://doi.org/10.1080/01496395.2016.1246569.
58. Morrow, CP, McGaughey, AL, Hiibel, SR, Childress, AE. Submerged or sidestream? The influence of module configuration on fouling and salinity in osmotic membrane bioreactors. *J Membr Sci* 2018;548:583–92. https://doi.org/10.1016/j.memsci.2017.11.030.
59. Ngo, HH, Guo, W, Xing, W. Evaluation of a novel sponge-submerged membrane bioreactor (SSMBR) for sustainable water reclamation. *Bioresour Technol* 2008;99:2429–35. https://doi.org/10.1016/j.biortech.2007.04.067.
60. Fortunato, L, Jeong, S, Wang, Y, Behzad, AR, Leiknes, T. Integrated approach to characterize fouling on a flat sheet membrane gravity driven submerged membrane bioreactor. *Bioresour Technol* 2016;222:335–43. https://doi.org/10.1016/j.biortech.2016.09.127.
61. Aftab, B, Khan, SJ, Maqbool, T, Hankins, NP. Heavy metals removal by osmotic membrane bioreactor (OMBR) and their effect on sludge properties. *Desalination* 2017;403:117–27. https://doi.org/10.1016/j.desal.2016.07.003.
62. Kunacheva, C, Soh, YNA, Stuckey, DC. Effect of feed pH on reactor performance and production of soluble membrane products (SMPs) in a submerged anaerobic membrane bioreactor. *Chem Eng J* 2017;320:135–43. https://doi.org/10.1016/j.cej.2017.03.018.

63. De Vrieze, J, Hennebel, T, Van den Brande, J, Bilad, RM, Bruton, TA, Vankelecom, IFJ, et al. Anaerobic digestion of molasses by means of a vibrating and non-vibrating submerged anaerobic membrane bioreactor. *Biomass Bioenergy* 2014;68:95–105. https://doi.org/10.1016/j.biombioe.2014.06.009.
64. Lu, F, Li, C, Wang, Z, Zhao, W, Chu, J, Zhuang, Y, et al. High efficiency cell-recycle continuous sodium gluconate production by Aspergillus niger using on-line physiological parameters association analysis to regulate feed rate rationally. *Bioresour Technol* 2016;220:433–41. https://doi.org/10.1016/j.biortech.2016.08.062.
65. Nguyenhuynh, T, Nithyanandam, R, Chong, CH, Krishnaiah, D. A review on using membrane reactors in enzymatic hydrolysis of cellulose. *J Eng Sci Technol* 2017;12:1129–52. https://doi.org/10.1016/j.bcab.2017.08.013.
66. Kantzow, C, Mayer, A, Weuster-Botz, D. Continuous gas fermentation by Acetobacterium woodii in a submerged membrane reactor with full cell retention. *J Biotechnol* 2015;212:11–18. https://doi.org/10.1016/j.jbiotec.2015.07.020.
67. Mayer, A, Schädler, T, Trunz, S, Stelzer, T, Weuster-Botz, D. Carbon monoxide conversion with Clostridium aceticum. *Biotechnol Bioeng* 2018;115:2740–50. https://doi.org/10.1002/bit.26808.
68. Brannock, MWD, De Wever, H, Wang, Y, Leslie, G. Computational fluid dynamics simulations of MBRs: inside submerged versus outside submerged membranes. *Desalination* 2009;236:244–51. https://doi.org/10.1016/j.desal.2007.10.073.
69. Brannock, M, Wang, Y, Leslie, G. Mixing characterisation of full-scale membrane bioreactors: CFD modelling with experimental validation. *Water Res* 2010;44:3181–91. https://doi.org/10.1016/j.watres.2010.02.029.
70. Vlaev, SD, Tsibranska, I. Shear stress generated by radial flow impellers at bioreactor-integrated membranes. *Theor Found Chem Eng* 2016;50:959–68. https://doi.org/10.1134/s004057951606018x.
71. Meng, L, Cheng, JC, Jiang, H, Yang, C, Xing, WH, Jin, WQ. Design and analysis of a submerged membrane reactor by CFD simulation. *Chem Eng Technol* 2013;36:1874–82. https://doi.org/10.1002/ceat.201300206.
72. Langer, ES, Rader, RA. Continuous bioprocessing and perfusion. *Bioprocess J* 2014:50–5.
73. Wagner, R. Value-adding perfusion-based biomanufacturing. *Eur Biotechnol* 2015;14:86–8.
74. Fane, AG, Li, NN, Fane, AG, Winston-Ho, WS, Matsuura, T. Submerged membranes. In: *Advance membrane technology and applications*. New Jersey: Wiley; 2008.
75. Bohm, L, Drews, A, Prieske, H, Berube, PR, Kraume, M. The importance of fluid dynamics for MBR fouling mitigation. *Bioresour Technol* 2012;122:50. https://doi.org/10.1016/j.biortech.2012.05.069.
76. Braak, E, Alliet, M, Scherite, S, Albasi, C. Aeration and hydrodynamics in submerged membrane bioreactors. *J Membr Sci* 2011;379:1–18. https://doi.org/10.1016/j.memsci.2011.06.004.
77. Chau, CCV, Berube, PR, Hall, ER. Relationship between types of surface shear stress profiles and membrane fouling. *Water Res* 2011;45:6403–16. https://doi.org/10.1016/j.watres.2011.09.031.
78. Meng, F, Zhang, S, Oh, Y, Zhou, Z, Shin, HS, Chae, SR. Fouling in membrane bioreactors: an updated review. *Water Res* 2017;114:151–80. https://doi.org/10.1016/j.watres.2017.02.006.
79. Du, X, Qu, F, Liang, H, Li, K, Chang, H, Li, G. Cake properties in ultrafiltration of TiO2 fine particles combined with HA: in situ measurement of cake thickness by fluid dynamic gauging and CFD calculation of imposed shear stress for cake controlling. *Environ Sci Pollut Res* 2016;23:8806–18. https://doi.org/10.1007/s11356-015-5984-3.

80. Qi, C, Wang, J, Lin, Y. New insight into influence of mechanical stirring on membrane fouling of membrane bioreactor: mixed liquor properties and hydrodynamic conditions. *Bioresour Technol* 2016;211:654–63. https://doi.org/10.1016/j.biortech.2016.03.143.
81. Vlaev, S, Tsibranska, I, Dzhonova-Atanasova, D, Popov, R. Structural anomalies in stirred submerged bioreactors relevant to immersed membrane use. *Food Sci Appl Biotechnol* 2018;1:56–62. https://doi.org/10.30721/fsab2018.v1.i1.12.
82. Banerjee, S, De, S. An analytical solution of Sherwood number in a stirred continuous cell during steady state ultrafiltration. *J Membr Sci* 2012;389:188–96. https://doi.org/10.1016/j.memsci.2011.10.029.
83. Koutsou, CP, Karabelas, AJ, Kostoglou, M. Fluid dynamics and mass transfer in spacer-filled membrane channels: effect of uniform channel-gap reduction due to fouling. *Fluid* 2018;3:1–20. https://doi.org/10.3390/fluids3010012.
84. Salcedo-Díaz, R, García-Algado, P, García-Rodríguez, M, Fernández-Sempere, J, Ruiz-Beviá, F. Visualization and modeling of the polarization layer in crossflow reverse osmosis in a slit-type channel. *J Membr Sci* 2014;456:21–30. https://doi.org/10.1016/j.memsci.2014.01.019.
85. Van den Berg, GB, Smolders, CA. Flux decline in ultrafiltration processes. *Desalination* 1990;77:101–33. https://doi.org/10.1016/0011-9164(90)85023-4.
86. Nagy, E, Kulcsár, E, Nagy, A. Membrane mass transport by nanofiltration: coupled effect of the polarization and membrane layers. *J Membr Sci* 2011;368:215–22. https://doi.org/10.1016/j.memsci.2010.11.046.
87. Xiao, K, Shen, Y, Huang, X. An analytical model for membrane fouling evolution associated with gel layer growth during constant pressure stirred dead-end filtration. *J Membr Sci* 2013;427:139–49. https://doi.org/10.1016/j.memsci.2012.09.049.
88. Cao, H, O'Rourke, M, Habimana, O, Casey, E. Analysis of surrogate bacterial cell transport to nanofiltration membranes: effect of salt concentration and hydrodynamics. *Separ Purif Technol* 2018;207:498–505. https://doi.org/10.1016/j.seppur.2018.06.072.
89. Tsibranska, I, Dzhonova-Atanasova, D, Panyovska, St. Effect of variable flux and rejection in membrane separation of polyphenols-containing natural extracts. *J of CTM* 2020;4. in press.
90. Praneeth, K, Moulik, S, Vadthya, P, Bhargava, SK, Tardio, J, Sridhar, S. Performance assessment and hydrodynamic analysis of a submerged membrane bioreactor for treating dairy industrial effluent. *J Hazard Mater* 2014;274:300–13. https://doi.org/10.1016/j.jhazmat.2014.04.030.
91. Ding, A, Liang, H, Li, G, Derlon, N, Szivak, I, Morgenroth, E, et al. Impact of aeration shear stress on permeate flux and fouling layer properties in a low pressure membrane bioreactor for the treatment of grey water. *J Membr Sci* 2016;510:382–90. https://doi.org/10.1016/j.memsci.2016.03.025.
92. Yang, J. *Prospects for flux enhancement in anaerobic membrane bioreactors treating saline wastewater* [Dissertation]. Delft: TU Delft, Delft University of Technology; 2013.
93. Geraldes, V, Semião, V, De Pinho, MN. Flow and mass transfer modelling of nanofiltration. *J Membr Sci* 2001;191:109–28. https://doi.org/10.1016/s0376-7388(01)00458-6.
94. Keir, GP. *Coupled modelling of hydrodynamics and mass transfer in membrane filtration* [PhD thesis]. Australia: Deakin University; 2012.
95 Jogdand, A, Chaudhuri, A. Modelling of concentration polarization and permeate flux variation in a roto-dynamic reverse osmosis filtration system. *Desalination* 2015;375:54–70. https://doi.org/10.1016/j.desal.2015.07.011.
96. Dzhonova-Atanasova, D, Tsibranska, I, Paniovska, S. CFD simulation of cross-flow filtration. *Chem Eng Trans* 2018;70:2041–6. https://doi.org/10.3303/CET1870341.

97. Ratkovich, N, Berube, PR, Nopens, I. Assessment of mass transfer coefficients in coalescing slug flow in vertical pipes and applications to tubular airlift membrane bioreactors. *Chem Eng Sci* 2011;66:1254–68. https://doi.org/10.1016/j.ces.2010.12.034.
98. Bérubé, PR, Afonso, G, Taghipour, F, Chan, CCV. Quantifying the shear at the surface of submerged hollow fiber membranes. *J Membr Sci* 2006;279:495–505. https://doi.org/10.1016/j.memsci.2005.12.043.
99. Reiss, LP, Hanratty, TJ. An experimental study of the unsteady nature of the viscous sublayer. *AIChE J* 1963;9:154–60. https://doi.org/10.1002/aic.690090204.
100. Böhm, L, Jankhah, S, Tihon, J, Bérubé, PR, Kraume, M. Application of the electrodiffusion method to measure wall shear stress: integrating theory and practice. *Chem Eng Technol* 2014;37:938–50. https://doi.org/10.1002/ceat.201400026.
101. Vlaev, SD, Dzhonova-Atanasova, D. Local velocity and shear deformation rate at model membranes immersed in a bioreactor agitated by curved-blade impeller: the effect of membrane position. *MMT- J Inter Sci Publicat* 2017;11:216–29.
102. Li, K, Kong, J, Tan, X. Design of hollow fibre membrane modules for soluble gas removal. *Chem Eng Sci* 2000;55:5579–88. https://doi.org/10.1016/s0009-2509(00)00193-7.
103. Hanratty, TJ, Campbell, JA. Measurement of wall shear stress. In: Goldstein, RJ, editor. *Fluid mechanics measurements*. Washington: Hemisphere; 1987.
104. Fimbres-Weihs, GA, Wiley, DE. Review of 3D CFD modeling of flow and mass transfer in narrow spacer-filled channels in membrane modules. *Chem Eng Process* 2010;49:759–81. https://doi.org/10.1016/j.cep.2010.01.007.
105. Tiraferri, A, Yip, NY, Straub, AP, Castrillon, SRV, Elimelech, M. A method for the simultaneous determination of transport and structural parameters of forward osmosis membranes. *J Membr Sci* 2013;444:523–38. https://doi.org/10.1016/j.memsci.2013.05.023.
106. Geraldes, V, Semião, V, de Pinho, MN. Flow and mass transfer modelling of nanofiltration. *J Membr Sci* 2001;191:109–28. https://doi.org/10.1016/s0376-7388(01)00458-6.
107. Koutsou, CP, Karabelas, AJ. Shear stresses and mass transfer at the base of a stirred filtration cell and corresponding conditions in narrow channels with spacers. *J Membr Sci* 2012;299-400:60–72. https://doi.org/10.1016/j.memsci.2012.01.029.
108. Becht, NO, Malik, DJ, Tarleton, ES. Evaluation and comparison of protein ultrafiltration test results: deadend stirred cell compared with a crossflow system. *Separ Purif Technol* 2008;62:228–39. https://doi.org/10.1016/j.seppur.2008.01.030.
109. Ghosh, R, Cui, ZF. Mass transfer in gas-sparged ultrafiltration: upward slug flow in tubular membranes. *J Membr Sci* 1999;162:91–102. https://doi.org/10.1016/s0376-7388(99)00126-x.
110. Saeed, A, Vuthaluru, R, Vuthaluru, HB. Investigations into the effects of mass transport and flow dynamics of spacer filled membrane modules using CFD. *Chem Eng Res Des* 2015;93:79–99. https://doi.org/10.1016/j.cherd.2014.07.002.
111. Cath, TY, Elimelech, M, McCutcheon, JR, McGinnis, RL, Achilli, A, et al. Standard methodology for evaluating membrane performance in osmotically driven membrane processes. *Desalination* 2013;312:31–8. https://doi.org/10.1016/j.desal.2012.07.005.
112. Cui, ZF, Chang, S, Fane, AG. The use of gas bubbling to enhance membrane processes. *J Membr Sci* 2003;221:1–35. https://doi.org/10.1016/s0376-7388(03)00246-1.
113. Zhang, K, Cui, Z, Field, RW. Effect of bubble size and frequency on mass transfer in flat sheet MBR. *J Membr Sci* 2009;332:30–7. https://doi.org/10.1016/j.memsci.2009.01.033.
114. Fimbres-Weihs, BGA, Wiley, DE. Numerical study of two-dimensional multi-layer spacer designs for minimum drag and maximum mass transfer. *J Membr Sci* 2008;325:809–22. https://doi.org/10.1016/j.memsci.2008.09.005.

115. Zamani, F, Chew, JW, Akhondi, E, Krantz, WB, Fane, AG. Unsteady-state shear strategies to enhance mass-transfer for the implementation of ultrapermeable membranes in reverse osmosis: a review. *Desalination* 2015;356:328–48. https://doi.org/10.1016/j.desal.2014.10.021.
116. Liang, YY, Chapman, MB, Fimbres Weihs, GA, Wiley, n DE. CFD modelling of electro-osmotic permeate flux enhancement on the feed side of a membrane module. *J Membr Sci* 2014;470:378–88. https://doi.org/10.1016/j.memsci.2014.07.039.
117. Rodrigues, C, Rodrigues, M, Semiao, V, Geraldes, V. Enhancement of mass transfer in spacer-filled channels under laminar regime by pulsatile flow. *Chem Eng Sci* 2015;123:536–41. https://doi.org/10.1016/j.ces.2014.11.047.
118. Du, X, Wang, Y, Leslie, G, Liang, H. Shear stress in a pressure-driven membrane system and its impact on membrane fouling from a hydrodynamic condition perspective: a review. *J Chem Technol Biotechnol* 2017;92:463–78. https://doi.org/10.1002/jctb.5154.
119. Koutsou, CP, Karabelas, AJ. A novel retentate spacer geometry for improved spiral wound membrane (SWM) module performance. *J Membr Sci* 2015;488:129–42. https://doi.org/10.1016/j.memsci.2015.03.064.
120. Shakaib, M, Hasani, SMF, Mahmood, M. CFD modeling for flow and mass transfer in spacer-obstructed membrane feed channels. *J Membr Sci* 2009;326:270–84. https://doi.org/10.1016/j.memsci.2008.09.052.
121. Amokrane, M, Sadaoui, D, Koutsou, CP, Karabelas, AJ, Dudeck, M. Study of flow field and concentration polarization evolution in membrane channels with two-dimensional spacers during water desalination. *J Membr Sci* 2015;477:139–50. https://doi.org/10.1016/j.memsci.2014.11.029.
122. Jaffrin, M. Dynamic crossflow filtration. In: Hoek, EMV, Tarabara, VV, editors. *Chapter in: Encyclopedia of membrane science and technology*. New Jersey: Wiley; 2013.
123. Jaffrin, MY, Ding, LH, Akoum, O, Brou, A. A hydrodynamic comparison between rotating disc and vibratory dynamic filtration systems. *J Membr Sci* 2004;242:155–67. https://doi.org/10.1016/j.memsci.2003.07.029.

Index

https://doi.org/10.1515/9783110678215-009