Computational protein crystallography

How to get the most out of your data

Bergmann, Justin

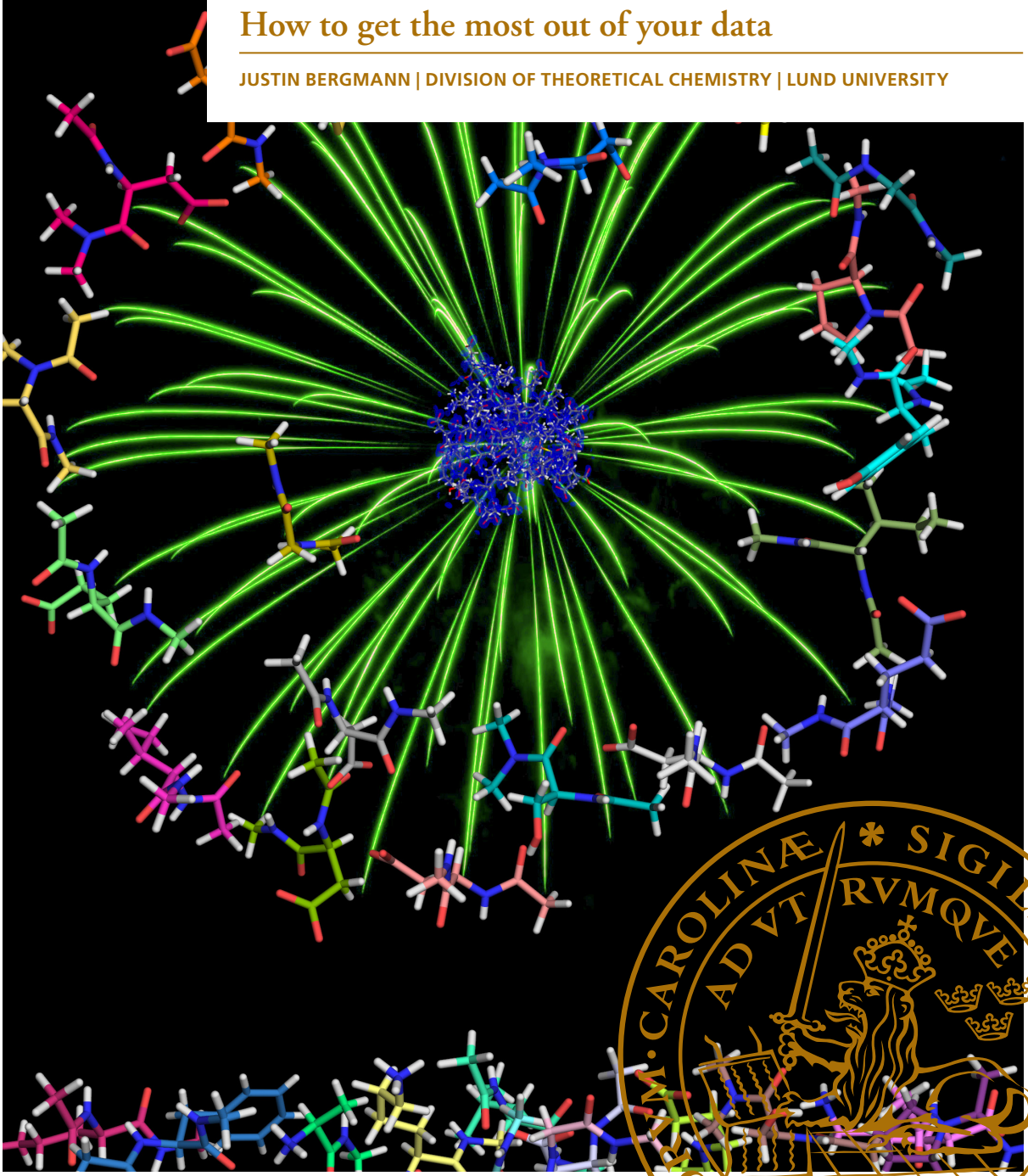
2022

Document Version:
Publisher's PDF, also known as Version of record

Computational protein crystallography

How to get the most out of your data

JUSTIN BERGMANN | DIVISION OF THEORETICAL CHEMISTRY | LUND UNIVERSITY



Computational protein crystallography

Computational protein crystallography

How to get the most out of your data

by Justin Bergmann

LUND
UNIVERSITY

Doctoral Dissertation
Thesis advisors: Prof. Ulf Ryde, Dr. Esko Oksanen
Faculty opponent: Prof. Paulina Dominiak

To be presented, with the permission of the Faculty of Science of Lund University, for public defense at the Centre for Chemistry and Chemical Engineering, Lecture Hall A, Naturvetarvägen 14, Lund, on Friday, 3rd June 2022 at 13:00.

Organization
LUND UNIVERSITY

Department of Chemistry
Box 124
SE–221 00 LUND
Sweden

Document name
DOCTORAL DISSERTATION

Date of disputation
2022-06-03

Sponsoring organization

Author(s)
Justin Bergmann

Title and subtitle
Computational protein crystallography: How to get the most out of your data

Abstract
It is important to obtain accurate three dimensional structures of molecules and proteins to understand and predict their function and behaviour. X-ray crystallography is the major technique to determine three dimensional structures of proteins. Although there have been major improvements on the experimental side in determining crystallographic data, only small progress has been made on the computational side to get a correct model and interpretation of this data.

In small-molecule crystallography, some of the shortcomings in the model have already been overcome, but in protein crystallography they still remain. Therefore, we have adapted the Hirshfeld atom refinement from small-molecule crystallography to make it available also to protein crystallography. This enables improved modelling of high-resolution protein data. To achieve this goal, we combined the molecular fractionation with conjugate caps approach with the Hirshfeld atom refinement. We call this combined method fragHAR. With fragHAR, we could perform the first Hirshfeld atom refinement of a metalloprotein.

Furthermore, we improved and applied the quantum refinement method, which employs quantum mechanics calculations to obtain a chemically and physically correct model for all parts of the protein, especially the active site. With quantum refinement, it is possible to distinguish between different interpretations of the structure, *e.g.* the elemental composition or the protonation state, even from medium-resolution crystallographic data. In this thesis, quantum refinement was improved for highly-charged systems by applying a continuum-solvent description of the surroundings in the quantum mechanics calculation. Furthermore, quantum refinement was applied to settle the nature of the unusual bidentate ligand in V-nitrogenase and the protonation state of the MoFe cluster in Mo-nitrogenase when inhibited by CO. For a recent structure of Mo-nitrogenase, we showed that there is no experimental support for the suggestion that N_2 is bound to the MoFe-cluster and presented a more likely model. We have also identified the most probable protonation states of the active site in acetylcholinesterase before and after inhibition by nerve agents. Finally, for triosephosphate isomerase we used a joint X-ray and neutron quantum refinement to investigate the hydrogen bond between an inhibitor and Lys-13.

Key words
Hirshfeld atom refinement (HAR), quantum refinement, protein crystallography, X-ray crystallography, neutron crystallography, nitrogenase, acetylcholin esterase, triose isomerase

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title

ISBN
978-91-7422-878-6 (print)
978-91-7422-879-3 (pdf)

Recipient's notes

Number of pages
171

Price

Security classification

Computational protein crystallography

How to get the most out of your data

by Justin Bergmann

LUND
UNIVERSITY

A PhD-thesis at a university in Sweden takes either the form of a single, cohesive research study (monograph) or a summary of research papers (compilation thesis), which the PhD student has written alone or together with one or several other author(s).

In the latter case the thesis consists of two parts. An introductory text puts the research work into context and summarises the main points of the papers. Then, the research publications themselves are reproduced, together with a description of the individual contributions of the authors. The research papers may either have been already published or are manuscripts at various stages (in press, submitted, or in draft).

Cover illustration front: Fragmentation of *Pyrococcus abyssi* rubredoxin by Justin Bergmann.

Cover illustration back: *Phalacrocoracidae* on a stone by Justin Bergmann.

Funding information: The thesis work was financially supported by the Swedish Research Council.

Faculty of Science, Department of Chemistry, Division of Theoretical Chemistry

ISBN: 978-91-7422-878-6 (print)
ISBN: 978-91-7422-879-3 (pdf)

Printed in Sweden by Media-Tryck, Lund University, Lund 2022

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

Printed matter 3041 0903

MADE IN SWEDEN

”A theory is something nobody believes, except the person who made it.
An experiment is something everybody believes, except the person who made it.”
Albert Einstein

Contents

List of publications

This thesis is based on the following publications, referred to by their Roman numerals:

I **fragHAR: towards ab initio quantum-crystallographic X-ray structure refinement for polypeptides and proteins**
J. Bergmann, M. Davidson, E. Oksanen, U. Ryde & D. Jayatilaka
IUCrJ, 7, 158–165, **2020**.

II **Linear-scaling aspherical refinements of proteins**
J. Bergmann, F. Kleemiss, J. Creutzberg, E. Oksanen & U. Ryde,
Manuscript

III **Can the results of quantum refinement be improved with a continuum-solvation model?**
J. Bergmann, E. Oksanen, & U. Ryde
Acta Cryst., B77(6), 906–918, **2021**.

IV **Hydrogen bonding in the active site of a triosephosphate isomerase E97Q variant studied by quantum refinement**
J. Bergmann, E. Oksanen, & U. Ryde
Manuscript

V **Quantum-refinement studies of the bidentate ligand of V-nitrogenase and the protonation state of CO-inhibited Mo-nitrogenase**
J. Bergmann, E. Oksanen, & U. Ryde
J. Inorg. Biochem., 219, 111426, **2021**.

VI **Critical evaluation of a crystal structure of nitrogenase with bound N_2 ligands**
J. Bergmann, E. Oksanen, & U. Ryde
J. Biol. Inorg. Chem., 26(2), 341–353, **2021**.

VII **Quantum-refinement study of protonation states in the active site of acetylcholinesterase**
J. Bergmann, O. Caldararu, E. Oksanen, F. Ekström, A. Linusson & U. Ryde
Manuscript

Publications not included in this thesis

VIII **Geometry and electronic structure of the P-cluster in nitrogenase studied by combined quantum mechanical and molecular mechanical calculations and quantum refinement**
L. Cao, M. C. Börner, **J. Bergmann**, O. Caldararu & U. Ryde
Inorg. Chem., 58(15), 9672–9690, **2019**.

IX **Combining crystallography with quantum mechanics**
J. Bergmann, E. Oksanen, & U. Ryde
Curr. Opin. Struct. Biol., 72, 18-26, **2022**.

X **Computational enhanced X-ray diffraction analysis of a gold(III) complex interacting with the human telomeric DNA G-quadruplex. Unravelling non-unique ligand positioning**
D. Cirri, C. Bazzicalupi, U. Ryde, **J. Bergmann**, F. Binacchi, A. Nocentini, A. Pratesi, P. Gratteri & L. Messori
submitted to *Int. J. Biol. Macromol.*

XI **Benchmark study of redox potential calculations for iron–sulfur clusters in proteins**
S. Jafari, Y.A. Tavares Santos, **J. Bergmann**, M. Irani & U. Ryde
Inorg. Chem., **2022**, DOI: 10.1021/acs.inorgchem.1c03422

Author contributions

Paper I: fragHAR: towards ab initio quantum-crystallographic X-ray structure refinement for polypeptides and proteins

I participated in developing and implementing the fragHAR approach in TONTO. I performed, analysed and evaluated all the crystallographic refinements. I participated in writing the manuscript.

Paper II: Linear-scaling aspherical refinements of proteins

I participated in developing and implementing the fragHAR approach in Olex2. I performed, analysed and evaluated the crystallographic refinements. I participated in writing the manuscript.

Paper III: Can the results of quantum refinement be improved with a continuum-solvation model?

I participated in developing the idea. I performed the crystallographic refinements, analyses and evaluation. I participated in writing the manuscript.

Paper IV: Hydrogen bonding in the active site of a triosephosphate isomerase E97Q variant studied by quantum refinement

I performed the crystallographic refinements, analyses and evaluation. I participated in writing the manuscript.

Paper V: Quantum-refinement studies of the bidentate ligand of V-nitrogenase and the protonation state of CO-inhibited Mo-nitrogenase

I performed the crystallographic refinements, analyses and evaluation. I participated in writing the manuscript.

Paper VI: Critical evaluation of a crystal structure of nitrogenase with bound N_2 ligands

I performed the crystallographic refinements, analyses and evaluation. I participated in writing the manuscript.

Paper VII: Quantum-refinement study of protonation states in the active site of acetylcholinesterase

I performed the crystallographic refinements, analyses and evaluation. I participated in writing the manuscript.

Acknowledgements

First of all I would like to thank **Ulf** for the fantastic four years of my PhD and the amazing supervision. Thank you very much for all your support, your trust in me and for giving me the freedom to work independently. Thank you for sharing all your experience with me and for enabling me to have great new experiences in Sweden, Australia and Denmark.

Next, I would like to thank **Esko** for the fantastic supervision, for answering all my questions concerning crystallography and for all the nice lunches we had together.

Thank you **Dylan** and **Julie** for the great time in Australia. You made sure I had a very pleasant stay over the three months. Thank you for all the fun adventures we had together. In this context, I would also like to thank **Max D.** and **Asja** for the warm welcome to Perth and for having a good time together.

Among all my current and former colleges, first I would like to thank **Majda** for the warm welcome to the group and for being a great office mate during my first year in Lund. Thank you **Vilhelm** for being such a pleasant office mate now. Thank you **Octav** for the introduction to quantum refinement and for all your help setting up my first project so that I could continue my work independently. Thank you **Magne**, **Hao** and **Kristoffer** for the nice atmosphere in the group even during the time when it was not possible to see each other in person. Thank you **Ernst**, **Stephanie**, **Ellen**, **Vidar**, **Victor**, **Samuel S.** and **Samuel L.**, **Eric**, **Erik** and all other members of theoretical and physical chemistry for all the nice coffee and lunch gatherings and the really enjoyable Friday "fika".

Thank you **Joel** for all the help and the fun hours in the gym, and that it is even possible to discuss quantum mechanics questions during weight lifting with you. Thank you **Dora**, **Jen**, **Iria** and **Marlisa** for being a fantastic hiking group and for all our fun activities during the weekends.

Thank you **Simon** for introducing me to HAR. Thank you **Florian** for implementing fragHAR in Olex2 with me.

I would also like to thank **Birger** for introducing me to quantum crystallography and for still giving valuable advice after all this time.

Thank you **Max S.**, **Laura**, **Marc**, **Eike**, **Dina**, **Victoria**, **Teresa**, and **Fabian** for the good time together at school and university back in Germany and for still being good friends even when I moved abroad.

A very special thank goes to my parents. Thank you **Mama** and **Papa** for all your

help and support. Without your efforts none of this would have been possible for me. Thank you for believing in me, even at times when no one else did. I would also like to say thank you to **Valerie** We have chosen some different paths in our lives but I can always count on you.

Last but not least, I would like to thank **Erna** for all your support. I am grateful that I could talk to you about everything, whether it was about work or something else. Last but not least, for always being there for me.

Abbreviations

ACh	acetylcholine
AChE	acetylcholinesterase
ADP	atomic displacement parameter
AO	atomic orbital
ATP	adenosine triphosphate
DFT	density functional theory
DHAP	dihydroxyacetone phosphate
DNA	deoxyribonucleic acid
ED	electron density
ELMO	extremely localized molecular orbital
GAP	D-glyceraldehyde 3-phosphate
GGA	generalised gradient approximation
GTO	Gaussian-type orbital
HAR	Hirshfeld atom refinement
HAR-ELMO	HAR combined with ELMO libraries
HF	Hartree–Fock
IAM	independent atom model
IUCr	International Union of Crystallography
KS	Kohn–Sham
LCAO	linear combination of atomic orbitals
LDA	local density approximation
LSCF	local self-consistent field
MAD	multiple-wavelength anomalous dispersion
MFCC	molecular fractionation with conjugate caps
MIR	multiple isomorphous replacement
MM	molecular mechanics
MO	molecular orbital
MR	molecular replacement
NMR	nuclear magnetic resonance
ONIOM	our own N-layer integrated molecular orbital molecular mechanics
PDB	protein data bank
PGA	2-phosphoglycolate
pMMO	particulate methane monooxygenase
RSCC	real-space correlation coefficient
RSR	real-space R factor
RSZD	real-space difference density Z score
SAD	single-wavelength anomalous dispersion
SCF	self-consistent field

TIM	triosephosphate isomerase
vdW	van der Waals
QM	quantum mechanics
QM/MM	quantum mechanics combined with molecular mechanics

Popular science summary

Chemical reactions are crucial for life on earth. For instance, metabolism and signalling in all organisms critically depend on chemical reactions. Of special interest are reactions that do not occur spontaneously. Over millions of years of evolution, nature found ways to make even these reactions feasible. To achieve this, nature uses large and complex molecules that are called proteins.

To understand how nature enables reactions, it is essential to study the proteins involved. One way to understand how proteins function is to look at their structures. To exemplify how the function depends on the structure, we can use a metaphoric fork. The function of a fork clearly depends on its form. Such a strong correlation between structure and function exist not only on a macroscopic level but also on an atomic level. An analogy to the fork can be drawn on an atomic scale for proteins and other molecules. Unfortunately, obtaining the structure of a molecule is not as straightforward as looking at a fork.

To obtain the three dimensional structure of a molecule, X-ray crystallography is the method of choice. For such an experiment, it is crucial to bring the molecule of interest into a crystalline form. This means generating a three dimensional periodic "solid" of the molecule. Techniques for crystallisation have undergone huge improvements during the last century and it is nowadays possible to crystallise entire proteins of increasing complexity. Once the crystal is generated, we can place it into a diffractometer. By shining X-rays on the protein crystal, the diffractometer acts like a super sensitive microscope, making it possible to visualise the structure. Unfortunately, it is not possible to obtain a direct image of the molecule in the crystal because some information of the picture is lost during the measurement. A macroscopic parable would be that we could only see the shadow of the fork but not the fork itself. By taking many "pictures" of this "shadow" and building a model that can reproduce this shadow we can obtain a picture of the object of interest.

While doing this for a molecule, we may encounter two specific problems that both relate to the resolution of the measurement. The first one is that we measure really good (high-resolution) data, but we are using a less sophisticated model to describe the data. Then, our data show more features of the molecule than we can represent in our model and an insufficient model of the structure is obtained. A technical example would be that we have a movie in 4K but only a TV from 1950 to show it. The solution in this case is to buy a new TV or, in case of the molecule, to use a more sophisticated model.

The other and unfortunately more common problem is that only insufficient (low-resolution) data can be measured. This would be equivalent to obtaining only a

blurred picture. If the picture of a fork is too blurry, it is indistinguishable from a blurred spoon. To still obtain a correct model it is possible to use prior knowledge in the model building. For example, we know that the fork has prongs and might include this information in the model.

Fortunately, we can not only measure structural properties of molecules, but we can also calculate them. However, in order to do this, we need to solve a complex mathematical problem (the Schrödinger equation). Nowadays, this can be done with help of modern computers. With these calculated properties we can improve our model for both problematic cases that were described above.

However, the problem is that these calculations become computationally prohibitive on today's computational resources for large systems, like proteins. Therefore, we applied two approaches to improve the model quality of crystallographic data using theoretical calculations.

To model high-resolution data (a 4K image), we need to do a calculation for the whole protein. Therefore, we split the system into smaller fragments and calculate the property of interest for each fragment. Then, we can obtain the property for the whole protein by recombining the properties from all fragments. This results in a description that is nearly as accurate as we would get from a single calculation of the whole system, but in a fraction of the time and computational cost. In the context of this thesis, we implemented such a fragmentation approach to obtain an advanced description of protein structures. We thoroughly tested this new method on small molecules and then applied it to proteins. In particular, we obtained the first structural model of a metalloprotein (*i.e.* a protein that contains a metal atom) with this approach. Thereby, we showed that this new strategy has benefits for the protein crystallography community.

In the other case, where we only have low-resolution data (the blurry picture), we are often especially interested in a small part of the protein, *e.g.* the active site. Then, we can divide the system in two regions: one part of special interest and one part of lower interest. Then, the latter part is treated with standard (cheap) crystallographic methods, whereas the part of special interest is treated with a highly accurate and more expensive method. This combination of different computational methods with the crystallographic data provides a deeper insight into the structure of the protein and therefore also into its function.

In this thesis, three proteins were in focus. The first is nitrogenase, which makes nitrogen bioavailable. Nitrogen is essential for all growing organism. On an industrial scale, we use the Haber–Bosch process, which has a huge energy consumption but feeds the world with fertilisers. The second protein is triosephosphate isomerase, which is involved in the abstraction of energy from sugar in organisms. The last pro-

tein is acetylcholinesterase, which is crucial for neuronal information transmission to muscles, to trigger proper movements of the muscles.

For these three proteins, we were able to clarify exactly what is seen in crystal structures, e.g. what atom types are involved (carbon or nitrogen) and the positions the atoms in the region of interest (we sharpened the picture so that we can distinguish between the fork and the spoon). These findings help to better understand the reactions performed by the three proteins.

In conclusion, with these new methods it is possible to use extract information from the experiment and to obtain more accurate structures of the systems of interest. This makes more complicated or even not feasible alternative experiments partially obsolete. With these benefits, the research process is sped up and we obtain more reliable information about the proteins. Therefore, we can better explain the functions of proteins, which brings us one step closer to understanding life on earth.

Introduction

In nature, a strong correlation between structure and function exists on a macroscopic scale, but also on an atomic scale [1–3]. For example, in enzymes, small modifications in the structure can have significant impact on the catalytic function [4, 5]. Therefore, it is necessary to know the atomic structure to understand how it functions.

Structure determination of molecules has become an important research area during the last century [6–8]. The main technique to obtain the three-dimensional structure of a protein is X-ray crystallography, although nuclear magnetic resonance (NMR), cryo-electron-microscopy, neutron crystallography and purely theoretical methods have also been used [9].

The fundamentals of X-ray crystallography were established already in 1913 by Bragg & Bragg [10] and in 1957, the first three dimensional structure of a protein was determined [6–8, 11]. Nowadays, structure determination with X-ray crystallography is a standard method and more than 100 000 protein structures [9] and over 1 million small-molecule structures [12] were determined with this technique.

To obtain a structure, a crystal of the protein is needed [13]. In the crystal, the protein is in a three dimensional periodic order. If this order is sufficient, a diffraction pattern can be observed when X-ray radiation is shined on the crystal [14]. This diffraction pattern is in a causal connection to the electron density of the molecule, but parts of the information about the electron density is not measurable (the phase problem) [15, 16]. Therefore, it is necessary to build a model, to calculate the theoretical diffraction pattern and to refine the model until the best fit to the measured data is obtained. The standard way to model X-ray crystallographic data is the independent atom model (IAM), in which each atom is treated as an individual spherical scatterer without interactions to its neighbouring atoms [17]. During the refinement, it is important to ensure that a chemically and physically meaningful model is obtained, even if this means that parts of the data are not explained by the model, to avoid overfitting and modelling of noise.

During this refinement process, three cases can occur. In the first case, the standard model is sufficient and the data are good enough that no external information and no modification of the model is needed. This is unfortunately unlikely for proteins.

In the second case, an insufficient amount of data are observed and the model has to be supported by additional knowledge, in form of restraints and constraints to ensure that a chemically and physically reasonable model is obtained [18]. This is usually the case for proteins. For standard amino acids in proteins, the parameters for restraints are well-defined from statistical analyses of high-resolution structures [19]. Unfortunately, for the active site, which often contains unusual ligands and metal clusters, the available empirical restraints are insufficient due to a lack of high-resolution data for these cases. To overcome this problem, the toolbox of quantum chemistry can be used to generate restraints for these more complicated parts of the protein [20]. In 2002, Ryde *et al.* developed a quantum mechanics combined with molecular mechanics (QM/MM)-like approach that is called quantum refinement [20, 21]. This approach is used in this thesis. In quantum refinement, restraints derived from a quantum mechanics (QM) calculation are used for the part of special interest. For the rest of the protein, empirical restraints are used. With this method, it is often possible to distinguish between different protonation states and element types, even for low-resolution data [22].

In the third case, ultrahigh-resolution data are obtained and it is possible to see chemically and physically meaningful features, which are not included in the standard model [23–25]. Due to improved infrastructure, this also occurs for some proteins. Therefore, it is necessary to replace the standard model of spherical electron densities to describe an atom with a more realistic aspherical electron density distribution that can be observed in molecules. This is mainly realised through two different refinement methods.

The first one is multipole refinement, in which multipoles are used to fit the aspherical electron density of an atom in a molecule [25, 26]. This method introduces more parameters to the refinement and is sensitive to overfitting. To overcome these problems, several multipole databases [27–31] have been assembled. Another method that does not have these problems is the Hirshfeld atom refinement (HAR) [32, 33]. In HAR, a wavefunction is calculated with standard QM methods and is partitioned into aspherical atomic fragments. From these, new aspherical atomic form factors are calculated, which can be used in the refinement without the introduction of additional parameters. The problem in this case is that wavefunction calculations scale with a power of greater than two with regards to the system size. With current computational software and hardware, it is very demanding to compute the wavefunction for a complete protein [34]. In this thesis, we developed an approach to make HAR feasible for proteins. To this end, a fragmentation approach for the QM calculations is used, in which the wavefunction is calculated only for single amino acids [35]. The atomic form factors are calculated individually from these fragment wavefunctions [36]. Since the computation of a wavefunction for the complete system is not neces-

sary, nearly linear scaling of the computational cost with increasing system size can be achieved.

Computational chemistry

2.1 Quantum chemistry

The electronic structure of a chemical system can be deduced by solving the time-independent Schrödinger equation (equation 2.1) [37]:

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

where E is the energy, Ψ the wavefunction and $\hat{H}$ is the Hamiltonian of the system. The Schrödinger equation represents an eigenvalue problem, where the eigenvalues are energies and the eigenfunctions are wavefunctions. This equation is too complicated to be solved analytically for all except very few simple systems. Instead, it can be solved numerically using some approximations. First, the Born–Oppenheimer approximation [38] is introduced, in which the electron movement is separated from the movement of the nuclei. This is a good approximation because the electrons are $\sim$2000 times lighter than the nuclei and move significantly faster. Therefore, it is possible to obtain two Hamiltonians, one for the nuclei and one for the electrons, which can be treated separately. Chemists are mostly interested in the electronic Hamiltonian. Furthermore, relativistic effects are neglected (to include these, the Dirac equation has to be solved). This is a good approximation for light elements but relativistic effects need to be taken into account for heavy elements [39]. Even with these approximations, the Schrödinger equation is still too complicated to be solved if more than one electron has to be described. Several further approximations are necessary for many-electron systems.

One approach to solve the Schrödinger equation is the Hartree–Fock (HF) method. This method [40–43] is a mean-field theory, in which each electron moves in the average field of all other electrons. Therefore, the complicated many-electron problem is reduced to an approximate one-electron problem that is expressed by the HF equation (equation 2.2):

$$\hat{f}(\boldsymbol{r})\psi_i(\boldsymbol{r}) = \epsilon_i\psi_i(\boldsymbol{r}) \tag{2.2}$$

where $\hat{f}$ is the one-electron Fock operator, which depends on the averaged field of all other electrons, ψ_i is the one-electron wavefunction, also called molecular orbital, and ϵ_i is the orbital energy.

The wavefunction in the HF method is described by a single Slater determinant [44], which is build out of molecular orbitals (MOs). These are the one-electron functions (ψ_i) in equation 2.2. These MOs are obtained by the linear combination of atomic orbitals (LCAO) approach (equation 2.3) [39]:

$$\psi_i(\boldsymbol{r}) = \sum_j c_{ij}\chi_j(\boldsymbol{r}) \tag{2.3}$$

where c_{ij} are expansion coefficients and χ_j are basis functions. A basis set is a collection of basis functions, which have similar shapes as atomic orbitals (AOs) [45]. They can be build out of Slater functions or Gaussian functions. Due to computational efficiency Gaussian functions are most frequently used, which results in Gaussian-type orbitals (GTOs). An infinite basis set would give a perfect representation of the MOs, but significantly smaller basis sets arenormally used, giving a better computational efficiency [39].

As previously mentioned, the Fock operator in equation 2.2 depends on all electrons in the system. To compute it, an initial guess is required for the coefficients (*e.g.* using extended Hückel theory). Then the Fock equation can be solved, which provides a new set of coefficients c_{ij}. However, the initial guess orbitals are not a perfect representation for the system and also the newly obtained coefficients have to be optimised. It can be shown that the best wavefunction gives the lowest energy (the variational principle). To obtain the best wavefunction, the results from optimising the coefficients can be used as new initial guess. This can be done iteratively until the energy converges, to obtain the best set of orbital coefficients for the system. This is called the self-consistent field (SCF) approach [39].

2.1.1 Density functional theory

The Hartree–Fock method neglects the instantaneous repulsion between the various electrons (electron correlation), which is a small contribution to the total energy of the system but really important to describe for example reaction energies. To take this into account, density functional theory (DFT) can be applied. DFT is grounded on the concept, that the ground-state energy of a system is uniquely determined by its electron density. This is the first Hohenberg–Kohn theorem [46]. This relates also to the second theorem, which states that the energy has a minimum for the correct electron density [46].

The advantage of DFT is that the density depends only on the three spatial coordinates, whereas wavefunctions depend on the $3N$ spatial and N spin coordinates for N electrons [47]. Unfortunately, the exact relation between energy and electron density is unknown. Therefore, the Kohn–Sham (KS)-DFT approximations are

normally applied. By introducing a non-interacting system that has the same density as the real system and by introducing orbitals, the energy functional can be divided into four terms, the non-interacting kinetic-energy functional ($T_s[\rho]$) [48], the nucleus–electron interaction ($v_{ne}(\mathbf{r})$), the Hartree energy functional ($E_H[\rho]$) and the exchange–correlation functional ($E_{ex}[\rho]$).

$$E[\rho] = T_s[\rho] + \int v_{ne}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + E_H[\rho] + E_{ex}[\rho] \tag{2.4}$$

The only term in equation 2.4 that is not exactly determinable is the exchange–correlation functional. To approximate this functional, several sophisticated approaches have been used. The easiest approximation is the local density approximation (LDA), where the exchange–correlation energy depends only on the density [48]. A more accurate approximation is the generalised gradient approximation (GGA), in which the gradient of the density is also included [49]. An even more sophisticated approach is the meta-generalised gradient approximation, in which also the second derivative of the density is taken into account [50]. In this thesis, we have mainly used the meta-GGA functional TPSS [51].

Unfortunately, DFT does not use the exact exchange energy. To address this problem, hybrid functionals were developed. In these, a certain amount the exact exchange from the HF method is added. One famous example is the B3LYP hybrid functional [52–55].

DFT and HF have in common that they cannot treat non-bonded dispersion. Therefore, empirical corrections for the dispersion have been developed [56, 57].

2.2 Molecular fractionation with conjugate caps

Most QM methods have a poor scaling of the computational cost with the number of atoms or basis functions. This makes these methods unfeasible for large systems with several hundreds or thousands of atoms [58–60]. One approach to solve this problem is to split the system into many smaller fragments [61–64]. QM calculations are then performed on each of these fragments and the results are recombined afterwards.

For large molecules, such a fragmentation approach faces the problem that it becomes necessary to break chemical bonds, which may lead to truncation problems if this is not properly handled. A simple way to address this problem is to cut only single bonds and saturate them with a hydrogen atom [65, 66].

Many properties are long-ranged and one hydrogen may not be sufficient to mimic the local environments. Then, it may be wise to include neighbouring and next neigh-

bouring atoms in the fragments, truncating with larger groups. These added atoms are called caps [35].

In the molecular fractionation with conjugate caps (MFCC) approach, these caps are recombined for each clipped bond and treated in a separate calculation. For global properties, *e.g.* for energies, the property of each fragment is summed up and the property of the recombined caps are subtracted afterwards (see figure 2.1) [35].

This leads to a linear or nearly linear scaling of the computational cost for systems if the fragment size is independent of the system size.

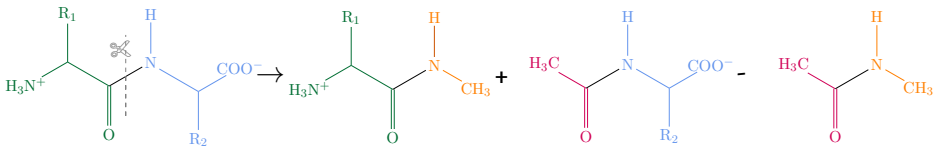


Figure 2.1: Example illustrating the MFCC procedure for cutting a dipeptide (left hand side) across the peptide bond (shown in black) producing two fragment molecules (right hand side) which are capped with $-COCH_3$ (red) and $-NHCH_3$ (orange) groups, comprised of the neighbour and next-neighbour non-hydrogen atoms and the recombination of these two caps.

2.3 The QM/MM approach

If only a small region of a system is of central interest, it is possible to describe different regions with different accuracy. In 1976, Warshel and Levitt [67] suggested an approach where the region of interest (also called system 1 or QM region) is described by a QM method and the surrounding by a cheaper [68] molecular mechanics (MM) method. This combination of the two methods is called the QM/MM approach [67, 69].

There are many variants of QM/MM methods. Two general approaches are common [69, 70]. The first one is an additive scheme where the total energy ($E_{\mathrm{QM/MM}}$) is build up from the MM energy for the surroundings (E_{MM2}), the QM of system 1 (E_{QM1}) and a coupling energy ($E_{\mathrm{QM1/MM12}}$) between the two subsystems (equation 2.5).

$$E_{\mathrm{QM/MM}} = E_{\mathrm{MM2}} + E_{\mathrm{QM1}} + E_{\mathrm{QM1/MM2}} \tag{2.5}$$

The additive approach requires special MM software [70], allowing the user to pick exactly what energy terms to include.

In the subtractive approach, a MM energy for the whole system (E_{MM12}) is calculated and the QM energy (E_{QM1}) is added. To avoid double-counting, a MM calculation is

also performed on system 1 (E_{MM1}) and this energy is subtracted from the combined MM and QM energies (equation 2.6).

$$E_{\mathrm{QM/MM}} = E_{\mathrm{MM12}} + E_{\mathrm{QM1}} - E_{\mathrm{MM1}} \tag{2.6}$$

For the subtractive approach no special QM or MM software is needed [70].

If it is not possible to build the QM region without cutting covalent bonds, it is important to minimise truncation errors [70, 71]. One possibility is to saturate the cut bonds with a hydrogen atoms. The atoms are added to the QM part at a standard distance to the remaining atom along the cut bond [70, 71].

An alternative approach is to place localised orbitals on the frontier bonds and keep these frozen during the SCF calculation, which is done for example in the local self-consistent field (LSCF) method [70, 71].

The subtractive scheme can be extended to several layers, as is done in the ONIOM (our own N-layer integrated molecular orbital molecular mechanics) approach [72, 73]. The software ComQum [74, 75] employs a subtractive QM/MM scheme and it has been adapted for quantum refinement (see section 4.2) which is used in this thesis [20, 21].

2.4 Continuum-solvent embedding

Usual QM calculations are performed in vacuum on a single system of interest. This does not provide an adequate description of the surroundings for chemical and physical processes in condensed phase. Different approaches have been developed to let the surroundings of the QM system also influence the QM calculation. One possibility to do this is to apply a continuum-solvent model [76]. In a continuum-solvent embedding, the surrounding is described implicitly, by placing the system of interest in a cavity surrounded by a dielectric continuum, described by a dielectric constant (relative permittivity, ε) [77].

The polarising dielectric continuum creates a surface charge distribution ($\sigma(\mathbf{r})$) on the interacting surface to counteract the charge distribution in the system of interest. The surface charge distribution or screening charge densities ($\sigma(\boldsymbol{r})$, see equation 2.7) depends on the chosen dielectric constant (ε), the surface normal vector ($\boldsymbol{n}(\boldsymbol{r})$) and $\boldsymbol{E}^{-}(\boldsymbol{r})$, which contributes to the electric field [78].

$$4\pi\varepsilon\sigma(\boldsymbol{r}) = (\varepsilon - 1)\boldsymbol{n}(\boldsymbol{r})\boldsymbol{E}^{-}(\boldsymbol{r}) \tag{2.7}$$

Unfortunately, equation 2.7 can not be solved analytically. Therefore, the surface is partitioned in small fragments with uniform charge in each fragment and it is solved with numerical methods in an iterative manner [77, 79].

Crystallography

The aim of crystallography is to obtain a three dimensional structural model by exposing a crystal of the investigated substance to short-wavelength radiation (in the order of Å). The importance of X-ray crystallography in obtaining protein structures is illustrated by the fact that ∼87% of the structures in the protein data bank (PDB) [9] are obtained by X-ray crystallography (January 2022 [9]). Crystallographic structure determination involves six steps: Protein purification, crystallisation, data collection (the diffraction experiment), data processing, phasing and model refinement. In this thesis, only published structures were used and therefore the first experimental steps are not further discussed. Instead, the aim of this thesis is to improve structures, for which refined model structures are already available.

3.1 Fundamental concepts

To understand the process of structure determination, it is important to understand what a crystal actually is. In 1992, the International Union of Crystallography (IUCr) updated the definition of a crystal. The new definition is "A material is a crystal if it has essentially a sharp diffraction pattern" [80–82] if it is exposed to radiation with a wavelength in the order of Å. To picture a crystal, the old definition is more descriptive. Before 1992, a crystal was defined as a solid, built out of three-dimensional periodically ordered atoms or molecules. The volume that is periodically repeated is called the unit cell. It has the dimensions a, b and c for the lengths in each direction and the angles α, β and γ between the sides.

To understand how to get the actual structure of a compound, it helps to look at both definitions. To understand diffraction, it is important to know that X-rays are scattered by the electrons and neutrons by the nuclei in the crystal. If these scatterers build up from layers with a distance (d) similar to the wavelength of the used radiation, under specific circumstances, constructive interference occurs (see figure 3.1). This relationship was discovered by William H. and William L. Bragg [10] and is now known as Bragg's law (equation 3.1).

$$n\lambda = 2d \sin\theta \tag{3.1}$$

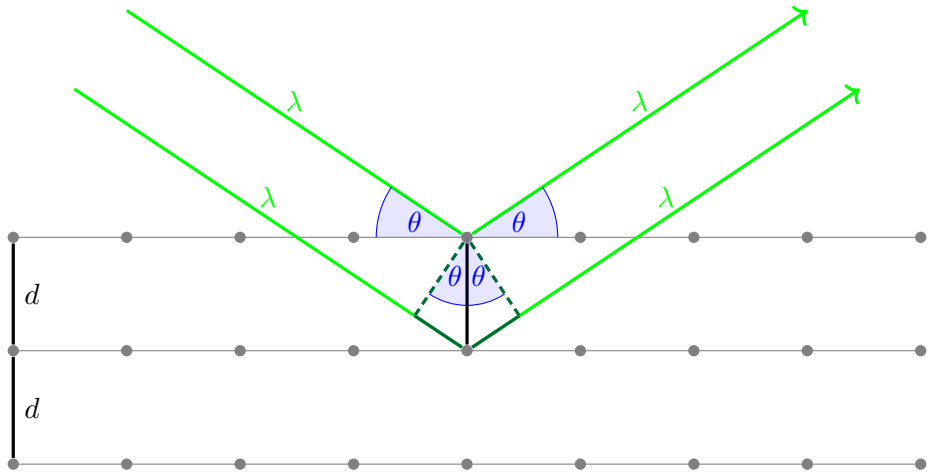


Figure 3.1: Schematic representation of Bragg's law.

According to Bragg's law, if the distance between two planes of scatterers times the sinus of the angle of the incoming wave is equal to a multiple of half of the wavelength, constructive interference occurs (see figure 3.1). This is observed as reflections on the detector. The orientation of the lattice planes is given in fractions of the unit cell parameters and are called Miller indices ($h = \frac{1}{a}$, $k = \frac{1}{b}$ and $l = \frac{1}{c}$). Each reflection is associated with one triad of Miller indices, hkl.

3.2 The structure factor

Owing to the diffraction, parts of the radiation beam change their direction. By placing the detector around the crystal these parts of the beam can be detected and the diffraction angle can be obtained. The position of the reflections depends on the positions of the atoms in the crystal, but the intensity (I_{obs}) of these reflections depends on the atom type and the local displacement of the atom (see equation 3.2).

$$I_{obs}(\mathbf{H}) \propto |F_{obs}(\mathbf{H})|^2 \tag{3.2}$$

where $\mathbf{H}$ is related to the Miller indices h, k and l. The structure factor (F) is the sum of all scatterers in the crystal (equation 3.3) [25, 26].

$$F = \sum_{j=1}^{N} f_j(\mathbf{H}) e^{2\pi i \mathbf{H}\cdot\mathbf{r}_j} T_j(\mathbf{H}, \mathbf{U}_j) \tag{3.3}$$

The structure factor (F) depends on the atomic form factor (f_j), the temperature factor $T_j(\mathbf{H}, \mathbf{U}_j)$ and the phase ($2\pi i\mathbf{H}$). The atomic form factor (f_j) depends on the element (see section 3.2.1). It is common to use tabulated atomic form factors for each element type. In more advanced refinements, individual atomic form factors are used for each atom. $T_j(\mathbf{H}, \mathbf{U}_j)$ takes into account that atoms are moving (vibrating) inside the crystal (see section 3.2.2). This movement limits the scattering power of the atom. It is important to note that the structure factor is a complex quantity that has a real part, which can be derived from the measured intensity (equation 3.2), and an imaginary part (the phase), which cannot be experimentally observed in a diffraction experiment.

To overcome this phase problem, it is mandatory to build a model, in which each atom (j) is assigned an atomic form factor and a displacement parameter [83]. With equation 3.3, it is possible to calculate the structure factor for each set of Miller indices (hkl) and refine the model to obtain an as good as possible fit to the experimental data (see section 3.4). The intensity of each reflection depends on all atoms in the unit cell of the crystal and therefore an as complete as possible model is needed to get a good representation of the experiment.

3.2.1 The atomic form factor

The atomic form factor (f_j) describes the scattering power for each atom (j) in the crystal [84]. Nuclei act as point scatterers and therefore the neutron scattering length has the form of a Dirac delta function, which is independent of the scattering angle [85]. More details about neutron crystallography are given in section 3.5. For X-rays, the atomic form factor depends on the electron density (ED) and is given by equation 3.4.

$$f_i(\mathbf{H}) = \int \rho_j(\mathbf{r}) e^{i\mathbf{H}\cdot\mathbf{r}} d\mathbf{r} \tag{3.4}$$

The atomic form factor does not only depend on the atomic ED (ρ) but also on the scattering angle. This is caused by the finite volume of an atom compared to a point scatterer [13]. For higher diffraction angles, the actual density has a larger offset from the point model. This offset results in a larger out-of-phase part of the scattering and in a reduced scattering power for the higher diffraction angles (seen in figure 3.2).

To obtain the atomic form factors, QM calculations can be performed (see section 2.1) and the calculated density can be transformed into the atomic form factor [83, 86]. For the standard IAM [17], Hartree–Fock calculations for light atoms and relativistic calculations for heavy atoms are used. The calculated density is afterwards spherically averaged [84] and normalised to the number of electrons for each element. Due to

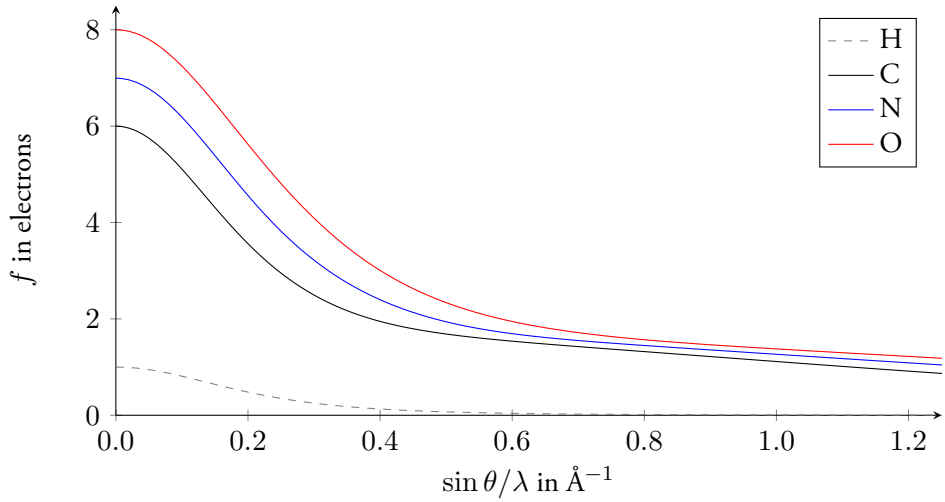


Figure 3.2: Atomic form factors for selected light elements.

errors in the experiments, it is often sufficient to take atomic form factors for the isolated neutral atom. The atomic form factors are then stored in tables [86] or directly in the program of use. However, for high-resolution data, notable shortcomings of the IAM can be observed and more sophisticated models are needed (see section 4.1).

Even with a perfect crystal (which does not exist) the dependence of the atomic form factor on the scattering angle limits the highest resolution that is possible to observe in a measurement. As seen in figure 3.2, hydrogen atoms have the weakest contribution to the overall scattering power and diffract X-rays only at low resolution. This explains why hydrogen atoms are hard to observe by X-ray crystallography. To see the small contribution of each hydrogen atom an overall good data set is needed [87].

3.2.2 Atomic displacement parameters

Atoms in a crystal are not fixed to their position in the lattice, but vibrate around their central position even at low temperatures. To take this movement into account, the displacement factor (the B-factor or, in small-molecule crystallography, the atomic displacement parameter (ADP)) is included in the calculation of the structure factor. The easiest way to describe it is an isotropic harmonic vibration around the central position. This vibration results in a spherical movement (equal in all three dimensions). For such an isotropic description, only a single displacement factor (U) is needed.

$$f' = f * e^{-8\pi^2 U \frac{sin^2\Theta}{\lambda^2}} \tag{3.5}$$

The factor $8\pi^2$ is normally combined with the displacement parameter (U) to get the Debye–Waller factor (B)

$$f' = f * e^{-B\frac{sin^2\Theta}{\lambda^2}} \tag{3.6}$$

If a more detailed description is needed, anisotropic atomic displacement parameters can be used, where U is described by a 3×3 tensor, consisting of six refined parameters, the diagonal elements U_{11}, U_{22} and U_{33} for the main axes and the off-diagonal elements U_{12}, U_{13} and U_{23}. Consequently, for an anisotropic description, five additional parameters are needed compared to the isotropic description of an atom.

3.3 The phase problem

In a diffraction experiment, unfortunately only the amplitude of the scatted wave ($|F(\mathbf{H})|^2$) can be observed and the phase ($\phi(\mathbf{H})$), which is needed to calculate the electron density (see equation 3.7), is lost.

$$\rho(\mathbf{r}) = \frac{1}{V}\sum_{hkl}|F(\mathbf{H})|e^{i\phi(\mathbf{H})}e^{-2\pi i\mathbf{H}\cdot\mathbf{r}} \tag{3.7}$$

For small molecules, direct methods are available to get an initial guess of the phases [13, 88–91]. In direct methods, some assumptions are made. The first is that the electron density at the atomic positions is always positive. The second assumption is that the scattering event only occurs at discrete atoms [16]. Furthermore, it is important that the electron density peaks for specific atoms are sufficiently larger than the average electron density so that they are clearly separated from each other. For low-resolution structures, application of direct methods is not possible, because the electron density cannot be obtained as discrete atoms anymore [15, 16, 92, 93].

Therefore, other methods have been developed to solve the structure of proteins. In molecular replacement (MR) a previously known similar structure is used to get an initial guess of the phases for the data [94, 95]. In multiple isomorphous replacement (MIR), heavy atoms are introduced to the structure before the measurement. Heavy atoms can be introduced by soaking the crystal in a solution with heavy metals, which bind to the protein at specific positions. These heavy atoms have well-separated and high electron density maxima above the average electron density and direct methods are therefore applicable. They can be used to obtain the positions of the heavy atoms and to calculate the phases afterwards. MIR has the disadvantage that it can cause changes to structure of the protein [96] and, even worse, the unit cell parameters can change, which significantly deteriorates the quality of the

phases. More novel methods are multiple-wavelength anomalous dispersion (MAD) and single-wavelength anomalous dispersion (SAD), where anomalous scattering of a native or foreign heavy atom is used as in MIR [97, 98]. Anomalous scattering is the scattering that does not follow Friedel's law [13].The anomalous scattering is increased if the wavelength is close to the absorption edge of the element [99].

3.4 Refinement

3.4.1 Small-molecule refinement

After solving the phase problem (see section 3.3) an initial model for the structure is obtained. This model covers only parts of the atoms in the unit cell. Therefore, this model needs to be improved and refined, by correcting initially mismatched elements, adding missing atoms and refining the positions and ADPs. Taking all this into account, a least-squares refinement between F_{obs} and F_{calc} can be performed. In this procedure, Q in equation 3.8 is minimised.

$$Q = \sum_{hkl} w(F_{obs}^2 - F_{calc}^2)^2 \tag{3.8}$$

Here, w takes the varying precision of individual reflections into account. After iterating this process, a final model is obtained with an optimised agreement to the experimental data.

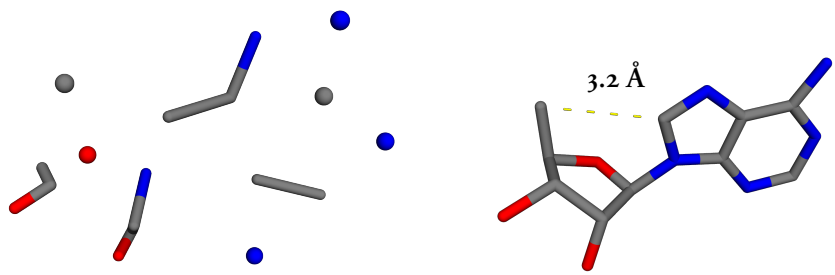
3.4.2 Protein refinement

In standard crystallographic refinement, the model (coordinates, ADPs, occupancies, etc.) is optimised by minimising the difference between structure factors observed experimentally and calculated from the model. Owing to the limited resolution of protein crystal structures, it is normally necessary to introduce restraints in the crystallographic refinement to ensure that the structure makes chemical sense (see figure 3.3) [18]. Restraints are previous knowledge, for example the expected values of chemical bonds and angles, which is given as known parameters to the refinement. Restraints could be understood as additional observations and increase the ratio between data and parameters. To avoid overfitting and to guarantee a stable refinement, a data-to-parameter greater than eight is recommended by the IUCr [100]. Restraints are usually derived from high-resolution structures [19] and in the language of computational chemistry, they represent a molecular-mechanics (MM) force field (see below).

Therefore, the refinement optimises an energy function with the form of equation 3.9.

$$E_{\text{cryst}} = w_{\text{A}} E_{\text{Xray}} + E_{\text{MM}} \tag{3.9}$$

Here, E_{Xray} is the crystallographic goodness-of-fit criterion, typically a maximum-likelihood function [101, 102], E_{MM} is the term for the empirical restraints and w_{A} is a weight factor determining the relative importance of the two terms.



(a) (b)

Figure 3.3: Importance of the empirical restraints for a group with poor electron density: The adenosyl group in methylmalonyl coenzyme A mutase at 2.2 Å resolution, refined without (a) and with (b) empirical restraints [22]. (Taken from reference [22], reprinted with permission from Elsevier.)

The empirical restraints typically contain terms for bond distances (E_{bond}), bond angles (E_{angle}), torsion angles ($E_{torsion}$), improper dihedral angles (E_{imp}) and van der Waals (vdW) interactions (E_{vdW}).

$$E_{MM} = E_{bond} + E_{angle} + E_{torsion} + E_{imp} + E_{vdW} \tag{3.10}$$

They are accurate for protein residues and nucleic acids, because much experimental data are available for these, stored in standard libraries [103, 104]. On the contrary, they are less accurate for non-standard residues, which can be better described using quantum refinement (see section 4.2). It is also possible to introduce other restraints, for example for the ADPs, which have to be similar for all atoms in one group. This is a reasonable assumption because covalently bonded atoms vibrate in a similar way and completely different vibration patterns for bonded atoms are not physically meaningful.

One advantage of the separate terms for each property in equation 3.10 is that it is possible to refine the target functions for bonds, angles, torsion angles etc. one by one and not all of them at the same time. This reduces the complexity of the refinement target. Such a strategy is for example implemented in the Phenix program [105].

In protein crystals 27–65% of the unit cell volume is solvent [106]. Therefore, it is not sufficient to consider only the protein. However, the solvent is highly disordered (not crystalline) and has to be treated separately. To compute the contribution of the

solvent to the structure factor, the following equations can be applied.

$$F_{calc} = F_{model} + F_{bulk} \tag{3.11}$$

$$F_{bulk} = k_{sol} e^{\left(-\frac{B_{sol} s^2}{4}\right)} F_{mask} \tag{3.12}$$

To calculate F_{bulk}, known bulk solvent parameters [107] are used (k_{sol} and B_{sol}, $s^2 = h'Gh$, where G is the reciprocal-space metric tensor, h is a column vector of the Miller indices and h' its transpose). The last parameter F_{mask} is the solvent mask, which is one at the disordered solvent region and zero at protein or discrete solvent molecules [107].

3.5 Neutron crystallography

The fundamental concepts of crystallography are independent of the type of radiation that is used, as long as the wavelength is similar to the atomic distances in the crystal. X-rays interact with the electron density. Therefore, it is difficult to discern hydrogen atoms (which contain only a single electron) and distinguish between neighbouring elements in the periodic table. Furthermore, it is impossible to distinguish between different isotopes [108]. In contrast, neutrons interact with the nuclei of the atoms, which leads to a direct correlation between the nuclear scattering-length density maps and the actual atom positions.

To handle neutron crystallographic data, some adjustments of the previously described procedures for X-ray data are needed. To calculate the structure factor for neutron diffraction data, it is necessary to replace the atomic form factor in equation 3.3 with the neutron scattering length, $\bar{b}_j$ [25, 86], which results in equation 3.13.

$$F_{calc} = \sum_{j=1}^{N} \bar{b}_j(\mathbf{H}) e^{2\pi i \mathbf{H}\cdot\mathbf{r}_j} T_j(\mathbf{H}, \mathbf{U}_j) \tag{3.13}$$

The neutron scattering lengths depend on the element and the isotope. They cannot be theoretically calculated, but have to be obtained by experiments [25, 26].

As seen in figure 3.4, hydrogen (^{1}H) has a negative scattering length, which can lead to cancellation of peaks if a hydrogen atom is close to an atom with a positive scattering length. Furthermore, it also has a high incoherent scattering contribution (see table 3.1), which contributes only to the background and not specifically to the diffraction peaks. Therefore, neutron crystallography experiments are normally performed on samples where hydrogen atoms have been replaced by deuterium (^{2}H or D), which has a similar scattering length as carbon and a low incoherent scattering.

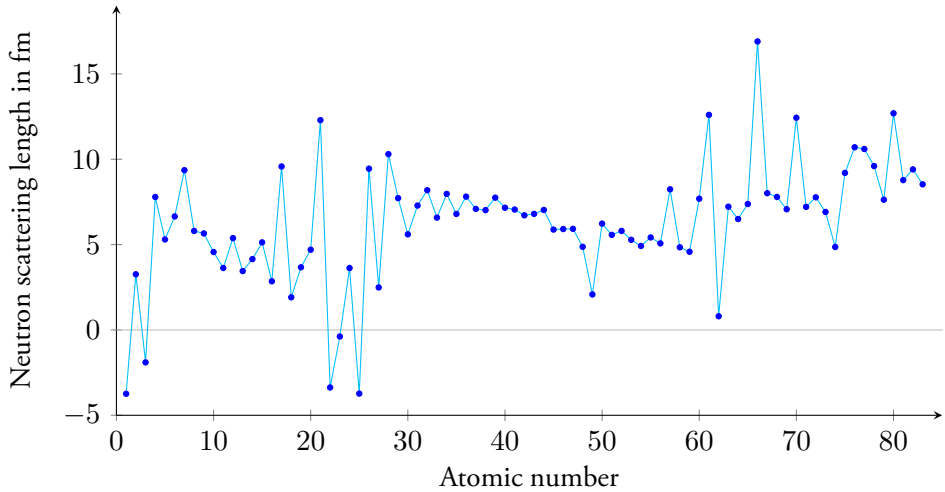


Figure 3.4: Coherent neutron scattering lengths [109] for the elements from hydrogen to bismuth in their natural isotope abundance.

Table 3.1: Coherent (b_{coh}) and incoherent (b_{incoh}) scattering lengths, spin and natural abundance of selected nuclei [110].

Nucleus	Spin	Natural abundance/ %	b_{coh}/fm	b_{incoh}/fm
$^{1}_{1}H$	1/2	99.985	-3.7390(11)	25.274(9)
$^{2}_{1}H$	1	0.015	6.671(4)	4.04(3)
$^{12}_{6}C$	0	98.90	6.6511(16)	0
$^{14}_{7}N$	1	99.63	9.37(2)	2.0(2)
$^{16}_{8}O$	0	99.762	5.803(4)	0
$^{31}_{15}P$	1/2	100	5.13(1)	0.2(2)
$^{32}_{16}S$	0	95.02	2.804(2)	0

All these aspects make neutron crystallography the gold standard to determine hydrogen positions, protonation states and even hydrogen displacement parameters [31, 71, 111–115].

Unfortunately, neutron crystallography also has its downsides. Due to the low flux of neutrons at neutron sources (compared to X-rays at synchrotrons), larger crystals and longer exposure times are needed during the measurement. The large cost to produce neutrons leads to expensive and scarce beamtime, which are additional limiting factors to use this method.

Furthermore, the treatment of hydrogen atoms in protein crystallography more than double the number of parameters to refine. To address this problem, joint X-ray and

neutron refinement [116] can be performed, as long as both datasets are isomorphous and measured at the same temperature. This results in the refinement term in equation 3.14

$$E_{\text{cryst}} = w_{\text{x}} E_{\text{X-ray}} + w_{\text{n}} E_{\text{neutron}} + E_{\text{MM}} \tag{3.14}$$

which contains two weight-factors w_{x} and w_{n}. With these, the influence of each set of data on the final model can be adjusted.

Computational protein crystallography

4.1 Aspherical refinement

With high-resolution X-ray crystallographic measurements, it is possible to capture features of the electron density that cannot be modelled by the traditional IAM. These differences occur since the electron density in a real molecule is not spherical around the atoms but also located in the bonds and lone-pairs. To take these effects into account, Hansen and Coppens [117, 118] introduced a multipole formalism. This description has the advantage that actual chemical features in the data are captured in the refinement, but it also introduces many more parameters to the refinement. More refined parameters lead to the risk of overfitting and require charge density resolution measurements. To overcome the problem of introducing more parameters, several multipole databases [27–31] have been developed, with the aim of transferring fixed multipoles to all atoms and refine only the positions and displacement parameters [119]. These databases have been applied to ultrahigh-resolution data sets of small proteins [120–123].

4.1.1 Hirshfeld atom refinement

In the Hirshfeld atom refinement (HAR) [32, 33], tailor made aspherical atomic form factors are obtained for each atom in the structure by QM calculations. These can be used without introducing more parameters to the refinement. To this end, first a QM single-point electron density of the structure is calculated. This electron density is then partitioned into the atoms using the Hirshfeld stockholder partitioning (equation 4.1) [124]:

$$\rho_A(\vec{r}) = w_A(\vec{r})\rho_{molecule}(\vec{r}) \tag{4.1}$$

where $w_A(\vec{r})$ is described by the density of atom A (ρ_A^0) at a certain point divided by the density contribution from all atoms ($\sum_B \rho_B^0$) at the same point (equation 4.2).

$$w_A(\vec{r}) = \frac{\rho_A^0(\vec{r} - \vec{r}_A)}{\sum_B \rho_B^0(\vec{r} - \vec{r}_B)} \tag{4.2}$$

Aspherical atomic form factors are calculated via Fourier transformation of the Hirshfeld-partitioned electron density.

In the second step, the coordinates and ADPs are refined in a classical least-squares refinement with the tailor-made aspherical atomic form factors. This leads to a new set of coordinates and the process is restarted beginning with the electron density step. Both steps are repeated iteratively until the new electron density does not lead to significant changes in the structure [33].

The outlined procedure results in a flexible technique that combines the advantage of the IAM regarding the small number of parameters with the accuracy of the multipole model. This better description gives HAR the ability to derive accurate and precise positions of hydrogen atoms, as demonstrated for a large set of small organic molecules [32, 33, 112, 125] (the IAM gives too short bond lengths to hydrogen atoms because the centre of the electron density does not coincide with the nucleus).

Unfortunately, standard QM calculations are impracticable for molecules as large as proteins. Since HAR depends on iterative QM calculations, refinement of large molecules is not feasible. This challenge is addressed in papers I and II of this thesis.

4.2 Quantum refinement

As outlined in section 3.4.2, in macromolecular crystallography, it is usually necessary to apply restraints. They are typically accurate for standard residues, but often less accurate for hydrogen atoms, unusual residues (substrates, cofactors, inhibitors or bound ligands, *e.g.* drug candidates) and metal sites. This results in an inaccurate and not necessarily chemically meaningful description of these regions, which often are found in the active site of enzymes.

To overcome these problems, the empirical restraints can be replaced for a small, but interesting, part of the protein (*e.g.* the active site) by restraints derived from a QM calculation in the same way as in standard subtractive QM/MM methods [20, 21, 126, 127]. By applying this QM/MM-like approach to the standard refinement function in equation 3.9, a new energy function is obtained (see equation 4.3).

$$E_{\text{cqx}} = w_{\text{MM}}(w_{\text{A}}E_{\text{X-ray}} + E_{\text{MM12}} - E_{\text{MM1}}) + E_{\text{QM1}} \tag{4.3}$$

where E_{QM1} is the QM energy of the region of interest called system 1 and E_{MM12} includes the standard empirical restraints for the whole protein. To avoid double-counting of energy terms, we need to subtract the corresponding MM energy of system 1, E_{MM1} from E_{MM12}. Furthermore, a separate weight factor w_{MM} is required to bring the QM and MM energy on the same energy scale, because the empirical restraints are normally in statistical units, whereas the QM energy is in energy units. This approach is called quantum refinement. It can be implemented for any macromolecular crystallographic refinement (*i.e.* X-ray, neutron or joint X-ray neutron) [22].

Such an energy function is implemented in the ComQum-X software [20–22, 128], which is an interface between the QM software Turbomole [129] and the crystallography and NMR system (CNS) [130, 131] software. Several other groups have implemented similar approaches, as is discussed in Ryde *et al.* (2022) [22]. For example, Merz *et al.* developed an approach in which they replaced the whole MM part by linear-scaling QM methods [132]. In the Q|R project[133], several QM programs were combined with the crystallographic software Phenix [105, 133].

Quantum refinement is computationally more expensive than a standard refinement. However, in the regions of interest, such as active sites, it is possible to distinguish similar elements and different protonation states using medium-resolution X-ray crystallographic data [22].

4.3 Model validation

Besides comparing to reference data, there are several ways to validate how well a model fits the crystallographic raw data. These validation methods can be divided into two classes. Global methods take into account the complete model, whereas local methods validate only one residue or one atom.

The most widely used quality measure is the R value (in macromolecular crystallography also called R_{work}), which quantifies the agreement between model and measured data (equation 4.4).

$$R = \frac{\sum_{hkl} ||F_{obs}| - |F_{calc}||}{\sum_{hkl} |F_{obs}|} \tag{4.4}$$

A R value of zero is a perfect fit.

In macromolecular crystallography (see section 3.4.2), even with empirical restraints, overfitting can occur. This means that a good fit between data and model, as well as a low R value could be obtained without being supported by the experimental data. To avoid this effect, a control measure has been introduced, the R_{free} value [134], where a small fraction of the reflections (usually around 5%) are not used for the refinement. This small data set is used afterwards to evaluate the fit of the model, which is always lower than for the refined data set. A significant deviation between R and R_{free} indicates overfitting. Therefore, R and R_{free} are not only quality measures to validate the final model, but should also be closely monitored during the refinement.

Especially for large molecules with several thousands of atoms, the global quality measures are not sensitive enough to distinguish local differences (*e.g.* correct element or protonation states). Therefore, local validation tools are needed. A visual way of validating a model is to look at Fourier transformed structure factors, which results in density maps. The most commonly used maps are the F_{obs} map and $F_{\mathrm{obs}} - F_{\mathrm{calc}}$ map in small molecular crystallography and the likelihood-weighted $2mF_o - DF_c$ and $mF_o - DF_c$ maps in macromolecular crystallography. The F_{obs} and the $2mF_o - DF_c$ maps are also called electron density maps in X-ray crystallography and nuclear scattering-length density maps in neutron crystallography. $F_{\mathrm{obs}} - F_{\mathrm{calc}}$ and $mF_o - DF_c$ maps are called difference density maps and represent the difference between the measured data and the model. Overmodelled regions are represented by negative residual densities (usually drawn in red) and undermodelled regions are represented by positive residue densities (usually drawn in green). For example, the difference density map for a shifted sulfur atom can be seen in figure 4.1.

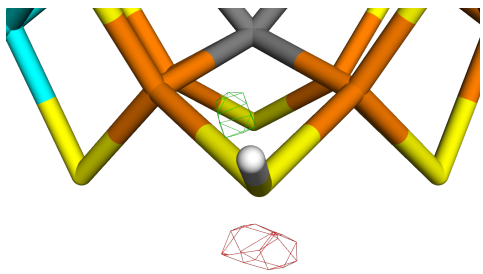


Figure 4.1: An incorrectly positioned sulfur atom. The $mF_{\mathrm{o}} - DF_{\mathrm{c}}$ map is contoured at $+3\sigma$ (green) and -3σ (red). (Taken from reference [135], reprinted with permission from Elsevier.)

To obtain a more quantitative measure, several local agreement statistics can be calculated. For example, the real-space R factor (RSR) [136] and real-space correlation coefficient (RSCC) [137] are correlated to the accuracy and precision of the model, while the real-space difference density Z score (RSZD) [138] is correlated only to the accuracy.

Quantum refinement has the advantage that quality measures can be obtained not only from the crystallographic part of the refinement but also from the QM calculations. The strain energy (ΔE_{QM}) [20, 139] gives a measure of how much the crystallographic data distorts the system from the optimised geometry. To compute the strain energy, two calculations are needed. The first is the normal quantum refinement with a w_A factor (see equation 4.3) greater than zero and the second calculation is a refinement with w_A equal to zero. This implies that no crystallographic data is employed and it becomes a QM/MM-like geometry optimisation (with the CNS force field). The strain energy is obtained by subtracting the two QM energies from each other (equation 4.5).

$$\Delta E_{QM} = E_{QM1}(w_A > 0) - E_{QM1}(w_A = 0) \tag{4.5}$$

A small strain energy indicates that the QM model fits the experimental data well. This is a robust method for the comparison between models with the same number of atoms and the same net charge, whereas for models with differing numbers of atoms or charges, a careful evaluation is advised.

Studied proteins

5.1 Nitrogenase

Nitrogen is crucial for the existence and growth of living organisms [140, 141], being as a key element of amino acids and DNA. Although the nitrogen content in the air is as large as 78%, it is not bioavailable owing to the strong and inert triple bond in N_2 [141]. Therefore, fixation of nitrogen in a reduced or oxidised form is needed to make it available to organisms.

On the industrial side, the Haber–Bosh process is used to make nitrogen bioavailable and the fixed nitrogen is mainly used as fertilisers [141]. In this process, nitrogen reacts with hydrogen to form ammonia. To make this reaction possible, high temperature, high pressure and an iron catalyst are required [141]. Currently, approximately 1% of the world energy consumption is used for the Haber–Bosh process [142].

Nitrogenase is the only enzyme that can catalyse the cleavage of the nitrogen triple bond [143, 144]. This highly energy-demanding reaction takes eight electrons and 16 molecules of adenosine triphosphate (ATP) to process one N_2 molecule (see equation 5.1). In contrast to the Haber–Bosch process, the reaction takes place at ambient temperature and pressure [144].

$$N_2 + 8\,e^- + 8\,H^+ + 16\,ATP \xrightarrow{\text{nitrogenase}} 2\,NH_3 + H_2 + 16\,ADP + 16\,P_i \quad (5.1)$$

The reaction has been studied extensively by both experimental and theoretical methods [143–146]. The reaction cycle is often described by the Lowe-Thorneley cycle, which consists of eight intermediates, $E_0 - E_7$, which differ in the number of added electrons and protons [143, 147]. E_0 is the ground state.

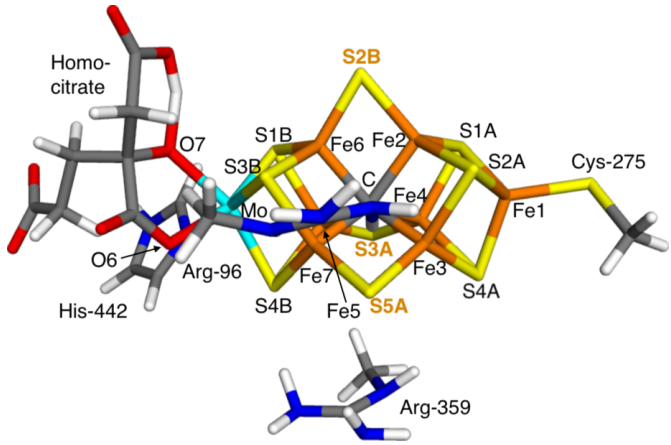


Figure 5.1: The FeMo cluster in nitrogenase illustrating the atom names used in this thesis. The three μ_2 bridging sulfide ions, which are of our special interest, are emphasised with bold orange text.

The active site of nitrogenase is the Mo/V/$FeFe_7S_9$ cluster. Molybdenum nitrogenase has the highest and iron nitrogenase the lowest reactivity [148]. The FeMo cluster in nitrogenase is shown in figure 5.1.

5.2 Acetylcholinesterase

Acetylcholinesterase (AChE) catalyses the cleavage of the ester bond in acetylcholine (ACh) (see figure 5.2). ACh is a neurotransmitter that is responsible for the activation of muscle cells [149, 150]. The cleavage of ACh terminates the synaptic transmission and reactivates the synapse [149]. Inhibition of AChE results in serious damage of the organism. Therefore, AChE it is the target of organophosphorus nerve agents.

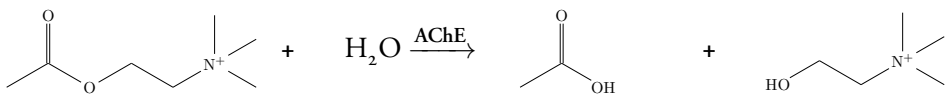


Figure 5.2: The hydrolysis reaction of ACh by AChE.

AChE is a dimer of two chains where both of them contain an active site in a deep and narrow gorge ending by Ser-203 (see figure 5.3) [150]. Organophosphorus nerve agents bind covalently to Ser-203.

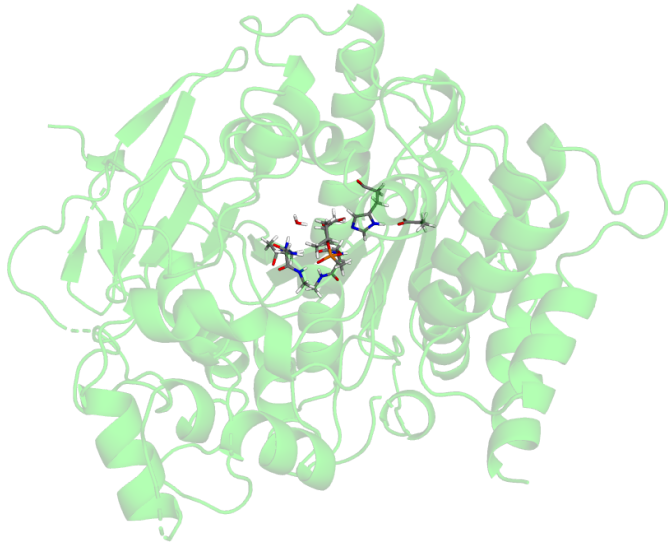


Figure 5.3: Structure of AChE with the active site highlighted [151].

5.3 Triosephosphate isomerase

Triosephosphate isomerase (TIM) is a highly efficient enzyme involved in glycolysis. It catalyses the 1,2 proton shift of dihydroxyacetone phosphate (DHAP) to D-glyceraldehyde 3-phosphate (GAP) (see figure 5.4).

Figure 5.4: 1,2 proton shift from DHAP to GAP.

TIM is one of the most efficient enzymes, with a reaction constant close to the diffusion limit and no further co-substrate is needed [152–154].

The catalytic activity is at its active site where Asn-11, Lys-13, His-95 and Glu-167 stabilise the substrate. Glu-167 is the catalytic base and if it is mutated to aspartate, TIM loses 80–98% of its catalytic activity [4]. If this mutation is on both deoxyribonucleic acid (DNA) strands (homozygous), the life expectancy is drastically reduced, *e.g.* to five years for humans [155, 156]. TIM has been intensively studied with experimental methods. Already in 1976, the first crystal structure was published [157] and today, several atomic resolution crystal structures are available [158]. Inhibitors that mimic

the substrate or reaction intermediates are used to study the reaction mechanism [159–161].

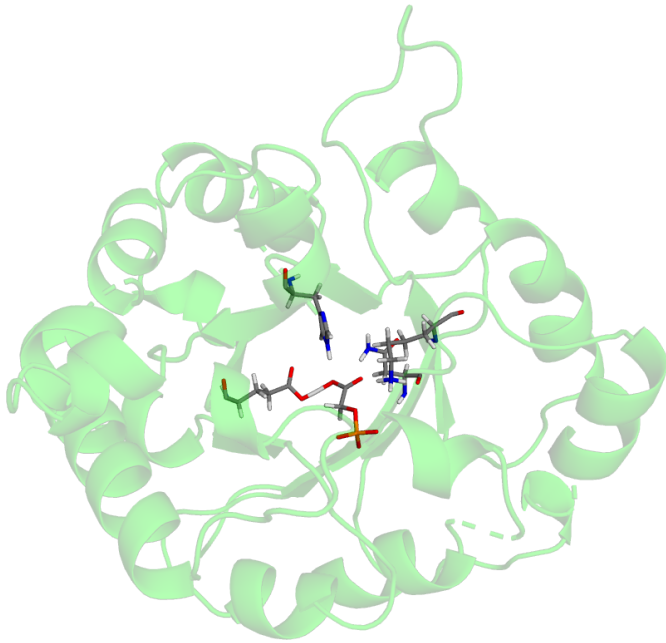


Figure 5.5: Structure of TIM with highlighted active site.

Scientific publications

Summary of the papers

The papers included in this thesis combine quantum mechanical methods with crystallographic refinement. They can be divided into two groups.

- In papers I and II, the problem of modelling high-quality data with an insufficient model is addressed with the adaption of the HAR method to polypeptides and proteins.
- In papers III to VII, we further developed and applied quantum refinement for systems for which only medium-resolution data are available.

Paper I

Hirshfeld atom refinement (HAR) is a crystallographic refinement technique that uses a tailor made aspherical structure factor for each atom in the structure [32]. HAR consists in the following steps: First, a wavefunction is calculated on the initial coordinates of the molecule. Second, from this wavefunction the electron density is computed and partitioned into atoms with the Hirshfeld stockholder partitioning scheme. Third, the Fourier transform of the Hirshfeld atomic ED is the atomic structure factor, which is finally used in the crystallographic least-squares refinement, yielding a new structure. These steps are iterated until the structure is converged [33].

For large molecules it becomes prohibitive to calculate the wavefunction in reasonable time with available computational resources. To solve this problem, a database approach has been developed, HAR combined with ELMO libraries (HAR-ELMO) [162]. However, with HAR-ELMO, only the 20 standard amino acids and water can be treated. Since most proteins contain metal centres or some non-protein ligands, this method is not applicable for most structures, besides the most simple polypeptides.

To avoid the inflexibility of database approaches, we combined a fragmentation approach with HAR. To this end, we used the MFCC approach, according to which a large molecule is divided into small fragments with conjugate caps to avoid truncation errors (see section 2.2). This new method is called fragHAR, and we tested it on three peptide model systems (*e.g.* the one in figure 6.1).

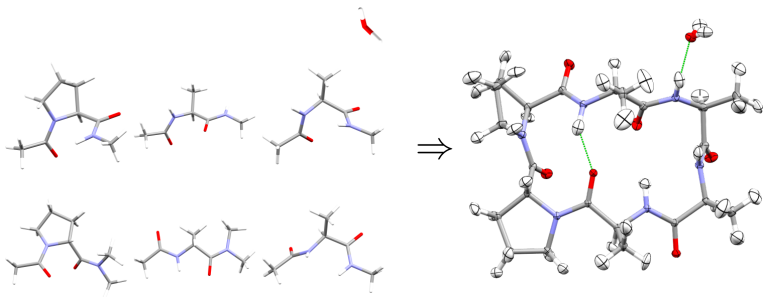


Figure 6.1: The fragments employed for cyclo-(Ala)$_4$-(d,l-Pro)$_2$ (A_4P_2) and the fragHAR structure. (Taken from reference [36], reprinted with permission from IUCr.)

For non-hydrogen atoms no significant difference between classical HAR and fragHAR was observed. For hydrogen atoms, most X—H bond lengths were in statistical agree-

ment between both refinement methods.[1] fragHAR takes only a fraction of the time compared to the traditional HAR because the QM calculation for the whole system is avoided. The only shortcoming of the fragmentation is observed for hydrogen atoms involved in hydrogen bonds. In those cases, fragHAR provides too short X−H bond lengths, because the effect of the hydrogen bond acceptor is neglected in the fragment calculation. To overcome this neglect, it is possible to join both fragments involved in the hydrogen bond to get X−H bond length in statistical agreement with the classical HAR.

[1]Equation 2 in the paper should be:

$$w\text{RMSD} = \sqrt{\left\langle \frac{(A_i - B_i)^2}{s.u.(A_i)^2 + s.u.(B_i)^2} \right\rangle}$$

Paper II

As described in paper I, Hirshfeld atom refinement (HAR) is challenging for large systems like proteins. To overcome problems in the refinement and the QM calculations, we implemented fragHAR [36] in the NoSpherA2 [163] interface in the Olex2 [164, 165] program. In this implementation, olex2.refine is combined with Orca 5.0 for the QM calculations [166]. In contrast to the previous implementation in TONTO, alternative conformations and restraints are available in the refinement. For the SCF step, a wide variety of QM methods and basis sets are available.

For alternative conformations, fragHAR provides a convenient solution. With fragHAR, it is not necessary to compute the wavefunction for the whole system multiple times, but only for fragments with alternative conformations. Furthermore, we have also implemented a new automatic capping approach for hydrogen bonds to overcome the shortcomings for hydrogen atoms involved in such interactions, as was described in paper I.

We first tested the new implementation of fragHAR on three small polypeptides, which were also used in paper I. On these small test cases, fragHAR in Olex2 reproduces HAR results with a similar performance as our previous implementation in TONTO. Moreover, the hydrogen-bond capping and treatment of alternative conformations are effective and accurate.

After the validation, we applied a first cycle of fragHAR (Hirshfeld atom fit [32]) to two proteins, crambin [167] and rubredoxin [168]. For crambin, it was possible to refine individual X−H distances, which results in statistical agreement with averaged X−H bond lengths from small-molecule neutron crystallography [169]. Furthermore, the overall R value is improved and the residual density describing bonding features are reduced, compared to refinement with the IAM.

For rubredoxin it was not possible to refine individual hydrogen positions. Nevertheless, the overall R value is improved and the description of the electron density around the iron–sulfur cluster is improved. The difference in the description compared to the IAM can be seen in figure 6.2 left.

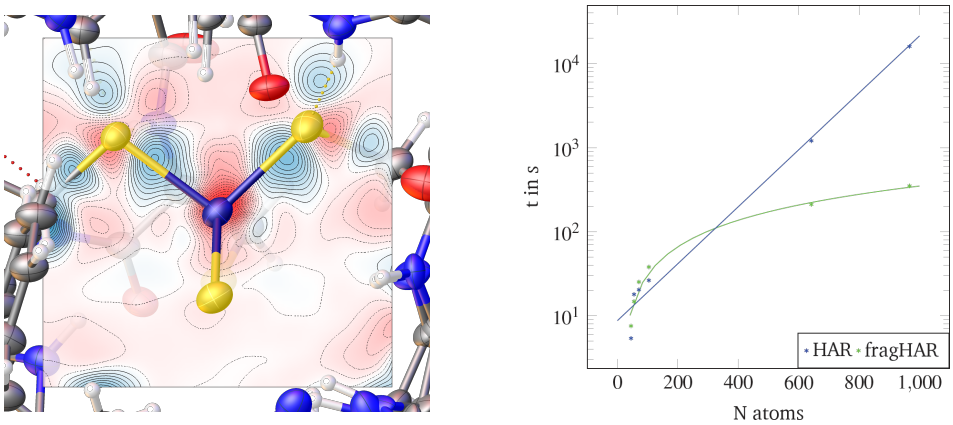


Figure 6.2: Deformation density in the iron–sulfur cluster of rubredoxin on the left (isovalues from $-0.2\ e\text{Å}^{-3}$, red, to 0.2 $e\text{Å}^{-3}$, blue). Timing for one cycle of HAR and fragHAR in Olex2 on the right.

The advantage of fragHAR is that the time consumption increases linearly with the system size instead of exponentially as for HAR. This advantage becomes significant only for large systems, whereas for the oligopeptides, the required time for HAR and fragHAR is similar. For rubredoxin with 967 atoms, the fragHAR calculation is 46 times faster than the HAR calculation. (see figure 6.2 right).

Thus, we can conclude that fragHAR provides a method that can perform *ab initio* aspherical refinements even of proteins.

Paper III

Quantum refinement replaces the standard empirical restraints used in a crystallographic refinement with more accurate restraints derived from QM calculations for a small but interesting part of the protein [20, 21]. This results in a dependency of the refinement on the QM method, basis set and description of the surrounding. Standard QM calculations are performed in vacuum at zero kelvin. This is clearly not a good representation of the actual environment in the crystal. In fact, the experimental crystallographic data are usually collected at 100 K and the QM system (see section 4.2) is surrounded by the protein.

One way to reduce this mismatch, is to increase the size of system 1, but this would make the calculation computationally more expensive. In paper III, we instead try to introduce a continuum solvent for the environment of the QM system [76]. In this way, the continuum solvent can compensate charges in the QM system and a description closer to the actual environment can be achieved without increasing the size of the QM region.

To compare quantum refinement calculations in vacuum with an increased QM system or with continuum solvation, we employed five different crystal structures. Three structures involved nitrogenase with highly negative charged QM systems. In addition, one structure of particulate methane monooxygenase (pMMO) and one structure of AChE were tested.

For the highly charged systems, the continuum-solvation approach gives a significant improvement in terms of the RSZD and the strain energy (see figure 6.3). However, for QM systems with a low charge, no significant improvement was observed, but also no significant deterioration.

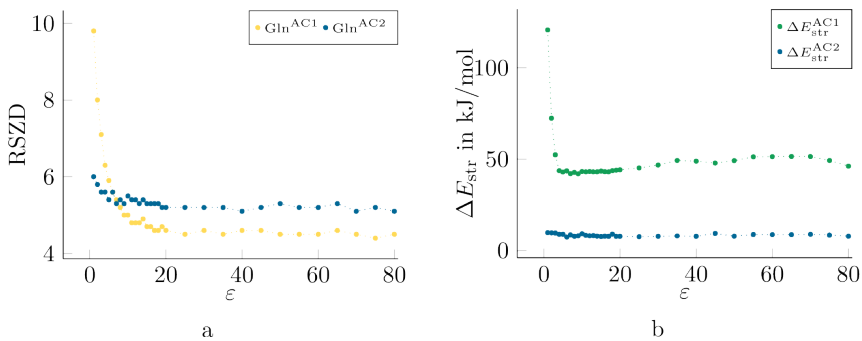


Figure 6.3: RSZD values (a) and strain energies (b) as a function of the dielectric constant (ε) for quantum refinement calculations of nitrogenase with a OH^- ligand. (Taken from reference [170], reprinted with permission from IUCr.)

We also studied how the results depended on the value of the dielectric constant (ε). For highly charged systems, significant improvements of the strain energy and the RSZD scores could be obtained already with small dielectric constants (see figure 6.3). With ε greater than 20, the improvement converged and no further effects were observed for larger values. For QM systems with a low charge, no significant differences were obtained, irrespective of the value of ε.

In conclusion, it is favourable to use solvation models for the QM calculations in quantum refinements with highly charged QM systems.

Paper IV

Triosephosphate isomerase (TIM) is a key enzyme in glycolysis, where it catalyses the internal reduction–oxidation from dihydroxyacetone phosphate (DHAP) to D-glyceraldehyde 3-phosphate (GAP). The turnover is so efficient that it has been argued that it is a catalytically perfect enzyme, with a turnover rate close to the diffusion limit [171, 172].

In this study, the Glu97Gln variant of TIM with the inhibitor 2-phosphoglycolate (PGA) bound to the active site is investigated by X-ray and neutron crystallography. In the wild-type enzyme, the carboxylic group of the PGA inhibitor shows two conformations. In one, PGA forms a hydrogen bond to His-95, whereas in the other, it instead forms a hydrogen bond to Lys-13. In the variant structure, only the latter conformation is seen. Moreover, a significant positive nuclear scattering-length difference density between PGA and Lys-13 can be observed (see figure 6.4 left). To explain this difference, we investigated the protonation state of the carboxylate group in PGA. We employed both refinements with standard crystallographic restraints and quantum refinement with different descriptions of the surrounding and alternative conformations.

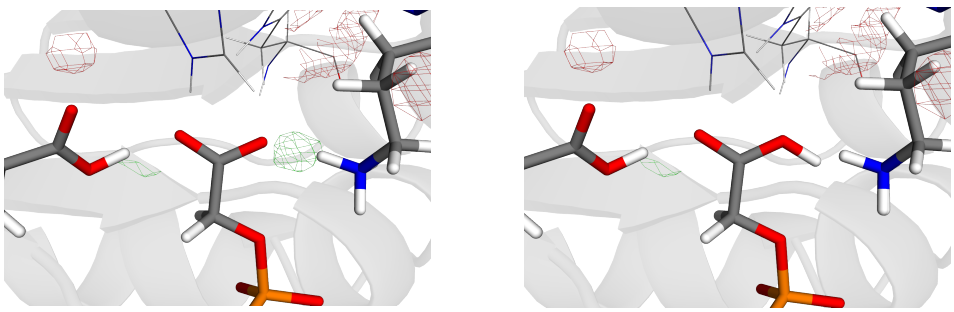


Figure 6.4: Nuclear scattering-length difference density in the active site of TIM. Left: 100% occupancy for the proton on Lys-13 gives a positive volume between Lys-13 and PGA. Right: best model with 30% occupancy of the proton on PGA. The $mF_o - DF_c$ nuclear scattering-length difference density maps are contoured at $+3\sigma$ (green) and -3σ (red).

In the quantum refinements, we used three different sizes for the QM system. The standard QM system consisted of PGA, Asn-11, Lys-13, Glu-167 and His-95. To avoid the large negative charge ($-2e$) of the QM system, two alternative approaches were tested. In one, we used a minimal QM system including only Lys-13, Glu-167 and PGA but excluding the phosphate group of PGA, with its double negative charge. The second approach was to enhance the QM system by all groups that form hydrogen bonds to the phosphate group of PGA, in order to counteract its negative charge. For the surroundings, five different descriptions were used. The standard approach in quantum refinement is to run the QM calculations in vacuum. To obtain a more

realistic description, a continuum-solvent model (as suggested in paper III [170]) with a dielectric constant of 2, 4 or 80 was used. Finally, since hydrogen positions are available in neutron crystallographic data, also a point-charge model of the surroundings can be applied.

With the minimal and small QM system in vacuum or with a low dielectric constant, the proton shared between PGA and Lys-13 is located at PGA. For higher dielectric constants, the point-charge model and the large QM system, this hydrogen atom moves to Lys-13.

To evaluate the degree of protonation at PGA and Lys-13, different occupancies for the two protonation states were tested. To do so, an approach to treat multiple conformations was implemented in joint neutron and X-ray quantum refinement. With this new implementation, refinements with alternative conformations of the shared hydrogen were performed. For comparison, joint neutron and X-ray refinement was also performed with phenix.refine [105]. From these refinements we conclude that the best fit to the experimental data is obtain if Lys-13 is protonated to 70% and PGA to 30% (see figure 6.4 right).

Paper v

Nitrogenase is the only enzyme that can cleave the triple bond in N_2, making nitrogen available to plants. Therefore, the mechanism of this enzyme is of high interest. In this paper, two structures of nitrogenase were investigated. The first structure is one of V-nitrogenase where we study the nature of the bidentate ligand replacing S3A (see figure 5.1 for atom labels).[2] For the second structure of Mo-nitrogenase, we suggest the protonation state of the MoFe-cluster after inhibition with CO.

In the structure of V-nitrogenase an unusual bidentate ligand replaces S3A (see figure 5.1 for atom label) in the FeV-cluster. This bidentate ligand could be carbonate, bicarbonate or nitrate (see figure 6.5). Quantum refinements with all three possible ligands were performed. The original crystallographers tentatively identified the unusual bidentate ligand as carbonate [173] and this was as confirmed by our quantum refinement calculations.

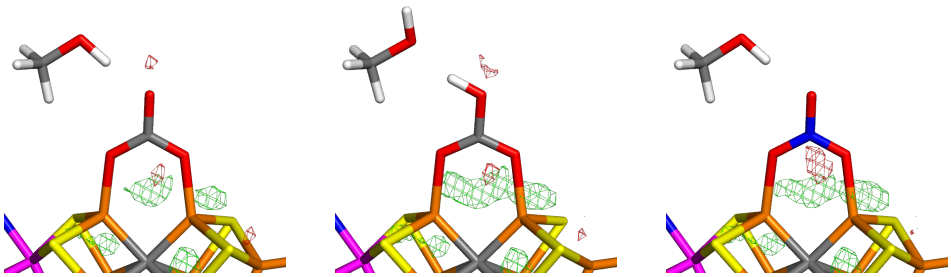


Figure 6.5: Active site of V-nitrogenase with (a) CO_3^{2-}, (b) HCO_3^- or (c) NO_3^- bound to the active site. The $F_o - F_c$ electron-density difference maps are shown in green ($+3\sigma$) and red (-3σ). (Taken from reference [135], reprinted with permission from Elsevier.)

In the second part of paper V, the CO inhibited FeMo cluster of Mo-nitrogenase was investigated (see figure 6.6). Experimentally, it is known that CO binds only to reduced states of nitrogenase (E_2–E_4) [174–177] replacing the S2B sulfide ligand (see figure 5.1). We assumed that CO binds to the doubly reduced E_2 state, which should be doubly protonated. The hope was to identify the positions of these protons, which could give valuable clues to the structures or other states in the reaction mechanism. However, it is possible that the sulfide ligand S2B takes one or even both protons with it when it dissociates from the cluster (as HS^- or H_2S).

[2] correction of the publication: S3A is replaced and not S2B

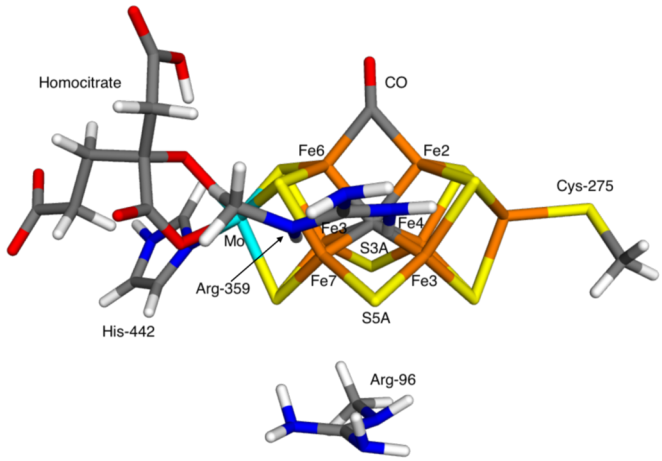


Figure 6.6: The active site of CO-inhibited Mo-nitrogenase, showing the QM system employed in our quantum-refinement calculations, as well as the name of the various atoms in the cluster. (Taken from reference [135], reprinted with permission from Elsevier.)

We tested 15 different protonation and charge states with quantum refinement, but it was not possible to find any evidence of protonation in the FeMo cluster. Therefore, the most likely interpretation is that the dissociating S2B ligand takes two protons with it. Alternatively, the cluster could also be deprotonated after the binding of CO, *e.g.* the hydrogen atoms dissociate as a H_2 molecule. A third explanation could be that the cluster actually is protonated but that the proton may move extensively around on the cluster, so that the structure is a mixture of several protonation states.

Paper VI

In Paper VI, we evaluated a structure of Mo-nitrogenase, solved by Kang *et al.* [178]. In this structure, it was suggested that all belt sulfide atoms (see figure 5.1, atoms S2B, S3A and S5A) could be replaced by the N_2 substrate during the catalytic cycle. This is a sensational suggestion, but it does not agree with previous mechanistic suggestions [179, 180]. The crystal structure was measured to 1.83 Å resolution. In chain A, S2B and in chain C, S3A and S5A are replaced by N_2. Furthermore, the authors suggested that homocitrate is bound in an asymmetric manner to the molybdenum ion of the MoFe cluster. They also presented anomalous electron density maps measured at 7100 eV, which selectively shows sulfur atoms, suggesting reduced densities for the dissociated ligands. Since this structure would provide a new view of the catalytic mechanism, we critically checked the data quality, evaluated the anomalous electron density and performed new refinements of the MoFe clusters in chain A and C with quantum refinement and different sets of ligands.

The integration of the anomalous electron density shows that the density around S2B in chain A has a deviation of 1.7 σ from the average of all sulfur atoms in the chain A cluster, but it still has a similar intensity as S1B. In chain C, S3A deviates by only 0.90 σ and S5A deviates by only 0.81 σ from the average density of all sulfur atoms in this cluster. Furthermore, both of them still have a higher intensity than S3B. Therefore, the anomalous electron density gives no convincing indication that these sulfide ions have dissociated.

Analysis of the X-ray data showed a high anisotropy [181]. It is only along the c^*-axis the crystal scatters to the reported resolution of 1.83 Å, whereas along the b^*-axis, it scatters only to ∼2.6 Å resolution. The anisotropy of the data may explain why the density of the sulfide ions is elongated, which was interpreted as a diatomic ligand. This anisotropy was further enhanced by the use of anisotropic B factors, which is highly questionable for X-ray data with a resolution of only ∼2.6 Å in one direction.

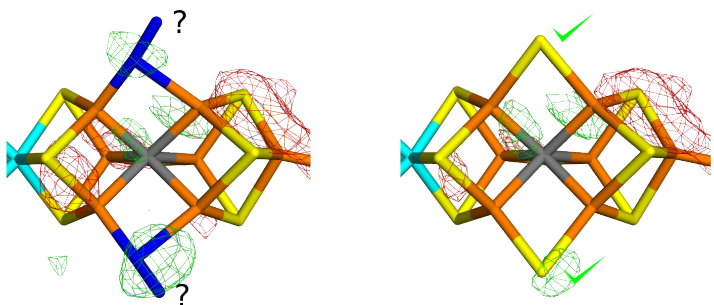


Figure 6.7: Possible interpretations of the bridging ligands in chain C of Mo-nitrogenase. (Taken from reference [182], reprinted with permission from Springer.)

The quantum refinements show that in all cases the residual density around the bridging ligand is lower if it is modelled as S^{2-}, rather than N_2 or N_2H_2. For S2B in chain A and S5A in chain C, the residual density completely disappears at a 3 σ map (see figure 6.7). Moreover, the binding of the homocitrate ligand is in all cases symmetric and no significant bond length differences between Mo–O6 and Mo–O7 (for labels of oxygen atoms compare figure 5.1) were observed.

In summary, we find no support for the suggestion that the belt sulfur atoms (S2B, S3A or S5A) are replaced with a N_2-derived ligand. Instead, a standard structure of the resting-state MoFe cluster fits the experimental data properly.

Paper VII

In Paper VII, we investigated the protonation state of the active site of acetylcholinesterase (AChE) with and without inhibition by organophosphorus nerve agents. AChE hydrolyses the neurotransmitter acetylcholine in cholinergic synapses. This reactivates the synapse and stops the signal [149, 150]. Common nerve agents inhibit AChE and block this reactivation of the synapses. To develop antidotes against these nerve agents, it is crucial to know the structure and protonation state of the resting (apo) enzyme and the enzyme inhibited by the nerve agent.

With this aim, we have performed quantum refinements of six different structures of AChE, of which one was in the apo state and the other five were inhibited with four different nerve agents [151, 183–187].

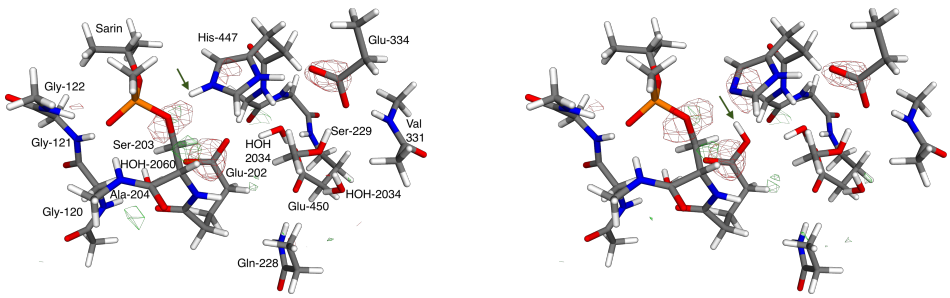


Figure 6.8: Quantum-refined structures of acetylcholinesterase phosphonylated by sarin in the a) P00 and b) P03 protonation states. The $mF_o - DF_c$ difference maps are contoured at $+3\sigma$ (green) and -3σ (red). The arrows indicate the position of the moved proton.

For human AChE inhibited by sarin (5FPQ [151]), 14 different protonation states were tested. In the starting structure (P00), HIS-447 is doubly protonated (see figure 6.8a). For eleven protonation states, one proton was moved along a hydrogen bond (P01–P11). In two states, $H_{\varepsilon 2}$ or $H_{\delta 2}$ of HIS-447 were deleted (P12 and P13). However, only P00, P03 (where $H_{\varepsilon 2}$ of HIS-447 was moved to GLU-202), P05 (where $H_{\delta 2}$ of HIS-447 was moved to GLU-334), P12 (where $H_{\varepsilon 2}$ of HIS-447 was deleted) and P13 (where $H_{\delta 2}$ of HIS-447 was deleted) resulted in stable conformations. Consequently, refinements with these five protonation states where performed on all six structures. The sum of the RSZD scores of all residues in the quantum system (see figure 6.9) were computed to decide which protonation state fits the experimental data best.

The differences in the RSZD score vary between all structures and sometimes they are rather small. In all cases except 5FPQ, the P03 state (see figure 6.8 b) gives the best fit to the experimental data. This is a strong indication that P03 is the best representation of the protonation of the active site of AChE.

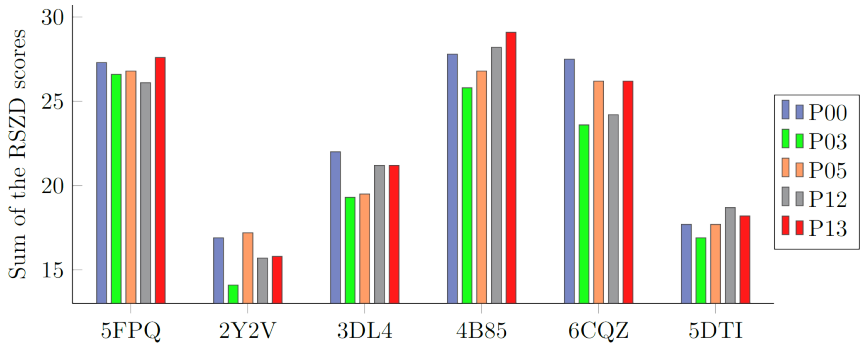


Figure 6.9: Sum of the RSZD scores for the QM system with various protonation states.

Conclusions and Outlook

The goal of this thesis was to improve the refinement of protein crystal structures by combining computational and crystallographic methods. To this end, we have followed two lines. In the first, we made further developments in the description of ultrahigh-resolution crystallographic data. In the second, we further developed the quantum refinement for medium- and low-resolution data. Furthermore, we applied quantum refinement to resolve ambiguities in existing protein structures.

In the first two publications (papers I and II), we introduced a linear-scaling fragmentation approach for the QM calculation in the Hirshfeld atom refinement. This new HAR variant is called fragHAR. With fragHAR, we developed a new way to obtain *ab initio* structure factors for proteins in reasonable time and at reasonable computational cost. We first applied fragHAR to small polypeptides and showed that structures obtained with fragHAR are in good statistical agreement to structures obtained with the classical HAR. The only shortcoming was observed for hydrogen atoms involved in hydrogen bonds. To overcome this shortcoming, we introduced a sophisticated capping for hydrogen bonds in paper II. This new capping is computationally only slightly more expensive than the original fragHAR but provides results in statistical agreement to the original HAR for all atoms. We were also able to present the first HAR for a metalloprotein.

In paper III, we improved the quantum refinement by applying a continuum solvent description for the surroundings in the QM calculation. This treatment compensates charges in the QM system and reduces errors introduced by electrostatics in the QM region. The continuum solvent description reduces the computational cost compared to enlarging the QM region, which would otherwise be necessary to correctly represent the electrostatics. We applied this new strategy to five protein structures and showed that this gives a significant improvement for highly charged systems and only minor effects for low charged systems.

In paper IV, we introduced a new approach for joint neutron and X-ray quantum refinement to treat alternative conformations. We used this new implementation to evaluate the protonation state in the active site of TIM inhibited by PGA. We concluded that it is most likely that Lys-13 is protonated to 70% and PGA to 30%.

In paper V, we used quantum refinement to identify a ligand and to study possible protonation states in the active sites of two types of nitrogenases. First, we proved from a crystallographic point of view that the unusual bidanted ligand in vanadium nitrogenase is CO_3^{2-}. Furthermore, we investigated the protonation state of the MoFe cluster of Mo-nitrogenase after inhibition by CO. We showed that the MoFe cluster is most likely not protonated when inhibited with CO.

In paper VI, we used quantum refinement to critically evaluate a recently published structure of Mo-nitrogenase that suggested a replacement of all belt sulfide atoms by N_2. We showed that this suggestion is not supported by the experimental data. Instead, all belt sulfide atoms seem to be present in the crystal structure and the structure is better described as the ground-state (E_0) of nitrogenase.

In the last paper VII, we refined six different structures of AChE with quantum refinement to evaluate the protonation state in the active site. Knowledge about the protonation state of AChE is crucial for the development of antidotes against nerve agents. We investigated one structure without any inhibitor and five structures inhibited by nerve agents. We concluded that Glu-202 is most likely protonated and His-447 is deprotonated on Nε2.

In conclusion, strategies for improved refinement of protein crystal structures were introduced and applied. We made HAR applicable for the refinement of ultrahigh-resolution protein crystal structures as demonstrated by the refinement of the metalloprotein rubredoxin. With this refinement, we showed that HAR has benefits also for the protein crystallographic community. Therefore, the next steps will be to apply it to more proteins of interest. Previous publications [27, 163] suggest that the aspherical description of metals gives an improved model. This leads to the question whether fragHAR could improve the description of metal clusters in proteins, like the one in nitrogenase. Furthermore, it would be advantageous to couple fragHAR with a more dedicated protein-refinement program, but also be important study more protein structures at a sufficiently high resolution. It will be interesting to see if HAR allows the identification of more hydrogen atoms at a lower resolution.

For quantum refinement, we have overcome some of the problems with electrostatics by applying a continuum-solvent method for the QM calculation. Furthermore, we have applied quantum refinement to identify ligands and protonation states in nitrogenase. For AChE and TIM we obtained deeper insights in the protonation states of the active site. The quantum refinements in this thesis proved again that this strategy provides valuable and deepened insights into structural features of proteins. There is a large number of interesting protein structures that are already published and for which quantum refinement could lead to a better interpretation of the data. Therefore, the next step for quantum refinement would be to implement it in a more modern pro-

tein crystallographic program. This is currently under development in our group and could make quantum refinement available for a broader user community.

Furthermore, it would be interesting to combine quantum refinement with Hirshfeld atom refinement. Quantum refinement currently uses spherical atomic form factors and, in a first step, they could be replaced by aspherical form factors for the QM region. Since the QM calculation is already part of quantum refinement, this would only add the computational cost for the aspherical partitioning of the electron density and the calculation of the atomic form factors. Note that this approach would imply a combined spherical and aspherical refinement for different regions of the system, which could be interesting on its own because it would be computationally more efficient than a Hirshfeld atom refinement for the complete system. To accomplish a better description of the whole system, all spherical form factors in quantum refinement could be replaced by aspherical form factors calculated using the fragHAR method.

References

[1] C. A. Orengo, A. E. Todd, and J. M. Thornton. From protein structure to function. *Curr. Opin. Struct. Biol.*, 9(3):374–382, **1999**.

[2] J. S. Fetrow and J. Skolnick. Method for prediction of protein function from sequence using the sequence-to-structure-to-function paradigm with application to glutaredoxins/thioredoxins and t1ribonucleases. *J. Mol. Biol.*, 281(5): 949–968, **1998**.

[3] C. Missale, S. R. Nash, S. W. Robinson, et al. Dopamine receptors: from structure to function. *Psychol. Rev.*, 78(1):189–225, **1998**.

[4] I. O. Daar, P. J. Artymiuk, D. C. Phillips, and L. E. Maquat. Human triosephosphate isomerase deficiency: a single amino acid substitution results in a thermolabile enzyme. *Proc. Natl. Acad. Sci. U.S.A.*, 83(20):7903–7907, **1986**.

[5] A. R. Fersht, A. Matouschek, and L. Serrano. The folding of an enzyme: I. theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.*, 224(3):771–782, **1992**.

[6] J. C. Brooks-Bartlett and E. F. Garman. The nobel science: One hundred years of crystallography. *Interdiscip. Sci. Rev.*, 40(3):244–264, **2015**.

[7] M. Jaskolski, Z. Dauter, and A. Wlodawer. A brief history of macromolecular crystallography, illustrated by a family tree and its n obel fruits. *FEBS J.*, 281 (18):3985–4009, **2014**.

[8] G. M. Sheldrick. A short history of shelx. *Acta Crystallogr. A*, 64(1):112–122, **2008**.

[9] H. M. Berman, J. Westbrook, Z. Feng, et al. The protein data bank. *Nucleic Acids Res.*, 28(1):235–242, **2000**. URL `www.rcsb.org`.

[10] W. H. Bragg and W. L. Bragg. The reflection of x-rays by crystals. *Proc. Math. Phys. Eng. Sci.*, 88(605):428–438, **1913**.

[11] J. C. Kendrew, G. Bodo, H. M. Dintzis, et al. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666, **1958**.

[12] C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward. The cambridge structural database. *Acta Crystallogr. B*, 72(2):171–179, **2016**.

[13] W. Massa. *Kristallstrukturbestimmung*, volume 6. Springer, **2007**.

[14] H. R. Powell. From then till now: changing data collection methods in single crystal x-ray crystallography since 1912. *Crystallogr. Rev.*, 25(4):264–294, **2019**.

[15] G. Taylor. The phase problem. *Acta Crystallogr. D*, 59(11):1881–1890, **2003**.

[16] G. M. Sheldrick. Shelxt–integrated space-group and crystal-structure determination. *Acta Crystallogr. A*, 71(1):3–8, **2015**.

[17] A. H. Compton. The distribution of the electrons in atoms. *Nature*, 95(2378): 343–344, **1915**.

[18] G. J. Kleywegt and T. A. Jones. [11] model building and refinement practice. *Meth. Enzymol.*, 277:208–230, **1997**.

[19] R. A. Engh and R. Huber. Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr. A*, 47(4):392–400, **1991**.

[20] U. Ryde, L. Olsen, and K. Nilsson. Quantum chemical geometry optimizations in proteins using crystallographic raw data. *J. Comput. Chem.*, 23(11):1058–1070, **2002**.

[21] U. Ryde and K. Nilsson. Quantum chemistry can locally improve protein crystal structures. *J. Am. Chem. Soc.*, 125(47):14232–14233, **2003**.

[22] J. Bergmann, E. Oksanen, and U. Ryde. Combining crystallography with quantum mechanics. *Curr. Opin. Struct. Biol.*, 72:18–26, **2022**.

[23] R. E. Cachau, J. Zhu, and M. C. Nicklaus. The upcoming subatomic resolution revolution. *Curr. Opin. Struct. Biol.*, 58:53–58, **2019**.

[24] T. Petrova and A. Podjarny. Protein crystallography at subatomic resolution. *Rep. Prog. Phys.*, 67(9):1565, **2004**.

[25] P. Coppens. *X-ray charge densities and chemical bonding*, volume 4. International Union of Crystallography, **1997**.

[26] C. Gatti and P. Macchi. A guided tour through modern charge density analysis. In *Modern Charge-Density Analysis*, pages 1–78. Springer, **2011**.

[27] B. Dittrich, C. Hübschle, K. Pröpper, et al. The generalized invariom database (gid). *Acta Crystallogr. B*, 69(2):91–104, **2013**.

[28] P. M. Dominiak, A. Volkov, X. Li, et al. A theoretical databank of transferable aspherical atoms and its application to electrostatic interaction energy calculations of macromolecules. *J. Chem. Theory Comput.*, 3(1):232–247, **2007**.

[29] K. N. Jarzembska and P. M. Dominiak. New version of the theoretical databank of transferable aspherical pseudoatoms, ubdb2011–towards nucleic acid modelling. *Acta Crystallogr. A*, 68(1):139–147, **2012**.

[30] S. Domagała, B. Fournier, D. Liebschner, et al. An improved experimental databank of transferable multipolar atom models–elmam2. construction details and applications. *Acta Crystallogr. A*, 68(3):337–351, **2012**.

[31] K. K. Jha, B. Gruza, P. Kumar, et al. Taam: a reliable and user friendly tool for hydrogen-atom location using routine x-ray diffraction data. *Acta Crystallogr. B*, 76(3):296–306, **2020**.

[32] D. Jayatilaka and B. Dittrich. X-ray structure refinement using aspherical atomic density functions obtained from quantum-mechanical calculations. *Acta Crystallogr. A*, 64(3):383–393, **2008**.

[33] S. C. Capelli, H.-B. Bürgi, B. Dittrich, et al. Hirshfeld atom refinement. *IUCrJ*, 1(5):361–379, **2014**.

[34] L. E. Ratcliff, S. Mohr, G. Huhs, et al. Challenges in large scale quantum mechanical calculations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 7(1):e1290, **2017**.

[35] D. W. Zhang and J. Zhang. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy. *J. Chem. Phys.*, 119(7):3599–3605, **2003**.

[36] J. Bergmann, M. Davidson, E. Oksanen, et al. fraghar: towards ab initio quantum-crystallographic x-ray structure refinement for polypeptides and proteins. *IUCrJ*, 7(2):158–165, **2020**.

[37] E. Schrödinger. Quantisierung als eigenwertproblem. *Ann Phys*, 385(13):437–490, **1926**.

[38] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Ann Phys*, 389(20):457–484, **1927**.

[39] A. Szabo and N. S. Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, **2012**.

[40] D. R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 89–132. Cambridge university press, **1928**.

[41] V. Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Z. Phys.*, 61(1):126–148, **1930**.

[42] C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23(2):69, **1951**.

[43] G. Hall. The molecular orbital theory of chemical valency viii. a method of calculating ionization potentials. *Proc. R. Soc. A*, 205(1083):541–552, **1951**.

[44] J. C. Slater. The theory of complex spectra. *Phys. Rev.*, 34(10):1293, **1929**.

[45] D. Feller. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.*, 17(13):1571–1586, **1996**.

[46] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B): B864, **1964**.

[47] R. A. Friesner. Ab initio quantum chemistry: Methodology and applications. *Proc. Natl. Acad. Sci. U.S.A.*, 102(19):6648–6653, **2005**.

[48] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133, **1965**.

[49] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. letters*, 77(18):3865, **1996**.

[50] J. P. Perdew, A. Ruzsinszky, L. A. Constantin, et al. Some fundamental issues in ground-state density functional theory: A guide for the perplexed. *J. Chem. Theory Comput.*, 5(4):902–908, **2009**.

[51] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria. Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.*, 91:146401, **Sep 2003**. doi: 10.1103/PhysRevLett.91.146401. URL `https://link.aps.org/doi/10.1103/PhysRevLett.91.146401`.

[52] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37(2):785, **1988**.

[53] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, **Sep 1988**. doi: 10.1103/PhysRevA.38.3098. URL `https://link.aps.org/doi/10.1103/PhysRevA.38.3098`.

[54] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58(8):1200–1211, **1980**.

[55] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem. A*, 98(45):11623–11627, **1994**.

[56] S. Grimme. Density functional theory with london dispersion corrections. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 1(2):211–228, **2011**.

[57] S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth. Dispersion-corrected mean-field electronic structure methods. *Chem. Rev.*, 116(9):5105–5154, **2016**.

[58] A. V. Akimov and O. V. Prezhdo. Large-scale computations in chemistry: a bird's eye view of a vibrant field. *Chem. Rev.*, 115(12):5797–5890, **2015**.

[59] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.*, 157(6):479–483, **1989**.

[60] M. Head-Gordon. Quantum chemistry and molecular processes. *J. Phys. Chem. A*, 100(31):13213–13225, **1996**.

[61] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko. Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev.*, 112(1):632–672, **2012**.

[62] M. A. Collins and R. P. Bettens. Energy-based molecular fragmentation methods. *Chem. Rev.*, 115(12):5607–5642, **2015**.

[63] K. Raghavachari and A. Saha. Accurate composite and fragment-based quantum chemical models for large molecules. *Chem. Rev.*, 115(12):5643–5677, **2015**.

[64] S. L. Dixon and K. M. Merz Jr. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.*, 104(17):6643–6649, **1996**.

[65] L. Huang, L. Massa, and J. Karle. Kernel energy method illustrated with peptides. *Int. J. Quantum Chem.*, 103(6):808–817, **2005**.

[66] L. Huang, L. Massa, and J. Karle. Kernel energy method: Basis functions and quantum methods. *Int. J. Quantum Chem.*, 106(2):447–457, **2006**.

[67] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103(2):227–249, **1976**.

[68] A. D. MacKerell Jr. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.*, 25(13):1584–1604, **2004**.

[69] H. M. Senn and W. Thiel. Qm/mm methods for biomolecular systems. *Angew. Chem.*, 48(7):1198–1229, **2009**.

[70] L. Cao and U. Ryde. On the difference between additive and subtractive qm/mm calculations. *Front. Chem.*, 6:89, **2018**.

[71] W. F. Sanjuan-Szklarz, M. Woińska, S. Domagała, et al. On the accuracy and precision of x-ray and neutron diffraction results as a function of resolution and the electron density model. *IUCrJ*, 7(5), **2020**.

[72] M. Svensson, S. Humbel, R. D. Froese, et al. 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. A*, 100(50):19357–19363, **1996**.

[73] L. W. Chung, W. Sameera, R. Ramozzi, et al. The oniom method and its applications. *Chem. Rev.*, 115(12):5678–5796, **2015**.

[74] U. Ryde. The coordination of the catalytic zinc ion in alcohol dehydrogenase studied by combined quantum-chemical and molecular mechanics calculations. *J. Comput. Aided Mol. Des.*, 10(2):153–164, **1996**.

[75] U. Ryde and M. H. Olsson. Structure, strain, and reorganization energy of blue copper models in the protein. *Int. J. Quantum Chem.*, 81(5):335–347, **2001**.

[76] B. Mennucci. Continuum solvation models: What else can we learn from them? *J. Phys. Chem. Lett.*, 1(10):1666–1674, **2010**.

[77] A. Klamt, C. Moya, and J. Palomar. A comprehensive comparison of the iefpcm and ss (v) pe continuum solvation methods with the cosmo approach. *J. Chem. Theory Comput.*, 11(9):4220–4225, **2015**.

[78] A. Klamt and G. Schüürmann. Cosmo: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2*, pages 799–805, **1993**.

[79] R. Bonaccorsi, P. Palla, and J. Tomasi. Conformational energy of glycine in aqueous solutions and relative stability of the zwitterionic and neutral forms. an ab initio study. *J. Am. Chem. Soc.*, 106(7):1945–1950, **1984**.

[80] I. U. of Crystallography. *Acta Cryst.*, A48:922–946, **1992**.

[81] A. Authier and G. Chapuis. *A Little Dictionary of Crystallography*. International Union of Crystallography Chester, **2014**.

[82] R. Lifshitz. Introduction to fourier-space crystallography lecture notes for the international school on quasicrystals may 13-20, 1995, balatonfüred, hungary. *Acta Cryst*, 48:928, **1992**.

[83] S. Grabowsky, A. Genoni, and H.-B. Bürgi. Quantum crystallography. *Chem. Sci.*, 8(6):4159–4176, **2017**.

[84] S. Grabowsky, A. Genoni, S. P. Thomas, and D. Jayatilaka. The advent of quantum crystallography: form and structure factors from quantum mechanics for advanced structure refinement and wavefunction fitting. In *21st Century Challenges in Chemical Crystallography II*, pages 65–144. Springer, **2020**.

[85] G. L. Squires. *Introduction to the theory of thermal neutron scattering*. Courier Corporation, **1996**.

[86] E. Prince and A. J. C. Wilson. *International tables for crystallography*, volume 100. Kluwer., **2004**.

[87] B. Dittrich, J. Lübben, S. Mebs, et al. Accurate bond lengths to hydrogen atoms from single-crystal x-ray diffraction by including estimated hydrogen adps and comparison to neutron and qm/mm benchmarks. *Chem. Eur. J.*, 23(19):4605–4614, **2017**.

[88] D. Harker and J. Kasper. Phases of fourier coefficients directly from crystal diffraction data. *Acta Cryst.*, 1(2):70–75, **1948**.

[89] J. t. Karle and H. Hauptman. The phases and magnitudes of the structure factors. *Acta Cryst.*, 3(3):181–187, **1950**.

[90] D. Sayre. The squaring method: a new method for phase determination. *Acta Cryst.*, 5(1):60–65, **1952**.

[91] G. L. Taylor. Introduction to phasing. *Acta Crystallogr. D*, 66(4):325–338, **2010**.

[92] I. Usón and G. M. Sheldrick. Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.*, 9(5):643–648, **1999**.

[93] G. M. Sheldrick. Phase annealing in shelx-90: direct methods for larger structures. *Acta Crystallogr. A*, 46(6):467–473, **1990**.

[94] M. Rossmann and D. M. Blow. The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.*, 15(1):24–31, **1962**.

[95] M. G. Rossmann. The molecular replacement method. *Acta Crystallogr. A*, 46 (2):73–82, **1990**.

[96] D. Green, V. M. Ingram, and M. F. Perutz. The structure of haemoglobin-iv. sign determination by the isomorphous replacement method. *Proc. R. Soc. A*, 225(1162):287–307, **1954**.

[97] W. A. Hendrickson. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, 254(5028):51–58, **1991**.

[98] W. A. Hendrickson, J. R. Horton, and D. M. LeMaster. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (mad): a vehicle for direct determination of three-dimensional structure. *EMBO J.*, 9 (5):1665–1672, **1990**.

[99] M. Cianci, J. Helliwell, M. Helliwell, et al. Anomalous scattering in structural chemistry and biology. *Crystallogr. Rev.*, 11(4):245–335, **2005**.

[100] P. Müller. Practical suggestions for better crystal structures. *Crystallogr. Rev.*, 15 (1):57–83, **2009**.

[101] N. S. Pannu and R. J. Read. Improved structure refinement through maximum likelihood. *Acta Crystallogr. A*, 52(5):659–668, **1996**.

[102] P. D. Adams, N. S. Pannu, R. J. Read, and A. T. Brünger. Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl. Acad. Sci. U.S.A.*, 94(10):5018–5023, **1997**.

[103] R. A. Engh and R. Huber. Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr. A*, 47(4):392–400, **1991**.

[104] D. E. Tronrud and P. A. Karplus. A conformation-dependent stereochemical library improves crystallographic refinement even at atomic resolution. *Acta Crystallogr. D*, 67(8):699–706, **2011**.

[105] P. D. Adams, P. V. Afonine, G. Bunkóczi, et al. Phenix: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D*, 66(2):213–221, **2010**.

[106] M. Chruszcz, W. Potrzebowski, M. D. Zimmerman, et al. Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter? *Protein Sci.*, 17(4):623–632, **2008**.

[107] J.-S. Jiang and A. T. Brünger. Protein hydration observed by x-ray diffraction: solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.*, 243(1):100–115, **1994**.

[108] P. Luger. *Modern X-ray analysis on single crystals*. de Gruyter, **2011**.

[109] V. F. Sears. Scattering lengths for neutrons. In E. Prince, editor, *International tables for crystallography*, volume C, pages 444–454. Wiley Online Library, **2006**.

[110] E. Price. *International Tables for Crystallography - Volume C: Mathematical, Physical and Chemical Tables*. Wiley, Chichester, **2011**.

[111] M. Woińska, S. Grabowsky, P. M. Dominiak, et al. Hydrogen atoms can be located accurately and precisely by x-ray crystallography. *Sci. Adv.*, 2(5):e1600192, **2016**.

[112] B. Dittrich, C. B. Hübschle, M. Messerschmidt, et al. The invariom model and its application: refinement of d, l-serine at different temperatures and resolution. *Acta Crystallogr. A*, 61(3):314–320, **2005**.

[113] P. Munshi, A. Ø. Madsen, M. A. Spackman, et al. Estimated h-atom anisotropic displacement parameters: a comparison between different methods and with neutron diffraction results. *Acta Crystallogr. A*, 64(4):465–475, **2008**.

[114] D. A. Myles. Neutron protein crystallography: current status and a brighter future. *Curr. Opin. Struct. Biol.*, 16(5):630–637, **2006**.

[115] W. B. O'Dell, A. M. Bodenheimer, and F. Meilleur. Neutron protein crystallography: A complementary tool for locating hydrogens in proteins. *Arch. Biochem. Biophys.*, 602:48–60, **2016**.

[116] P. V. Afonine, M. Mustyakimov, R. W. Grosse-Kunstleve, et al. Joint x-ray and neutron refinement with phenix. refine. *Acta Crystallogr. D*, 66(11):1153–1163, **2010**.

[117] N. K. Hansen and P. Coppens. Testing aspherical atom refinements on small-molecule data sets. *Acta Crystallogr. A*, 34(6):909–921, **1978**.

[118] T. S. Koritsanszky and P. Coppens. Chemical applications of x-ray charge-density analysis. *Chem. Rev.*, 101(6):1583–1628, **2001**.

[119] J. M. Bąk, S. Domagała, C. Hübschle, et al. Verification of structural and electrostatic properties obtained by the use of different pseudoatom databases. *Acta Crystallogr. A*, 67(2):141–153, **2011**.

[120] C. Lecomte, C. Jelsch, B. Guillot, et al. Ultrahigh-resolution crystallography and related electron density and electrostatic properties in proteins. *J. Synchrotron Radiat.*, 15(3):202–203, **2008**.

[121] M. Elias, D. Liebschner, J. Koepke, et al. Hydrogen atoms in protein structures: high-resolution x-ray diffraction structure of the dfpase. *BMC Res. Notes.*, 6(1): 1–7, **2013**.

[122] K. Pröpper, J. Holstein, C. Hübschle, et al. Invariom refinement of a new monoclinic solvate of thiostrepton at 0.64 å resolution. *Acta Crystallogr. D*, 69 (8):1530–1539, **2013**.

[123] M. Malinska and Z. Dauter. Transferable aspherical atom model refinement of protein and dna structures against ultrahigh-resolution x-ray data. *Acta Crystallogr. D*, 72(6):770–779, **2016**.

[124] F. L. Hirshfeld. Bonded-atom fragments for describing molecular charge densities. *Theor. Chem. Acc.*, 44(2):129–138, **1977**.

[125] M. Woińska, D. Jayatilaka, and B. Dittrich *et al.* Validation of x-ray wavefunction refinement. *ChemPhysChem*, 18(23):3334–3351, **2017**.

[126] H. M. Senn and W. Thiel. QM/MM methods for biomolecular systems. *Angew. Chem.*, 48(7):1198–1229, **2009**.

[127] U. Ryde. QM/MM calculations on proteins. *Meth. Enz.*, 577:119–158, **2016**.

[128] U. Ryde, L. Olsen, and K. Nilsson. Quantum chemical geometry optimizations in proteins using crystallographic raw data. *J. Comput. Chem.*, 23(11):1058–1070, **2002**.

[129] F. Furche, R. Ahlrichs, C. Hättig, et al. Turbomole. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 4(2):91–100, **2014**.

[130] N. Pannu, R. Read, L. Rice, et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D*, 54:905–21, **1998**.

[131] A. T. Brunger. Version 1.2 of the crystallography and NMR system. *Nat. Protoc.*, 2(11):2728, **2007**.

[132] N. Yu, H. P. Yennawar, and K. M. Merz. Refinement of protein crystal structures using energy restraints derived from linear-scaling quantum mechanics. *Acta Crystallogr. D*, 61(3):322–332, **2005**.

[133] M. Zheng, J. R. Reimers, M. P. Waller, and P. V. Afonine. Q| r: quantum-based refinement. *Acta Crystallogr. D*, 73(1):45–52, **2017**.

[134] A. T. Brünger. Free r value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355(6359):472–475, **1992**.

[135] J. Bergmann, E. Oksanen, and U. Ryde. Quantum-refinement studies of the bidentate ligand of v-nitrogenase and the protonation state of co-inhibited mo-nitrogenase. *J. Inorg. Biochem.*, 219:111426, **2021**.

[136] T. A. Jones, J.-Y. Zou, S. t. Cowan, and M. Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, 47(2):110–119, **1991**.

[137] T. Jones and M. Kjeldgaard. [10] electron-density map interpretation. *Meth. Enzymol.*, 277:173–208, **1997**.

[138] I. J. Tickle. Statistical quality indicators for electron-density maps. *Acta Crystallogr. D*, 68(4):454–467, **2012**. doi: 10.1107/S0907444911035918.

[139] Z. Fu, X. Li, and K. M. Merz Jr. Accurate assessment of the strain energy in a protein-bound drug using qm/mm x-ray refinement and converged quantum chemistry. *J. Comput. Chem.*, 32(12):2587–2597, **2011**.

[140] S. J. Ferguson. Nitrogen cycle enzymology. *Curr. Opin. Chem. Biol.*, 2(2): 182–193, **1998**.

[141] J. W. Erisman, M. A. Sutton, J. Galloway, et al. How a century of ammonia synthesis changed the world. *Nat. Geosci.*, 1(10):636–639, **2008**.

[142] M. Capdevila-Cortada. Electrifying the haber–bosch. *Nat. Catal.*, 2(12):1055–1055, **2019**.

[143] B. K. Burgess and D. J. Lowe. Mechanism of molybdenum nitrogenase. *Chem. Rev.*, 96(7):2983–3012, **1996**.

[144] B. M. Hoffman, D. Lukoyanov, Z.-Y. Yang, et al. Mechanism of nitrogen fixation by nitrogenase: the next stage. *Chem. Rev.*, 114(8):4041–4062, **2014**.

[145] K. M. Lancaster, M. Roemelt, P. Ettenhuber, et al. X-ray emission spectroscopy evidences a central carbon in the nitrogenase iron-molybdenum cofactor. *Science*, 334(6058):974–977, **2011**.

[146] R. Bjornsson, F. A. Lima, T. Spatzal, et al. Identification of a spin-coupled mo (iii) in the nitrogenase iron–molybdenum cofactor. *Chem. Sci.*, 5(8):3096–3103, **2014**.

[147] P. Wilson, A. Nyborg, and G. Watt. Duplication and extension of the thorneley and lowe kinetic model for klebsiella pneumoniae nitrogenase catalysis using a mathematica software platform. *Biophys. Chem.*, 91(3):281–304, **2001**.

[148] R. R. Eady. Structure- function relationships of alternative nitrogenases. *Chem. Rev.*, 96(7):3013–3030, **1996**.

[149] P. Tiwari, S. Dwivedi, M. P. Singh, et al. Basic and modern concepts on cholinergic receptor: A review. *Asian Pacific J. Trop. Dis.*, 3(5):413–420, **2013**.

[150] I. Silman and J. L. Sussman. Acetylcholinesterase: how is structure related to function? *Chem.-Biol. Interact.*, 175(1-3):3–10, **2008**.

[151] A. Allgardsson, L. Berg, C. Akfur, et al. Structure of a prereaction complex between the nerve agent sarin, its biological target acetylcholinesterase, and the antidote hi-6. *Proc. Natl. Acad. Sci. U.S.A.*, 113(20):5514–5519, **2016**.

[152] W. J. Albery and J. R. Knowles. Free-energy profile for the reaction catalyzed by triosephosphate isomerase. *Biochemistry*, 15(25):5627–5631, **1976**.

[153] W. J. Albery and J. R. Knowles. Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry*, 15(25):5631–5640, **1976**.

[154] S. C. Blacklow, R. T. Raines, W. A. Lim, et al. Triosephosphate isomerase catalysis is diffusion controlled. *Biochemistry*, 27(4):1158–1165, **1988**.

[155] F. Orosz, J. Olah, and J. Ovadi. Triosephosphate isomerase deficiency: facts and doubts. *IUBMB Life*, 58(12):703–715, **2006**.

[156] F. Orosz, J. Oláh, and J. Ovádi. Triosephosphate isomerase deficiency: new insights into an enigmatic disease. *Biochim. Biophys. Acta*, 1792(12):1168–1174, **2009**.

[157] D. Banner, A. Bloomer, G. Petsko, et al. Atomic coordinates for triose phosphate isomerase from chicken muscle. *Biochem. Biophys. Res. Commun.*, 72(1): 146–155, **1976**.

[158] R. Wierenga, E. Kapetaniou, and R. Venkatesan. Triosephosphate isomerase: a highly evolved biocatalyst. *Cell. Mol. Life Sci.*, 67(23):3961–3982, **2010**.

[159] R. Wolfenden. Transition state analogues for enzyme catalysis. *Nature*, 223 (5207):704–705, **1969**.

[160] E. Lolis and G. A. Petsko. Transition-state analogues in protein crystallography: probes of the structural source of enzyme catalysis. *Annu. Rev. Biochem.*, 59(1): 597–630, **1990**.

[161] E. Lolis and G. A. Petsko. Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5-. ang. resolution: implications for catalysis. *Biochemistry*, 29(28):6619–6625, **1990**.

[162] L. A. Malaspina, E. K. Wieduwilt, J. Bergmann, et al. Fast and accurate quantum crystallography: from small to large, from light to heavy. *J. Phys. Chem. Lett.*, 10(22):6973–6982, **2019**.

[163] F. Kleemiss, O. V. Dolomanov, M. Bodensteiner, et al. Accurate crystal structures and chemical properties from nosphera2. *Chem. Sci.*, 12(5):1675–1692, **2021**.

[164] O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, et al. Olex2: a complete structure solution, refinement and analysis program. *J. Appl. Crystallogr.*, 42(2):339–341, **2009**.

[165] L. J. Bourhis, O. V. Dolomanov, R. J. Gildea, et al. The anatomy of a comprehensive constrained, restrained refinement program for the modern computing environment–olex2 dissected. *Acta Crystallogr. B*, 71(1):59–75, **2015**.

[166] F. Neese. The orca program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2(1):73–78, **2012**.

[167] C. Jelsch, M. M. Teeter, V. Lamzin, et al. Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proc. Natl. Acad. Sci. U.S.A.*, 97(7):3171–3176, **2000**.

[168] H. Bönisch, C. L. Schmidt, P. Bianco, and R. Ladenstein. Ultrahigh-resolution study on *Pyrococcus abyssi* rubredoxin. I. 0.69Å X-ray structure of mutant W4L/R5S. *Acta Crystallogr. D*, 61(7):990–1004, **Jul 2005**.

[169] F. H. Allen and I. J. Bruno. Bond lengths in organic and metal-organic compounds revisited: X—h bond lengths from neutron diffraction data. *Acta Crystallogr. B*, 66(3):380–386, **2010**.

[170] J. Bergmann, E. Oksanen, and U. Ryde. Can the results of quantum refinement be improved with a continuum-solvation model? *Acta Crystallogr. B*, 77(6), **2021**.

[171] W. J. Albery and J. R. Knowles. Efficiency and evolution of enzyme catalysis. *Angew. Chem. Int. Ed.*, 16(5):285–293, **1977**.

[172] X. Zhai, T. L. Amyes, and J. P. Richard. Role of loop-clamping side chains in catalysis by triosephosphate isomerase. *J. Am. Chem. Soc.*, 137(48):15185–15197, **2015**.

[173] D. Sippel and O. Einsle. The structure of vanadium nitrogenase reveals an unusual bridging ligand. *Nat. Chem. Biol.*, 13(9):956–960, **2017**.

[174] R. C. Pollock, H. I. Lee, L. M. Cameron, et al. Investigation of co bound to inhibited forms of nitrogenase mofe protein by 13c endor. *J. Am. Chem. Soc.*, 117(33):8686–8687, **1995**. doi: 10.1021/ja00138a033.

[175] H. I. Lee, L. M. Cameron, B. J. Hales, and B. M. Hoffman. CO Binding to the FeMo Cofactor of CO-Inhibited Nitrogenase: ^{13}CO and ^{1}H Q-Band ENDOR Investigation. *J. Am. Chem. Soc.*, 119(42):10121–10126, **1997**. doi: 10.1021/ja9715096.

[176] S. J. George, G. A. Ashby, C. W. Wharton, and R. N. F. Thorneley. Time-resolved binding of carbon monoxide to nitrogenase monitored by stopped-flow infrared spectroscopy. *J. Am. Chem. Soc.*, 119(27):6450–6451, **1997**. doi: 10.1021/ja971088s.

[177] B. J. Hales. Ethylene glycol quenching of nitrogenase catalysis: An electron paramagnetic resonance spectroscopic study of nitrogenase turnover states and co bonding. *Biochemistry*, 54(27):4208–4215, **2015**. doi: 10.1021/acs.biochem.5b00426.

[178] W. Kang, C. C. Lee, A. J. Jasniewski, et al. Structural evidence for a dynamic metallocofactor during N_2 reduction by mo-nitrogenase. *Science*, 368(6497):1381–1385, **2020**.

[179] J. Varley, Y. Wang, K. Chan, et al. Mechanistic insights into nitrogen fixation by nitrogenase enzymes. *Phys. Chem. Chem. Phys.*, 17(44):29541–29547, **2015**.

[180] L. Cao and U. Ryde. "quantum refinement with multiple conformations; applications to the p-cluster in nitrogenase". *Acta Crystal. D*, 76:1145–1156, **2020**.

[181] I. J. Tickle, A. Sharff, C. Flensburg, et al. *The STARANISO/PDBpeep Server*, **2020 (accessed December 15)**. URL `http://staraniso.globalphasing.org/172cgi-bin/PDBpeep.cgi?crit=tism&tval=1.20&ID=6ug0`.

[182] J. Bergmann, E. Oksanen, and U. Ryde. Critical evaluation of a crystal structure of nitrogenase with bound n2 ligands. *J. Biol. Inorg. Chem.*, 26(2):341–353, **2021**.

[183] C. Akfur, E. Artursson, and F. Ekströmm. Methylphosphonate adducts of acetylcholinesterase investigated by time correlated single photon counting and x-ray crystallography. *To be published.*, .

[184] E. Carletti, H. Li, B. Li, et al. Aging of cholinesterases phosphylated by tabun proceeds through o-dealkylation. *J. Am. Chem. Soc.*, 130(47):16011–16020, **2008**.

[185] C. D. Andersson, N. Forsgren, C. Akfur, et al. Divergent structure–activity relationships of structurally similar acetylcholinesterase inhibitors. *J. Med. Chem.*, 56(19):7615–7624, **2013**.

[186] F. S. Katz, S. Pecic, T. H. Tran, et al. Discovery of new classes of compounds that reactivate acetylcholinesterase inhibited by organophosphates. *ChemBioChem*, 16(15):2205–2215, **2015**.

[187] S. M. Bester, M. A. Guelta, J. Cheung, et al. Structural insights of stereospecific inhibition of human acetylcholinesterase by vx and subsequent reactivation by hi-6. *Chem. Res. Toxicol.*, 31(12):1405–1417, **2018**.

Printed by Media-Tryck, Lund 2022 NORDIC SWAN ECOLABEL 3041 0903

ISBN: 978-91-7422-878-6

Theoretical Chemistry
Department of Chemistry
Faculty of Science
Lund University