Exploring non-covalent interactions between drug-like molecules and the protein acetylcholinesterase

Lotta Berg
“Life is a relationship between molecules, not a property of any one molecule”

Emile Zuckerkandl and Linus Pauling

1962
# Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>vi</td>
</tr>
<tr>
<td>Sammanfattning</td>
<td>vii</td>
</tr>
<tr>
<td>Papers covered in the thesis</td>
<td>viii</td>
</tr>
<tr>
<td>List of abbreviations</td>
<td>x</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1. Drug discovery</td>
<td>1</td>
</tr>
<tr>
<td>1.2. The importance of non-covalent interactions</td>
<td>1</td>
</tr>
<tr>
<td>1.3. Fundamental forces of non-covalent interactions</td>
<td>2</td>
</tr>
<tr>
<td>1.4. Classes of non-covalent interactions</td>
<td>3</td>
</tr>
<tr>
<td>1.5. Hydrogen bonds</td>
<td>3</td>
</tr>
<tr>
<td>1.5.1. Definition of the hydrogen bond</td>
<td>3</td>
</tr>
<tr>
<td>1.5.2. Classical and non-classical hydrogen bonds</td>
<td>4</td>
</tr>
</tbody>
</table>
|     1.5.3. CH\(
|∙∙∙
|Y hydrogen bonds                                                       | 5    |
|   1.6. Interactions between aromatic rings                             | 6    |
|   1.7. Thermodynamics of binding                                       | 7    |
|   1.8. The essential enzyme acetylcholinesterase                      | 8    |
|     1.8.1. The relevance of research focused on AChE                  | 8    |
|     1.8.2. The structure of AChE                                      | 9    |
|     1.8.3. AChE in drug discovery                                     | 10   |
| 2. Scope of the thesis                                                | 12   |
| 3. Background to techniques and methods                                | 13   |
|   3.1. Structure determination by X-ray crystallography               | 13   |
|     3.1.1. The significance of crystal structures of proteins         | 13   |
|     3.1.2. Data collection and structure refinement                   | 13   |
|     3.1.3. Electron density maps                                      | 14   |
|     3.1.4. Modelling of ligands                                      | 15   |
|   3.2. Molecular mechanics                                            | 16   |
|     3.2.1. Force fields                                               | 16   |
|     3.2.2. Molecular docking                                          | 17   |
|   3.3. Quantum mechanics                                              | 17   |
|     3.3.1. Quantum mechanics to study non-covalent interactions       | 17   |
|     3.3.2. Quantum mechanical methods                                 | 18   |
|     3.3.3. The basis set                                              | 19   |
|     3.3.4. Dispersion correction                                      | 19   |
|     3.3.5. Benchmarking methods                                       | 20   |
|     3.3.6. Symmetry-adapted perturbation theory                       | 20   |
|     3.3.7. Electrostatic potential maps                               | 21   |
|   3.4. Biochemical characterization                                   | 21   |
|     3.4.1. The Ellman assay                                           | 21   |
|     3.4.2. Isothermal titration calorimetry                           | 22   |
4. Targeting acetylcholinesterase 24
  4.1. Identification of chemical leads 24
  4.2. Bioactive conformations of AChE inhibitors 25
  4.3. Molecular docking to AChE 25
  4.4. Contribution and significance of the results 26
  4.5. Further reading 26

5. X-ray crystal structures of acetylcholinesterase-ligand complexes 27
  5.1. Available crystal structures of mAChE 27
  5.2. From the protein’s point of view 27
  5.3. From the ligands’ point of view 30
  5.4. Contribution and significance of the results 32
  5.5. Further reading 33

6. Processing of X-ray crystal structures using density functional theory 34
  6.1. Why X-ray crystal structures need additional processing 34
  6.2. The quantum chemical cluster approach 35
    6.2.1. Pre-processing of the protein-ligand complexes using molecular mechanics 35
    6.2.2. Construction of reduced model systems 35
    6.2.3. Geometry optimizations of reduced models using DFT 37
  6.3. Analysis of non-covalent interactions 37
  6.4. Evaluation of multiple conformations 39
  6.5. Contribution and significance of the methodology 39
  6.6. Further reading 40

7. Similar but different: Enantiomeric inhibitors of acetylcholinesterase 41
  7.1. From the racemate to the pure enantiomers 41
  7.2. Biochemical characterization of (R)- and (S)-C5685 42
  7.3. Structural interpretation of enthalpy-entropy compensation 43
  7.4. Contribution and significance of the results 45
  7.5. Further reading 46

8. The nature of activated CH···Y hydrogen bonds 47
  8.1. Exploring non-covalent interactions in AChE-ligand complexes 47
  8.2. Characterization of individual non-covalent interactions using DFT 49
    8.2.1 Interaction energies in vacuum 49
    8.2.2. Distance dependence 49
    8.2.3. Effect of the environment 51
    8.2.4. Effect of the formal charge 51
  8.3. The fundamental forces of activated CH···Y hydrogen bonds 51
  8.4. Interaction energies obtained by force fields 52
  8.5. Contribution and significance of the results 53
  8.6. Further reading 54

9. On the reactivation of nerve-agent inhibited acetylcholinesterase by antidotes 55
  9.1. Elucidating the reactivation mechanism 55
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.2. Crystallographic and DFT refinement of HI6-sarin-μAChE</td>
<td>55</td>
</tr>
<tr>
<td>9.3. Validation and interpretation of the HI-6-sarin-μAChE structure</td>
<td>58</td>
</tr>
<tr>
<td>9.4. Contribution and significance of the results</td>
<td>60</td>
</tr>
<tr>
<td>9.5. Further reading</td>
<td>61</td>
</tr>
<tr>
<td>10. Concluding remarks</td>
<td>62</td>
</tr>
<tr>
<td>11. Acknowledgements</td>
<td>65</td>
</tr>
<tr>
<td>12. References</td>
<td>68</td>
</tr>
<tr>
<td>Appendix I</td>
<td>1</td>
</tr>
</tbody>
</table>
Abstract

The majority of drugs are small organic molecules, so-called ligands, that influence biochemical processes by interacting with proteins. The understanding of how and why they interact and form complexes is therefore a key component for elucidating the mechanism of action of drugs. The research presented in this thesis is based on studies of acetylcholinesterase (AChE). AChE is an essential enzyme with the important function of terminating neurotransmission at cholinergic synapses. AChE is also the target of a range of biologically active molecules including drugs, pesticides, and poisons. Due to the molecular and the functional characteristics of the enzyme, it offers both challenges and possibilities for investigating protein-ligand interactions. In the thesis, complexes between AChE and drug-like ligands have been studied in detail by a combination of experimental techniques and theoretical methods. The studies provided insight into the non-covalent interactions formed between AChE and ligands, where non-classical CH···Y hydrogen bonds (Y = O or arene) were found to be common and important. The non-classical hydrogen bonds were characterized by density functional theory calculations that revealed features that may provide unexplored possibilities in for example structure-based design. Moreover, the study of two enantiomeric inhibitors of AChE provided important insight into the structural basis of enthalpy-entropy compensation. As part of the research, available computational methods have been evaluated and new approaches have been developed. This resulted in a methodology that allowed detailed analysis of the AChE-ligand complexes. Moreover, the methodology also proved to be a useful tool in the refinement of X-ray crystallographic data. This was demonstrated by the determination of a prereaction conformation of the complex between the nerve-agent antidote HI-6 and AChE inhibited by the nerve agent sarin. The structure of the ternary complex constitutes an important contribution of relevance for the design of new and improved drugs for treatment of nerve-agent poisoning. The research presented in the thesis has contributed to the knowledge of AChE and also has implications for drug discovery and the understanding of biochemical processes in general.

Keywords: acetylcholinesterase, drug discovery, density functional theory, hydrogen bond, nerve-agent antidote, non-covalent interaction, protein-ligand complex, structure-based design, thermodynamics, X-ray crystallography
Sammanfattning

*En studie av icke-kovalenta interaktioner mellan läkemedelslika molekyler och proteinet acetylkolinesteras*


Med hjälp av både experimentella och teoretiska metoder har olika aspekter som är relevanta för uppkomsten av ett protein-ligand komplex studerats. Detta har bland annat resulterat i en ökad förståelse för egenskaperna och drivkraften hos en speciell klass av icke-kovalenta interaktioner, så kallade icke-klassiska vätebindningar, som visade sig vara vanligt förekommande och viktiga mellan ligander och acetylkolinesteras. Icke-kovalenta interaktioner skiljer sig från de kovalenta bindningar som binder samman atomer till molekyler, och kan generellt anses vara svagare och kan därför bildas och brytas lättare än kovalenta bindningar. Icke-kovalenta interaktioner är viktiga i många kemiska processer, inte minst när det gäller hur biomolekyler som proteiner ser ut, hur de fyller sin funktion, och hur de interagerar med andra molekyler. Som del av forskningen har befintliga beräkningsmetoder utvärderats och nya tillvägagångssätt har utvecklats. Resultaten har gett nya insikter i, och potentiella möjligheter för, både läkemedelsutveckling och studier av biokemiska processer i kroppen.
Papers covered in the thesis


*The authors contributed equally to the work

Reprints were made with permission from the publishers
List of abbreviations

ACh  acetylcholine
AChE  acetylcholinesterase
Asp  aspartic acid
cf.  confer (“compare”)
CAS  catalytic site
CBS  complete basis set
CCSD(T)  coupled cluster singles doubles with perturbative triples
DFT  density functional theory
DMF  dimethylformamide
dRMSD  directional root-mean-square deviation
DMSO  dimethyl sulfoxide
DTNB  dithiobisnitrobenzoate
e.g.  exempli gratia (“for example”)
\(E_D\)  dispersion correction energy
\(E_{DFT}\)  density functional theory energy
\(E_{DFT-D}\)  dispersion corrected density functional theory energy
\(E_{\text{disp}}\)  dispersion energy
\(E_{\text{elec}}\)  electrostatic energy
\(E_{\text{exch-rep}}\)  exchange-repulsion energy
\(\Delta E_{\text{gas}}\)  interaction energy in gas phase
\(E_{\text{ind}}\)  induction energy
ESP  electrostatic potential
\(\Delta E_{\text{solv}}\)  solvated interaction energies
et al.  et alii (“and others”)
\(E_{\text{tot}}\)  total energy
\(F_c\)  calculated reflection amplitudes
\(F_o\)  observed reflection amplitudes
\(\Delta G\)  change in Gibbs free energy
Glu  glutamic acid
Gly  glycine
\(\Delta G_{\text{solv}}\)  free energy of solvation
\(h\text{AChE}\)  \(Homo\ sapiens\) acetylcholinesterase
\(\Delta H\)  change in enthalpy
\(\Delta HF\)  higher order terms
His  histidine
HTS  high throughput screening  
*i.e.*  *id est* ("that is")  
$IC_{50}$  half maximal inhibitory concentration  
ITC  isothermal titration calorimetry  
IUPAC  International Union of Pure and Applied Chemistry  
$K_a$  equilibrium association constant  
L  ligand  
LCBU  Laboratory for Chemical Biology Umeå  
mAChE  *Mus musculus* acetylcholinesterase  
MD  molecular dynamics  
MM  molecular mechanics  
NMR  nuclear magnetic resonance  
OP  organophosphorous  
P  protein  
PAS  peripheral anionic site  
PC  principal component  
PCA  principal component analysis  
PDB  protein data bank  
Phe  phenylalanine  
PL  protein-ligand  
QM  quantum mechanics  
QSAR  quantitative structure-activity relationship  
R  ideal gas constant  
RMSD  root-mean-square deviations  
$\Delta S$  change in entropy  
SAPT  symmetry-adapted perturbation theory  
Ser  serine  
T  absolute temperature  
Trp  tryptophan  
Tyr  tyrosine  
WFT  wave function theory  
Å  Ångström
1. Introduction

- The aim of this chapter is to give a general introduction to aspects of medicinal chemistry and related fields that are relevant to the research presented in chapters 4-9 and the appended papers.
- Acetylcholinesterase as a research topic is presented and some important features of the enzyme are described.

1.1. Drug discovery

The development of a drug is a long and costly process that is typically initiated by the identification of a therapeutically relevant target. Once the target has been validated, compounds that interact with the target need to be identified. This can for example be achieved by screening large libraries of compounds in vitro (i.e. high throughput screening; HTS). Drug candidates that are suitable for clinical trials can thereafter be obtained by optimizing the pharmacodynamic and pharmacokinetic properties of a number of lead compounds identified in the HTS. A drug discovery program typically involves experts from many disciplines, including organic chemists to synthesize new compounds and biochemists to evaluate them. In addition, computational methods are routinely applied at different stages of the drug development process.\[1\] Computational approaches to design new compounds include both structure- and ligand-based design. If the three-dimensional structure of the target is known, structure-based design (e.g. molecular docking\[2\]) can be applied whereas ligand-based approaches can be used if it is unknown (e.g. establishment of quantitative structure-activity relationships; QSAR\[3\]).

1.2. The importance of non-covalent interactions

Drugs produce their effect by interacting (either covalently or non-covalently) with their biological target in the body. The majority of the approved drugs are small organic molecules (herein denoted ligands) that alter biochemical processes by interacting with proteins, the most common being G-protein coupled receptors, nuclear receptors, and ligand- or voltage-gated ion channels.\[4,5\] Elucidating the mechanism behind the binding of small organic compounds to proteins is therefore highly relevant both for drug discovery and for understanding numerous biochemical processes that depend on the binding of a ligand to a protein.
A fundamental understanding of the ligand-binding event requires insight into e.g. the non-covalent interactions that stabilize a protein-ligand complex as well as the dynamics and thermodynamics of the system. Non-covalent interactions differ from covalent bonds in that no electrons are shared between the participating atoms. Non-covalent interactions are therefore generally weaker than covalent bonds. They are nonetheless specific, attractive, and (importantly) reversible and can be formed and broken without being associated with a large energy cost. Non-covalent interactions play a key role in biological systems by impacting the structure, dynamics and function of biomolecules.[6,7] In addition, they also influence pharmacokinetic properties of molecules such as solubility, partitioning, distribution and permeability, that are highly important parameters in drug development.[8]

In a protein-ligand complex, non-covalent interactions may be formed both intramolecularly between the amino acids of the protein, and intermolecularly between the protein and the ligand. Simultaneously occurring interactions in a complex may influence each other, and the total interaction energy of a number of non-covalent interactions can be either greater or less than the sum of the interaction energies of the individual interactions. Non-covalent interactions can therefore interact in either a “positive cooperative” or “negative cooperative” manner.[9,10]

1.3. Fundamental forces of non-covalent interactions

The adopted geometry of a protein-ligand complex will represent the most energetically favorable complex resulting from a compromise between both attractive and repulsive forces. The main forces that contribute to the interaction energies of a non-covalent interaction arise as a result of interacting multipoles (either permanent, instantaneous, or induced) and can be described as:

· Electrostatic: an attractive or repulsive force that arises as a result of the interaction between two permanently charged molecules, two polar molecules, or a permanently charged molecule and a polar molecule

· Induction: an attractive force that arises between an instantaneous dipole induced in a nonpolar molecule by an electric field caused by a permanently charged or polar molecule

· Dispersion: an attractive force that arises as a result of correlated electron fluctuations in two nonpolar molecules, also referred to as London dispersion forces
· Exchange-repulsion: a short range repulsive force that arises between two molecules as a result of overlapping electron densities

The interaction energies of a non-covalent complex can be computationally decomposed into the above listed energy terms by symmetry-adapted perturbation theory (SAPT) calculations, as described in section 3.3.6 and presented in chapter 8.

1.4. Classes of non-covalent interactions

Non-covalent interactions are typically categorized according to e.g. their geometrical preferences, their interaction strength, and the composition of the fundamental forces that contribute to their interaction energies. As there are no strict rules for categorizing non-covalent interactions, there may be discrepancies (not necessarily contradictions) in the literature. The classification presented in this thesis might, in other words, differ from the classification made by others. One should therefore keep in mind that it is not the name used to denote the interaction that is of importance, but rather the properties of the interaction itself that matters.

In this thesis, the focus has been on different classes of hydrogen bonds and aromatic interactions. These interactions are briefly described in the following paragraphs. There are, however, other types of interactions that are important in protein-ligand complexes, for example ionic bonds, halogen bonds, and hydrophobic interactions.[11]

1.5. Hydrogen bonds

1.5.1. Definition of the hydrogen bond

The hydrogen bonds are an extensively studied class of non-covalent interactions that is highly important in proteins and protein-ligand complexes. To be characterized as a hydrogen bond according to the International Union of Pure and Applied Chemistry (IUPAC) definition, the interaction needs to fulfill a number of criteria.[12,13] Some key features of a hydrogen bond are presented in the following sections.

The hydrogen bond is an attractive interaction between a hydrogen bound to an atom more electronegative than H (donor), and a second electronegative atom (acceptor). Hydrogen bonds have clear geometrical preferences (i.e. they are directional), where the strength of the interaction is affected by both the angle and distance between the donor and the acceptor. Moreover, hydrogen bonds often reinforce each other in a positive cooperative manner. The forces
that contribute to the interaction energy in a hydrogen bond include
electrostatics, induction, and dispersion. In addition, the hydrogen bond
exhibits a partial covalent character as a result of charge transfer between the
donor and acceptor atoms. Experimental evidence of hydrogen bond
formation can be obtained by, for example, infrared spectroscopy or nuclear
magnetic resonance (NMR).

Depending on the nature of the donor and acceptors, hydrogen bonds can vary
in strength between ~0.5-40 kcal/mol. Examples of different types of
hydrogen bonds considered in this thesis are presented in Figure 1.

![Figure 1. Examples of hydrogen bonds considered in this thesis, including “classical” hydrogen bonds and “non-classical” aromatic or CH···Y hydrogen bonds.](image)

**1.5.2. Classical and non-classical hydrogen bonds**

Hydrogen bonds can be classified as “classical” or “non-classical” depending
on the participating donor and acceptor groups. Classical hydrogen bonds
typically involve strong acceptors and strong donors (e.g. N or O and H bound
to either N or O) and are usually strong, highly directional, and dominated by
electrostatic forces. Non-classical hydrogen bonds, on the other hand, involve either a weak donor and a strong acceptor (e.g. CH···O), a strong donor and a weak acceptor (e.g. OH···arene) or a weak donor and a weak acceptor (e.g. CH···arene). The non-classical hydrogen bonds typically have a larger dispersion component compared to the classical hydrogen bonds.

Non-classical hydrogen bonds are generally considered weak (<4 kcal/mol). In fact, they are sometimes referred to as “weak hydrogen bonds” in the literature. Despite their lower interaction strengths, it has been suggested that the non-classical hydrogen bonds may be as important as their classical analogues in protein-ligand complexes by being associated with lower desolvation costs and/or by cooperative effects.

**1.5.3. CH···Y hydrogen bonds**

It has been clearly demonstrated that CH groups can participate in hydrogen bonds (so-called CH···Y type hydrogen bonds, where Y may represent O or arene). The ability of CH to act as a hydrogen bond donor is dependent on the polarization of the CH that can be affected by e.g. the hybridization of the carbon atom or by the number and strength of electron withdrawing groups bound to it. The polarization of the CH will ultimately be reflected in the directionality, the electrostatic component, and the strength of the resulting hydrogen bond. With decreasing polarization of the CH, the interactions might represent van der Waals interactions rather than hydrogen bonds. Similar to the hydrogen bond donors, the hydrogen bond acceptor abilities of arenes may be influenced by the electronic properties of substituents on the aromatic ring, where electron donating substituents have been shown to enhance the strength of the hydrogen bond.

The introduction of a positively charged functional group (e.g. a charged amine) in the hydrogen-bond donor molecule has shown to have a significant effect on the strength of both CH···O and CH···arene hydrogen bonds. The introduction of a charged will result in more electron deficient CH-donors (activated CH), and the resulting activated CH···Y hydrogen bond will generally be stronger and have a larger electrostatic component compared to the corresponding neutral CH···Y hydrogen bond (Figures 1 and 2). Activated CH···arene hydrogen bonds are sometimes categorized as cation-π interactions, which have been extensively studied in model systems and proteins. Interactions between ligands containing a positively charged ammonium group and an aromatic ring are herein categorized as hydrogen bonds as it has been shown that it is often the CH that facilitates the interaction with the arene.
Figure 2. Calculated interaction energies (\(\Delta E_{\text{gas}}\)) in kcal/mol of non-covalent interactions between the side chain of Tyr (phenol) and either a charged or neutral amine (A or B, respectively), or an alkane (C). The interaction energies were calculated using the BLYP-D3/aug-cc-pVTZ method (paper 3).

Non-classical CH···Y hydrogen bonds have been shown to be abundant in both proteins and in protein-ligand complexes.\(^{[7,28-30]}\) Although the role of CH···Y hydrogen bonds in biological systems has not been fully elucidated, they are believed to be important and that they can be utilized in drug development.\(^{[7,17,28,31,32]}\)

1.6. Interactions between aromatic rings

Interactions between aromatic rings are common and important in biological systems where they influence the structure of both proteins and protein-ligand complexes.\(^{[25,33]}\) In a protein-ligand complex, aromatic interactions can be formed either intramolecularly between aromatic residues in the protein or intermolecularly to aromatic rings in ligands. Aromatic interactions are typically observed in one of the three geometries presented in Figure 3, where the parallel-displaced and T-shaped edge-to-face represent the preferred geometries. As an example, the parallel-displaced and T-shaped edge-to-face geometries of the benzene dimer are equally attractive (~2.8 kcal/mol) whereas the face-to-face geometry is less attractive (~1.8 kcal/mol).\(^{[34]}\) Substituents on the aromatic ring can influence the interaction energies. The aromatic interactions can become either stronger or weaker depending on the adopted geometry of the aromatic systems and the electronic properties of the substituent.\(^{[35,36]}\)
In addition to the aromatic interactions described above, aromatic rings can also act as both hydrogen bond acceptors and donors in non-classical hydrogen bonds (see section 1.5.2.[15,38]). The CH of an aromatic system can, for example, act as a hydrogen bond donor and the T-shaped edge-to-face interaction can thereby also be categorized as a CH···arene hydrogen bond (see section 1.5.3.).

1.7. Thermodynamics of binding

The formation of a non-covalent protein-ligand complex is a reversible event where the free protein (P) and ligand (L) states are at equilibrium with the protein-ligand bound state (PL; Eq.1). The equilibrium is determined by the Gibbs free energy (ΔG) and the related equilibrium association constant (K_a) according to Eq. 1-3.

\[
[P] + [L] \rightleftharpoons [PL] \quad \text{Eq.1}
\]

\[
K_a = \frac{[PL]}{[P][L]} \quad \text{Eq. 2}
\]

\[
\Delta G = -RT\ln K_a \quad \text{Eq. 3}
\]

Where R is the ideal gas constant and T is the absolute temperature. A change in ΔG of 1.4 kcal/mol is equivalent to a 10-fold change in K_a.[38]

In order for a non-covalent complex to be formed, the process must be energetically favorable (i.e. the ΔG needs to negative). As shown in Eq. 4, the ΔG is the sum of an enthalpic (ΔH) and an entropic term (-TΔS). These terms can be obtained from isothermal titration calorimetry experiments (ITC; see section 3.4.2).

\[
\Delta G = \Delta H - T\Delta S \quad \text{Eq. 4}
\]
The thermodynamics of protein-ligand binding is influenced by several factors, for example non-covalent interactions, desolvation and dynamics.[39] The enthalpic term is directly related to the presence and strength of non-covalent interactions[40,41] while the entropic term is related to the change in order of the system e.g. conformational degrees of freedom[42] or the displacement of water from the binding surface (i.e. desolvation)[39,43-44].

The binding of a ligand to a protein can be either entropically or enthalpically dominated depending on the contribution of $\Delta H$ and $-T\Delta S$ to the $\Delta G$.[45] In drug discovery, it has been found that more potent compounds are more often obtained as a result of an improved entropic rather than enthalpic term.[46] One issue that needs to be overcome to improve the binding affinity of a ligand is the commonly observed “enthalpy-entropy compensation phenomenon”,[39,45-47] Although the origins of enthalpy-entropy compensation have not been fully elucidated, it has been found that an enthalpic gain resulting from the formation of e.g. additional and/or stronger hydrogen bonds between the ligand and the protein is often associated with an entropic penalty as the complex will be more restricted (i.e. have fewer accessible arrangements). The stronger and more directional the interaction, the more it opposes motion, and the larger the penalty to the entropic term.[10]

1.8. The essential enzyme acetylcholinesterase

1.8.1. The relevance of research focused on AChE

Acetylcholinesterase (AChE) is an essential enzyme with the important function of terminating neurotransmission by hydrolysis of the neurotransmitter acetylcholine at cholinergic synapses (ACh; Figure 4). Due to the molecular and functional characteristics of the enzyme, AChE represents a system that offers both challenges and possibilities for investigating a wide range of aspects relevant to e.g. enzymatic catalysis, protein-ligand recognition, molecular modeling, and drug discovery. Key features of AChE is the deep and highly aromatic active site gorge[48] (see section 1.8.2.), the high catalytic turnover rate[49], the strong electrostatic dipole moment[50], and the complex substrate trafficking[51,52].

![Enzymatic hydrolysis of acetylcholine into choline and acetic acid by AChE.](image)

**Figure 4.** Enzymatic hydrolysis of acetylcholine into choline and acetic acid by AChE.
AChE is also the target of a range of biologically active molecules including drugs, pesticides, and poisons. One class of drugs that target AChE are the so-called nerve-agent antidotes, the design of which poses a challenge in drug discovery. Nerve-agent antidotes are reactive molecules that exercise their physiological effect by removing a covalently bound nerve-agent adduct in the active site of AChE by a chemical reaction. They are, importantly, required to do so in a specific and non-toxic fashion. Although the reactivators were first discovered in the 1950s,[53] the detailed reactivation mechanism is still unknown. Exploring their mechanism of action is of interest from both an academic and a drug discovery perspective.

1.8.2. The structure of AChE

The determination of the X-ray crystal structure of AChE from the Pacific electric ray (Torpedo californica) revealed several important features of the enzyme.[48,54] AChE is an α/β protein belonging to the serine hydrolase family. The catalytic triad of the enzyme, consisting of the catalytic residue Ser203, Glu334 and His447, was found close to the base of a ~20 Å deep and highly aromatic gorge (Figure 5). The active site of AChE consists of two subsites, the peripheral anionic site (PAS) located at the entrance of the gorge and the catalytic site (CAS) at the bottom of the gorge. The CAS can in turn be divided into four subsites related to their role in the catalytic reaction, i.e. the anionic site that interacts with the quaternary ammonium moiety of ACh, the esteratic site containing the catalytic triad, the oxyanion hole that stabilizes the transition state, and the acyl pocket that confers substrate specificity.[53]

Figure 5. The three-dimensional structure of AChE determined by X-ray crystallography with the PAS and the CAS illustrated by a dark grey and light grey molecular surface, respectively.
Within this thesis, mainly structures of *Mus musculus* acetylcholinesterase (mAChE) have been studied and the numbering of the residues of other acetylcholinesterases referred to in the text have been altered to the corresponding to mAChE position. The sequence identity of mAChE compared to *Homo sapiens* acetylcholinesterase (hAChE) is ~88% and all residues in the binding site are identical.

1.8.3. **AChE in drug discovery**

From a medicinal point of view, inhibitors of AChE are used for symptomatic treatment of *e.g.* Alzheimer’s disease and myasthenia gravis.[56] Both covalent and non-covalent drugs have been approved for treatment of cholinergic deficiencies. One example is donepezil (marketed under the trade name Aricept) that was approved by the United States Food and Drug Administration in 1996.[57,58]

AChE is also the target of both natural toxins and man-made poisons, for example the snake venom fasciculin and the nerve agent sarin. Nerve agents are highly poisonous organophosphorus (OP) compounds that inhibit the function of AChE by covalently binding to the catalytic Ser203 residue (OP-AChE; Figure 6). Currently available treatment of nerve-agent intoxication involves the use of oxime-based antidotes (*e.g.* HI-6, 2-PAM and obidoxime).[59] These nerve-agent antidotes are able to restore the enzymatic activity of OP-AChE by cleaving the bond between the nerve agent adduct and the Ser203 residue by nucleophilic attack of the oxime functional group. The therapeutic value of reactivators of inhibited AChE is not limited to the treatment of victims of nerve-agent exposure. Nerve-agent antidotes may also be used to treat poisoning caused by other toxic organophosphorus compounds, *e.g.* pesticides that bind to AChE.[59] Pesticide poisoning is unfortunately a well-recognized problem that is associated with health problems and a large number of fatalities worldwide.[60]
Figure 6. Schematic overview of the inhibition, aging, and reactivation of AChE exemplified by the nerve agent sarin and the antidote HI-6.

Due to a number of limitations associated with the currently used nerve-agent antidotes, there is a need for new antidotes with improved pharmacokinetic and reactivating properties. The limitations include poor blood-brain barrier permeability and inability to reactivate aged OP-AChE (Figure 6). Moreover, the current antidotes have been shown to be highly dependent on the nerve agent bound to Ser203 and are therefore not regarded as "broad-spectrum".

2. Scope of the thesis

The research presented in this thesis aims to provide insight into aspects of the formation of a protein-ligand complex at a molecular level. This has been achieved by studying complexes between the essential enzyme AChE and small drug-like molecules by a combination of biochemistry, X-ray crystallography, and computational chemistry. Special attention has been paid to the non-covalent interactions in X-ray crystal structures of AChE-ligand complexes that have been analyzed and subsequently characterized using quantum mechanical methods. The performance of currently used structure-based design methods has been evaluated as part of the research and the potential significance of the results for drug discovery is discussed.
3. Background to techniques and methods

- A brief introduction to the theoretical background of the experimental techniques and computational methods that have been utilized in this thesis is provided in this chapter.

- The aim is to provide the non-expert reader with some key aspects relevant to the context of the research presented in chapters 4-9 and the appended papers.

3.1. Structure determination by X-ray crystallography

3.1.1. The significance of crystal structures of proteins

X-ray crystallography has made significant contributions to many fields of science and has played an important role in as many as 29 awarded Nobel Prizes.[65] It is currently the most common technique for determining three-dimensional structures of macromolecules such as proteins.[66] Since the first protein structure was determined in the 1950s,[67] X-ray crystallography has provided important insights not only to the structure, but also the function of proteins. The possibility to study protein-ligand complexes at a molecular level has also provided an understanding of many aspects of ligand binding.[68] X-ray crystal structures have also been successfully utilized in the drug discovery process, where structure-based design efforts have resulted in a number of drugs and clinical candidates, e.g. HIV-protease inhibitors.[69]

In this thesis, crystal structures of mAChE-ligand complexes have been studied to investigate different aspects of the ligand binding event (chapters 4-9 and papers 1-4).

3.1.2. Data collection and structure refinement

The determination of a structural model using X-ray crystallography is made in several consecutive steps (Figure 7). The process starts with the collection of a large number of reflection intensities in so-called diffraction patterns. These constitute the experimental data and arise as a result of the scattering of X-rays by the electrons in the studied molecules. As the data from individual molecules is too weak to measure, it has to be amplified by using crystals of the protein of interest.
The reflections are converted to structure factor intensities or amplitudes. In addition, the phase of each structure factor has to be determined. For the mAChE structures studied in this thesis, initial phases were obtained by using a previously determined mAChE structure (PDB code: 1J06) in a “rigid-body refinement”. The initial coordinates and structure factors are thereafter refined in several cycles to improve the phase information of the structure factors, which in turn allows for more accurate molecular models to be built. The refinement of the structure involves both optimizations by a refinement program (e.g. within the Phenix software suite\cite{70}) and manual building of the model by the crystallographer based on the interpretation of the electron density maps (see section 3.1.3). In the refinement process, geometrical restraints are used to ensure that the model is reasonable in terms of bond lengths, angles and torsions, where lower resolution structures are typically more dependent on geometric restraints than high resolution structures. The resolution of the structure relates to the level of details that can be distinguished in the electron density maps.\cite{71}

The $R$-factor and the $R_{\text{free}}$\cite{72} are statistical metrics of the agreement between the observed structure factors ($F_o$) and those calculated from the model ($F_c$), and are commonly used global quality measures of the final model. Although these measures relate to the overall quality of the data, they do not contain information about the local quality. To evaluate the local quality of a ligand modeled into a protein, the electron density maps are commonly used.

3.1.3. Electron density maps

The experimental diffraction data is visualized as electron density maps that are interpreted during the model building and refinement process.\cite{71} The most commonly used electron density maps are the $2F_o-F_c$ and $F_o-F_c$ maps. These maps are constructed by subtracting the calculated (model) structure factors ($F_c$) from the observed (experimental) structure factors ($F_o$). A region where the $2F_o-F_c$ map covers the atoms of the structural model agrees with the experimental data, while electron density without any modeled atoms
indicates areas where atoms are missing in the model. Conversely, atoms that are not covered by electron density are incorrectly placed in the model or are not defined by the experimental data. The interpretation of the $2F_o-F_c$ map can be aided by visualization of the $F_o-F_c$ map that highlights where the model and experimental data differ (e.g. missing or incorrectly placed atoms).

### 3.1.4. Modelling of ligands

Structure determinations of the protein-ligand complexes presented in this thesis have been achieved by soaking the protein crystals with dissolved ligand prior to the diffraction experiment. If an electron density that corresponds to the shape of the ligand was observed in the initial model, the ligand was modelled to fit the density. The modeling of a ligand in the electron density is exemplified in Figure 8 by the (S)-C5685-mAChE complex that is presented and analyzed in chapter 7 and paper 2.

![Figure 8](image)

**Figure 8.** Illustration of crystallographic modeling of a ligand. **A.** Initial electron density maps of active site residues obtained after rigid-body refinement. The $2F_o-F_c$ (blue) and $F_o-F_c$ (positive electron density close to the presumed ligand and Tyr337 is shown in green) are shown at a contour level of 1 and 3 $\sigma$, respectively. **B.** The refined (S)-C5685-mAChE X-ray crystal structure featuring a new conformation of Tyr337 and (S)-C5685 (carbons colored yellow) refined to an occupancy of 0.91. The $2F_o-F_c$ map is shown in blue at a contour level of 1 $\sigma$. Water molecules are omitted in the figure for clarity.

Modeling and refining the coordinates of the ligand in a protein-ligand complex is not always straightforward. In comparison to the protein, the electron density associated with the ligand is often not as well-defined (may
be due to multiple binding modes, higher thermal motion or conformational disorder of the ligand) and the ligand geometries are therefore often less certain.[73,74] Also, obtaining chemically correct ligand geometries from the refinement can be challenging as it is more difficult to define geometrical restraints for small molecules than proteins as a result of their larger chemical diversity compared to the amino acids (i.e. in terms of their composition of atoms and conformational flexibility).[66] The modeling of the ligand usually does not have a dramatic effect on the global quality indicators such as the $R$-factor or $R_{free}$.[75] Instead, the crystallographer’s interpretation of the ligand model is usually supported by an omit map. The omit map is a $F_o-F_c$ difference map that has been calculated for the final model after several cycles of “simulated annealing” refinement in which the modeled ligand has been omitted from the coordinates. The omit map gives an indication of the local quality of the ligand model by representing an “unbiased” density corresponding to the omitted atoms.

### 3.2. Molecular mechanics

#### 3.2.1. Force fields

Molecular mechanics (MM) are used in many applications of computational chemistry, e.g. conformational search methods, docking and scoring, and molecular dynamics (MD) simulations. In comparison to quantum mechanics (QM see section 3.3.), MM uses a simplified representation of atoms and bonds where the electrons are neglected. Molecules are instead considered as a collection of balls (atoms) connected by springs (the chemical bonds; Figure 9). The MM methods thereby allow fast calculations on large systems, but are unable to account for quantum effects such as polarization and charge transfer.

In MM, force fields are used to determine the energy of a system. A force field is a mathematical function that includes terms for stretching of bonds, bending of angles and rotation of torsion angles (Figure 9). Non-bonded interactions are typically considered by a Coulombic term representing the electrostatic interactions and a Lennard-Jones function for the van der Waals interactions. Force fields are empirical and may be parametrized using e.g. experimental or QM data.
3.2.2. Molecular docking

Molecular docking is commonly used in structure-based drug design, both for lead identification and lead optimization.\(^2\) The aim of molecular docking is to predict bioactive conformations of ligands in the binding site of a protein and to subsequently estimate the binding affinity of the docked ligand. During the docking, many different orientations of the ligand are generated (docking poses) by a search algorithm. A scoring function is thereafter used to identify the pose with the best fit in the active site and finally to estimate the binding affinity of the ligand to the protein. The scoring functions are relatively simple energy functions that are a compromise between speed and accuracy, and can be classified as either empirical, force-field based, or knowledge-based.\(^76\)

There are a number of different available docking software that performs the docking in different ways, for example in terms of how they treat protein and ligand flexibility during the docking and the algorithms that are used for generating the ligand conformations. In this thesis, the potential use of molecular docking in structure-based design of AChE inhibitors has been investigated (chapter 4 and paper 1).

3.3. Quantum mechanics

3.3.1. Quantum mechanics to study non-covalent interactions

QM methods are valuable tools not only in chemistry, but also in other fields such as materials science and nanoscience.\(^77\) In computational chemistry, QM has been integrated in for example docking and ligand-affinity prediction methods and QSAR.\(^78,79\) QM is also commonly used for parameterization of MM force fields (see section 3.2.1).
In this thesis, QM has been used to study non-covalent interactions in terms of their interaction energies and fundamental forces (chapter 8 and paper 3). As isolated model systems can be studied in the absence of competing interactions or solvation effects, the use of QM calculations allows non-covalent interactions to be characterized beyond experimental methods.[80] The interaction energies ($\Delta E_{\text{gas}}$) of a non-covalent interaction can be estimated by subtracting the calculated energies of the interacting monomers ($E_{\text{monomer1}}$ and $E_{\text{monomer2}}$) from the total energy of the complex ($E_{\text{complex}}$) according to Eq. 5.

$$\Delta E_{\text{gas}} = E_{\text{complex}} - E_{\text{monomer1}} - E_{\text{monomer2}}$$  Eq. 5

QM calculations have also been applied both to fine-tune X-ray crystal structures (described in chapter 6) and has been integrated into the refinement of X-ray crystallographic data (chapter 9 and paper 4).

In the following sections, a brief introduction to the QM methods of relevance to this thesis is provided. For further details, the provided references as well as relevant textbooks[81,82] are recommended.

### 3.3.2. Quantum mechanical methods

The quantum behavior of atoms and molecules is described by the Schrödinger equation (Eq. 6).

$$\hat{H}\psi = E\psi$$  Eq. 6

where $\hat{H}$ is the Hamiltonian operator that returns the ground state energy $E$ of the system, and $\psi$ is the wave function from which all measurable properties of a system can be derived. As the Schrödinger equation can only be solved for a one-electron system, methods to approximate solutions for larger systems have been developed and are generally categorized as either wave function theory methods (WFT; e.g. HF or MP2) or density functional theory methods (DFT; e.g. B3LYP or M06-2X). As the names reveal, WFT uses the wave function to determine the ground state energy of the system whereas DFT uses the electron density.

One of the main issues in both WFT and DFT is to approximate the term that describes how electrons interact with each other (exchange-correlation functional). In the DFT functionals, the exchange-correlation is approximated in different ways. An accurate approximation of electron-electron interactions is crucial for describing the dispersion forces that are important in non-covalent complexes.
Standard DFT functionals are unable to account for long-range electron correlation and they generally produce poor results for non-covalent complexes.\cite{80,83} A functional suitable for studying non-covalent complexes is M06-2X, which has been applied in geometry optimizations of reduced protein-ligand systems (chapter 6). M06-2X implicitly accounts for medium-range electron correlations as a result of the parameterization of the functional,\cite{84} and has been shown to provide improved results for non-covalent complexes compared to many other DFT functionals.\cite{83,85}

3.3.3. The basis set

In QM calculations, the quality of the results depends on both the theory used (e.g. the DFT functional) and the size of the basis set (i.e. the number of included basis functions). To obtain reliable results, a sufficiently accurate functional and a sufficiently large basis set is needed. The use of a small basis set may be correlated with an overestimation of the binding energy (basis set superposition error) that can be corrected for by counterpoise corrections.\cite{86}

A basis set is a collection of functions representing one-electron atomic orbitals that can be linearly combined to create molecular orbitals. In addition, diffuse and polarization functions can be added to both the hydrogens and/or the heavy atoms of the system. Polarization functions (denoted by e.g. p or *) are important for a correct representation of bonding while diffuse functions (denoted by e.g. aug or +) are important to obtain accurate energies for anionic and weakly bonded systems (i.e. non-covalent interactions). Two examples of popular basis sets that have been used in this thesis are Pople’s 6-31G** and Dunning’s aug-cc-pVTZ.

3.3.4. Dispersion correction

To address the fact that the currently available DFT functionals cannot accurately account for long-range electron correlations (i.e. the dispersion forces), a number of dispersion-accounting methods have been developed. These methods provide an additional energy term for the dispersion ($E_D$) that can be added to the calculated energies by DFT ($E_{\text{DFT}}$), the sum of which is the dispersion corrected energy ($E_{\text{DFT-D}}$) according to Eq. 7.\cite{87,88}

$$E_{\text{DFT-D}} = E_{\text{DFT}} + E_D \quad \text{Eq. 7}$$

The energy term obtained by these methods is typically empirically derived and can be calculated at a low computational cost. It has been shown that the dispersion accounting methods significantly improve the performance of several DFT functionals, including results for non-covalent complexes.\cite{83} In
this thesis, the D3 method in conjunction with the BLYP functional was found to be as a suitable method for estimating interaction energies of non-covalent interactions (chapter 8 and paper 3).

3.3.5. Benchmarking methods

The common practice to assess the accuracy of a DFT method is to compare the results to those obtained by a higher level of theory. The current method of choice for benchmarking methods for non-covalent complexes is the coupled cluster singles doubles with perturbative triples (CCSD(T)) method with extrapolation to the complete basis set (CBS) limit. This method has been shown to produce chemically accurate results (i.e. within 1 kcal/mol).[6,90]

In this thesis, the CCSD(T)/CBS method has been used to benchmark methods (both with and without dispersion correction) to estimate $\Delta E_{gas}$ of non-covalent complexes (Table 1, chapter 8 and paper 3).

Table 1. Calculated interaction energies ($\Delta E_{gas}$) in kcal/mol for a set of non-covalent interactions

<table>
<thead>
<tr>
<th>Interaction type</th>
<th>CCSD(T) CBS</th>
<th>M06-2X aug-cc-pVTZ</th>
<th>B3LYP aug-cc-pVTZ</th>
<th>BLYP-D3 aug-cc-pVTZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH····H2O</td>
<td>-7.59</td>
<td>-9.90</td>
<td>-5.67</td>
<td>-7.76</td>
</tr>
<tr>
<td>CH····O</td>
<td>-3.12</td>
<td>-2.95</td>
<td>-0.19</td>
<td>-3.78</td>
</tr>
<tr>
<td>CH····arene (Trp)</td>
<td>-15.06</td>
<td>-14.45</td>
<td>-6.89</td>
<td>-16.06</td>
</tr>
<tr>
<td>CH····arene (Tyr)</td>
<td>-10.58</td>
<td>-10.70</td>
<td>-3.36</td>
<td>-11.44</td>
</tr>
<tr>
<td>Classical H-bond</td>
<td>-12.69</td>
<td>-15.23</td>
<td>-10.84</td>
<td>-13.43</td>
</tr>
<tr>
<td>Classical H-bond</td>
<td>-6.26</td>
<td>-5.90</td>
<td>-4.90</td>
<td>-6.59</td>
</tr>
</tbody>
</table>

[a] The results are presented in paper 3.

3.3.6. Symmetry-adapted perturbation theory

SAPT[91,92] is a WFT method where the total interaction energy ($E_{tot}$) can be computationally decomposed into electrostatic ($E_{elec}$), dispersion ($E_{disp}$), induction ($E_{ind}$), exchange-repulsion ($E_{exch-rep}$) and higher order ($\Delta HF$) terms according to Eq. 8.

$$E_{tot} = E_{elec} + E_{disp} + E_{ind} + E_{exch-rep} + \Delta HF$$  \hspace{1cm} \text{Eq. 8}

SAPT is an accurate, but computationally expensive, method where the intermolecular interactions are treated as small perturbations of the system. The computational cost can, however, be reduced if the monomers are treated by DFT and only the intermolecular interactions are treated perturbatively.
e.g. the DFT-SAPT method\cite{93,94}. DFT-SAPT thus allows larger systems to be considered in the calculations. In this thesis, DFT-SAPT calculations have been applied to decompose the interaction energies of non-covalent interactions (chapter 8 and paper 3).

### 3.3.7. Electrostatic potential maps

The electrostatic potential (ESP) is defined as the work required to bring a positive test charge from infinity to a given point in space, and thus describes the attraction or repulsion of a positive charge at that point. ESP maps can be constructed for atoms or molecules by mapping the calculated ESP on the surface of the molecular electron density (i.e. electron distribution). An ESP map thereby gives information about the shape and size of the molecule as well as the electron distribution.

In this thesis, ESP maps have been constructed from ESP and electron density surfaces calculated by the M06-2X/6-31G** method. In the presented maps, the electron density surfaces are typically shown at an isovalue of 0.01 electrons/Bohr$^3$ and the ESP (in kcal/mol) is visualized by a color-spectrum ranging from red (electron rich; i.e. negative ESP), orange, yellow, green, blue, to purple (electron deficient; i.e. positive ESP). The ESP maps have been used as a tool to analyze non-covalent interactions, where overlapping electron densities have been visually inspected and the ESP at the close contacts have been considered. More details regarding the interpretation of the ESP maps are found in section 6.3.

### 3.4. Biochemical characterization

#### 3.4.1. The Ellman assay

The half maximal inhibitory concentration ($IC_{50}$) values presented for the AChE inhibitors in this thesis have been determined using the well-established Ellman assay\cite{95}. The Ellman assay is a colorimetric method that utilizes the reagents acetylthiocholine (an analogue of the natural substrate acetylcholine) and dithiobisnitrobenzoate (DTNB). These reagents will give rise to a yellow colored product resulting from the hydrolysis of acetylthiocholine (Figure 10). The enzymatic activity can thus be monitored by measuring the increase of yellow color over time ($\Delta$Absorbance at 412 nm). By adding an inhibitor at different concentrations, dose-response curves can be constructed from which the $IC_{50}$-value can be determined (Figure 11). A lower $IC_{50}$-value is equivalent to a more potent inhibitor.
Figure 10. The reagents used in the Ellman assay including acetylthiocholine and DTNB.

Figure 11. Example of a dose-response curve used for determining the IC$_{50}$ values of AChE inhibitors. In the plot, the enzymatic activity (%) is plotted against the logarithm of the concentration of the inhibitor (M). The IC$_{50}$-value of the inhibitor is obtained from the fitted sigmoidal curve, and is indicated by a dotted line crossing the concentration axis.

3.4.2. Isothermal titration calorimetry

ITC is a technique that allows the thermodynamic parameters of binding to be determined for a given system,[96] for example of a compound that inhibits the enzymatic activity of AChE as presented in this thesis (chapter 7 and paper 2). In an ITC experiment, the heat absorption or emission is measured during the titration of two molecules with known concentrations (e.g. small molecule (R)-C5685 to the enzyme mAChE; Figure 12A) at a constant temperature and constant pressure. From the heat signal, the $\Delta H$ and the $K_a$ can be obtained that subsequently allow the $\Delta G$ to be calculated according to Eq. 3. As $\Delta G$ and $\Delta H$ are experimentally determined, $\Delta S$ can be calculated according to Eq. 4.
**Figure 12.** ITC experiments of (R)-C₅₆₈₅ (A) and (S)-C₅₆₈₅ (B) to mAChE, respectively. The upper panels show the heat change upon titration of the ligand to the cell containing the protein (raw data). The lower panels show the integrated heat data points plotted against the molar ratio of the added ligand. The fitted sigmoidal curve gives the equilibrium association constant ($K_a$) and the reaction enthalpy ($\Delta H$).
4. Targeting acetylcholinesterase

- 124 drug-like inhibitors of acetylcholinesterase with diverse physicochemical properties were identified in a HTS.
- The bioactive conformations of six inhibitors to mAChE were determined by X-ray crystallography.
- Reproducing the bioactive conformation of the AChE inhibitors using molecular docking proved to be a challenging task.

4.1. Identification of chemical leads

Our research efforts focused on AChE was initiated by a HTS performed at the screening platform of the Laboratory for Chemical Biology Umeå (LCBU)[97]. In the screen, 17 500 drug-like compounds were assessed as inhibitors of hAChE by the Ellman assay[95] resulting in 124 compounds that were identified as “hits”. Examples of the identified inhibitors and their determined IC_{50}-values are found in Figure 13.

![Figure 13. Structure and potency of AChE inhibitors identified as hits in the HTS.](image)

The identified hits are chemically diverse and vary in terms of size, hydrophobicity, flexibility, charge and other electronic properties. Their structures spanned a large “chemical space” established by principal component analysis (PCA) of their physicochemical features as described by calculated 2D descriptors. Importantly, many of the hits were structurally different from a number of previously known inhibitors (e.g. tacrine and HLö-7) and may be interesting chemical leads in future medicinal chemistry projects.
4.2. Bioactive conformations of AChE inhibitors

The bioactive conformations of six of the identified hits (Figure 13) to mAChE were determined using X-ray crystallography at resolutions ranging from 2.3 to 2.8 Å. The overall conformations of the main chain of the protein were similar in the determined complexes, while some of the side chain conformations in the active site gorge deviated. Five of the ligands span the entire active site gorge and participates in non-covalent interactions with residues in both the PAS and the CAS whereas one ligand only interacts with residues in the PAS. An extensive analysis of the non-covalent interactions formed between the HTS hits and mAChE in the determined crystal structures is presented and further discussed in chapter 5.

4.3. Molecular docking to AChE

The possibility to use molecular docking in structure-based design of AChE inhibitors was investigated by a (re-)docking study. The aim of the study was to establish a general docking protocol that could be applied to predict the bioactive conformations of new inhibitors. The six inhibitors with known binding modes were used in the study, and were all docked to the same mAChE structure including 12 conserved water molecules. Three commonly used docking software were used; FRED[98], GOLD[99-101], and Glide[102,103]. The output from the dockings was evaluated by calculating the root-mean-square deviations (RMSD) of the heavy atoms of the docking poses compared to the crystallographic pose. The results clearly showed that the default settings resulted in poor pose predictions using all three software. Glide gave best results, generating acceptable poses (RMSD-value < 2 Å) for two of the ligands.

In attempts to improve the docking accuracy, the parameter settings used for the docking in Glide were altered where both the number of poses subjected to post-docking energy minimization and the number of output poses were increased. Using this modified protocol, acceptable poses were extracted for all ligands, but they were unfortunately not recognized by the scoring function (i.e. they were not found among the top ranked by Glidescore SP). Re-scoring of the extracted poses using a total of 11 scoring functions available from different sources improved the results, but the predictions of the crystallographic poses were still not satisfactory. The best performance was obtained with scoring functions available in FRED where the scoring function Chemgauss3 accurately identified acceptable poses for five out of seven ligands and PLP accurately predicted four out of seven ligands.
4.4. Contribution and significance of the results

The structural features and potency of the identified HTS hits make them interesting to pursue in medicinal chemistry projects aimed at designing drug candidates with improved properties compared to those in use today. This is a relevant task as the current treatments of cholinergic deficiencies and nerve-agent intoxication are associated with limited efficacy and/or adverse effects. The identified hits in the HTS do not contain nucleophilic functional groups required to reactivate inhibited AChE (e.g. oximes), and can therefore be considered useful as scaffolds in the design of novel nerve-agent antidotes. In addition to the potential use in drug discovery applications, the diverse properties of the hits make them interesting as probes to study the non-covalent interactions between ligands and AChE.

Using molecular docking to predict the bioactive conformations of AChE inhibitors proved to be challenging. Despite our efforts, exploring several docking software, parameter settings and scoring functions, the accuracy of the docking was still unsatisfactory. The shortcomings observed in the docking study might be due to several reasons. It may be directly related to the fact that the active site gorge is relatively large and that the two subsites (PAS and CAS) have similar properties. In fact, there are symmetric ligands that have been shown to bind simultaneously to both the PAS and the CAS (e.g. ortho-7 PDB code: 2GYV and bis-tacrine PDB code: 5EI5). The similar properties of the subsites clearly poses a challenge for the docking software, as many of the generated poses are flipped compared to the crystallographic pose. Obtaining a general protocol for docking to AChE is also challenged by the fact that the side chains in the active site gorge can adopt different conformations in complexes with ligands. These conformational changes are difficult to predict and may be critical for obtaining satisfactory accuracy in the dockings. It can be concluded that structure-based design of AChE inhibitors by methods that rely on molecular docking needs to be critically evaluated and shall preferably be accompanied by experimental data (e.g. X-ray crystallography) to produce reliable results.

4.5. Further reading

Further details regarding the HTS, the determination of the crystal structures and the docking study can be found in paper 1. The crystal structures including the HTS hits are further analyzed in chapter 5.
5. X-ray crystal structures of acetylcholinesterase-ligand complexes

- Several residues in the active site gorge display significant changes in their adopted conformations in complexes with different ligands.
- AChE interacts with non-covalent ligands by aromatic interactions as well as both classical and non-classical hydrogen bonds.

5.1. Available crystal structures of mAChE

To date, there are approximately 80 X-ray crystal structures of mAChE deposited in the Protein Data Bank\cite{106,107} (PDB; survey 2016-09-23). The available structures include the apo form of mAChE, binary complexes with either covalent or non-covalent ligands, and ternary complexes including both covalent and non-covalent ligands. In this chapter, an overview of the protein conformations adopted in crystal structures of mAChE is presented. Moreover, the non-covalent interactions formed in the complexes between mAChE and drug-like ligands are analyzed and reported.

5.2. From the protein’s point of view

To explore differences in the mAChE conformations that are related to the binding of ligands, the side chain conformations adopted by the active site residues were compared to the apo form of mAChE (Figure 14). A total of 60 mAChE structures were analyzed, including 50 structures obtained from the PDB and 10 in house structures (Table A1). The backbone conformations of these structures were similar, with RMSD-values for all Cα between 0.1 and 0.7 Å. Different conformations were, however, adopted by the loop close to the rim of the gorge in two of the complexes (Leu289-Phe297).

The side chain conformations of the active site residues were characterized by calculating RMSD-values compared to the apo form of mAChE (PDB code: 1J06). RMSD-values can be used to quantify differences in conformations of molecules or, in this case, side chains. RMSD-values are, however, dependent on the reference state that is used and an obvious problem is that two conformers that deviate to similar extents from the reference (but in different directions) cannot be differentiated. To increase the information content of the analysis, the RMSD-values were further resolved by implementing a method to assign a direction to the RMSD-values of each conformer (“directional RMSD”; dRMSD). By this method, two equally deviating
conformers could be distinguished and the dRMSD-values thus provides a better representation of the conformational flexibility of the residue (Figure 14B-C). The method for assigning dRMSD-values is described in detail in Appendix 1.

Figure 14. A. The residues included in the analysis with at least one atom within 4.5 Å of the active site gorge of mAChE (illustrated by a molecular surface in magenta). B. Adopted Tyr337 conformations. C. Adopted Phe338 conformations. In B and C, the reference (apo mAChE, PDB code: 1J06) is shown with black carbons. Conformations that have been assigned a positive dRMSD-value are shown in white carbons and a negative dRMSD-value with grey carbons.

The dRMSD-values calculated for the active site residues in the mAChE crystal structures were analyzed by PCA. PCA is a suitable method for this analysis as it is an unsupervised projection method that extracts systematic variation in a dataset by uncorrelated variables: the principal components (PCs). The results from the PCA are shown by the score and loading plots in Figure 15. As similar backbone conformations are a prerequisite for detailed analysis of the side chain conformation using this approach, the two complexes with alternative loop conformations (2WU4 and 5DTJ) were excluded in the presented model. The score plot shows how the crystal structures relate to each other in terms of their adopted side chain conformations (i.e. crystal structures that are located in the same region in the score plot adopt similar conformations; Figure 15A). The loading plot shows which residues that are important for describing the differences between the crystal structures that are apparent in the score plot. The residues that display large differences in their conformations in the complexes are therefore equivalent to the ones that have a high loading value in the loading plot (highlighted in red in Figure 15B).
The presented model comprises three PCs and describes 63% of the original data ($R^2_X$; Table A2).

**Figure 15.** PCA of the dRMSD-values calculated for 38 residues in the mAChE crystal structures. **A.** Score values for the first three PCs. The complexes are color coded according to; apo: green, non-covalent ligands: blue, covalent ligands: red, ternary complexes: yellow. **B.** Loading values for the first three PCs. The residues with large differences in their adopted conformations are highlighted in red.

Two clear groups were identified in the PCA (Figure 15A). One group consisted of the complexes that include pyridinium oximes (regardless if they are part of a binary or ternary complex). The “oxime group” was formed by complexes 2GYU, 2GYV, 2GYW, 2JEY, 2JEZ, 2WHP, 2WHR, 2WU3, 5PPP, and AL042. This group was mainly formed due to the adopted conformation of Trp286. The other group consisted of complexes including covalently bound nerve agents. This group was formed by the complexes 2Y2U, 3DL4, 3DL7, 3ZLU, and 3ZLT and was mainly a result of the conformations of Phe338 and His447. The majority of the binary complexes involving non-covalent ligands were clustered with the apo 1J06. Some complexes displayed unique conformations, mainly 1Q83, 5EHN, 5EIE, and 5EIH all of which contain the TZ2PA5 or TZ2PA6 inhibitors or their precursors. In addition, the ternary complex ortho-7-tabun-mAChE (2JF0) also adopted a conformation that was significantly different from the other complexes.

It is apparent that the main differences in the analyzed complexes were a result of different conformations adopted for Tyr72, Asp74, Leu76, Trp286, Tyr337, Phe338, Tyr341 and His447 (Figure 15B). The conformations of the remaining residues appeared to be similar and only minor conformational changes were observed.
5.3. From the ligands’ point of view

The interactions formed between mAChE and a range of non-covalent ligands were analyzed in 26 binary complexes. These complexes represent a range of bioactive conformations; some of the ligands bind at the top of the gorge (PAS specific), some closer to the bottom of the gorge (CAS specific), while some span the entire gorge. Moreover, the ligands represent different compound classes and the majority contain positively charged functional groups (tertiary amines or pyridinium rings). Prior to the analysis, the complexes were processed by adding and energy minimizing hydrogen atoms using a MM force field (cf. section 6.2.1). The non-covalent interactions were thereafter analyzed by using a combination of visual inspection and the criteria listed in Table 2.

Table 2. Geometric criteria used to identify and categorize non-covalent interactions

<table>
<thead>
<tr>
<th>Interaction type</th>
<th>Abbreviation</th>
<th>Geometric criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical H-bond</td>
<td>HB</td>
<td>donor-H--acceptor ≤2.4 Å and ≥120°</td>
</tr>
<tr>
<td>Aromatic hydrogen bond</td>
<td>OH--Ar</td>
<td>OH--centroid ≤2.5 Å and ≥120°</td>
</tr>
<tr>
<td>Aromatic interactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parallel displaced (p)</td>
<td>Ar (p)</td>
<td>centroid-centroid ≤4.0 Å</td>
</tr>
<tr>
<td>Edge-to-face (ef)</td>
<td>Ar (ef)</td>
<td>centroid-centroid ≤5.0 Å</td>
</tr>
<tr>
<td>CH--arene</td>
<td>CH--Ar</td>
<td>carbon-centroid ≤4.0 Å</td>
</tr>
<tr>
<td>COO--arene</td>
<td>COO--Ar</td>
<td>O--arene-CH ≤2.8 Å</td>
</tr>
</tbody>
</table>

The non-covalent interactions in the different complexes are presented in Table 3 (more details regarding the complexes and the ligands included are found in Table A1). In Table 3, the residues that participate in interactions with at least three ligands are presented. An interesting observation is that there is no clear correlation between the affinity of the ligands and either the number or nature of the non-covalent interactions.
### Table 3. Non-covalent interactions in binary mAChE-ligand complexes

<table>
<thead>
<tr>
<th>PDB code</th>
<th>Asp</th>
<th>Trp</th>
<th>Tyr</th>
<th>Glu</th>
<th>Trp</th>
<th>Phe</th>
<th>Tyr</th>
<th>Phe</th>
<th>Tyr</th>
</tr>
</thead>
<tbody>
<tr>
<td>2GYU</td>
<td>74</td>
<td>86</td>
<td>124</td>
<td>283</td>
<td>286</td>
<td>295</td>
<td>337</td>
<td>338</td>
<td>341</td>
</tr>
<tr>
<td>2GYV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2GYW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2GEY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4ARA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4ARB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4B7Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4B80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4B81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4B82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4B83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4B84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4B85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4BTL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5FOQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5FUM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5231a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7491a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7643a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL037a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL042a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL157a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL158a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL166a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL167a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL237a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[a] in house structures that are not deposited in the PDB
[b] sulfur hydrogen bond acceptor, donor-H···S distance 2.7 Å
The residues that were identified as key interactions sites with ligands span the entire active site; Asp74, Tyr124, Glu285, Trp286, Phe295 and Tyr341 are located in the PAS and Trp86, Tyr337 and Phe338 are located in the CAS (Figure 16). Trp86 and Trp286 participate in aromatic or CH···arene interactions to many of the ligands. Classical hydrogen bonds are mainly formed by the side chain hydroxyl of Tyr124 and the backbone N of Phe295. Residues Asp74, Glu285, Tyr337, Phe338 and Tyr341 are also frequently involved in different types of interactions in many of the complexes. In addition to the direct interactions between the protein and ligands, several water mediated hydrogen bonds were observed. The water mediated hydrogen bonds are mainly formed by Gly121, Gly122 and Ser203, Ser293 and Phe295. The analysis of water is complicated by the fact that the modelling of water molecules is often dependent on the resolution of the crystal structure and it is therefore difficult to make conclusions regarding their potential role based on the presented data.

Figure 16. Key residues for non-covalent interactions with ligands. Residues that display large deviations in their conformations in the complexes are highlighted in cyan.

5.4. Contribution and significance of the results

According to the presented analysis, the side chains of a number of active site residues adopt significantly different conformations in the investigated AChE-ligand complexes. Interestingly, five of the nine key residues for mediating non-covalent interactions with ligands also display large conformational changes. As crystal structures represent static models of a protein-ligand complexes, they contain limited information regarding the dynamics of the system over time. The conformational changes observed in the analyzed complexes can therefore not be interpreted as dynamic properties of the
residue. To obtain information regarding the motions of a molecular system, techniques such as NMR or MD simulations are more appropriate.

Many of the ligands that have been determined in complex with mAChE contain positively charged tertiary amines, most of which participate in non-classical hydrogen bonds to mAChE. As many top selling drugs contain amines (approximately one third of the top 100 drugs[^11]), understanding the types of interactions that they may form with proteins is highly relevant for drug development. Furthermore, the analysis of the binary complexes provides insight into the general binding mode of non-covalent ligands that may be useful in future medicinal chemistry projects focusing on AChE inhibitors.

Based on the presented analysis of the crystal structures, it can be concluded that AChE constitutes a useful and relevant system for studying several aspects of ligand binding in detail. This includes the significance of conformational changes observed upon binding of ligands, the importance of water-mediated interactions in protein-ligand complexes, and the nature and role of non-classical hydrogen bonds. Moreover, the multivariate approach (including the calculation of the dRMSD-values) for characterizing the protein conformation may be used to improve the results from structure-based design. The method can, for example, be used to select representative protein conformations for use in molecular docking.

### 5.5. Further reading

A list of the included mAChE-ligand complexes and a description of the method for assigning the dRMSD-values can be found in Appendix 1.
6. Processing of X-ray crystal structures using density functional theory

- A methodology for fine-tuning protein-ligand complexes derived from X-ray crystallography has been developed.
- The methodology allows evaluation of the geometries modeled in the crystal structures.
- The fine-tuned structures allow detailed analysis of the non-covalent interactions including those involving CH groups.

6.1. Why X-ray crystal structures need additional processing

We have established a procedure for fine-tuning protein-ligand complexes to improve the geometries and interpretation of X-ray crystal structures. The additional processing of the X-ray coordinates has been carried out using a quantum chemical cluster approach\cite{112,113} and includes geometry optimizations of truncated structures of the protein-ligand complexes (hereafter referred to as “reduced models”). The objective of the methodology is to address the following concerns that are related to the interpretation of X-ray crystal structures determined at ~2.5 Å resolution.

- The uncertainties in the X-ray coordinates in the studied mAChE-ligand complexes may be as high as ~0.4 Å.\cite{68}

- The geometries of the modelled ligands may be poor as a result of disordered electron densities and/or incorrect chemical restraints applied during the refinement process.\cite{66,73}

- The X-ray data does not support the modeling of hydrogen atoms. Protonation or tautomer states can therefore not be extracted from the crystal structures.\cite{75}

- The geometrical arrangement of ligands’ functional groups can be crystallographically indistinguishable (e.g. inversion of a tertiary amine).

An overview of the methodology adopting the quantum chemical cluster approach is presented in Figure 17. More details are provided in the following sections and an illustrative example where the methodology has been utilized is provided at the end of the chapter.
6.2. The quantum chemical cluster approach

6.2.1. Pre-processing of the protein-ligand complexes using molecular mechanics

The initial processing of the protein-ligand complexes included addition of hydrogen atoms, assignment of protonation states of acidic/basic groups, and assignment of tautomeric states. The preparation of all atom structures was followed by a constrained energy minimization (MMFF94s\cite{114,115}) where all added hydrogen atoms were freely minimized while a force was applied to the heavy atoms after a deviation of 0.2 Å from their initial position. The processing of the protein-ligand complexes by MM (Figure 17B-C) resulted in initial correction of bond angles and bond lengths as well as a reasonable initial placement of the added hydrogen atoms.

6.2.2. Construction of reduced model systems

To facilitate geometry optimizations using DFT, the size of the protein-ligand complexes has to be carefully reduced due to the computational cost of the method. The key objective was that the resulting reduced model should be a
good representation of the studied system and has thus been uniquely constructed for each considered protein-ligand complex (Figure 17D and 18).

Figure 18. Example of a reduced model of the (R)-C5685·mAChE complex including 192 atoms. The atoms that have been freely geometry optimized are shown as balls (67 atoms; carbons colored purple and hydrogens colored light blue), with the truncation sites indicated by an asterisk. Atoms shown as sticks represent the protein environment and have been fixed in their initial positions (carbons colored grey).

In all of the systems considered in this thesis, the reduced models have been centered on the ligand. In some cases, the entire ligand has been included whereas other reduced models have focused on part of the ligand. The residues and water molecules adjacent to the ligand atoms have been included to represent the protein environment and their positions have been fixed during the geometry optimizations to avoid artificial movements. The electron density maps ($2F_o-F_c$ and $F_o-F_c$) have been consulted during the setup of the reduced models to identify dynamic residues and/or atoms with uncertain geometries in the X-ray structure (i.e. disordered or weak densities). Atoms that have been categorized as poorly defined or dynamic from the X-ray data have been allowed to be flexible in the DFT calculations to improve and/or validate their geometries. The flexible residues have been fixed at the truncation sites (indicated by an asterisk in Figure 18). Water molecules in direct contact with the ligand have, in most cases, been either fully or partially flexible (hydrogens only).
6.2.3. Geometry optimizations of reduced models using DFT

The reduced models have been geometry optimized by the M06-2X/6-31G** method (Figure 17E-F). The output from the geometry optimizations have been evaluated by comparing the converged geometries of the reduced models with the X-ray coordinates that were used as input in the analysis. Converged geometries with deviations >1 Å have been deemed incorrect. Incorrect geometries could be traced back to inappropriate reduction of the system resulting in artificial movements, incorrectly assigned protonation state of the ligand or residues, high energy conformation ligand geometry modelled in the X-ray model, or other errors in the X-ray model (as a result of missing or questionable features in the electron density). The optimized reduced models were further evaluated by assessing their agreement with the experimental X-ray data. This was achieved by comparing the converged geometries with relevant omit maps and by calculating new 2F₀-Fc maps corresponding to a structural model featuring the converged geometries.

The processing of the X-ray crystal structures using the quantum chemical approach has been an iterative process. Many different reduced models have typically been explored (both in terms of the size and the subset of flexible atoms) before a model that generated reliable results has been obtained (i.e. geometries of the reduced models that converged in agreement with the crystallographic data). The optimized reduced models have been used for detailed interpretations of non-covalent interactions (chapters 7-9 and papers 2-4) and as input in further calculations (chapter 8 and paper 3). The methodology has also been used to guide the crystallographic refinement of a ternary complex and to validate the final structural model (Figure 17G; chapter 9 and paper 4).

6.3. Analysis of non-covalent interactions

In this thesis, the structures of the protein-ligand complex that have been processed using the methodology outlined in this chapter have been subjected to detailed analysis of the non-covalent interactions (Figure 17H). The non-covalent interactions have been visually analyzed and classified based on ESP maps (described in section 3.3.7.) in combination with geometric criteria (Figure 19 and Table 4). As exemplified in Figure 19B-D, the non-covalent interactions indicated by dashed lines correspond to overlapping electron densities in the calculated ESP maps. As a result of the electrostatic nature of the hydrogen bonds, the surfaces representing the acceptors and donors reflect the differences in the ESP of the participating atoms, i.e. a relatively electron rich acceptor versus a relatively electron deficient donor.
Figure 19. Analysis of non-covalent interactions based on a reduced model processed using the quantum chemical cluster approach. A. ESP map constructed for a reduced model. B-D. Stick models and ESP maps representing a classical hydrogen bond (B), an activated CH···O hydrogen bond (C) and an activated CH···arene hydrogen bond (D). The hydrogen bonds are indicated by dashed lines in the stick models and the donor-H···acceptor angles distances are provided.

Table 4. The geometric criteria used in the assignment of classical and non-classical hydrogen bonds

<table>
<thead>
<tr>
<th>Interaction type</th>
<th>Geometric criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical H-bond</td>
<td>donor-H···acceptor ≤2.2 Å and ≥140°</td>
</tr>
<tr>
<td>CH···O[^8]</td>
<td>donor-H···acceptor ≤2.4 Å and ≥140°</td>
</tr>
<tr>
<td>CH···arene (Trp)[^18,116]</td>
<td>donor-H···acceptor ≤3.0 Å and ≥140°</td>
</tr>
<tr>
<td>CH···arene (Tyr)[^8,116]</td>
<td>donor-H···acceptor ≤2.6 Å and ≥150°</td>
</tr>
</tbody>
</table>

[^8]: The distances and angles of the CH···arene interactions were measured to the contact point perpendicular to the plane of the aromatic ring.

It should be noted that more stringent geometric criteria have been used to assign the non-covalent interactions compared to the complexes analyzed in chapter 5 that had not been subjected to additional processing by DFT.
6.4. Evaluation of multiple conformations

In the (R)-C5685-mAChE crystal structure (PDB code: 4ARA), four possible geometrical arrangements are plausible as both the active site residue Tyr337 and the ligand was modelled in two conformations each. The different geometries of the system were evaluated using the methodology outlined in this chapter.

Reduced models representing the four geometries were constructed focusing on the N-ethyl pyrrolidine part of (R)-C5685 and the side chain of Tyr337 (Figure 18). The ligand and Tyr337 were flexible in the geometry optimizations, as well as the adjacent residue Phe338 and a number of water molecules. The results indicated that all of the modelled geometries are plausible, as the converged structures were both geometrically and energetically comparable (Figure 20). The processed (R)-C5685-mAChE complex allowed for structural interpretation of the thermodynamic parameters of the binding of (R)-C5685 as described in chapter 7 and paper 2.

![Figure 20](image)

**Figure 20.** The conformations of Tyr337 and the ligand converged in similar geometries in the four possible arrangements of the (R)-C5685-mAChE complex. A. Converged geometries for A-Tyr337 and B-Tyr337 B. Converged geometries for A-(R)-C5685 and B-(R)-C5685.

6.5. Contribution and significance of the methodology

X-ray crystallography allows accurate and highly useful molecular models of protein-ligand complexes to be determined. There is, nonetheless, a degree of uncertainty in the X-ray coordinates that users must be aware of and address appropriately to obtain reliable results in detailed studies. In fact, it has been suggested that the lack of understanding of the limitations of the X-ray crystal structures is one of the major pitfalls in structure-based design.
To improve the detailed analysis of intermediate resolution X-ray crystal structures, a methodology employing the quantum chemical cluster approach has been developed. The quantum chemical cluster approach is well-established and has been applied to many other systems, e.g. for binding affinity estimations\cite{79} and for studying enzymatic reactions\cite{112}. As the structures processed according to the methodology outlined in this chapter are assumed to represent low energy conformations, they can be used to establish the preferred geometries and energetic profiles of non-covalent interactions. This is not necessarily true for X-ray crystal structures. In fact, it has been estimated that as many as 70 % of the ligands included in complexes deposited in the PDB have geometrical errors, many of which can be corrected by minor adjustments in the atomic positions.\cite{73} In addition, the methodology also allows the modelled geometries to be evaluated and potential steric clashes between the protein and the ligand to be relieved.

One important aspect that has been explored by this methodology (and that has large implications on the non-covalent interactions) is to establish likely protonation and tautomer states of species in studied systems. This is far from a trivial task and can rarely be supported by crystallographic data as the intermediate resolution structures does not allow reliable modelling of hydrogen atoms.

An alternative method to improve the local quality of protein crystal structures is to incorporate QM methods in the refinement of the X-ray data (i.e. “quantum refinement”).\cite{118,119} Although great care has been taken to evaluate the optimized geometries by comparing them with experimental data, the methodology outlined herein will likely not result in converged structures that are equivalent to those obtained by quantum refinement. As the methodology adopted herein is disconnected from the experimental data, it can instead be regarded as an external validation of the X-ray model. Validation of the X-ray models may be very valuable as the crystallographer’s interpretation of the electron density map can be supported or in some cases challenged.

Finally, the methodology presented herein may also be used in combination with conventional crystallographic refinement both to guide the refinement process and to validate the resulting structural models (chapter 9 and paper 4).

### 6.6. Further reading

The methodology described in this chapter has been implemented in chapter 7 and paper 2 ((R)- and (S)-C5685), chapter 8 and paper 3 (AL032, C7653 and (R)-donepezil) and chapter 9 and paper 4 (HI-6-sarin-mAChE).
7. Similar but different: Enantiomeric inhibitors of acetylcholinesterase

- Two enantiomeric inhibitors of AChE were identified with nearly identical affinity but with different thermodynamic signatures.
- The existence of several possible geometrical arrangements of the (R)-enantiomer is consistent with a more favorable entropy of binding.
- A single non-classical CH···arene hydrogen bond formed by the (S)-enantiomer may be related to the more favorable enthalpy of binding.

7.1. From the racemate to the pure enantiomers

The racemate of C5685 was identified as an inhibitor of hAChE in the HTS (chapter 4). As a result of the subsequent structure determination campaign, the structure of the racemic-C5685 was determined in complex with mAChE. In the crystal structure, the substituted aromatic ring was clearly defined in the PAS with the nitro group forming a hydrogen bond to Phe295N, while the electron density corresponding to the chiral N-ethyl pyrrolidine moiety was disordered (Figure 21A). In addition to the disordered electron density corresponding to the missing parts of the ligand, the nearby residue Tyr337 adopted two conformations.

Molecular docking was applied to generate plausible bioactive conformations of both the (R)- and the (S)-C5685 enantiomers (paper 1). Neither the X-ray crystal data nor the docking results indicated any preferred binding of either of the enantiomers. Efforts to synthesize enantiomerically pure C5685 for further biochemical characterization were therefore made. The pure C5685-enantiomers could be synthesized from commercially available 1 as outlined in Scheme 1.
Scheme 1. Synthetic route towards enantiomerically pure (R)-C5685 and (S)-C5685. a) CH₃OH, H₂SO₄, reflux, 96%; b) CH₃I, NaH, DMSO, room temperature, quantitative; c) NaOH, dioxane, 55 °C, 93%; d) (COCl)₂, DMF, CH₃CN, 0 °C → room temperature; e) (R)-2-aminomethyl-1-ethylpyrrolidine, Et₃N, CH₂Cl₂, room temperature, 72% over two steps; f) (S)-2-aminomethyl-1-ethylpyrrolidine, Et₃N, CH₂Cl₂, room temperature, 74% over two steps.

7.2. Biochemical characterization of (R)- and (S)-C5685

Biochemical characterization by the Ellman assay[95] and ITC experiments revealed equal potency but different thermodynamic signatures of the enantiomers to both the mouse and the human form of AChE (Table 5). The (R)-C5685 and (S)-C5685 are, in other words, examples of inhibitors that display enthalpy-entropy compensation.[39,45-47] The results show that the binding of (R)-C5685 is more entropically favorable while the binding of (S)-C5685 is more enthalpically favorable, resulting in nearly identical ΔG for the two enantiomers.
Table 5. IC₅₀-values and thermodynamic parameters for the binding of (R)-C₅₆₈₅ and (S)-C₅₆₈₅ to *Mus musculus* and *Homo sapiens* AChE

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Species</th>
<th>IC₅₀ [a] (µM)</th>
<th>ΔG (kcal/mol)</th>
<th>ΔH (kcal/mol)</th>
<th>ΔS (cal/mol·K)</th>
<th>-TΔS (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R)-C₅₆₈₅</td>
<td>mAChE</td>
<td>0.7</td>
<td>-8.1</td>
<td>-5.6</td>
<td>8.5</td>
<td>-2.5</td>
</tr>
<tr>
<td>(S)-C₅₆₈₅</td>
<td>mAChE</td>
<td>0.7</td>
<td>-7.9</td>
<td>-7.5</td>
<td>1.4</td>
<td>-0.4</td>
</tr>
<tr>
<td>(R)-C₅₆₈₅</td>
<td>hAChE</td>
<td>1.3</td>
<td>-7.8</td>
<td>-7.4 [b]</td>
<td>1.5 [b]</td>
<td>-0.4</td>
</tr>
<tr>
<td>(S)-C₅₆₈₅</td>
<td>hAChE</td>
<td>1.4</td>
<td>-7.7</td>
<td>-11.3 [b]</td>
<td>-12.4 [b]</td>
<td>3.6</td>
</tr>
</tbody>
</table>

[a] mean values from triplicate dose response curves  
[b] mean values from two ITC experiments

7.3. Structural interpretation of enthalpy-entropy compensation

The bioactive conformations of (R)-C₅₆₈₅ and (S)-C₅₆₈₅ to mAChE were determined by X-ray crystallography at 2.5 Å and 2.25 Å resolution, respectively (Figure 21B-C). The PAS binding moiety (i.e. the substituted aromatic ring) adopted similar conformations in all three determined complexes (i.e. racemic-, (R)-, and (S)-C₅₆₈₅·mAChE) and the major differences were observed in the CAS. Interestingly, two conformations of Tyr337 was observed in the (R)-C₅₆₈₅·mAChE complex while only one conformation was observed in (S)-C₅₆₈₅·mAChE. Also, in the (R)-C₅₆₈₅·mAChE, two conformations of the ligand were modelled representing both inversions of the tertiary amine.

![Figure 21. The bioactive conformations of the racemic, (R)-, and (S)-C₅₆₈₅·mAChE. A. racemic-C₅₆₈₅·mAChE with carbons colored cyan (PDB code: 4A23). B. (R)-C₅₆₈₅·mAChE with carbons colored green (PDB code: 4ARA). C. (S)-C₅₆₈₅·mAChE with carbons colored yellow (PDB code: 4ARB). The final 2Fo-Fc electron density maps are shown in blue contoured at 1σ. In A the Fo-Fc map (positive electron density corresponding to the chiral N-ethyl pyrrolidine moiety) is shown in green contoured at 3σ.](image-url)
To improve the interpretation of the X-ray crystal structures, the X-ray coordinates were fine-tuned using the methodology described in chapter 6. The methodology allowed the possible geometric arrangements of the (R)-C5685-mAChE to be evaluated, where all four geometries (resulting from alternative conformations of Tyr337 and (R)-C5685) of the complex were considered chemically plausible (described in section 6.4.). The non-covalent interactions were subsequently analysed in the complexes, revealing both classical and non-classical CH···O and CH···arene hydrogen bonds between the enantiomers and mAChE (Figure 22).

**Figure 22.** The non-covalent interactions in the processed X-ray crystal structures of (R)-C5685-mAChE and (S)-C5685-mAChE. A-D. The four conformations of the (R)-C5685-mAChE including the A- and B-conformation of Tyr337 (A-B, or C-D, respectively) and the A- and B-conformation of (R)-C5685 (A and C, or B and D, respectively). E. The conformation of (S)-C5685-mAChE. In the figure, red and blue dashed lines indicate hydrogen bonds (classical, CH···O or CH···arene) and van der Waals/electrostatic interactions, respectively.

Attempts to rationalize the observed thermodynamic parameters on a structural basis were made based on the adopted conformations and the analysis of the non-covalent interactions in the complexes. We propose that the population of several, energetically comparable, conformations indicate a higher degree of residual mobility of the (R)-C5685-mAChE complex that is entropically favorable (Figure 22A-D). The single conformation of (S)-C5685, on the other hand, is stabilized by a CH···arene hydrogen bond to the side
chain of Tyr337 (Figure 22E). This non-classical hydrogen bond is activated by the proximal charged amine of the N-ethyl pyrrolidine, and may be the reason for the more favorable enthalpic component of the binding of (S)-C5685 compared to (R)-C5685. Furthermore, the restriction of the complex by the activated CH···arene hydrogen bond may limit the conformational freedom of the (S)-C5685-mAChE complex that may be associated with an entropic penalty compared to (R)-C5685-mAChE complex.

7.4. Contribution and significance of the results

The C5685-enantiomers display the enthalpy-entropy compensation phenomenon, with nearly identical affinity for AChE but with different thermodynamic signatures. These results were unexpected as previous studies have shown that AChE exhibits stereochemical preference. The (R)-enantiomer of donepezil is, for example, ~5 times more potent than the (S)-enantiomer.[120] Enthalpy-entropy compensation is not unique for this particular system, it has been reported in the literature many times before.[45,47] The fact that two enantiomers exhibit this phenomenon provides a rare opportunity to rationalize this effect on a structural basis as the ligand flexibility and desolvation effects can be regarded similar for the two enantiomers.

By analyzing the non-covalent interactions in the X-ray crystal structures, the experimentally observed differences in the thermodynamic signatures could be attributed to specific features of the complexes. The more favorable binding entropy of (R)-C5685 may be related to a higher degree of conformational freedom in the (R)-C5685-mAChE complex, while the more favorable binding enthalpy of (S)-C5685 may be due to the formation of an activated CH···arene hydrogen bond. The potentially important role of CH···arene hydrogen bonds is not unique to the systems studied here. In fact, CH···arene hydrogen bonds have been reported to be important for e.g. the stability of the photoactive yellow protein[121] and the substrate specificity of AChE.[32]

Establishing the link between structural features and the thermodynamic parameters, where the enthalpy and entropy contributions are further related to e.g. specific interactions or conformational flexibility, can have a significant impact on future ligand design efforts.[39,41,122,123] Exploring these effects is not a trivial task and further studies are warranted. In this respect, the enantiomeric ligands constitute important systems that can facilitate further studies.
7.5. Further reading

Further details regarding the determination of the racemic-C5685·mAChE crystal structure and the docking of the enantiomers can be found in paper 1. The synthesis of (R)- and (S)-C5685, the determination of their bioactive conformations and thermodynamic signatures are presented in paper 2. The methodology used to fine-tune the X-ray crystal structures and analyze the non-covalent interactions is described in chapter 6.
8. The nature of activated CH···Y hydrogen bonds

- Classical and non-classical CH···Y hydrogen bonds were characterized by DFT calculations on isolated acceptor-donor fragments derived from X-ray crystal structures.

- CH···Y hydrogen bonds were found to be equally strong or stronger than classical hydrogen bonds while displaying different properties, e.g. in terms of the composition of the fundamental forces that contribute to their interaction energies.

- The results may have implications for both structure-based design and drug discovery.

8.1. Exploring non-covalent interactions in AChE-ligand complexes

Although the differences in the thermodynamic parameters of the C5685 enantiomers could be rationalized on a structural level, several fundamental questions remained unanswered. What is the interaction energy of an individual non-covalent interaction in a protein-ligand complex? What are the forces that contribute to the overall interaction energies of a non-covalent interaction? How do the properties of non-classical CH···Y hydrogen bonds relate to other more extensively studied interactions such as the classical hydrogen bonds?

Studying individual non-covalent interactions experimentally is challenging as advanced and highly sensitive methods are needed.[6] For this task, theoretical methods such as QM can be used.[80] QM methods have been used to study non-covalent complexes, often represented by model system in their optimal geometries.[6] In protein-ligand complexes, the observed bioactive conformation corresponds to the overall most energetically favorable conformation that arise as a result of several simultaneous non-covalent interactions. It is therefore not certain that the estimated interaction energies calculated for the model systems in optimal geometries can be directly transferred to describe interaction in a “biological setting” (i.e. in a protein-ligand complex with suboptimal geometries and a protein and water environment) as the strengths of non-covalent interactions are dependent on both the environment and geometrical parameters.[164]
To gain deeper insight into the nature of non-covalent interactions in experimentally determined geometries, a set of crystal structures of AChE-ligand complexes featuring presumably positively charged amines were fine-tuned and analyzed according to the methodology described in chapter 6 (Figure 23A). The processing of the crystal structures allowed for a more detailed analysis of the non-covalent interactions compared to that reported in chapter 5.

**Figure 23.** A. The chemical structure and name of the ligands in the studied AChE-ligand complexes referred to by their PDB entry code. B. Schematic representation of the different classes of hydrogen bonds considered in the analysis.
Both classical and activated CH···Y type hydrogen bonds (Y = O or arene) are formed between AChE and the ligands in the investigated complexes. Due to the different geometrical preferences and properties of the hydrogen bond acceptors, the non-covalent interactions were sub-categorized (and later analyzed) as CH···H₂O, CH···O, CH···arene (Trp), CH···arene (Tyr), and classical hydrogen bonds (Figure 23B).

Non-covalent interactions were extracted from the fine-tuned AChE-ligand complexes, where the donors and acceptors were represented by fragment pairs (Figure 24). Some important characteristics of the studied hydrogen bonds were established by DFT calculations and the results are briefly presented in the following sections.

8.2. Characterization of individual non-covalent interactions using DFT

8.2.1 Interaction energies in vacuum

The interaction energies of the hydrogen bonds were estimated using dispersion-corrected DFT by subtracting the energies of the protein and ligand monomers from the energy of the complex (Eq. 5, see section 3.3.1). The results showed that all CH···Y hydrogen bonds were attractive in vacuum, with interaction energies ranging from -19.8 to -3.8 kcal/mol (Figure 24). The CH···arene class represent the most energetically favorable interactions, where the CH···arene (Trp) interactions is somewhat stronger than the CH···arene (Tyr). The CH···O and CH···H₂O are comparable in terms of their interaction energies. The interaction energies of the CH···Y interactions are comparable in strength or stronger than the classical hydrogen bonds included as references in the study (-13.4 – -4.4 kcal/mol). The geometries representing the non-classical hydrogen bonds extracted from the protein-ligand complexes differed by approximately ~2.4 kcal/mol from the optimal geometry adopted by the fragment pairs optimized in vacuum.

8.2.2. Distance dependence

The distance dependence of the hydrogen bonds was investigated in a distance scan where the interaction energies were estimated at increasing separations of the donor and acceptor fragments. The distance scans showed that the optimal separation of the interacting monomers for the CH···Y interactions appear to be roughly 20% longer (~2.4 Å) compared to the classical hydrogen bonds (~2.0 Å).
Figure 24. The fragment pairs representing individual hydrogen bonds extracted from the fine-tuned AChE-ligand complexes. The $\Delta E_{\text{gas}}$ were calculated using the BLYP-D3/aug-cc-pVTZ method and are reported in kcal/mol. The donor···H-acceptor angles and distances are provided in the figure.
8.2.3. Effect of the environment

To study the hydrogen bonds in biologically relevant settings (i.e. in aqueous solution and the binding pocket of a protein), the free energies of solvation (ΔGsolv) were calculated and added to the interaction energies estimated in vacuum (ΔEgas) according to Eq. 9. The ΔGsolv were calculated using both the dielectric constant corresponding to water (ε = 78.4) and the protein environment (approximated by chloroform, ε = 4.7).

\[
\Delta E_{\text{solv}} = \Delta E_{\text{gas}} + \Delta G_{\text{solv}} \quad \text{Eq. 9}
\]

The results showed that the all CH···Y interactions except for the CH···H₂O class were attractive in biologically relevant environments. The solvated interaction energies (ΔEₘₙₐₜ) ranged between -8.2 and -2.4 kcal/mol in water and -8.5 to -1.1 kcal/mol in chloroform. The CH···H₂O class was weakly attractive or in some cases repulsive in water. Comparatively, the solvated interaction energies of the classical hydrogen bonds were reduced to a similar extent as the CH···Y interactions in both water and chloroform.

8.2.4. Effect of the formal charge

Calculations on neutral fragments were performed to investigate the effect of the formal charge (i.e. the “activation”) on the interaction energies of the hydrogen bonds. These calculations were performed on fragment pairs that were mutated in silico to represent the neutral amine or carbon analogues, respectively (Figure 2, see section 1.5.3). The results showed that the interaction energies of the CH···arene and CH···O interactions were reduced by on average 60% when the formal charge was removed (similar results were obtained for the neutral amine and carbon analogue). The strength of the CH···H₂O class was even more reduced than the CH···arene or CH···O classes, where the neutral analogues were only weakly attractive. As the effect of the formal charge on the CH hydrogen bonds was found to be significant, the activated CH···Y hydrogen bonds are expected to stabilize protein-ligand complexes to a higher degree than the neutral amine or carbon analogue would.

8.3. The fundamental forces of activated CH···Y hydrogen bonds

Decomposition of the interaction energies into electrostatic, exchange-repulsion, dispersion, induction and higher order terms was achieved by DFT-SAPT\(^{[93,94]}\) calculations (see section 3.3.6). The results showed that the activated CH···Y hydrogen bonds are mainly governed by dispersion forces
\( (E_{\text{disp}} = 43-50\% \text{ of their total attractive force} \), but electrostatics \( (E_{\text{elec}} = 27-36 \% \) \) and induction \( (E_{\text{ind}} = 15-19\% \) \) also contribute to the overall interaction energy (Figure 25). The results can be compared with the charged classical hydrogen bonds (e2 and e4) that are mainly stabilized by electrostatics (57\%) with smaller contributions from dispersion (21\%) and induction (14\%). Decomposition of the interaction energies at increasing separation of the fragments showed that the dispersion component of the CH···arene interactions declined rapidly and the electrostatic component became dominant at distances above ~3.5 \( \text{Å} \). This effect was not observed for the CH···O interactions that became very weak at increasing distances.

**Figure 25.** The energy components that contribute to the interaction energies obtained by DFT-SAPT calculations using the PBE0/aug-cc-pVDZ method. **A.** The energy components of a representative fragment pair from each interaction class in kcal/mol. **B.** The average contributions from the attractive energy components to the total energy in percent for all studied fragment pairs in each class, respectively. The energy terms are color coded in the figure, where red represents \( E_{\text{elec}} \), green \( E_{\text{disp}} \), light grey \( E_{\text{ind}} \), white \( \Delta HF \), black \( E_{\text{exch-rep}} \) and dark-grey \( E_{\text{tot}} \).

8.4. Interaction energies obtained by force fields

The abilities of commonly used force fields (MMFF94s\(^{[114,115]} \) and OPLS2005\(^{[123,126]} \)) to reproduce the DFT-D3 interaction energies of the hydrogen bonds were assessed. The results showed that the force field methods perform poorly for most of the investigated non-classical hydrogen bonds (Table 6).
Table 6. Calculated interaction energies ($\Delta E_{\text{gas}}$) in kcal/mol for the fragment pairs using DFT-D3 or MM force fields$^a$

<table>
<thead>
<tr>
<th>Code</th>
<th>Class</th>
<th>$\Delta E_{\text{gas}}$ (BLYP-D3/aug-cc-pVTZ)</th>
<th>$\Delta E_{\text{gas}}$ (MMFF94s)</th>
<th>$\Delta E_{\text{gas}}$ (OPLS2005)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a4</td>
<td>CH···H₂O</td>
<td>-7.8</td>
<td>-6.3</td>
<td>-6.4</td>
</tr>
<tr>
<td>b1</td>
<td>CH···O</td>
<td>-3.8</td>
<td>-0.1</td>
<td>-1.4</td>
</tr>
<tr>
<td>c1</td>
<td>CH···arene (Trp)</td>
<td>-16.1</td>
<td>-9.6</td>
<td>-11.9</td>
</tr>
<tr>
<td>d1</td>
<td>CH···arene (Tyr)</td>
<td>-11.4</td>
<td>-5.4</td>
<td>-7.7</td>
</tr>
<tr>
<td>e2</td>
<td>classical (charged)</td>
<td>-13.4</td>
<td>-15.9</td>
<td>-16.4</td>
</tr>
<tr>
<td>e3</td>
<td>classical</td>
<td>-6.6</td>
<td>-6.6</td>
<td>-5.3</td>
</tr>
</tbody>
</table>

[a] The results are presented in full in paper 3.

8.5. Contribution and significance of the results

The results presented herein clearly demonstrate the significance of activated CH···Y hydrogen bonds in the investigated AChE-ligand complexes; they are abundant and display geometrical preferences in the experimentally determined structures. Importantly, the non-classical hydrogen bonds are comparable to (or stronger than) the classical hydrogen bonds. The geometrical preferences and the composition of the fundamental forces that contribute to their interaction energies are, however, different compared to the classical hydrogen bonds. The non-classical hydrogen bonds are mainly governed by dispersion forces, but the electrostatic contribution is still significant. Due to the different nature of the non-classical hydrogen bonds, they may provide unique opportunities in drug discovery. The presented results are not limited to AChE as these types of interactions have been identified in many other biological systems as well, for example in the nicotinic ACh receptor$^{[127]}$.

The results also reveal that two currently available force fields are not able to accurately describe the non-classical hydrogen bonds. It is therefore likely that force fields will produce incorrect results (in terms of both geometries and interaction energies) for systems where these types of interactions are present, and will ultimately influence the accuracy of e.g. molecular docking and molecular dynamics simulations. This needs to be considered in structure-based design efforts focused on systems where these interactions are important, and is likely an important factor for the poor pose-predictions observed in chapter 4.
8.6. Further reading

Additional details regarding the crystal structures and the analysis of the non-covalent interactions are found in paper 2 (4ARA: (R)- and 4ARB: (S)-C5685 mAChE), paper 3, and the respective original references (4B7Z\textsuperscript{128}: AL032, 5FOQ: C7653 and 4EYz\textsuperscript{129}: (R)-donepezil). The methodology used to fine-tune the AChE-ligand complexes is described in chapter 6. The theoretical methods used to characterize the non-covalent interactions are briefly introduced in chapter 3 and all details regarding the DFT calculations on the fragment pairs are found in paper 3.
9. On the reactivation of nerve-agent inhibited acetylcholinesterase by antidotes

- A prereaction conformation of the HI-6-sarin-mAChE complex could be determined as a result of the combination of diffusion-trap cryocrystallography, DFT calculations, and kinetic studies.
- The structure provided insight into the role of a key residue in reactivation of nerve-agent inhibited AChE by oxime-based antidotes.
- The determined structure of HI-6-sarin-mAChE can be useful in the development of improved nerve-agent antidotes, e.g. by facilitating detailed mechanistic studies of the reactivation reaction.

9.1. Elucidating the reactivation mechanism

Obtaining structural data suitable for studying the reactivation mechanism in detail is not trivial as the system preferably should be determined in a conformation relevant to the reactivation reaction (i.e. a prereaction conformation). In a previously determined ternary complex between HI-6 and sarin-mAChE (PDB code: 2WHP), the position of the nucleophilic oxime of HI-6 could not be modeled and the sarin adduct adopts a conformation that shields the phosphorous atom. Detailed mechanistic studies by e.g. QM methods starting from this crystal structure are therefore challenging as it does not contain information on how the oxime approaches the sarin adduct. To determine the structure of this ternary complex in a prereaction conformation, a methodology including time-resolved diffusion-trap cryocrystallography experiments and the incorporation of DFT calculations in the refinement process was developed.

9.2. Crystallographic and DFT refinement of HI6-sarin-mAChE

Crystals representing the ternary HI-6-sarin-mAChE complex were obtained by treating crystals of mAChE with sarin that were subsequently soaked with a solution of HI-6. Diffraction data was thereafter collected after termination of the reactivation reaction at fixed time points. The efforts resulted in two refined structures where interconnecting electron density was observed between HI-6 and the sarin adduct. The structural refinement was focused on
one dataset and allowed modelling of the ternary HI6-sarin- mAChE complex at 2.4 Å resolution. The model obtained by conventional crystallographic refinement contained some ambiguities as the electron density maps did not define the entire system satisfactorily (Figure 26A). This was especially pronounced for the sarin adduct and the oxime of HI-6.

To improve the initial crystallographic model, the coordinates were processed according to the methodology described in chapter 6. Here the aim of the processing was different compared to the examples presented in chapters 7 and 8. As the experimental data was challenging to interpret in some regions, the aim of the calculations was to evaluate the modelled geometry of the complex rather than fine-tuning it. The focus in the calculations was on HI-6, sarin, and the nearby residues and water molecules (Figure 27). As the electron density map could not distinguish between the two isomers of HI-6 (i.e. the E- and Z-oxime), both isomers were explored in the calculations. The results obtained from the DFT calculations were used to guide the conventional refinement and to identify regions with questionable geometries. One of the issues that was revealed by the DFT calculations was, for example, a high energy conformation of the oxime functionality of HI-6. The high energy conformation was likely due to the disordered electron density map in combination with incorrect or insufficient geometrical restraints implemented during the refinement process.

The refinement of the structure (conventional and DFT) was repeated in an iterative way until the results converged between the crystallographic refinement and the DFT calculations. This resulted in a final geometry of the complex that was chemically plausible and (importantly) in agreement with the experimental electron density. The final structure that was deposited in the PDB was obtained after a final round of conventional crystallographic refinement and included both isomers of the oxime (Figure 26B; PDB code: 5FPP).
Figure 26. A. The initial crystallographic model of HI-6-sarin-mAChE B. The final model obtained by a combination of X-ray crystallographic refinement and DFT calculations including both isomers of HI-6 (PDB code: 5FPP). The 2Fo-Fc electron density map of HI-6 and Ser203-sarin is shown in blue contoured at $1\sigma$.

Figure 27. The reduced model of HI-6 (represented by the protonated E-isomer), sarin and the adjacent residues and water molecules that were evaluated by DFT geometry optimizations (281 atoms). Carbons colored purple and hydrogen atoms colored light blue were flexible in the geometry optimizations (59 atoms). The positions of the remaining atoms were fixed in their initial coordinates.
9.3. Validation and interpretation of the HI-6-sarin-mAChE structure

To facilitate detailed analysis of the structure of the HI-6-sarin-mAChE complex, the deposited coordinates were fine-tuned according to the methodology described in chapter 6. In the calculations, both isomers and protonation states of the oximes were included (i.e. geometries of four unique reduced model systems were considered).

Analysis of the structure showed that the conformation of HI-6 in the PAS is similar to the conformation observed in the previously determined binary and ternary complexes (PDB codes: 2GYU and 2WHP). The carboxy-amino substituted pyridinium ring of HI-6 forms stacking interactions with Tyr124 and Trp286, a direct hydrogen bond to Ser298, and a water-mediated hydrogen bond to Glu285. In the structure, the pyridinium oxime binds in the vicinity of the sarin adduct in the CAS.

The interaction between HI-6 and the Ser203-sarin adduct was further investigated by dispersion-corrected DFT (in analogy with the estimation of the interaction energies of the hydrogen bonds presented in chapter 8). These calculations were performed on reduced fragment pairs that were extracted from the converged reduced models (Figure 28). The results showed that the interactions between all four states of the oxime and the sarin adduct were favorable (-10.5 – -3.1 kcal/mol), where the protonated forms of the oximes were more attractive than the deprotonated forms.
Figure 28. A-D. Stick models of the fragment pairs representing the pyridinium oxime of HI-6 and the Ser203-sarin adduct. The hydrogen bonds are indicated by dashed lines in the stick models and the donor-H···acceptor angles distances and are provided. The \( \Delta E_{\text{gas}} \) values were calculated using the BLYP-D3/aug-cc-pVTZ method. E-H. ESP maps for the HI-6···Ser203-sarin fragment pairs. A and E corresponds to the protonated E-oxime, B and F the deprotonated E-oxime, C and G the protonated Z-oxime, and D and H the deprotonated Z-oxime.

An interesting feature in the structure that differed from the previously determined binary and ternary complexes was the conformation of the sarin adduct. Instead of shielding the phosphorous atom of sarin (as in e.g. the 2WHP complex), the isopropyl group replaces a conserved water molecule and forms a CH···O hydrogen bond to Glu202. This conformational change of the sarin adduct may be important for the reactivation reaction as it exposes the phosphorous to the oxime (Figure 29). The interaction between the carboxylate of Glu202 and the sarin adduct was favorable according to DFT calculations (\(-15\) kcal/mol where \(-7\) kcal/mol corresponds to the CH···O hydrogen bond; paper 4).
The important role of Glu202 in the reactivation of sarin-mAChE by HI-6 was further supported by studying the Glu202Asn-mAChE mutant. Measurements of reactivation kinetics showed that the reactivation rate was severely reduced in the mutant (~40-fold reduced compared to the wild type; paper 4). The affinity for HI-6 was, however, only marginally reduced. Glu202 has previously been identified as a key residue in AChE, e.g. for the catalytic function of AChE[133] and in the aging reaction of nerve agents[132].

Figure 29. A. The geometry adopted by HI-6-sarin-mAChE in the previously determined ternary complex including a conserved water molecule (PDB code: 2WHP). B. The geometry adopted by HI-6-sarin-mAChE featuring the conformational change of the sarin isopropyl group (PDB code: 5FPP). C. Stick model of the fragment pair representing the of Glu202 and Ser203-sarin with the CH···O hydrogen bond indicated with a red dashed line. The $\Delta E_{\text{gas}}$ was estimated using the BLYP-D3/aug-cc-pVTZ method. D. ESP map for the HI-6···Ser203-sarin fragment pair.

9.4. Contribution and significance of the results

The determined structure of HI-6-sarin-mAChE represents a previously unknown conformation of the sarin adduct. In this conformation, the phosphorous atom of sarin is exposed to the nucleophilic oxime (~3.3 Å) and therefore constitutes a good starting point for detailed studies of the reactivation mechanism. As such, the deposited HI-6-sarin-mAChE might be the most relevant experimentally determined structure for detailed studies by e.g. QM methods that is available to date. In mechanistic studies, both the important role of Glu202 and the possible isomer and protonation states of the oxime can be further investigated at a molecular level. As the development of new nerve-agent antidotes has been suggested to be hampered by the limited knowledge of the reactivation mechanism,[61] the determined structure constitutes an important contribution that may provide insight that is useful.
in the design of new and improved treatments for OP poisoning. Finally, the results are not limited to detail studies of the reactivation reaction at a molecular level. They may also have implications for the kinetic models that are used to investigate and describe reactivation reaction.

9.5. Further reading

The methodology integrated into the refinement of the HI-6-sarin-mAChE structure is presented in chapter 6. Further details are found in paper 4.
10. Concluding remarks

Small organic molecules influence numerous biochemical processes by forming complexes with proteins. Elucidating how and why these molecules interact with their targets is essential, not only for the fundamental understanding of the processes themselves, but also for drug discovery, biochemistry, medicine and other related fields. To obtain a comprehensive understanding of the interactions between a ligand and a protein, many aspects need to be investigated and data from different sources are needed. Research in this field is thus by its nature multidisciplinary, often involving synthetic organic chemistry, biochemistry, structural biology, medicinal chemistry, and computational chemistry to name a few. The studies presented in this thesis have used a multidisciplinary approach, which can be exemplified by the identification of chemical leads by using a combination of biochemistry, structural biology and molecular modelling (cf. paper 1). In the thesis, the discovered inhibitors have been used to probe the non-covalent interactions in AChE and have provided detailed insight into binding of drug-like ligands to AChE (cf. chapter 5 and paper 2 and 3). The discovered inhibitors may, however, also be used as starting points in the design of new drugs for treatment of cholinergic deficiencies and nerve-agent poisoning.

The structural interpretation of the thermodynamic signatures of the enantiomeric inhibitors of AChE was also facilitated thanks to a successful collaboration between disciplines (cf. paper 2). The establishment of how specific interactions and conformational flexibility affect the thermodynamic signature of ligand binding is perhaps as challenging as it is important and may provide unexplored opportunities in drug discovery.[39,41,122,123] To date, these effects are not understood well enough to be generally applied in the design of new ligands. The enthalpy-entropy compensation observed for enantiomeric inhibitors of AChE constitutes an important discovery and can provide opportunities also in future studies. It would, for example, be very interesting to study these complexes by NMR as this technique can give additional insight into conformational flexibility of the systems and their properties in aqueous solution.[123]

The study of the enantiomeric inhibitors also highlighted the role of non-classical hydrogen bonds in AChE-ligand complexes. Supported by studies on additional complexes, it was concluded that non-classical CH···Y hydrogen bonds are abundant and important for the binding of drug-like ligands to AChE (cf. paper 3). In order to facilitate detailed analysis of the non-covalent interactions based on X-ray crystal structures, a methodology for fine-tuning
the crystallographic coordinates was developed that can be applied to study virtually any protein-ligand complex (c.f. chapter 6).

So how can the non-classical hydrogen bonds be used as tools in drug discovery? In order to utilize interactions successfully in for example structure-based design, knowledge of the properties that the non-covalent interactions confer in a protein-ligand complex is needed. In this thesis, individual non-classical hydrogen bonds have been characterized by using DFT calculations. Our studies showed that the non-classical hydrogen bonds are equally strong (or in some cases even stronger) than classical hydrogen bonds (cf. paper 3). The importance of a non-covalent interaction is, however, not necessarily equivalent to its interaction energy as both strong and weak interactions operate simultaneously in a protein-ligand complex and may confer different properties (e.g. affinity, selectivity, on-off rate, thermodynamic or pharmacokinetic properties). The DFT calculations also revealed features of the non-classical hydrogen bonds, for example the composition of their fundamental forces, that may provide unexplored opportunities if utilized in for example structure-based design. Another important discovery was that two commonly used force fields are not able to accurately describe these non-classical hydrogen bonds, and better parameters for this class of interactions are needed. At present, structure-based design is likely to generate poor results and needs to be applied with caution to both AChE as well as to other complexes where these interactions are present (cf. paper 1 and paper 3).

The cooperative effects of multiple CH···Y interactions observed in the AChE-ligand complexes remain to be studied. It would be interesting to investigate how the strength and properties of a CH···arene hydrogen bond is affected by other simultaneous non-covalent interactions of different classes. The study of cooperative effects is highly relevant in protein-ligand complexes as the adopted bioactive conformation will be a result of all participating interactions and may reveal important aspects relevant to the ligand binding event. Studies on cooperative effects are available in the literature,(9) but as each interaction is dependent on their unique geometry and environment, more sampling is called for.

The determination of the prereaction structure of the HI-6-sarin-mAChE provided important insight that is relevant for the design of new reactivators of AChE. The previously unknown conformation of the system provides unique opportunities to study the reactivation mechanism at a molecular level by computational approaches. The implications of detailed mechanistic insight for the design of new reactivators is difficult to foresee. From the results obtained in this thesis it is, however, obvious that a fundamental
understanding of the molecular properties of AChE (e.g. in terms of important non-covalent interactions and conformational flexibility) is crucial to obtain reliable results in such studies. Moreover, the integration of DFT calculations into the refinement of crystallographic data provides possibilities to study dynamic and/or reactive systems that are challenging by conventional crystallographic approaches. The developed methodology thereby offers great opportunities for studying previously inaccessible systems of this category (c.f. paper 4).

Although the contribution of this research has been presented in the context of drug discovery, the results have implications in many other fields of science. As non-covalent interactions affect the properties of most molecules, the importance of understanding the ways in which they interact with each other goes far beyond the scope of this thesis.
11. Acknowledgements

Så är det nu äntligen dags att skriva de sista sidorna i denna avhandling! Jag känner mig oerhört stolt över att mitt namn står skrivet på omslaget till denna bok och att jag varit delaktig i den forskning som den innefattar. Ni är många som på olika sätt bidragit till denna avhandling, till er vill jag passa på att rikta ett stort och innerligt tack!

Speciellt tack till...

Min handledare Anna som valde att anställa mig som doktorand för många år sedan, och som efter det har låtit mig utforska det jag tycker är allra mest intressant inom kemi. Jag har sedan första dagen känt att min åsikt är viktig och att jag fått vara delaktig i alla moment i den forskning som jag bidragit till. Du är en förebild för mig, både som kvinnlig forskare och som chef!

Min biträddande handledare Fredrik, för att du alltid tagit dig tid att svara på mina otaliga frågor under årens lopp och för att du alltid tycker att våra forskningsresultat är intressanta (och av världsklass)! Att samarbeta med dig har kanske inte gjort mig till en mer ödmjuk, men säkerligen en bättre och mer entusiastisk forskare!

The talented co-authors of the papers that are appended to the thesis. Weixing for synthesizing my favorite molecules and Moritz for determining their thermodynamic signatures. Brijesh for making DFT less intimidating by patiently explaining important concepts. Anders för att du delat med dig av ditt perspektiv på kristalografi och gjort mig uppmärksam på såväl stora som små detaljer.

All previous and current members of the Linusson group; David, Ida, Brijesh, Urszula, Sofie, Cecilia E, Cecilia L, Tomas, and Rashmi for sharing your knowledge and for contributing to a creative research environment.

Min femstjärniga kontorskamrat David, som fått utstå allt från djupa suckar till höga hurra-rop genom åren. Tack för alla peppande och lugnande ord och goda råd, denna avhandling är tveklöst bättre tack vare dem!

Mina nuvarande doktorandkollegor i gruppen, Cecilia E och Cecilia L, för att det inte finns några dumma frågor. Tillsammans med er har jag lärt mig massor!
Min storsyster och största idol **Sofie** för att du banat min väg genom livet både på ett privat och professionellt plan. Inte minst för att du visade hur en avhandling skulle skrivas och försvaras på ett otroligt imponerande sätt i våras!

**Ida** för förstklassig handledning under både projekt- och examensarbete.

Studenter som på olika sätt bidragit med sina projekt och examensarbeten, speciellt **Karl** och **Cecilia L**.

**Andreas** och **Anna-Karin** som figurerat som biträde handledare i ett tidigare skede av doktorandstudierna.

Mina externa granskare **Elisabeth** och **Patrik** för att ni tagit er tid till och engagerat er vid mina uppföljningsmöten.

All the previous and current employees at the department and especially the organic chemistry corridor!

**Frippe** för att du en gång i tiden inspirerade mig till att plugga kemi.

**Barbro, Carina, Lars-Göran, Rosita, Sara** och **Maria** på institutionen som har hjälp mig på olika sätt.

Alla duktiga forskare på FOI som jag interagerat med på ett eller annat vis; **Anders, Nina, Christine, Elisabet** och **Susanne**.

Supporten på HPC2N. Speciellt tack till **Åke** som hjälpt mig att lösa både stora och små problem.

Mina goda vänner som sätter guldkant på tillvaron. Speciellt tack till bästa **Sandra**, mitt ständiga sällskap under grundutbildningen, för att du alltid tar dig tid att lyssna och ge goda råd.

Mina fina **mor-** och **farföräldrar** för många lärdomar och minnesvärda stunder.

**Mamma** och **pappa** för att ni alltid kavlar upp ärmarna och hjälper till, oavsett om det gäller att gräva upp buskar på tomten, skrapa gammal tapet från väggarna eller vakta barn.

Mina ovärderliga syskon **Sofie, Emil, och Robin** för att ni alltid ställer upp när det behövs. Jag är också glad och tacksam för att jag genom er utökat
familjen med Michael, Elin, Malin och det senaste tillskottet finaste lilla Eden! Tillsammans med er har jag aldrig tråkigt!

Min kära man Patrik för din ovillkorliga omtanke och ditt ständiga stöd. Jag känner mig otroligt lyckligt lottad som får dela mitt liv med dig!

Slutligen vill jag tacka de mest fantastiska personerna i mitt liv, mina två underverk Annie och Alfred, som har lärt mig att jag klarar så mycket mer än jag tror och som motiverar mig att göra mitt allra bästa oavsett vad det gäller. Jag hoppas att ni kommer att bli en bråkdel så stolta över mig som jag är över er!
12. References


Laboratories for Chemical Biology Umeå

FRED (Fast Rigid Exhaustive Docking), version 2.2.5, software available from OpenEye Scientific Software, Santa Fe, New Mexico.

GOLD (Genetic Optimization for Ligand Docking), version 4.1.2, software available from the Cambridge Crystallographic Datacenter, Cambridge, U.K.


Glide (Grid-based Ligand Docking with Energetics), version 5.5, software available from Schrödinger, LLC, New York, NY.


SIMCA-P+ v. 13.0.3. Umetrics AB, Umeå, Sweden.
Appendix I

A1. Method to assign directional RMSD-values

A1.1. AChE-ligand complexes included in the analysis

All structures of mAChE deposited in the PDB (2016-09-23) were considered in the analysis. Structures where one or more residues were mutated, structures determined at a resolution >2.8 Å, structures including other proteins than mAChE, and structures that were determined in different crystal forms were excluded from the analysis. In addition to the structures deposited in the PDB, ten in house structures were included in the study (Table A1). If alternative conformations were modelled in the active site, each conformation was represented as a unique complex. The included complexes are listed in Table A1.

A1.2. Preparation of the protein structures

The protein sequences were aligned and the Cα of the A-chains of all the mAChE structures were superposed to 1J06 using the protein superpose panel in MOE. The conformations of all residues with at least one atom within 4.5 Å of the binding site were included in the analysis. The included residues were: Gln71, Tyr72, Val73, Asp74, Leu76, Thr83, Trp86, Asn87, Pro88, Tyr119, Gly120, Gly121, Gly122, Tyr124, Ser125, Gly126, Ala127, Leu130, Tyr133, Glu202, Ala204, Trp236, Trp286, Leu289, Ser293, Ile294, Phe295, Arg296, Phe297, Tyr341, Gly342, Trp349, His447, Gly448, Tyr449, and Ile451. Since covalent ligands are found in the complexes, the catalytic Ser203 was excluded from the analysis.
Table A1. PDB and in house mAChE-ligand structures included in the analysis of the protein conformations and the mapping of non-covalent interactions.

<table>
<thead>
<tr>
<th>PDB code</th>
<th>Class</th>
<th>Resolution (Å)</th>
<th>Ligand</th>
<th>Alternative conformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1J06</td>
<td>apo</td>
<td>2.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5DTI</td>
<td>apo</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3DL4</td>
<td>covalent</td>
<td>2.5</td>
<td>Tabun</td>
<td></td>
</tr>
<tr>
<td>3DL7</td>
<td>covalent</td>
<td>2.5</td>
<td>Tabun (aged)</td>
<td></td>
</tr>
<tr>
<td>2JGE</td>
<td>covalent</td>
<td>2.6</td>
<td>Methamidophos</td>
<td></td>
</tr>
<tr>
<td>2JGF</td>
<td>covalent</td>
<td>2.5</td>
<td>Fenamiphos</td>
<td></td>
</tr>
<tr>
<td>2JGJ</td>
<td>covalent</td>
<td>2.5</td>
<td>Methamidophos (aged)</td>
<td></td>
</tr>
<tr>
<td>2JGL</td>
<td>covalent</td>
<td>2.6</td>
<td>VX/sarin (aged)</td>
<td></td>
</tr>
<tr>
<td>2H9Y</td>
<td>covalent</td>
<td>2.4</td>
<td>m-(N,N,N-trimethylammonio) trifluoroacetophenone</td>
<td></td>
</tr>
<tr>
<td>2HA0</td>
<td>covalent</td>
<td>2.2</td>
<td>4-ketoamyltrimethylammonium</td>
<td></td>
</tr>
<tr>
<td>2HA2</td>
<td>covalent</td>
<td>2.05</td>
<td>succinylcholine</td>
<td></td>
</tr>
<tr>
<td>2Y2U</td>
<td>covalent</td>
<td>2.6</td>
<td>VX</td>
<td>His447</td>
</tr>
<tr>
<td>2Y2V</td>
<td>covalent</td>
<td>2.45</td>
<td>Sarin</td>
<td>Tyr337</td>
</tr>
<tr>
<td>3ZLT</td>
<td>covalent</td>
<td>2.6</td>
<td>RVX</td>
<td>Phe338, His447</td>
</tr>
<tr>
<td>3ZLU</td>
<td>covalent</td>
<td>2.6</td>
<td>Cyclosarin</td>
<td>Phe338, His447</td>
</tr>
<tr>
<td>5HCU</td>
<td>covalent</td>
<td>2.4</td>
<td>DFP</td>
<td></td>
</tr>
<tr>
<td>2JEY*</td>
<td>non-covalent</td>
<td>2.7</td>
<td>HLö-7</td>
<td></td>
</tr>
<tr>
<td>2GYU*</td>
<td>non-covalent</td>
<td>2.2</td>
<td>HI-6</td>
<td></td>
</tr>
<tr>
<td>2GYV*</td>
<td>non-covalent</td>
<td>2.5</td>
<td>Ortho-7</td>
<td></td>
</tr>
<tr>
<td>2GYW*</td>
<td>non-covalent</td>
<td>2.4</td>
<td>Obidoxime</td>
<td></td>
</tr>
<tr>
<td>4ARA*</td>
<td>non-covalent</td>
<td>2.5</td>
<td>(R)-C5685</td>
<td>Tyr337</td>
</tr>
<tr>
<td>4ARB*</td>
<td>non-covalent</td>
<td>2.25</td>
<td>(S)-C5685</td>
<td></td>
</tr>
<tr>
<td>4B7Z*</td>
<td>non-covalent</td>
<td>2.3</td>
<td>AL032</td>
<td></td>
</tr>
<tr>
<td>4B80*</td>
<td>non-covalent</td>
<td>2.5</td>
<td>AL036</td>
<td></td>
</tr>
<tr>
<td>4B81*</td>
<td>non-covalent</td>
<td>2.8</td>
<td>AL011</td>
<td></td>
</tr>
<tr>
<td>4B82*</td>
<td>non-covalent</td>
<td>2.1</td>
<td>AL034</td>
<td></td>
</tr>
<tr>
<td>4B83*</td>
<td>non-covalent</td>
<td>2.4</td>
<td>AL041</td>
<td></td>
</tr>
<tr>
<td>4B84*</td>
<td>non-covalent</td>
<td>2.6</td>
<td>AL040</td>
<td></td>
</tr>
<tr>
<td>4B85*</td>
<td>non-covalent</td>
<td>2.1</td>
<td>AL028</td>
<td>Phe338</td>
</tr>
<tr>
<td>4BTL*</td>
<td>non-covalent</td>
<td>2.5</td>
<td>AL151</td>
<td></td>
</tr>
<tr>
<td>Table A1. continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5F0Q* non-covalent 2.3 C7653</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5FUM* non-covalent 2.5 AL200 Asp74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2WLS non-covalent 2.6 AMTS13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2WHR non-covalent 2.55 K027</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2HA3 non-covalent 2.25 Choline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1Q83 non-covalent 2.65 TZ2PA6 (syn)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1Q84 non-covalent 2.45 TZ2PA6 (anti)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1J07 non-covalent 2.35 decidium</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1N5R non-covalent 2.25 propidium</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5EHN non-covalent 2.6 TZ2PA5 (syn)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5EHQ non-covalent 2.5 TZ2PA5 (anti)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5EIE non-covalent 2.1 TZ2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5EIH non-covalent 2.7 TZ2 and PA5 His447</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5231* non-covalent 2.6 C5231</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7491* non-covalent 2.8 C7491</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7643* non-covalent 2.4 C7643</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL037* non-covalent 2.8 AL037</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL042* non-covalent 2.5 AL042</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL157* non-covalent 2.6 AL157</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL158* non-covalent 2.4 AL158</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL166* non-covalent 2.4 AL166</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL167* non-covalent 2.4 AL167</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL237* non-covalent 2.1 AL237</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2WU3 ternary 2.7 HI-6 and fenamiphos</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2WU4 ternary 2.4 Ortho-7 and fenamiphos</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2WHP ternary 2.2 HI-6 Asp74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2JEZ ternary 2.6 HLs-7 and tabun</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2JF0 ternary 2.5 Ortho-7 and tabun</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5DTJ ternary 2.7 SP-134 and DFP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5FPP ternary 2.4 HI-6 and sarin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Included in analysis of non-covalent interactions
A.1.3. Calculation of directional RMSD-values

For each individual residue (including all heavy atoms, both backbone and side chain), RMSD-values compared to the conformation observed in the reference structure (1J06) were calculated in MOE\textsuperscript{[133,134]}. The deviations of the C\textsubscript{\alpha} compared to 1J06 were monitored. In order to assign the dRMSD-values, the average coordinates of all heavy atoms of the side chains of each residue was calculated. For Gly residues, the average coordinates were calculated for all the backbone atoms. PCA was performed one residue at a time. The data was centered prior to extraction of the PCs. The score values for PC1, corresponding to the direction of the largest deviation in the average x-, y-, z-coordinates, were used to assign a direction to the calculated RMSD-values. A residue with a lower score value than 1J06 was assigned a negative dRMSD value, whereas residue with a larger score value was assigned a positive dRMSD value (Figure A1).

**Figure A1.** PCA score plot of the average x-, y-, z-coordinates (represented by spheres) for each residue was used to assign directions to the calculated RMSD-values relative to the reference 1J06. The residues represented by a higher score than 1J06 is assigned a positive dRMSD-value (white spheres) while a lower score is assigned a negative dRMSD-value (black spheres).
A.1.4. Analysis of the protein conformations

PCA was performed on the dRMSD-values in SIMCA\cite{135}. The data was centered prior to the analysis. The model statistics of the PCA presented in chapter 5 are found in Table A2.

Table A2. Model statistics for the first three components of the PCA model

<table>
<thead>
<tr>
<th>PC</th>
<th>Eigenvalue</th>
<th>$R^2_X$ (%)</th>
<th>$R^2_X$ (cumulative; %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.0</td>
<td>31.7</td>
<td>31.7</td>
</tr>
<tr>
<td>2</td>
<td>8.0</td>
<td>21.0</td>
<td>52.7</td>
</tr>
<tr>
<td>3</td>
<td>3.7</td>
<td>9.8</td>
<td>62.5</td>
</tr>
</tbody>
</table>