Computational Protein-Ligand Modelling of the Enzymes DNA gyrase and IcaB

Ahmad Alaa-El-Din Abd-El-Aziz

Master of Science by Research

University of York

Biology

March 2015

Abstract

Computational modelling of proteins and their interactions with small molecule ligands is a growing field of research. Such studies provide an understanding of how protein structure relates to mechanism and function as well as informing drug discovery and design.

This thesis had two main aspects: computational modelling of ciprofloxacin derivatives binding to DNA gyrase and homology modelling of the protein IcaB based on sequence alignment with a related protein, PgaB.

The inhibitory activity of synthetic ciprofloxacin derivatives (with various linkage to citrate groups) was experimentally assessed by gel electrophoresis to examine the effect on DNA gyrase binding to a target DNA strand. Overall, the derivative which possessed the greatest inhibition compared to the unmodified ciprofloxacin was the c-gly-ciprofloxacin derivative, which had a 2 atom linker between the ciprofloxacin and citrate groups. This correlated with the change in interactions seen between ciprofloxacin derivatives as computationally modelled by molecular mechanics methods.

The second aspect of the thesis was to generate a model of the protein IcaB to test the hypothesis that it is a deactylase of poly-N-acetyl-glucosamine (PNAG) during maturation of the poly-glycan in the extracellular matrix responsible for biofilm generation for bacteria. An initial review of deacetylase enzyme structures identified the conserved features required for activity. A homologous protein, Pga,B was then used as a template to generate a homology model of IcaB. The model maintained the orientation and positioning of the metal-binding and catalytic residues critical for proper deacetylase function. However, the PNAG binding groove, believed to be involved in the transport of the PNAG to the active site of PgaB, was not properly replicated in the IcaB model. Further modelling would require improved characterization of the binding groove of IcaB.

Table of Contents

List of Figures

List of Tables

List of Schemes

List of Equations

Acknowledgements

I would like to thank my supervisors, Professor Rod Hubbard and Professor Jen Potts for all their help and support during my MSc project. I would also like to acknowledge the guidance and advice that I received from Dr. Gavin Thomas, Dr. Anne-Katrin Duhme-Klair and Dr. Anne Routledge with the ciprofloxacin work. Thanks to all the members of the York Structural Biology Laboratory (YSBL) for general help and support with my project, in particular Tim Kirk for his help with some of the computational hardware setup. The opportunity to conduct my research in the YSBL is greatly appreciated. Finally I would like to thank the CIDCATS (Combating Infectious Disease: Computational Approaches in Translational Science) programme and the Wellcome Trust for their funding and the opportunity to do this project.

Author's Declaration

I declare that the work presented in this document is my own, unless explicitly indicated. This research was undertaken by myself, under the supervision of Professor Roderick E. Hubbard and Professor Jennifer Potts in the York Structural Biology Laboratory, Department of Chemistry, University of York. This work has not been submitted for any other award or qualification at any institute.

Results in chapter 2 are published in:

Milner, S. J.; Snelling, A. M.; Kerr, K. G.; Abd-El-Aziz, A.; Thomas, G. H.; Hubbard, R. E.; Routledge, A.; Duhme-Klair, A.-K. Bioorg. Med. Chem. 2014, 22 (16), 4499–4505.

Chapter 1: Introduction and Methods

1.1 Background

Proteins are involved in almost all biological processes including, but not limited to: enzyme catalysed reactions, gene expression, transcription, and cell signalling[1]. The study of ligands bound to proteins, particularly small-molecule ligands, has been of interest in order to control these reactions. As a result, drug discovery and design has developed through in-depth study of the interactions between the drug and potential target proteins both experimentally and computationally[2,3].

The field of computational protein-ligand interaction modelling has grown thanks to advances in modelling techniques and computational power over the past two decades[4]. The modelling of proteins interacting with ligands allows for virtual screening of a large number of compounds to identify potential candidate compounds with lower costs and improved speed, compared to traditional experimental screening[5]. In addition, this field has been assisted by improved experimental techniques, which have provided a multitude of structural data for proteins and their interactions that can be used to improve the computational models[6]. There have been a number of review articles that have given a comprehensive summary of the research done in this field. One of the earliest computational methods for modelling these interactions was protein docking, a review of which was published in 1996 by Lengauer and Rarey[7]. A review by Dill and McCallum from 2012 described the modern paradigm of protein structure modelling as divided into *ab initio* methods, which predict protein structure based only on the amino acid sequence, and methods such as homology modelling or protein threading, which rely on the protein's similarity with proteins or fragments of known structure[8–12]. Examples of each method are given by Yan et al., 2015 for use of an *ab initio* secondary structure prediction method and Semblat et al. 2015 for comparison with an established homology model[11,12]. Recently, a review published by Ghitti et al. in 2014 gave a summary of research done with Nuclear Magnetic Resonance (NMR), as well as protein docking and molecular dynamics simulations, to characterize interactions between ligands and target proteins in drug discovery[13].

In order to understand the results of the models of DNA gyrase and IcaB in this thesis, a brief summary of protein structure, protein-ligand interactions, and computational modelling of proteins is discussed below.

1.2. Protein Structure

A protein is composed from one or more chains of amino acid residues. Individual amino acids are linked together by peptide bonds, amide bonds formed from dehydration, which connect the acid carbon of the first amino acid to the amino group of the second amino acid[14]. In many proteins, amino acids are modified to allow for interactions that the canonical amino acids would be incapable of performing, or as part of an induction/repression mechanism. These modifications are referred to as post-translational modifications as they occur after translation of mRNA into peptide chains. Examples of post-translational modifications include phosphorylation, disulphide bonds, acetylation, methylation and glycosylation[15–20]. Removal of acetylated lysine residues on histone proteins is performed by histone deacetylases and will be discussed in greater detail in Chapter 3[18].

Hydrogen bonding between amino acid residues is involved in the structure of proteins, as well as in interactions of a protein with the bound ligand[21]. Hydrogen bonding is defined as an electrostatic attraction between a *hydrogen* atom bound to a highly electronegative atom such as nitrogen (N) or oxygen (O) and another electronegative atom[22]. In addition, hydrophobicity, the nature of non-polar groups to cluster together to minimize the surface area exposed to a polar solvent, causes amino acids with non-polar side chains to tend to cluster together and away from the solvent water which surrounds the protein. These interactions work cooperatively to form the three-dimensional protein structure.

There are four levels of protein structure connected by various interactions. The primary structure is the linear sequence of amino acids transcribed from mRNA. Hydrogen bonding within the main chain of the amino acid residues leads to the secondary structure motifs such as α-helices, β-sheets, loops, and turns. The α-helix structure formed by hydrogen bonding between main chain carbonyl oxygens and hydrogens from

the main chain amide group was first proposed by Pauling and Corey in 1951, and yields a right-handed helical structure with 3.6 amino acid residues per turn of the helix[23]. In the same year, β-sheet secondary structure was also proposed as a mechanism for satisfying the hydrogen bonding potential of the main chain as the protein is folded. In this case, the polypeptide main chain is in an extended conformation (or strand) with hydrogen bonds between the strands which run either parallel or anti-parallel[24]. The term random coil is used to refer to a generally unstructured region. Turns, portions of the structure in which the protein sequence reverses direction, are usually associated with proline and glycine residues. In the most common turn structure, β-hairpin turns, proline adds a kink to the protein sequence. Glycine, due to having only a hydrogen atom as a side chain, allows for greater rotational freedom assisting in the bending of the polypeptide chain[25]. An example of helix, sheet, and turn secondary structures are shown in Figure 1.1.

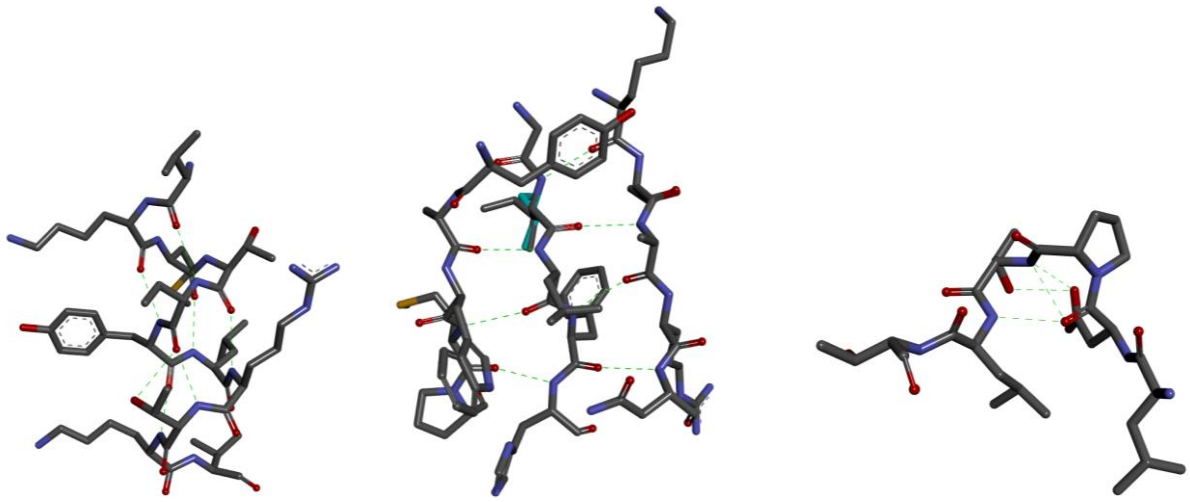


Figure 1.1: Peptide α-helix (right), β-sheet (middle), and β-turn (left) in stick diagrams. Hydrogen bonds are shown as dotted lines.

Side chain hydrogen bonding, van der Waals, and hydrophobic interactions result in the tertiary structure, also commonly referred to as folds[26]. Hydrophobic amino acid side chains such as phenylalanine, valine, and isoleucine tend to be oriented away from the solvent, while hydrophilic side chains such as those of serine, aspartate, and arginine tend to be oriented towards the solvent. This results in a hydrophobic core within a majority of globular proteins (i.e. non-membrane proteins)[27]. The tertiary structure, along with elements of the secondary structure, forms the overall conformation of a peptide

monomer. In some proteins, the monomer is functionally active. However, in others, the quaternary structure is formed by interactions of individual protein chains to form oligomers. This is necessary for many proteins, which require a full oligomer in order to be functional. There are proteins and domains of proteins that do not possess this structural organization, named intrinsically disordered proteins, which perform many biological functions including cell division, DNA condensation, and cell signalling. These proteins have emerged recently as a new area for investigation[28].

The different levels of protein structure can be illustrated by the structure of the protein haemoglobin. Haemoglobin is a protein that binds oxygen molecules within the bloodstream to allow for highly efficient transport between the lungs and the body's cells[29]. The transcribed sequence for a single peptide domain is the primary structure. Main-chain hydrogen bonding then forms a series of α-helices connected by loops to form the secondary structure. This secondary structure is then folded by interactions between side chains. As this folding occurs, a haem cofactor consisting of a porphyrin ring bonded to Fe^{2+} is added to the structure [30]. The iron atom in this haem is necessary for the function of the protein, as the oxygen molecule binds to this iron to be carried by haemoglobin[29]. The final protein monomer is then added to three other similar peptides to form the active tetramer, as shown in Figure 1.2[31].

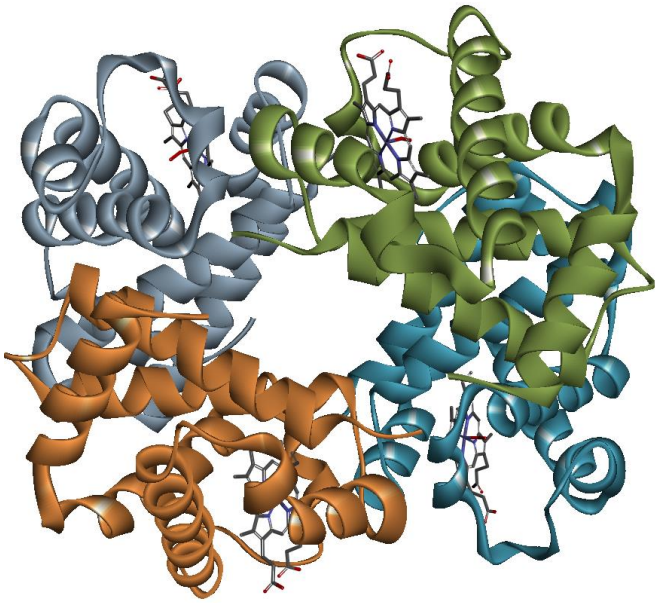


Figure 1.2: Structure of haemoglobin (colours represent individual monomers and haem prosthetic groups are shown in stick format) pdbID: 1GZX[31]

As the oxygen molecules bind to a bound haem cofactor, the haemoglobin conformation changes to a state with higher binding affinity, resulting in increased binding of oxygen at the other haems. The reverse is also true, as when an oxygen molecule dissociates from a haem, the haemoglobin reverts to the original conformation allowing for easier removal of the other oxygens. This change in conformation by binding at a region other than the active site is an allosteric transition, first described by name in 1965 by Monod et al. regarding the haemoglobin transition[32]. Generally, these sites cause a change in the conformation of the protein to enhance or inhibit the binding of the ligand in the active site[33].

The majority of proteins whose structures have been determined by x-ray crystallography or NMR methods (discussed below) are globular proteins that are stable in solution. There are several other types of proteins that play a major role in biological systems such as integral membrane proteins, membrane glycoproteins and peripheral membrane proteins. In the case of gram-negative bacteria, the cell wall consists of an inner membrane composed of peptidoglycan, a polymer of alternating N-acetylglucosamine and N-muramic acid, and an outer membrane composed of a phospholipid bilayer. Therefore, there is a periplasmic space between the two membranes, in which proteins involved in transport and biofilm formation, among others, have been shown to be localized[34,35].

Many protein structures have been identified by X-ray crystallography and Nuclear Magnetic Resonance (NMR)[36]. However, all structures referenced in this thesis were identified by means of X-ray crystallography. X-ray crystallography utilizes an X-ray beam directed at a crystallized molecule. By the angle of X-ray scattering, it is possible to determine the electron density at various positions, which can then be used to infer the atom positions and the bonding interactions in the overall molecule. The relationship between the angle between the incident ray and the crystal plane, the wavelength of X-rays, and the spacing between the planes in the crystal is described by Bragg's Law (Equation 1.1).

$$n\lambda = 2d\sin\theta$$

Equation 1.1: Bragg's Law, in which n is an integer, λ is the wavelength of the X-ray beam, d is the distance between two planes in the crystal lattice, and θ is the angle between the incident X-ray and the crystal.

In a single crystal diffraction, a pattern of spots known as a diffraction pattern will emerge as a result of this scattering. Using diffraction patterns from multiple angles, mathematical methods known as Fourier transforms can be employed to form a map of electron density, which may then be mapped using experimental data or fit to a theoretical structure[37,38]. The first published use of X-ray crystallography was to determine the protein structure of sperm whale myoglobin in 1958 by John Kendrew and coworkers[39]. X-ray and NMR structures of proteins are collected in the Protein Data Bank, a crystallographic database that includes proteins, DNA, and RNA structures and contained 110 206 biological macromolecular structures as of July 11, 2015[40].

An important tool for identifying the activity of the protein experimentally is the isoelectric point (pI or pH(I)): the point at which a molecule carries no overall charge. The pI gives a simple description of the overall charge surface of the molecule expressed as an integer. At low pI, a higher concentration of H^+ ions is required to neutralize the large negative charge on the protein surface and conversely, as the pI increases, the protein gains more positive charge[41]. The theoretical pI was calculated for this thesis using the online tool ProtParam from the ExPASy server at the Swiss Institute of Bioinformatics[42]. This method sets pKa for certain ionisable groups in proteins, which allows for a theoretical calculation of pI if the protein sequence is known[43].

1.3 Protein-Ligand Interactions

There are several types of binding interactions between proteins and small-molecule ligands, including enzyme substrates and protein inhibitors. Enzymes are protein catalysts for various biochemical reactions. For example, the enzyme lactase, which is found in the human small intestine, catalyzes the breakdown of the disaccharide lactose into the simple sugars galactose and glucose for absorption into the bloodstream[44].

Metalloproteins are proteins containing a metal ion as a cofactor. While some proteins contain a structural metal ion to maintain a fold of the protein, certain enzymes use the metal in a catalytic mechanism. For example, Cytochrome P450, a monooxygenase enzyme found in the human liver, uses a haem cofactor (a porphyrin ring bound to a Fe^{2+}) to perform a monooxygenase reaction to add an alcohol group to an aliphatic chain as the first step in breakdown of toxins or drugs in the body[45].

In addition to binding to the native substrate, enzymes as well as other proteins may bind inhibitors. An inhibitor is any molecule that binds to an enzyme and results in lowered activity. The activity of the inhibitor *in vitro* is generally measured by IC_{50} (inhibitory concentration 50), the concentration at which activity of the protein is reduced by 50% (originally referred to as I_{50})[46].

Binding sites in the protein some distance away from the active site can result in a change to the overall protein conformation. These are known as allosteric sites when this change in conformation affects the thermodynamics and/or kinetics of the activity of the system. For example, O_2 can act as an allosteric regulator in haemoglobin by enhancing the affinity of a separate subunit for another O_2 molecule through a change in the protein conformation. Therefore, a single O_2 molecule binding to a haemoglobin subunit can lead to haemoglobin fully saturated with oxygen. In addition, allostery allows for non-competitive inhibition, in which the inhibitor does not compete with the native substrate in order to inhibit the protein (competitive inhibition), but rather binds to an allosteric site in order to alter the active site and prevent binding[31].

The Allosteric Database (ASD) provides descriptions of structure, function, and disease caused by compounds that bind to allosteric sites as well as annotated allosteric sites. The ASD is updated regularly and is currently available online[47].

Proteins such as DNA gyrase bind to DNA to affect transcription of genes, which results in a change in gene expression (the proteins or nucleic acids produced from the original gene in DNA). DNA, deoxyribonucleic acid, is a double stranded nucleic acid polymer

formed from individual nucleotides[48]. An example of a nucleotide (adenosine) is shown in Figure 1.3.

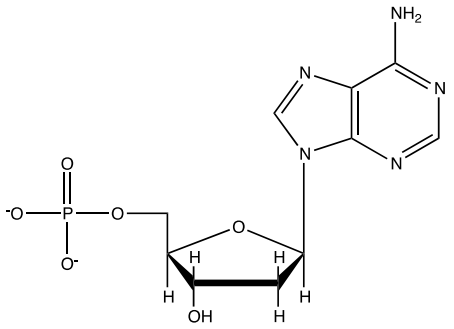


Figure 1.3: Structure of adenosine, a nucleotide consisting of a deoxyribose sugar, phosphate, and nitrogenous base (adenine)[48]

The sugar-phosphate bonds between the 5' –OH and phosphate oxygen form the backbone of DNA. Hydrogen bonds between the nitrogenous bases (2 for A-T, 3 for G-C) connect two strands of DNA. DNA forms a double helix due to a rotation within the sugar-phosphate backbone to prevent steric clashes between the nitrogenous bases. In addition, π-stacking interactions between the nitrogenous bases stabilize the adjacent purine/pyrimidine rings.

DNA within the cell is usually packaged during a non-dividing phase of the cell in the form of chromatin. Chromatin is formed by the DNA wrapping around histone proteins to form structures known as nucleosomes. Generally chromatin may be separated into heterochromatin, which is tightly bound to the histone proteins, and euchromatin, which is more loosely bound and therefore transcription related proteins are able to bind to the DNA[49]. Due to modifying the chromatin structure, post-translational modification of histone proteins greatly affects DNA transcription. For instance, histone methylation has been shown to be involved in the silencing of genes. An example of this is the histone methyltransferase EZH2, which methylates Lys27 of histone H3. Mutation of the Tyr641 residue in EZH2 has been identified from B-cell lymphomas and has been shown to increase the protein's activity, leading to repression of a tumour suppressor gene[50].

1.4 Computational Modelling of Protein Structure

One of the earliest methods for providing insight into the shape and energetics of a protein active site was the program GRID[51]. The basis for the algorithms used by GRID was a series of probes used to interact with the active site surface. These probes were a number of small functional groups including water, a methyl group, an amine nitrogen, a carboxyl oxygen, and a hydroxyl group. The energy contours formed by interaction of the probe with the protein surface were calculated. These contours showed regions of attraction with differing groups and could be shown using computer graphics. Therefore, drugs could be designed to maximize the number of contacts with the protein surface [51].

Homology modelling is based on the requirement that the function of a protein is conserved from one generation to the next. This requires conservation of structure, which in turn places constraints on the conservation of sequence. This means that proteins of similar sequence have similar structure. The method works well for proteins with greater than about 40% sequence similarity and is discussed further in section 1.6. A technique known as threading, or fold-recognition, is used if no protein structure is available with >40% sequence identity to the target. This technique relies on two assumptions: 1. the number of unique protein folds in nature is much smaller than the number of unique proteins (this is already clear, for example the number of TIM barrel structures adopted by proteins of very different evolutionary backgrounds) and 2. the protein structures in the PDB sample the possible protein folds [52]. There are a number of variants of the technique which test whether a sequence for a new protein fits into a known fold, including using empirical scoring functions (see below) or knowledge-based scoring functions derived from the analysis of the interactions seen in known structures. Finally, there are *ab initio* techniques (see below), in which a fold for the protein is derived from basic principles. There has recently been some success with using molecular mechanics/dynamics calculations to fold small proteins from sequence alone[53–55]. In addition, an overall approach called Rosetta has been developed where small segments of secondary structure are assessed for their fit to the sequence which are then annealed together to generate a final model[56–58]. None of these *ab initio* approaches are particularly robust, as shown by the most recent CASP (Critical Assessment of protein Structure Prediction) competitions, where the methods are

applied to new proteins before the structures are published, demonstrating that it remains a real challenge to predict the structure[59,60]. The most recent CASP competition, CASP X, has shown sustained improvement in refining models using molecular dynamics. Also, for the first time, models that use 3d contact information from new experimental techniques were tested. Although new inter-residue contact prediction methods may potentially be useful to provide scaffolds to build protein structures on, they remain a challenge to model [61]. Overall, the accuracy of protein structure prediction models is improving in non-template based modelling[60].

1.5 Computational Modelling of Protein-Ligand Interactions

Over the past 40 years, a number of different computational modelling techniques emerged for modelling the interactions between proteins and small molecule ligands including molecular mechanics and dynamics, virtual screening / docking, and quantum mechanics/molecular mechanics [62–64]. The binding of protein P to ligand L can be described by the equation PL <-> P + L**.** The equilibrium constant (K) can be used to approximate the experimental binding affinity as 1/K and is also related to the on and off rates (i.e. the rates of binding and dissociation, k_{on} and k_{off}) that describe the kinetics of the system as follows: $K = k_{off}/k_{on}$. The equilibrium constant is the ratio of products/reactants and is dictated by the difference in free energy between the products and reactants ($\Delta G = -RT\ln K$), which in turn is related to the enthalpy (energy of the system) and entropy (disorder in the system) as $\Delta G = \Delta H - T\Delta S$. Most empirical computational methods (such as those used in virtual screening) only consider the interactions made between protein and ligand as a sum of van der Waals, electrostatic, hydrophobic, and hydrogen bonding interactions[3]. More sophisticated approaches (such as in molecular mechanics) can explore the contributions of enthalpy and entropy on both sides of the equilibrium through the contributions from the free protein and ligand in solution as well as the complex[63,64].

Treatment of the solvent can be dealt with by techniques that handle the solvent implicitly as a continuum instead of explicit molecules, as these offer reasonable accuracy while remaining less computationally intensive[65]. An example of an implicit solvent technique is the Generalised Born surface area (GBSA) method, which treats the solvent

as a whole as a continuum in a neighbourhood of individual solute centres possessing different dielectric constant from the solvent[66,67]. Equation 1.2 describes the derivation of the free energy of solvent in this method.

$$\Delta G_{water} = E_{MM} + \Delta G_{solvation} - T\Delta S$$

Equation 1.2: Calculation of free energy of solvent water; E_{MM} is the molecular mechanics calculated energy for the solvent, $\Delta G_{solvation}$ is the difference in free energy caused by dissolving the solute into the solvent, and T is the temperature in Kelvin and S the entropy of the system

The molecular mechanics energy is taken from the energies of individual interaction types as calculated in Equation 1.3.

$$E_{MM} = E_{bonds} + E_{angle} - E_{torsion} + E_{electrostatic} + E_{vdW}$$

Equation 1.3: Molecular mechanics energy of water molecules

With advancements in modelling techniques as well as improved computational hardware, the idea of virtual screening, that is, screening a series of ligands against a protein target computationally, became increasingly practical[68]. An example of the techniques used was the design of renin inhibitors by Boger et al., which used both homology modelling and docking techniques (both of which will be described below[69]). Virtual screening has advantages over traditional experimental screening as it is cheaper and allows for a larger ligand library for an initial pass[70]. Docking programs such as GOLD and FLEXX have been employed for virtual screening. Molecular docking programs attempt to fit a flexible molecule into a rigid receptor by minimization of the ligand energy within restraints set by angles and distances from individual atoms from the ligand to the receptor[71,72]. Docking programs treat the protein structure as fixed in place and then search for conformations of each ligand that will fit into the binding site. The conformational search can use various computational methods such as genetic algorithms or Monte Carlo; the assessment of the quality of fit uses varying levels

of sophistication of energy evaluation[72,73]. These are either knowledge-based or based on classical physics-based energies such as Lennard-Jones potentials and Coulomb energies[74–76]. Docking has been used to investigate both protein-ligand and protein-protein interactions[77–79]. Once an initial set of so-called "poses" is generated, more sophisticated calculations can be used to analyse the results.

A range of more sophisticated methods have been developed to simulate the details of protein structure and motion. Molecular dynamics (MD) uses the molecular mechanics description of a molecule and allows for simulated physical motion over time within the system using classical or Newtonian physics for the entire structure. These calculations can simulate the motion in a molecular system. The timescale of MD simulations started as picoseconds, but now microsecond to millisecond timescales are possible to simulate, which can be used to model movement of an enzyme complex as well as the catalytic reaction[80].

In addition to comparative modelling such as homology modelling and protein threading, there has been increased research in the area of *de novo*, or *ab initio*, protein structure prediction. *Ab initio* techniques predict the position and energy of molecules based on quantum mechanics calculations. These models are derived from first principles or semiempirical methods, neither of which requires a template protein structure. Therefore, *Ab initio* methods allow for modelling of novel families of proteins with few related structures. In addition, quantum mechanics calculations can be used to model active sites with high precision. As no template is required to compare the structure with, a model produced may not match the actual 3D structure. The model may then be compared with a comparative model or molecular dynamics may be used to test the stability of the predicted conformation. In addition, quantum mechanics-based methods are computationally intensive, especially for larger systems such as complete proteins or protein-ligand complexes. A hybrid quantum mechanics/molecular mechanics (QM/MM) model may be used to counteract this, as the active site may be modelled with quantum mechanics methods for high precision while treating the remainder of the protein with less intensive classical physics-based molecular mechanics[64,81,82].

Using binding site mapping by programs such as GRID, programs such as HOOK may be used to build novel ligands[83]. HOOK connects potential functional groups that could bind to regions of the binding site with molecular skeletons. These skeletons maximise the groups' contact with the protein and are kept in a database. Once a skeleton has been selected, the resulting molecule's fit, referred to as the "overlap score", within the binding site is calculated based on a combination of van der Waals interactions. This process is then repeated with a number of potential molecules and then the results are ranked based on the score[83].

Metal and water interactions with the protein-ligand complex may require special parameterization. In particular, the treatment of solvent by methods such as GBSA is of great interest as modelling individual solvent molecules may be insufficient and lack overall solvent effects[66]. The model may be further complicated by the presence of flexible loop regions with less defined structure. Several programs have techniques to attempt to more accurately model these regions through *ab initio* modelling of loops in comparative modelling programs or extending short fragments into full loop domains by searching databases for the ends of the loop in *ab initio* programs [84–86].

1.6 Molecular Mechanics Modelling

Molecular mechanics use calculation of bond angles, distances, dihedral angles, and charges based on classical (or Newtonian) mechanics to set restraints on the target molecule. The structure is then fit within these restraints using molecular mechanics to minimize the potential energy of the system. This project used such molecular mechanics methods as an optimization step to test the stability of the final structures [87]. An alternate technique would be to use Monte Carlo methods to sample a series of random inputs (such as atom positions in a structure) to achieve an overall probability distribution[73]. The modelled structure may become trapped in local energy minima i.e. the structure is not globally preferred. This can be dealt with by varying the starting position to determine whether this has an effect on the resulting structure. In addition, there is an implicit assumption that for the overall protein structure, quantum physical calculations may be approximated by classical physics calculations. For the work in

Chapter 2, the solvent was not modelled, as there were no explicit solvent atoms in the crystal structure (PdbID: 2xct) and the models were meant to be a basic comparison as opposed to an in-depth description of the bound ligands[88]. MM calculations are unable to simulate motion without using MD, which becomes exponentially difficult as the number of atoms in the system increases. Similarly, MM may not be highly accurate, and QM also becomes difficult as the number of potential interactions increases. This can be ameliorated by QM/MM as this allows for more precise QM calculations at the active site alone. QM/MM was not used for the MM model in Chapter 2, as this would be more complex and computationally intensive for a short project that was simply used to support the experimental findings and provide a comparison between the DNA gyrase inhibitors. Protein-ligand binding is commonly simulated using molecular dynamics and Monte Carlo techniques[89].

1.7 Homology Modelling

Homology modelling is a computational method that attempts to model the three-dimensional structure of a protein based on another protein with which it shares a high level of sequence homology. Sequence homology is calculated as a percentage of amino acid residues conserved between the two (or more) proteins. It has been shown that proteins with sequence homology tend to have conserved tertiary structure. Therefore, the homology model uses a related protein (referred to as the template protein) with sequence homology to set restraints for the protein sequence with unknown structure[90]. These restraints consist of: ϕ and ψ dihedral angles, side chain dihedral angles, main chain C_α hydrogen bond distances, main chain N-O distances, and any other stereochemical restraints. >40% sequence homology is required for a trustworthy model, as it has been shown that above 40% homology, the structures tend to share a majority of the protein folds[91].

The target protein's sequence must be aligned to the template. Afterwards, the aligned target sequence is fit within the restraints set up for the template. The fit is scored using a probability density function, which attempts to minimize the entropy of the system in order to find the native protein structure [90].

Due to the need for high homology between the target and template proteins, homology modelling is less useful for proteins in newly discovered families or modelling those in families with few available crystal structures. When there is a well-conserved family of proteins, multiple sequence alignment is useful to provide a more accurate alignment with respect to the conserved regions of the protein. The difference between homology modelling and protein threading is the template used; although both require a template structure, homology modelling requires homologous regions of primary structure between the template structure and the target sequence while threading requires homologous regions of tertiary structure ("folds"). Homology modelling was performed for IcaB in chapters 3 and 4 to attempt to improve the model. Modeller was used for the homology modelling aspects of this thesis as it was one of the first homology modelling packages developed and has been well described and updated [90,91]. Automated modelling pipelines such as SWISS-MODEL facilitate the use of homology models for use in drug design or virtual screening by automating the entire process and providing a free and user-friendly web interface[92]. In addition, databases of stored models in SWISS-PROT (UniProt) and MODBASE may be searched to find existing homology models for the target protein[93,94]. Flexibility, particularly in protein side chains, is still a challenge for computational modelling in general and specifically homology modelling, despite techniques such as addition of molecular dynamics modelling of highly variable loop domains[86,95]**.** Although homology models with low sequence homology (<40%) are generally considered poor overall, highly homologous regions around binding sites are key and can lead to accurate site models despite poor global quality[95]. This will be further explored in this thesis through the homology model of IcaB in Chapter 4.

1.8 Outline of Thesis Contents

The focus of this thesis is on the computational modelling of a number of protein targets for structure-based drug design. While the first chapter gave a review of protein structure and function as a basis for molecular modelling in protein-ligand interactions, Chapter 2 is dedicated to the study of energy minimization of ciprofloxacin derivatives bound to DNA gyrase using CHARMm force fields. These derivatives were synthesized from ciprofloxacin and citrate molecules using linkers of various sizes. Chapter 3

provides a comprehensive review of deacetylase enzymes to elucidate features common to this enzyme family. The deacetylase family can be broadly categorized into two groups: lysine deacetylases and carbohydrate esterases. The conserved active site and substrate binding pocket residues of the carbohydrate esterase family 4 were key to producing a functionally relevant homology model of a poly-N-acetylglucosamine deacetylase, IcaB, by alignment with a related protein, PgaB. A full description and analysis of the homology model is described in Chapter 4.

Chapter 2: Modelling of Citric Acid-Ciprofloxacin Conjugates

2.1 Cystic Fibrosis and Opportunistic Infections

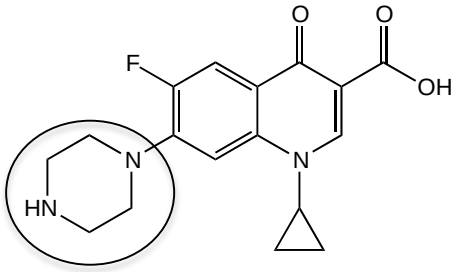
Cystic fibrosis is a congenital disease that is recessively inherited and is commonly diagnosed by increased sodium levels in sweat[96]. In cystic fibrosis, mucous builds up on the respiratory passages due to a decrease in chloride ion excretion, which is most often caused by a malfunctioning chloride channel encoded by the cystic fibrosis transmembrane conductance regulator (CFTR) gene found on human chromosome 7[97]. Biofilms of opportunistic pathogenic gram-negative bacteria such as *Pseudomonas aeruginosa* and *Burkholderia cepacia* may form in this layer and are difficult to treat even with long-term antibiotic therapy[98]. Chronic infection by *P. aeruginosa* is the primary cause of mortality and morbidity in patients with cystic fibrosis[99,100]. It has been well documented that bacterial resistance is forming to many antibacterial agents currently in use, particularly aminoglycosides, such as tobramycin, and fluoroquinolones, such as ciprofloxacin, in bacteria like *P. aeruginosa* that form biofilms[101–103].

One of the factors causing this resistance is the formation of bacterial biofilms. A biofilm is an aggregation of bacterial cells within an extracellular matrix formed by bacterial polymers and proteins[104]. The biofilm provides a mechanical barrier which prevents antibody binding and protects against innate host defences such as phagocytosis and antimicrobial peptides, including complement fragments[105]. Additionally, the presence of persister cells, cells that have entered a non-dividing dormant stage within the biofilm, can protect against antibiotics that inhibit DNA replication and processes of cell division, even those that diffuse through the biofilm[106]. These cells can then re-enter an active dividing stage, resulting in an antibiotic resistant population of cells[107]. As a result of these factors, the Centers for Disease Control and Prevention (CDC) estimated biofilms to be involved in 65% of bacterial infections in 1999[108].

In addition to the resistance mechanisms provided by the biofilms, there are a number of other methods including protein efflux pumps that remove antibiotics from the cell, hydrolytic enzymes (for example, β-lactamases which catalyze the breakdown of penicillins), or modification of antibiotic targets (such as methylation of an adenine residue in the small ribosome subunit, which lowers the binding affinity of erythromycin

and related antibiotics without affecting protein synthesis by the ribosome)[109]. *P. aeruginosa* utilizes all these mechanisms by use of four different multi-protein efflux complexes that remove almost all classes of antibiotics from the cell by at least one complex, synthesis of β-lactamases to protect against β-lactam based antibiotics such as penicillin, and mutation of the DNA gyrase subunit A, a common target of quinolone antibiotics[110].

Modification of previously existing antibiotics has shown to be a novel alternative strategy for overcoming resistance of cystic fibrosis-associated infections, such as *P. aeruginosa* [96]. Fluoroquinolones are effective antibiotics for many gram-positive and gram-negative bacterial infections including *P. aeruginosa* and *B. cepacia*, both commonly associated with cystic fibrosis[111]. However, fluoroquinolones are not highly localized in the bacterial cells due to low uptake across bacterial cell membranes caused by a number of protein factors[112]. The lower concentration of fluoroquinolone may promote the development of antibiotic-resistant strains by allowing these bacterial species to become more resistant[113]. In 2003, Pitt et al. reported that 30% of *P. aeruginosa* strains identified from patients with cystic fibrosis in the UK were found to be resistant to ciprofloxacin, a commonly prescribed fluoroquinolone[102]. The chemical structure of ciprofloxacin is shown below:



Piperazinyl ring

(**1**): Ciprofloxacin

A more recent study by Emerson et al. comparing drug resistance of *P. aeruginosa* between studies performed in the United States in 1995 and 2008 did not show an increase in resistance to ciprofloxacin. However, this may be due to the use of aminoglycoside antibiotics such as tobramycin as the primary therapy for cystic fibrosis, to which the bacterial strains did show an increase in resistance[114]. This antibiotic resistance has led to investigation into alternative routes of administration, such as inhalation of aerosolized antibiotics[100,115].

Conjugation of the piperazinyl ring of fluoroquinolones to a small molecule ligand has been used to produce a compound which maintained a similar level of antibiotic activity as the original fluoroquinolone[116]. As a result of this modification, conjugated fluoroquinolone derivatives with siderophore (small Fe^{3+}-chelating compound) ligands such as citrate, hydroxamate, and catecholate groups have previously been explored in order to increase the intracellular concentration in bacteria[117–119]. In addition, monosaccharide ligands such as glucose and galactose have been conjugated to ciprofloxacin to utilize bacterial carbohydrate transporter proteins[120]. Examples of these proteins include galactose permease (GalP), which forms trimers that transport sugars into the cell using the proton gradient produced during cellular respiration, and MglBAC, an ATP-dependent transport protein[121,122].

2.2. DNA Gyrase and Ciprofloxacin Activity

Fluoroquinolones such as ciprofloxacin bind to DNA gyrase, a type II topoisomerase enzyme, to inhibit negative supercoiling of DNA by preventing ligation of dsDNA breaks. The structure of the DNA gyrase-ciprofloxacin complex, which was used as a starting point for modelling, was taken from the Protein Data Bank pdbID: 2xct [88]. The structure of DNA gyrase from *Staphylococcus aureus* is shown in Figures 2.1 and 2.2 below and the binding site of ciprofloxacin in DNA gyrase is shown in Figure 2.3.

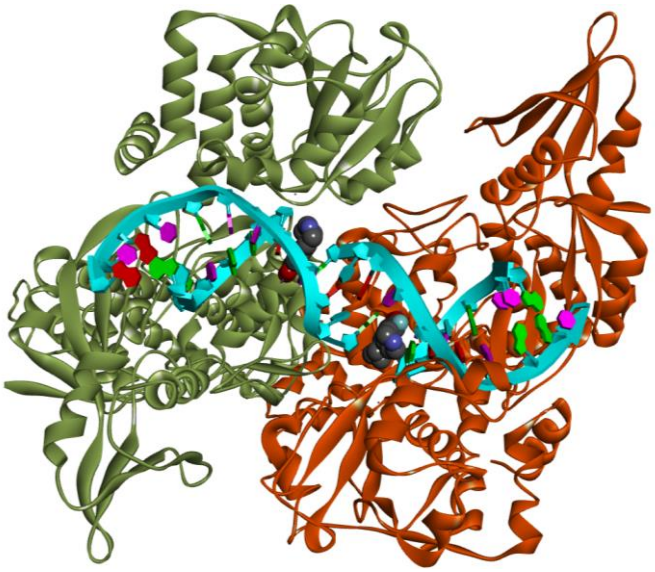


Figure 2.1

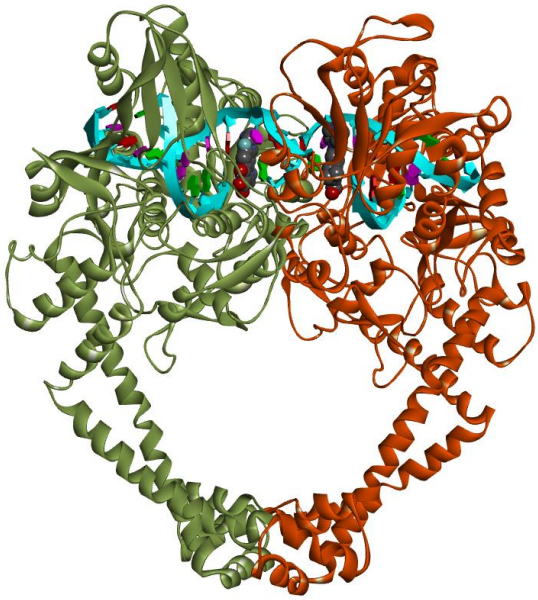


Figure 2.2

Top (Figure 2.1) and side (Figure 2.2) view of structure of *Staphylococcus aureus* DNA gyrase complexed to ciprofloxacin (ciprofloxacin represented as a space filling model, pdbID: 2xct)[88]. The red and green chains represent individual protein monomers of a gyrase A-gyrase B fusion. The turquoise helix represents a DNA segment.

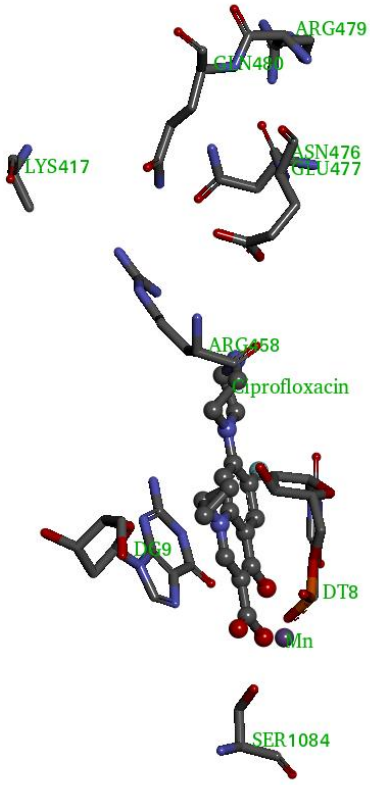


Figure 2.3: The ciprofloxacin-binding site with annotated adjacent amino acid and nucleotide residues

Type II topoisomerases use ATP to introduce negative supercoiling into DNA strands by creating double stranded breaks in a segment of DNA known as the G, or gate, strand through which the T, or transfer, strand of DNA may then be passed[88,123]. This inhibition occurs due to intercalation of the ciprofloxacin into DNA, which prevents binding of exposed phosphate groups in both strands of the broken DNA backbone to Tyr122, a tyrosine residue in DNA gyrase[111]. This double strand broken DNA prevents DNA synthesis by preventing the introduction of negative supercoils into the DNA backbone. Without DNA synthesis, apoptosis of the cell occurs through the fragmentation of bacterial chromosomes[124].

In order to enhance the gyrase inhibition of ciprofloxacin, conjugates of ciprofloxacin and citrate have been prepared[117,125]. An advantage of citrate as a ligand is that citrate has been shown to be used as a siderophore by bacteria. This has been best explored in *Escherichia coli* through the FecA membrane transporter[126,127]. The membrane porin OmpF (outer membrane protein F), an integral membrane protein that acts as a non-specific channel, has been shown to be an important factor in the uptake of ciprofloxacin and other fluoroquinolones[128]. A method of antibiotic resistant adaptation in *E. coli* is to reduce levels of OmpF transcription, which in turn reduces permeability to and intracellular concentration of fluoroquinolones[112]. However, iron uptake using a siderophore and FecA is unaffected and may then be exploited for fluoroquinolone uptake[129]. Thus, the presence of a citrate group on the target antibiotic may mediate transport of this antibiotic across cellular membranes even in the presence of low levels of OmpF. In addition, ciprofloxacin has been shown to retain its antimicrobial activity when conjugated to citrate[117].

While citrate's activity as a siderophore is best understood in *Escherichia coli*, no complete DNA gyrase X-ray crystallographic structure for *E. coli* exists in the Protein DataBank. As a result, the homologous *Staphylococcus aureus* DNA gyrase was selected as a model for the *E. coli* gyrase (43.2% sequence identity)[125]. A sequence alignment between the full *S. aureus* sequence and the *E. coli* sequence of a fusion protein combining the GyrA and GyrB subunits is shown in Figure 2.4.

```
S. aureus       ------------------------------------------------------------
E. coli Fusion  GLIAVVSVKVPDPKFSSQTKDKLVSSEVKSAVEQQMNELLAEYLLENPTDAKIVVGKIID


S. aureus       --------------------MDVASLPGKLADCSSKSPEECEIFLVEGDSAGGSTKSGRD
E. coli Fusion  AARAREAARRAREMTRRKGALDLAGLPGKLADCQERDPALSELYLVEGDSAGGSAKQGRN
                            :*:*.********..:.*  .*::**********:*.**:

S. aureus       SRTQAILPLRGKILNVEKARLDRILNNNEIRQMITAFGTGIG-GDFDLAKARYHKIVIMT
E. coli Fusion  RKNQAILPLKGKILNVEKARFDKMLSSQEVATLITALGCGIGRDEYNPDKLRYHSIIIMT
                 :.******:**********:*::*..:*:  :***:* *** .:::  * ***.*:***

S. aureus       DADVDGAHIRTLLLTFFYRFMRPLIEAGYVYIAQPPTGYKGLGEMNADQLWETTMNPEHR
E. coli Fusion  DADVDGSHIRTLLLTFFYRQMPEIVERGHVYIAQPP------------------------
                ******:************ *  ::* *:*******

S. aureus       ALLQVKLEDAIEADQTFEMLMGDVVENRRQFIEDNAVYANLDFAELPQSRINERNITSEM
E. coli Fusion  -LYKVKKG-------------------KQEQYIKD---------------------------
                 * :**                    ::.*:*:*

S. aureus       RESFLDYAMSVIVARALPDVRDGLKPVHRRILYGLNEQGMTPDKSYKKSARIVGDVMGKY
E. coli Fusion  DEAMDQYQISIALD------------------------GATLHTNASKSARVVGDVIGKY
                 *:: :* :*: :                         * * ... .****:****:***

S. aureus       HPHGDSSIYEAMVRMAQDFSYRYPLVDGQGNFGSMDGDGAAAMRFTEARMTKITLELLRD
E. coli Fusion  HPHGDSAVYDTIVRMAQPFSLRYMLVDGQGNFGSIDGDSAAAMRYTEIRLAKIAHELMAD
                ******::*:::***** ** ** **********:***.*****:** *::**: **: *

S. aureus       INKDTIDFIDNYDGNEREPSVLPARFPNLLANGASGIAVGMATNIPPHNLTELINGVLSL
E. coli Fusion  LEKETVDFVDNYDGTEKIPDVMPTKIPNLLVNGSSGIAVGMATNIPPHNLTEVINGCLAY
                ::*:*:**:*****.*: *.*:*:::****.**:******************:*** *:

S. aureus       SKNPDISIAELMEDIEGPDFPTAGLILGKSGIRRAYETGRGSIQMRSRAVIEERGG-GRQ
E. coli Fusion  IDDEDISIEGLMEHIPGPDFPTAAIINGRRGIEEAYRTGRGKVYIRARAEVEVDAKTGRE
                 .: ****  ***.* *******.:* *: **..**.****.: :*:** :*  .  **:

S. aureus       RIVVTEIPFQVNKARMIEKIAELVRDKKIDGITDLRDETSLRTGVRVVIDVRKDANASVI
E. coli Fusion  TIIVHEIPYQVNKARLIEKIAELVKEKRVEGISALRDE-SDKDGMRIVIEVKRDAVGEVV
                 *:* ***:******:********::*:::**: **** * : *:*:**:*::** ..*:

S. aureus       LNNLYKQTPLQTSFGVNMIALVNGRPKLINLKEALVHYLEHQKTVVRRRTQYNLRKAKDR
E. coli Fusion  LNNLYSQTQLQVSFGINMVALHHGQPKIMNLKDIIAAFVRHRREVVTRRTIFELRKARDR
                *****.** **.***:**:** :*:**::***: :. ::.*:: ** *** ::****:**

S. aureus       AHILEGLRIALDHIDEIISTIRESDTDKVAMESLQQR-----------------------
E. coli Fusion  AHILEALAVALANIDPIIELIRHAPTPAEAKTALVANPWQLGNVAAMLERAGDDAARPEW
                *****.* :** :** **. **.: *   *  :*  .

S. aureus       -----------FKLSEKQAQAILDMRLRRLTGLERDKIEAEYNELLNYISELETILADEE
E. coli Fusion  LEPEFGVRDGLYYLTEQQAQAILDLRLQKLTGLEHEKLLDEYKELLDQIAELLRILGSAD
                           : *:*:*******:**::*****::*:  **:***: *:**  **.. :

S. aureus       VLLQLVRDELTEIRDRFGDDRRTEIQLG---------------------------------
E. coli Fusion  RLMEVIREELELVREQFGDKRRTEITANSADINLEDLITQEDVVVTLSHQGYVKYQPLSE
                 *::::*:**  :*::***.*****  .

S. aureus       ------------------------------------------------------------
E. coli Fusion  YEAQRRGGKGKSAARIKEEDFIDRLLVANTHDHILCFSSRGRVYSMKVYQLPEATRGARG


S. aureus       ------------------------------------------------------------
E. coli Fusion  RPIVNLLPLEQDERITAILPVTEFEEGVKVFMATANGTVKKTVLTEFNRLRTAGKVAIKL



S. aureus       ------------------------------------------------------------
E. coli Fusion  VEGDELIGVDLTSGEDEVMLFSAEGKVVRFKESSVRAMGCNTTGVRGIRLGEGDKVVSLI
```

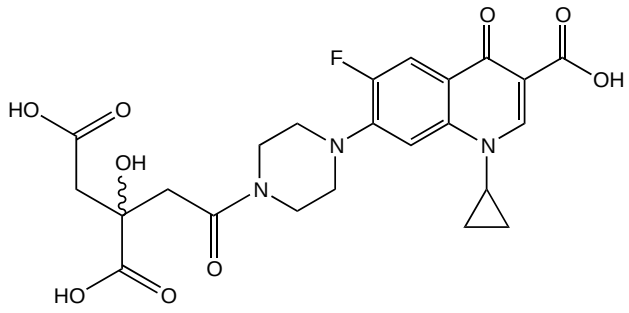
```
S. aureus       ------------------------------------------------------------
E. coli Fusion  VPRGDGAILTATQNGYGKRTAVAEYPTKSRATKGVISIKVTERNGLVVGAVQVDDCDQIM

S. aureus       ------------------------------------------------------------
E. coli Fusion  MITDAGTLVRTRVSEISIVGRNTQGVILIRTSEDENVVGLQRVAEPVDEEDLDTIDGSAA

S. aureus       ----------------------------------------------------
E. coli Fusion  EGDDEIAPEVDVDDEPEEE-----------------------------------
```

Figure 2.4: Sequence alignment, derived using ClustalX 2.1, of *S. aureus* DNA gyrase with *E. coli* DNA gyrase (protein sequence is formed from a construct between the GyrA and GyrB subunits)[130] (* = identical, : = conserved, .= semiconserved)

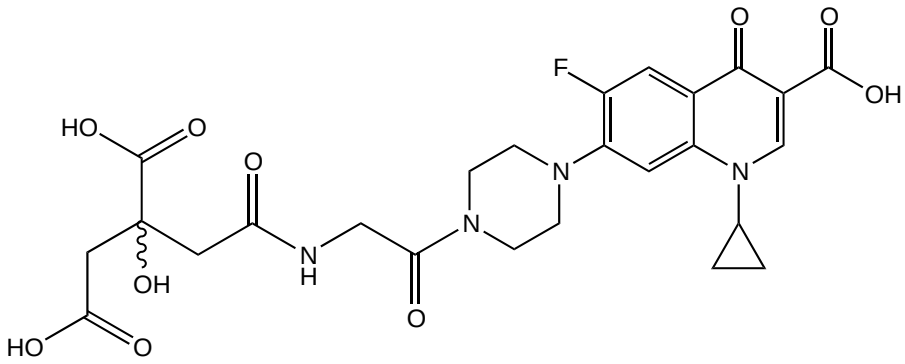
Conservation in Clustal is defined by the Gonnet Pam250 scoring matrix in which the conserved groups have a score above 0.5 and semiconserved groups have a score of 0.5 and below[130].

2.3 Structure of Ciprofloxacin Derivatives

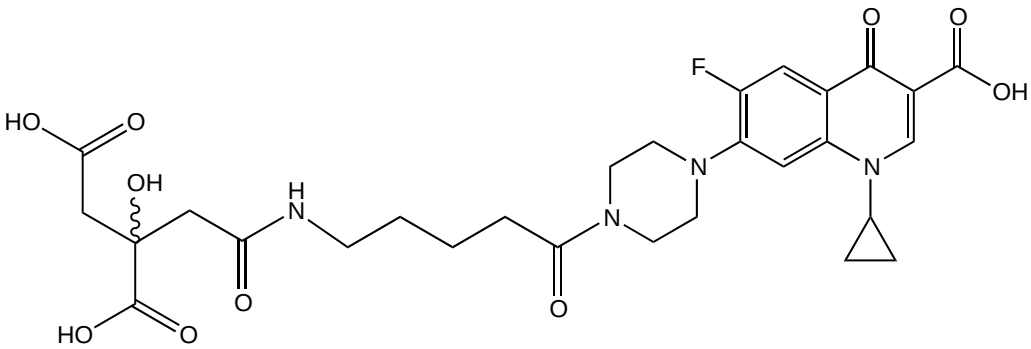
Derivatives of ciprofloxacin, **1,** were prepared with citrate conjugated to the C7-piperazinyl ring **2**[117]. To determine whether the length of the C7-piperazinyl attached group affected the DNA gyrase inhibition, a spacer group was added between the piperazinyl ring and the citrate group to yield c-gly-ciprofloxacin **3** and c-ava-ciprofloxacin **4**. All compounds were prepared by the research group of Dr. Anne-Katrin Duhme-Klair and Dr. Anne Routledge, Department of Chemistry, The University of York[125].



(**2**): c-ciprofloxacin



(**3**): c-gly-ciprofloxacin



(**4**): c-ava-ciprofloxacin

The synthesis of c-ciprofloxacin from the conjugate of citrate and ciprofloxacin is shown in Md-Saleh et al., 2009[117]. The synthesis of the c-gly-ciprofloxacin, **3**, and the c-ava-ciprofloxacin, **4**, are similarly described in Milner et al., 2006[131]. All syntheses were performed by the research group of Drs. Anne-Katrin Duhme-Klair and Anne Routledge in the Department of Chemistry at the University of York.

2.4 DNA Gyrase Inhibition Assays

Electrophoresis gel assays of DNA gyrase inhibition were performed for all compounds as described in the Inspiralis DNA gyrase inhibition assay protocol[132].

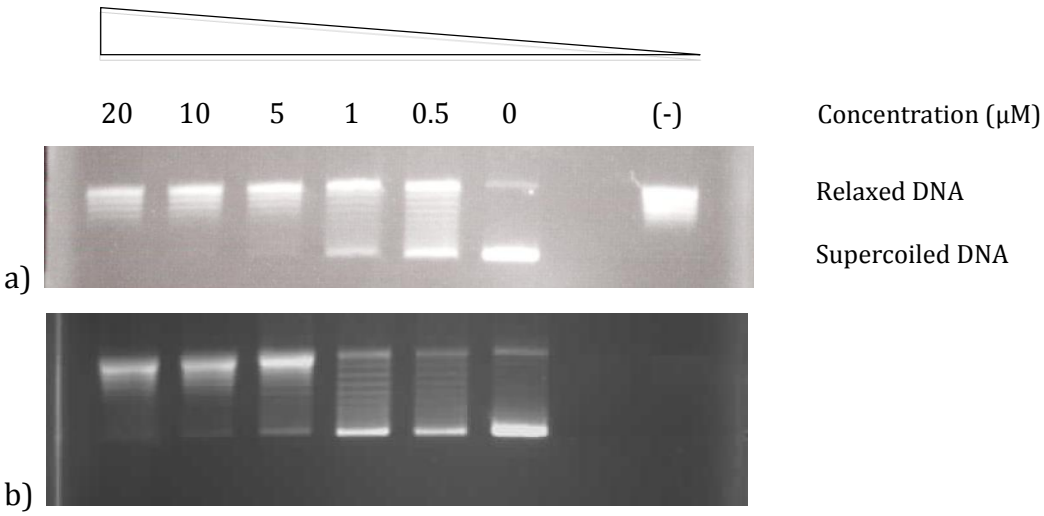
2.4.1 Methodology

All compounds were synthesized by the research group of Drs. Anne-Katrin Duhme-Klair and Anne Routledge in the Department of Chemistry at the University of York. All compounds were soluble at a concentration of 500 μM in 50mM Tris/HCl at pH 8.

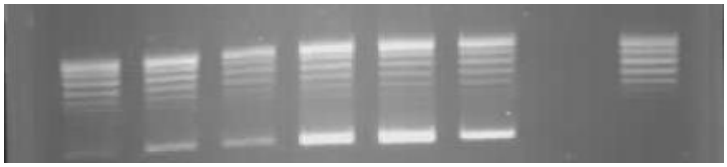
500 μM stock solutions in 50mM Tris/HCl at pH 8 were prepared for all compounds. These solutions were diluted to 30 μM in the same solvent before use in the gyrase inhibition assay. 1% TAE (Tris-acetate-EDTA) gels were made to test compounds using concentrations of ciprofloxacin or ciprofloxacin derivatives between 20 and 0.5 μM, as described in the Inspiralis DNA gyrase inhibition assay protocol[132]. One lane for the c-ava-ciprofloxacin gel was lost during the experiment due to damage to the gel.

2.4.2 Results

DNA gyrase assays were performed for a range of concentrations between 20 and 0.5μM for each of the test compounds using the Inspiralis *Escherichia coli* DNA gyrase assay kit 1[132]. The results of these assays are shown in Figure 2.5. A second set of assays was performed for all compounds except ciprofloxacin and the average was used for further analysis.



c)



d)

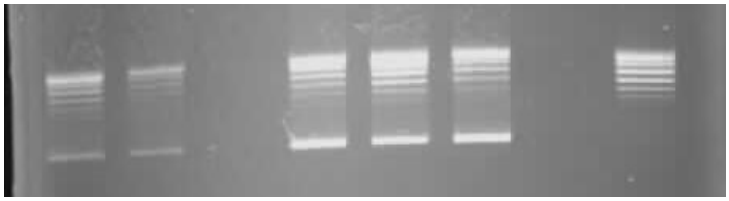


Figure 2.5: DNA gyrase assay electrophoresis gel a) ciprofloxacin; b) c-ciprofloxacin; c) c-gly-ciprofloxacin; d) c-ava-ciprofloxacin. The negative control lane contained only the DNA without any added gyrase

For each compound, the intensity of the supercoiled bands was measured using the ImageJ image-processing program[133]. The maximum intensity of these bands in each gel was set to 100% and all band intensities were normalized to the background intensity of each lane, which was then plotted with respect to the concentration of the compounds (Figures 2.6-2.9). To calculate the IC_{50} (the concentration in μM at which 50% inhibition is achieved) from each graph, Equation 2.1 was minimized using the Solver add-in from Microsoft Excel.

$$y = \mathrm{Bot} + \left(\frac{\mathrm{Top} - \mathrm{Bot}}{1 + 10^{k(IC_{50} - x)}}\right)$$

Equation 2.1: Curve to which data was fit for Figures 2.6-2.9. Top and Bot yield the limits on the y-axis of the curve, k represents the slope of the graph, and the IC_{50} is as stated above.

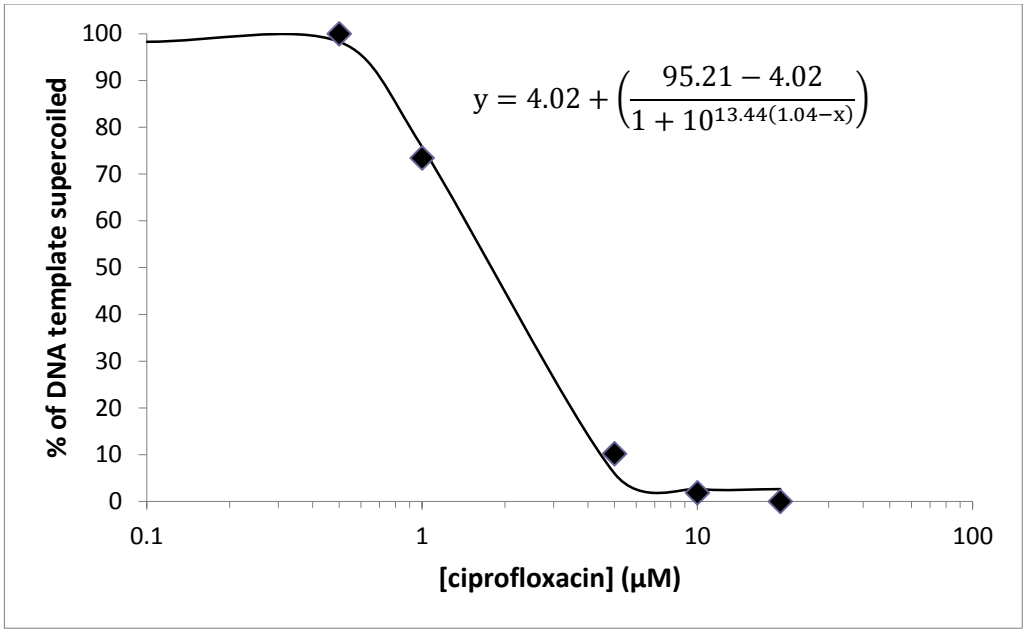


Figure 2.6: Effect of ciprofloxacin concentration on gyrase activity

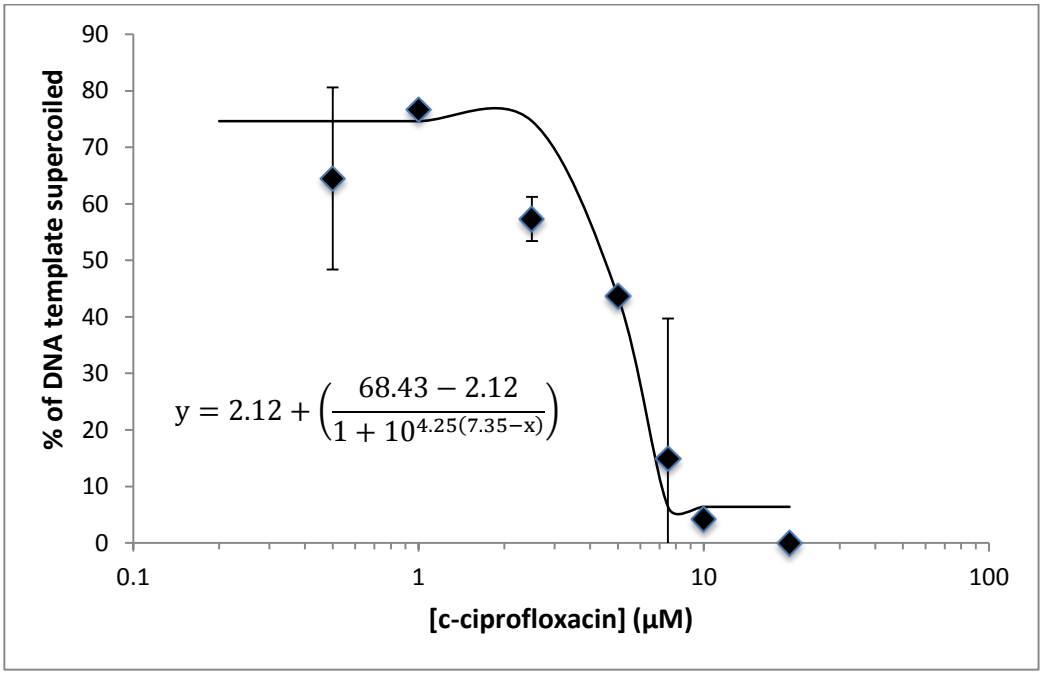


Figure 2.7: Effect of c-ciprofloxacin concentration on gyrase activity

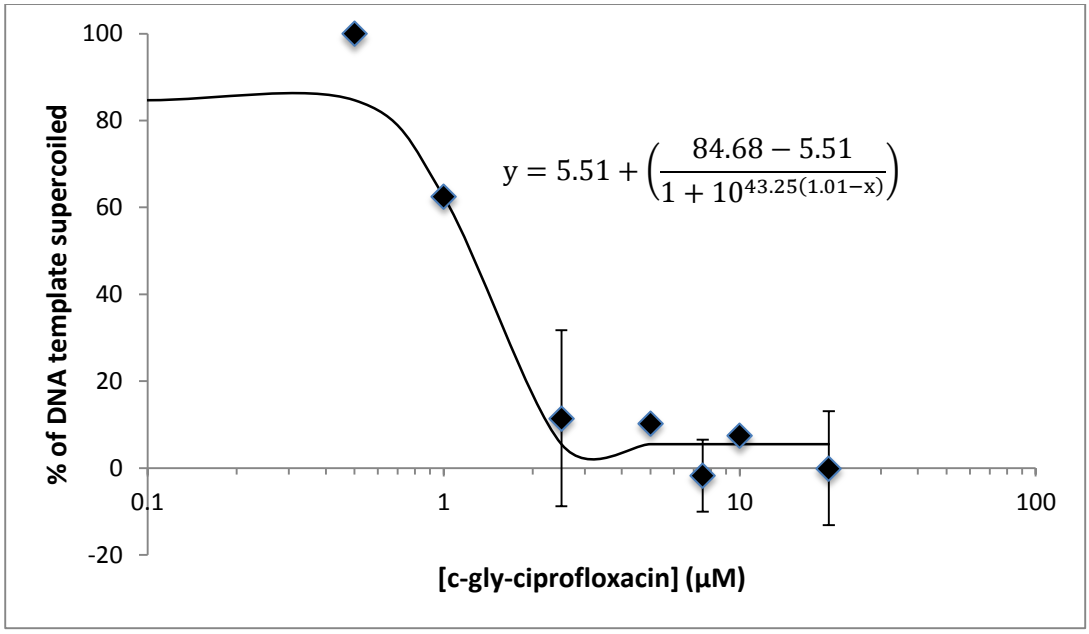


Figure 2.8: Effect of c-gly-ciprofloxacin concentration on gyrase activity

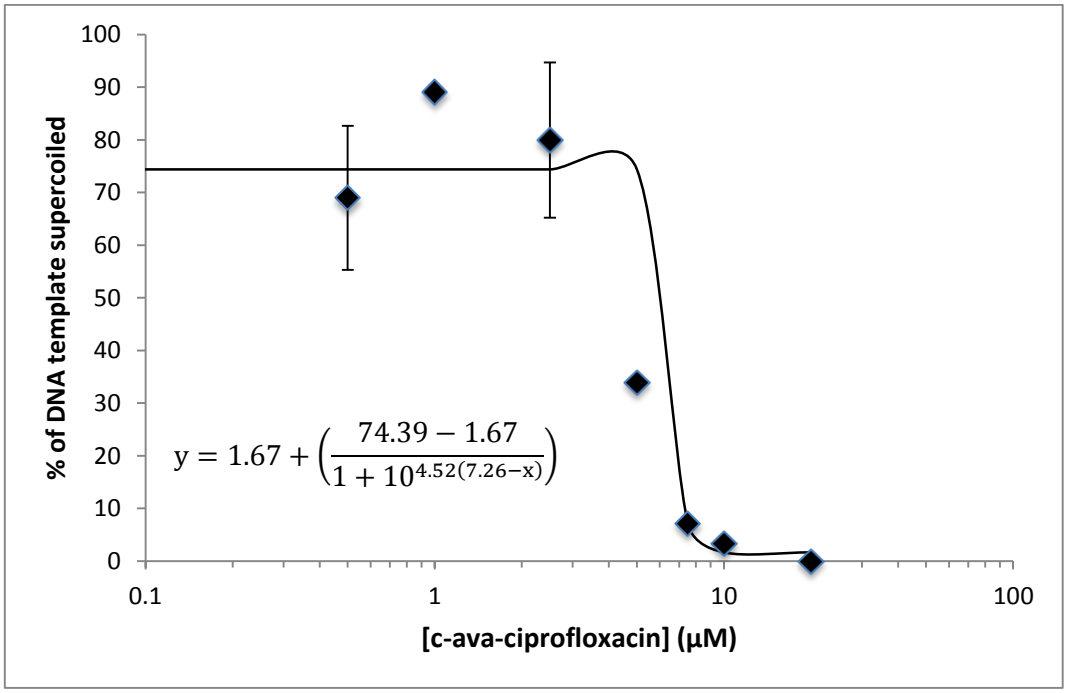


Figure 2.9: Effect of c-ava-ciprofloxacin concentration on gyrase activity

The steepest curve, which represents the compound that had the greatest effect at a lower concentration, was c-gly-ciprofloxacin. The IC_{50} for each compound was calculated (Table 2.1).

Table 2.1: IC_{50} calculated for each ciprofloxacin derivative based on electrophoresis gel assay data

Compound	IC50 (μM)
Ciprofloxacin	1.04
c-ciprofloxacin	7.35
c-gly-ciprofloxacin	1.01
c-ava-ciprofloxacin	7.26

2.4.3 Discussion

As the intensity of the supercoiled DNA increased with concentration of all ciprofloxacin derivatives, and DNA gyrase would have introduced negative supercoils to form the relaxed DNA structure, all compounds showed a concentration-dependent inhibitory effect on the DNA gyrase. This suggests that the ciprofloxacin derivatives have a similar mechanism to that of ciprofloxacin. Also, it can be shown qualitatively that, as the length of the linker increases, inhibition of DNA gyrase decreases. While one lane for the c-ava-ciprofloxacin gel was lost during the experiment, the overall pattern of the gel represented concentration-dependent inhibition.

IC_{50} values were lowest for ciprofloxacin and c-gly-ciprofloxacin and therefore these compounds yielded the greatest DNA gyrase inhibition. As a result of the small dataset and only two repeats of the experiment, these IC_{50} values are only useful for comparison between the various derivatives. More repeats of the gel assay would yield more accurate IC_{50} values.

2.5 Energy Minimization of Gyrase-Ciprofloxacin Complexes

Accelrys Discovery Studio 3.0 was used to compare these results to the computational energy of binding in the ciprofloxacin-gyrase complex.

The forcefield used in this thesis was the CHARMm forcefield and Momany-Rone partial charges to minimize structures in the Accelrys Discovery Studio 3.0 software package[134,135]. CHARMm (Chemistry at HARvard Molecular mechanics) is a

commercially available application of the CHARMM forcefield[134]. The potential energy of intermolecular and intramolecular interactions is calculated from this forcefield, which is a set of parameters governing the potential energy of various interactions, including electrostatics, Van der Waals, and hydrogen bonding as shown in Equation 2.2.

$$U(\vec{R}) = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{Urey-Bradley}} K_{UB}(S - S_0)^2 + \sum_{\text{dihedral}} K_\varphi(1 + \cos(n\varphi - \delta)) + \sum_{\text{impropers}} K_\omega(\omega - \omega_0)^2 + \sum_{\substack{\text{non-bonded} \\ \text{pairs}}} \left\{ \varepsilon_{ij}^{\min} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \right\} + \sum_{\text{residues}} U_{CMAP}(\varphi, \psi)$$

Equation 2.2: CHARMm energy calculation equation[134].

The first three terms represent the effect of bond distances (b), bond angles (θ), and dihedral angles (φ) on the overall potential energy (U). The Urey–Bradley term is a quadratic function of the distance between atoms A and C in a system of three bonded atoms A-B-C (S). The improper dihedral angle term represents a pseudo-dihedral angle at a branchpoint in which atoms B, C and D are all bound to the central atom A expressed as a quadratic function defined by B-C-A-D (ω). The improper and Urey-Bradley terms are used to account for vibrational spectra as well as out-of-plane motions in the molecule observed empirically through spectroscopy. The nonbonded term is a combination of the Coulombic interactions between point charges and Lennard-Jones potentials, which take into account Van der Waals attraction and core-core repulsion interactions between atoms. The point charges are represented as q_i and q_j and the distance is r_{ij}. The CMAP term is a 2-dimensional dihedral energy correction term used to optimize the protein backbone using the dihedral angles φ and ψ[134].

The structures of the ciprofloxacin derivatives bound to DNA gyrase were minimized in this forcefield using fixed atom constraints for all protein and DNA atoms by the Smart Minimizer algorithm in Accelrys Discovery Studio 3.0[135]. The algorithm consists of two aspects: the Steepest Descent method, a simple method that uses only the first derivative of the potential energy surface and saves only the positions between individual iterations for a quick removal of unfavourable positions, followed by the Conjugate Gradient method, an iterative method that records all previous minimization steps as well as the energy gradient, in order to come to a faster convergence than using the Steepest Descent Method alone.

Fixed atom constraints fix the positions of selected atoms during energy minimization or a molecular dynamics simulation. As a result, the CHARMm forcefield calculation will not cover these atoms. While the terms constraint and restraint have been used synonymously in some programs, a constraint as used by this algorithm completely fixes the positions of the atoms, while a restraint applies an additional penalty function to the calculated energy if the fixed atoms move from the starting positions.

2.5.1 Methodology

The ciprofloxacin and *Staphylococcus aureus* DNA gyrase complex structure (pdb: 2xct) was modeled computationally using Accelrys Discovery Studio 3.0[135]. To prepare the citrate derivatives, individual ligands were manually added to the ciprofloxacin one atom at a time. Appropriate hydrogen atoms and formal charges for pH 7 were added to simplify calculations. If necessary for future work, the effect of pH would then be varied using the structure yielded. The ciprofloxacin and ciprofloxacin derivative structures were parameterized using CHARMm forcefields and associated Momany-Rone partial charges[134].

All water molecules were removed to simplify calculations and avoid impact of solvent on the metal or hydrogen bonding interactions, as the resolution of this structure was 3.35Å. To be fully confident in the positions of the water molecules, a minimum resolution of ~2Å would be needed; therefore the water positions may not be highly reliable to use for modelling. The structure was minimized using fixed atom constraints

for all protein and DNA atoms. The Smart Minimizer algorithm of Discovery Studio 3.0 was employed using a maximum of 200 steps. Molecular dynamics simulations were then performed with no fixed atom positions, in order to determine the stability of the minimized structure[134]. The simulations were set to reach equilibrium by a maximum of 1000 steps of 0.001s at a target temperature of 300 K, followed by another minimization step, to ensure this structure was stable. The final structures derived for each protein-ligand complex were within an all-atom RMSD <1.0 Å. The minimized energy was calculated based on the overall CHARMm energy of the complex. The manganese ion coordinated to the ciprofloxacin was not fixed during minimization for simplicity as a full parameterization of the metal ion was beyond the scope of this project.

2.5.2 Results

The ciprofloxacin derivative-DNA gyrase complex was minimized using the CHARMm force field with Momany-Rone partial charges and optimized using molecular dynamics calculations in Accelrys Discovery Studio 3.0[135].

The manganese ion coordinated to the ciprofloxacin was not fixed during minimization and was found to change its position during minimization, which would not be expected due to coordination with the acid group of the quinoline ring as well as solvent water molecules. This may cause incorrect positioning of the metal due to incorrect parameterization of the manganese ion. This ion may need to maintain bonding with residues and/or water molecules. In addition, water molecules binding to the metal can be treated as fixed molecules or as a water bath. If the ion was fixed, a distance constraint could be used to maintain the bonding geometry. However, a full parameterization of all minimized groups was beyond the scope of this project.

The minimized structures of ciprofloxacin and its derivatives are shown in Figures 2.10-2.19. Note that the ciprofloxacin was minimized but retained its structure within 1Å RMSD of the crystal structure 2xct. c-ava-ciprofloxacin had two differing conformations within the same protein and both are described.

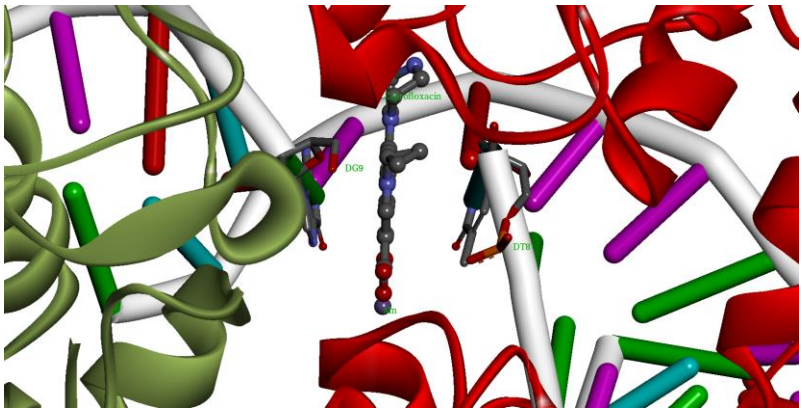


Figure 2.10: Computationally derived structure of the ciprofloxacin-DNA gyrase complex. The green and orange ribbons represent Gyrase A and B fusion protein subunits of DNA gyrase. The DNA sugar-phosphate backbone is in white while the bases are as follows: red = adenine, green= guanine, cyan = thymine, and magenta = cytosine.

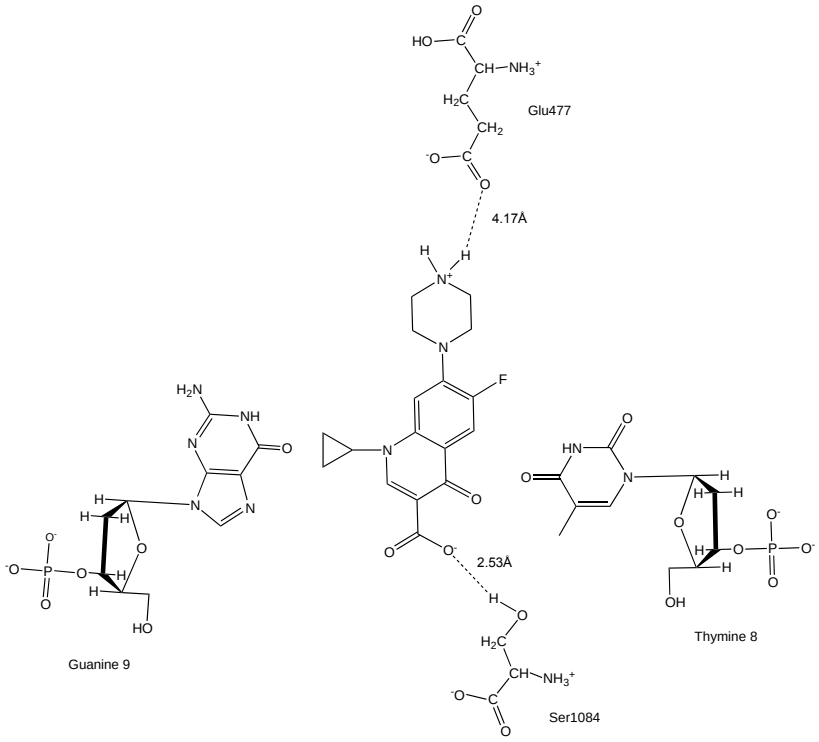


Figure 2.11: Hydrogen bonding interactions of ciprofloxacin-DNA gyrase complex. Note that the fluoroquinolone ring of all ciprofloxacin derivatives remains intercalated between the Thymine 8 and Guanine 9 residues of the DNA segment (Nucleotide residue numbers taken from the crystal structure).

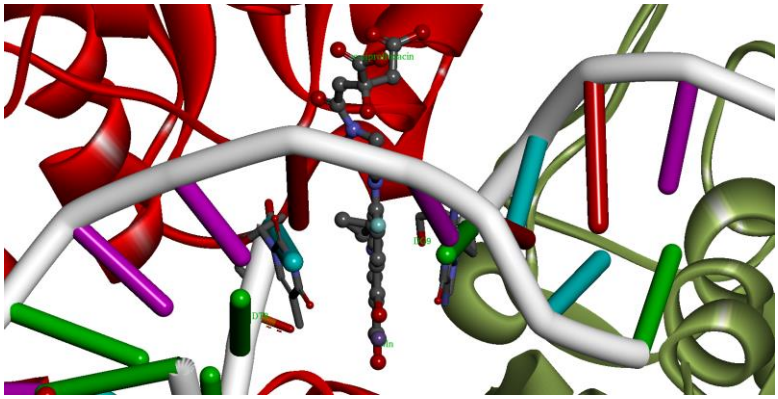


Figure 2.12: Computationally derived structure of the c-ciprofloxacin-DNA gyrase complex. The green and orange ribbons represent Gyrase A and B fusion protein subunits of DNA gyrase. The DNA sugar-phosphate backbone is in white while the bases are as follows: red = adenine, green= guanine, cyan = thymine, and magenta = cytosine.

The citrate attached to the ciprofloxacin piperazinyl ring caused increased hydrogen bonding between c-ciprofloxacin and arginine, lysine, and asparagine residues as shown in Figure 2.13.

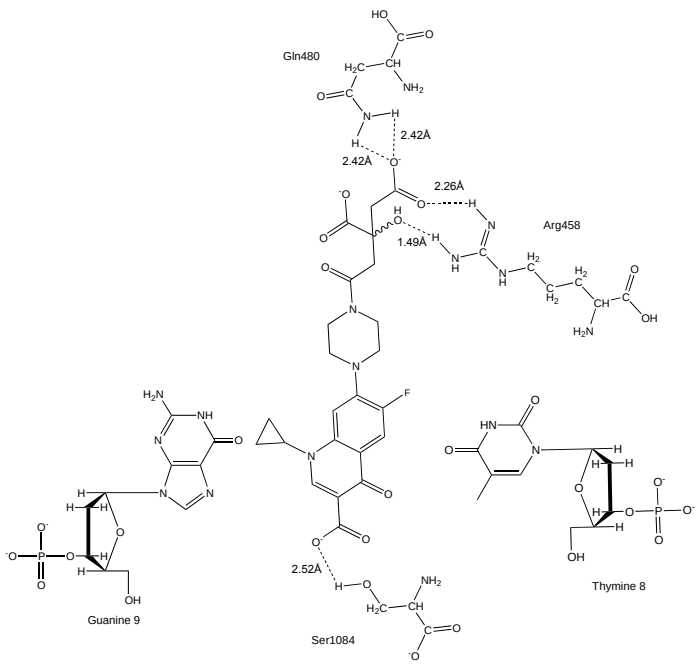


Figure 2.13: Hydrogen bonding interactions of the c-ciprofloxacin-DNA gyrase complex

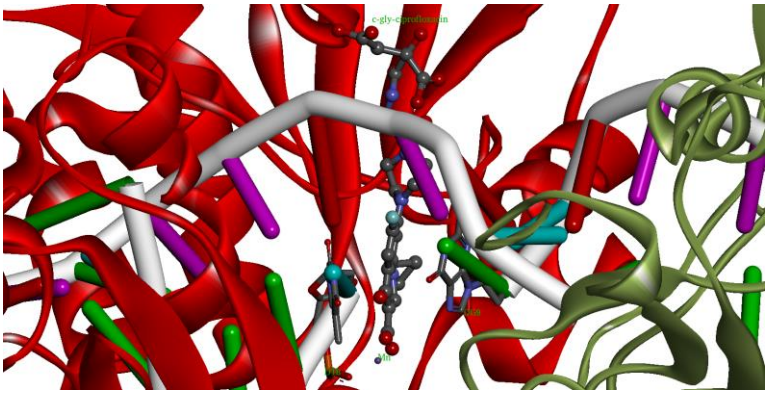


Figure 2.14: Computationally derived structure of the c-gly-ciprofloxacin-DNA gyrase complex. The green and orange ribbons represent Gyrase A and B fusion protein subunits of DNA gyrase. The DNA sugar-phosphate backbone is in white while the bases are as follows: red = adenine, green= guanine, cyan = thymine, and magenta = cytosine.

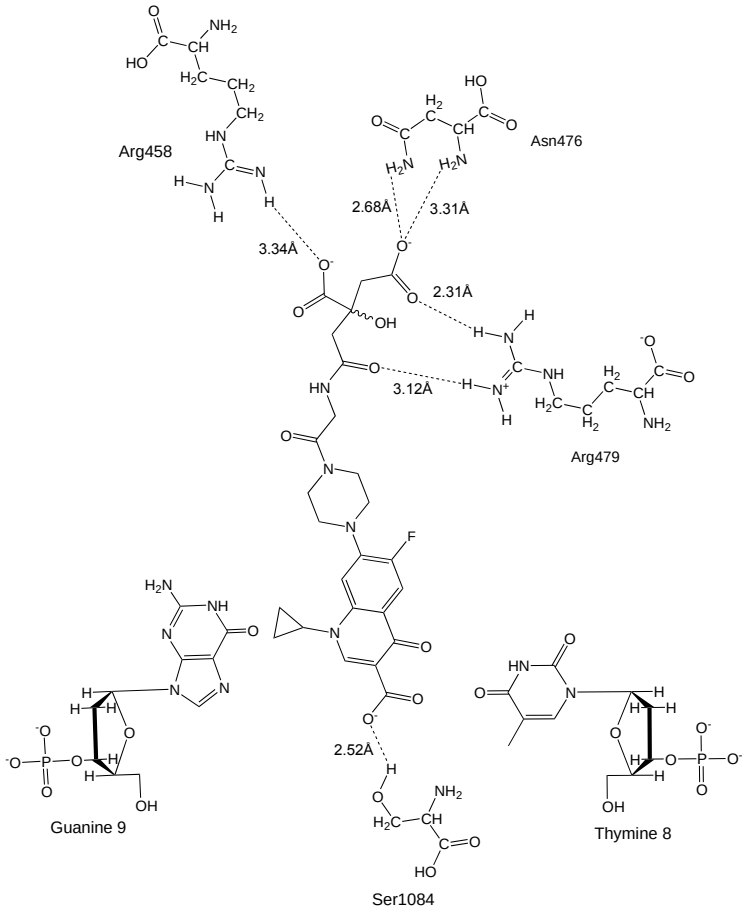


Figure 2.15: Hydrogen bonding interactions of the c-gly-ciprofloxacin-DNA gyrase complex

As shown in Figure 2.15, the c-gly-ciprofloxacin maintains the hydrogen bonding with the Arg458. In addition, Arg479 and Asn476 residues form hydrogen bonds with the c-gly-ciprofloxacin to further stabilize this structure.

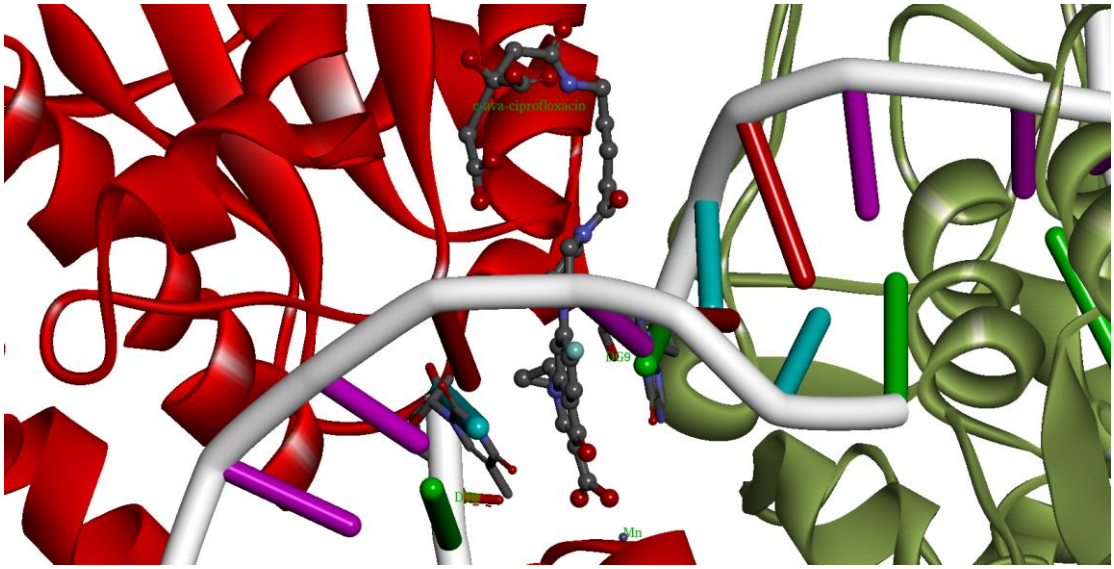


Figure 2.16: Computationally derived structure of conformation 1 of the c-ava-ciprofloxacin-DNA gyrase complex. The green and orange ribbons represent Gyrase A and B fusion protein subunits of DNA gyrase. The DNA sugar-phosphate backbone is in white while the bases are as follows: red = adenine, green= guanine, cyan = thymine, and magenta = cytosine.

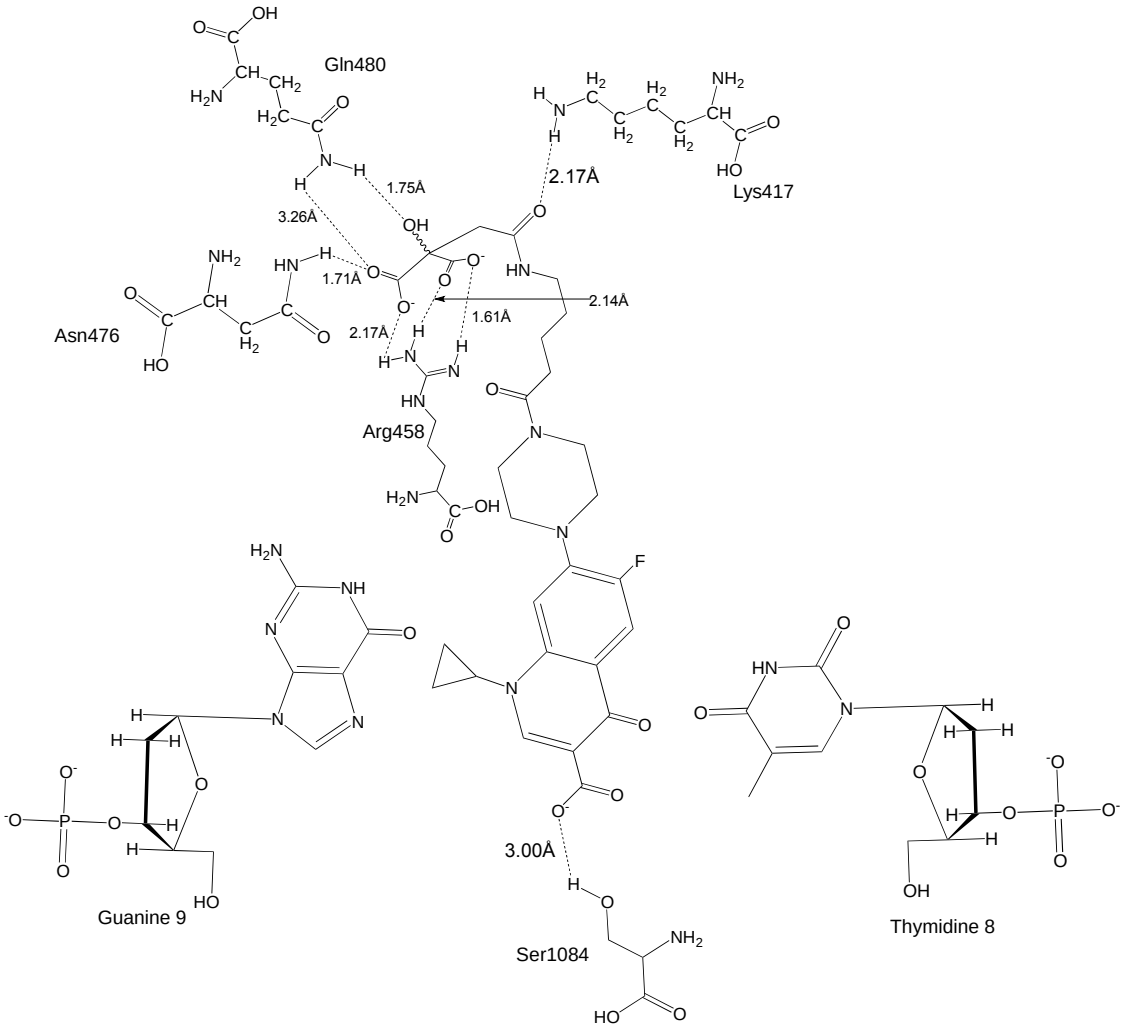


Figure 2.17: Hydrogen bonding interactions of the conformation 1 of c-ava-ciprofloxacin-DNA gyrase complex

This conformation of c-ava-ciprofloxacin shown in Figure 2.17 is directed away from the solvent towards the surface of the DNA gyrase due to hydrogen bonding with an arginine, Arg458, which bends the spacer group to allow maximum contact.

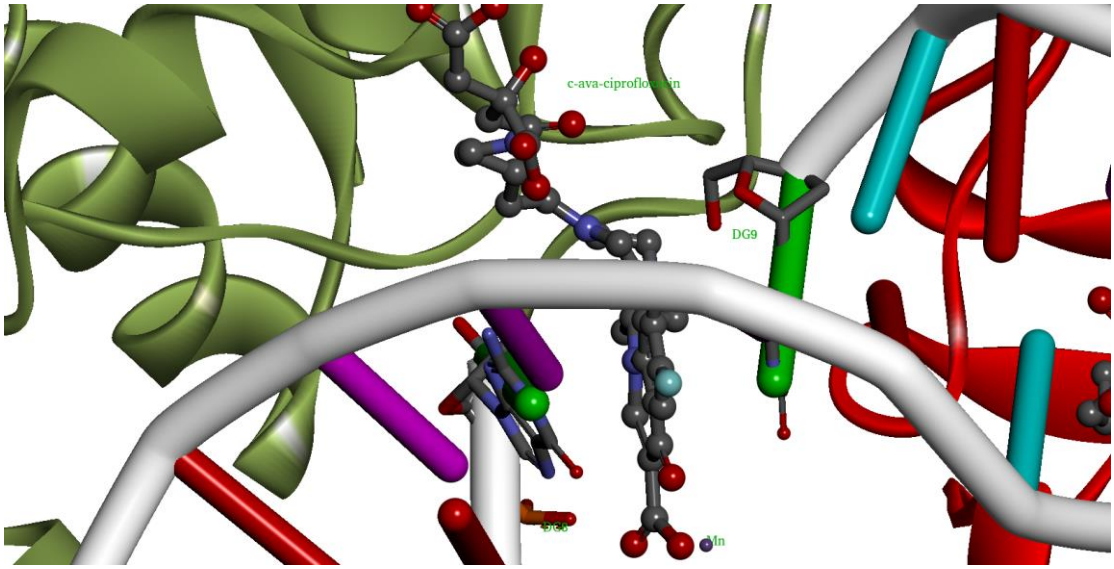


Figure 2.18: Computationally derived structure of conformation 2 of the c-ava-ciprofloxacin-DNA gyrase complex. The green and orange ribbons represent Gyrase A and B fusion protein subunits of DNA gyrase. The DNA sugar-phosphate backbone is in white while the bases are as follows: red = adenine, green= guanine, cyan = thymine, and magenta = cytosine.

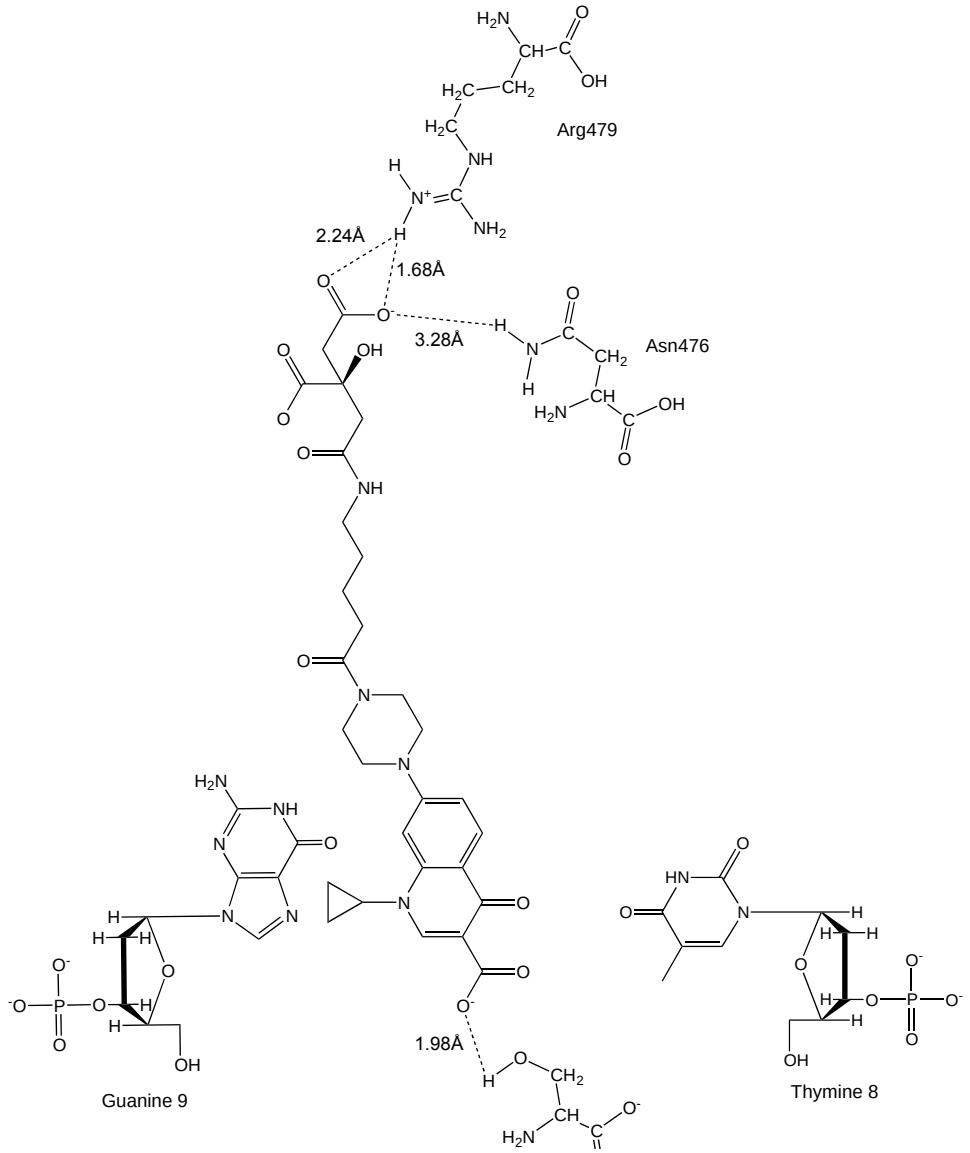


Figure 2.19: Hydrogen bonding interactions of the conformation 2 of c-ava-ciprofloxacin-DNA gyrase complex

The group attached to the piperazinyl ring of this conformation of c-ava-ciprofloxacin shown in Figure 2.19 remained solvent-oriented, as it did not form the hydrogen bond with the arginine residue bound to c-ava-ciprofloxacin conformation 1 (Arg458).

The ciprofloxacin core of all compounds remains intercalated with one of the DNA strands, while the C7 group attached to the piperazinyl ring is directed towards the solvent-exposed surface of the DNA gyrase. Hydrogen bonding between the atoms of the spacer and citrate groups attached to C7 and arginine, and other amino acids, stabilized

the interactions of these ciprofloxacin derivatives with DNA gyrase. The c-gly-ciprofloxacin complex had the greatest number of hydrogen bonding interactions. The stabilizing effects of this hydrogen bonding may explain the smaller change in energy between c-ciprofloxacin and c-gly-ciprofloxacin compared to the difference between c-gly-ciprofloxacin and c-ava-ciprofloxacin (Table 2.2). Despite the increase in spacer size from c-ciprofloxacin to c-gly-ciprofloxacin, the additional hydrogen bonding may stabilize the binding, resulting in a lower energy minimized structure for c-gly-ciprofloxacin than expected. Some of these hydrogen-bonding interactions are lost when the length of the spacer increases to c-ava-ciprofloxacin, which causes a larger increase in energy.

Table 2.2: Computationally determined energies of minimized DNA gyrase-ciprofloxacin derivative complexes

Compound	**Minimized Energy (kcal/mol)**
Ciprofloxacin	-2385
c-ciprofloxacin	-2273
c-gly-ciprofloxacin	-2218
c-ava-ciprofloxacin	-1750

2.5.3 Discussion

The compounds with groups conjugated to the piperazinyl ring had a higher energy than ciprofloxacin. The compound with the highest energy of the ciprofloxacin derivatives was the c-ava-ciprofloxacin, which also had the longest carbon chain linker. Therefore, the length of the linker is inversely proportional to the stability of the complex with DNA gyrase. Note that this was a crude modelling study and therefore, these energy values would require additional parameterization and correction to be accurate. However, as a qualitative comparison, this energy data reflects the experimental results of increasing energy of binding, and therefore decreased binding affinity, with length of the spacer.

Molecular dynamics simulations were performed on the minimized structures of the test compounds (Figures 2.10-2.19) using default parameters for CHARMm forcefields with

the positions of all atoms unfixed[134]. However, the structures derived from minimization were fairly stable and the structures achieved were almost identical (within RMSD of 1Å) to the original minimized structures.

c-ava-ciprofloxacin has two possible conformations due to a difference in binding, as shown in Figures 2.16-2.20. Conformation 2 forms fewer hydrogen bonds with the protein than conformation 1. The other ciprofloxacin derivatives do not show multiple conformations, likely due to the length of the aminovaleric acid spacer group introducing flexibility into the conjugated ligand, which may make the c-ava-ciprofloxacin more difficult to bind to the protein.

The length of the spacer group attached to the piperazinyl ring was found to be proportional to the energy of binding to the DNA gyrase. The c-gly-ciprofloxacin possessed the greatest number of hydrogen bonding interactions and added three atoms to the length of the spacer group. As a result, the optimal spacer group should not contain more than 1-3 atoms. This is further supported by comparison to an experimental screen of reference and clinical bacterial isolates against the derivatives to test for inhibition of growth by the research group of Drs. Anne-Katrin Duhme-Klair and Anne Routledge in the Department of Chemistry at the University of York[125]. In five out of the seven Gram-negative species tested, c-gly-ciprofloxacin was found to be more effective than c-ciprofloxacin. In Gram-positive species, both c-ciprofloxacin and c-gly-ciprofloxacin caused similar inhibition. In all strains, c-ava-ciprofloxacin possessed lower inhibitory activity or the strain was completely resistant to this compound. In all cases, the compounds tested still had less DNA gyrase inhibitory activity than ciprofloxacin[136].

Future work could be performed using other spacer groups to determine whether differing conformations of the same compound also have a difference in energy. It would also be useful to produce a ciprofloxacin derivative with a highly sterically hindered conjugate ligand to provide a negative control for further DNA gyrase assays with ciprofloxacin derivatives, as the steric hindrance at the solvent exposed region should prevent binding with high affinity to the DNA gyrase. Finally, proper parameterization of the manganese ion coordinated to the ciprofloxacin may result in a more accurate structure. A restraint on the distance between the manganese ion and the active site

residues as well as any solvent water molecules bound to the manganese may be required in order to maintain this distance during construction of the model. An important aspect of modelling the metal ion is to properly model the charge of the ion, which would require a higher resolution structure containing bound solvent atoms[137]. The solvent was not modelled, as there were no explicit solvent atoms in the crystal structure (PdbID: 2xct) and the models were meant to be a basic comparison as opposed to an in-depth description of the bound ligands.

Using other modelling techniques such as protein docking could provide a more precise model than using simple molecular mechanics, as well as a method to screen the DNA gyrase for inhibitory action of other ciprofloxacin derivatives. Alternatively, a QM/MM model could be used to continue the modelling performed, while increasing the precision by inclusion of quantum mechanical terms for the ciprofloxacin active site[82].

Unaltered ciprofloxacin was found to be the most effective inhibitor in the experimental assays as well as to possess the lowest energy structure in a computational model of DNA gyrase-ligand binding. The compound containing the longest spacer group, c-ava-ciprofloxacin, had the highest energy conformation, suggesting a proportional relationship between the size of the spacer and the energy of binding. The energy difference transition from the c-gly-ciprofloxacin to the c-ava-ciprofloxacin was much greater than from the c-ciprofloxacin to the c-gly-ciprofloxacin in the computational model. This was most likely due to additional stability in the c-gly-ciprofloxacin generated by increased hydrogen bonding interactions with DNA gyrase. This supports the experimental result that c-gly-ciprofloxacin had the lowest IC_{50} and therefore the greatest *in vitro* inhibition of the derivatives.

Chapter 3: Review of deacetylase enzymes

3.1 Introduction

IcaB was shown to have metal-dependent deacetylase activity from *Staphylococcus epidermidis* and was under study in the Potts lab at the University of York[138]. The description of deacetylase enzymes in the literature is divided into individual enzymes as opposed to a general summary. As a result, this chapter begins with an overview of bacterial deacetylase enzymes to summarize the structural and mechanistic features of such enzymes, with a particular focus on carbohydrate esterase family 4, of which IcaB is a member[139]. These features will then be used to evaluate the homology model produced in Chapter 4.

Deacetylase enzymes catalyze the removal by hydrolysis of an acetyl group from a substrate[140,141]. The removal of such acetyl groups is widely used for affecting many different biological mechanisms and processes including gene regulation via packing of DNA into heterochromatin, the breakdown of substrates for metabolic pathways, and production of cell surface carbohydrates involved in the formation of biofilms[140,142,143].

Many bacteria including *Streptococcus pneumoniae* and *Staphylococcus epidermidis* form carbohydrate-dependent biofilms using partially deacetylated exopolysaccharides (biofilms have been previously described in chapter 2). Synthesis of these functional polysaccharides includes partial deacetylation by deacetylase enzymes [144,145]. In addition, deacetylation of peptidoglycan in the bacterial cell wall has evolved as a mechanism for evading the host immune response, by conferring resistance to lysozyme. The removal of the N-acetyl group prevents binding with specific amino acids in the substrate binding site[146,147]. Deacetylation of lysine residues of histone proteins is essential for silencing of nuclear signalling because acetylated lysine residues are less positive, which reduces interactions with the phosphodiester backbone, thereby forming heterochromatin[148]. As described in Chapter 1, heterochromatin is a more condensed form of chromatin, which prevents binding of proteins to initiate DNA transcription. While the original pattern of acetylation is required for DNA attachment to the histone

proteins, its removal by deacetylase enzymes is required for proper formation of heterochromatin[149].

There are a number of deacetylase families based on substrate and sequence similarity[139,140]. The first deacetylase enzymes discovered were histone deacetylases, the structures of which were first reported in 1996[150]. The two major categories can be classified as carbohydrate esterases, which act on acetylated carbohydrates, and histone deacetylases including sirtuins. Sirtuins were initially believed to act on acetyl-lysine residues of histone proteins, similar to histone deacetylases. However, recent evidence has suggested an alternative function of hydrolysis of fatty acid chains from lysine residues Lys19 and Lys20 of TNFα to regulate TNFα secretion[151].

Table 3.1 summarizes the number of sequences and structures available for the families of deacetylase enzymes. The carbohydrate esterase protein families were found in CAZy, the Carbohydrate Active Enzymes database (http://www.cazy.org, Updated 2014-11-13)[152,153]. The sequences were listed in CAZy, and full descriptions were linked to GenBank, (Version 204.0)[154].

Table 3.1: Sequences and structures of known families of deacetylase enzymes as of June 8, 2015[152]

Family	Class	Sequences	Structures
Lysine Deacetylases	Histone Deacetylases	2443	91
	Sirtuins	81	31
	Citrulline deacetylase	2	2
Carbohydrate esterases	CE-1	3886	9
	CE-2	186	4
	CE-3	146	1
	CE-4	8582	18
	CE-5	868	0
	CE-6	165	1
	CE-7	495	4
	CE-8	932	4
	CE-9	3242	4
	CE-10	N/A	N/A
	CE-11	2585	7
	CE-12	552	1
	CE-13	1	0
	CE-14	2571	2
	CE-15	105	0
	CE-16	89	0

3.2 Histone Deacetylases and Sirtuins

The first major category of deacetylase enzymes is the family containing both the histone deacetylase and sirtuin enzymes, which show deacetylase activity against acetylated lysine residues.

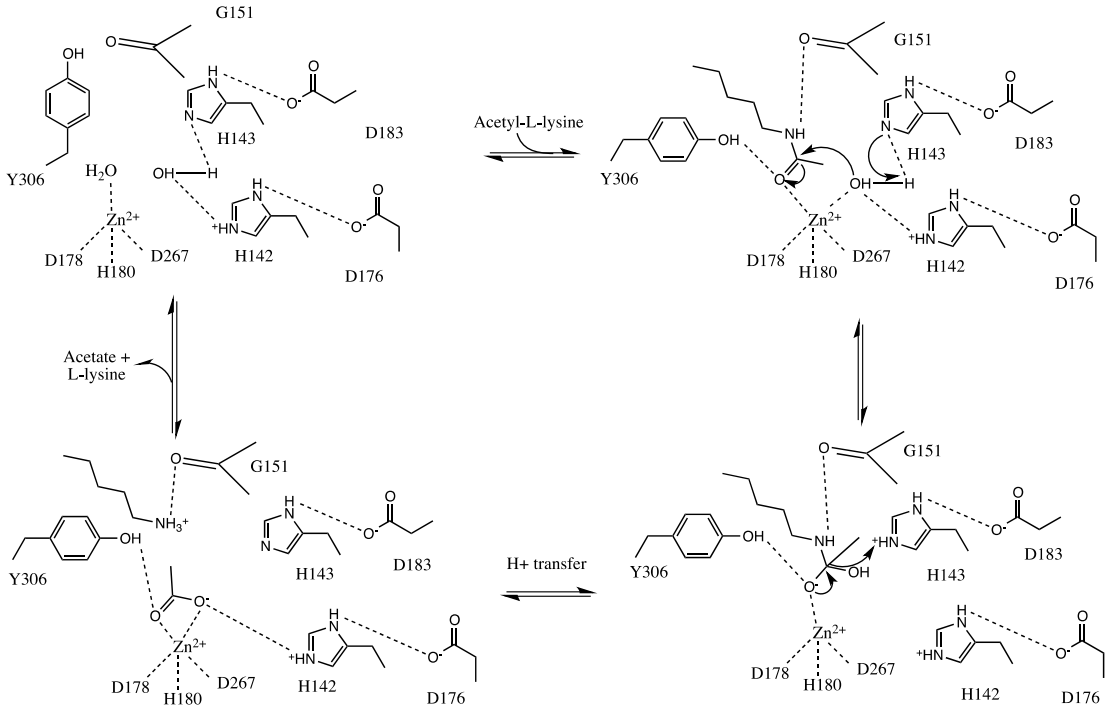
3.2.1 Histone Deacetylases

Histone deacetylation, specifically deacetylation of lysine residues in histones H3 and H4, is associated with reduced levels of transcription[18,155]. As a result, histone deacetylase enzymes (HDACs) have been investigated as targets for cancer treatment[156,157]. HDAC class I is found in a majority of cell types, whereas class II has shown higher levels of expression in certain cell types such as brain, ovary, pancreas and heart cells than others. The difference in the levels of class I and II expression in different tissues suggests a role of HDACs in cellular differentiation[140]. The catalytic domain of histone deacetylase 4 is an eight-stranded parallel β-sheet fold, as seen in Figure 3.1.

Figure 3.1: Histone deacetylase 4 bound to a triflouromethylketone inhibitor (the inhibitor is shown in stick form; pdbID: 2vqj)[158]

Conserved residues in loops L4 and L7 coordinate to a single Zn^{2+} ion at the active site. The surface residues leading to the active site vary between different HDAC isoforms, but the Zn^{2+} ion and attached residues are generally conserved[158,159]. The general histone deacetylase mechanism involves a divalent metal ion bound to a triad of negative

residues (Asp178, His180, and Asp267) as well as a catalytic acid residue and catalytic base residue. These residues, in addition to a hydrophobic tunnel for binding of the substrate, are common in most deacetylases. A proposed general catalytic mechanism for histone deacetylases is shown in Scheme 3.1.



Scheme 3.1: A generic histone deacetylase Zn^{2+}-dependent mechanism as reproduced from the description of Lombardi et al. with amino acid residue numbers taken from human HDAC8[159]

The catalytic base (His143) attacks a water molecule to form a nucleophile, which results in the release of an acetate molecule and a lysine residue with a free amine. In order to stabilize the histidine positions in the active site, these residues are hydrogen-bonded to aspartate residues (Asp176 and Asp183) elsewhere in the protein[159].

3.2.2 Sirtuins

In bacteria, sirtuins target a variety of DNA regulatory proteins and metabolic enzymes. Similar to histone deacetylases, sirtuins deacetylate lysine residues by a Zn^{2+}-dependent

mechanism but also use NAD^+ (nicotinamide adenine dinucleotide) as a cofactor. The catalytic core of sirtuin proteins is composed of a Zn^{2+}-binding domain, a larger Rossmann-fold domain and four loops that form the catalytic cleft. Within this cleft is a hydrophobic tunnel in which the lysine residue and NAD bind. Several intermediates and transition states along the reaction mechanism have been proposed; however, the exact catalytic mechanism for formation of the 2'-O-acetyl-ADP-ribose reaction product is structurally and biochemically unknown[160]. However, Hirsch et al. suggested a water molecule may act as the nucleophile activated by the residue Asp184[161]. Recent research has suggested that some sirtuins may not be deacetylases but instead remove fatty acid groups from Lys19 and Lys20 of Tumour Necrosis Factor-α in order to promote its secretion[151].

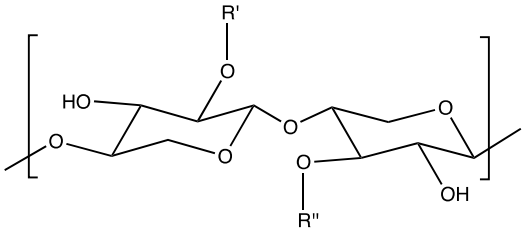
3.3 Carbohydrate Esterases

The second major group of deacetylase enzymes are carbohydrate esterases; a group of 16 families of deacetylase enzymes based upon amino acid sequence similarities[153]. These enzymes are responsible for deacetylation of cell surface carbohydrates such as chitin, peptidoglycan, and poly-N-acetylglucosamine[162]. The general active site for carbohydrate esterases is a catalytic triad of polar or negatively charged residues, which binds divalent ions with a preference for Zn^{2+}[152]. The esterase mechanism is started by nucleophilic attack from a catalytic base residue and a catalytic acid residue provides a free H^+ for the reaction, similar to the histone deacetylase mechanism. The other active site commonly used by carbohydrate esterases involves a Ser-His-Asp triad and the mechanism is similar to that used by serine proteases using a histidine as a catalytic acid/base residue while a serine performs a nucleophilic attack on the acetyl carbon[163].

Many of these proteins contain a hydrophobic tunnel region leading the substrates to the active site, as will be shown in sections 3.3.1-3.3.12. Families CE-1, CE-2, CE-3, CE-5, and CE-6 are all acetylxylan esterases with a Ser-His-Asp triad in the active site[164].

3.3.1 CE-2

Cellvibrio japonicus acetylxylan esterase (*Cj*CE2B) is a CE2 family enzyme involved in the breakdown of xylan from plant matter[165]. Xylan is a polymer of the monosaccharide xylose as shown below:



(1): generic xylan structure

R groups may include functional groups such as acetyl groups or other saccharide units in branched xylan chains. The structure of the *Cj*CE2B enzyme is an N-terminal β-barrel domain of around 130 residues linked to a C-terminal domain of approximately 220 residues. The C-terminal domain consists of an α/β-hydrolase fold. This fold, in turn, consists of repeating β-α-β motifs to form a five-stranded parallel β-sheet, (β2, β1, β3, β4, and β5). In addition, this sheet is located next to two α-helix regions on one side of the sheet and three α-helices. All helices run antiparallel to the β-strands. The C-terminal catalytic domain performs the deacetylation, whereas the N-terminal domain is believed to play a role in substrate recognition by forming part of the binding cleft[166]. The structure of *Cj*CE2B is shown in Figure 3.2[166].

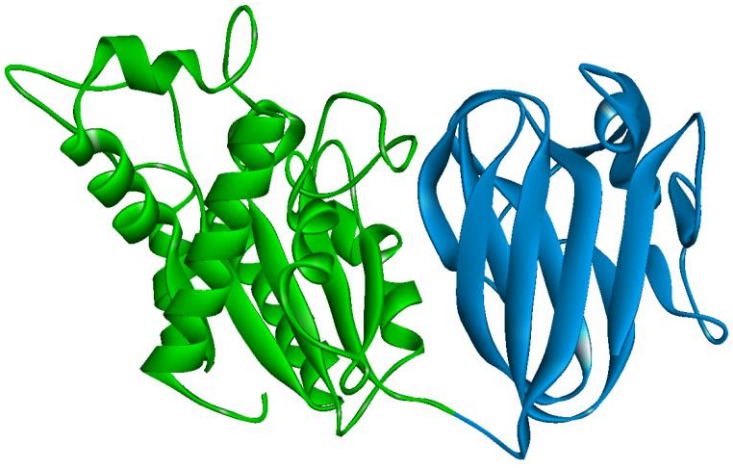


Figure 3.2: Structure of *Cj*CE2B at 2.0Å resolution (pdbID: 2W9X). Blue represents the N-terminal domain and green the C-terminal domain[166].

As with CE1 and CE6 enzymes, *Cj*CE2B contains a Ser-His-Asp triad. Ser151 acts as a nucleophile, His336 has a role as an acid/base residue, and the main chain carbonyl of Cys333 makes a hydrogen bond with His335 to stabilize and activate it. This triad resembles the Ser-His-Asp triad in serine proteases and therefore suggests a serine protease-like mechanism for this enzyme.

3.3.2 CE-3

The acetylxylan esterase *Ct*Ces3 from *Clostridium thermocellum* catalyzes the degradation of plant cell wall xylan as a nutrient source. The breakdown products have been investigated for use as biofuels[167]. The crystal structure shows a dimer interface near the active site that takes up 9% of the overall surface area. This suggests the enzyme may be a homodimer in solution, which shields the active site from the solvent. The monomer displays an α/β hydrolase fold, consisting of a central deformed five-stranded parallel β-sheet surrounded by six α-helices (Figure 3.3).

Figure 3.3: Structure of the *Ct*Ces3 N-terminal domain at 2.0Å resolution (pdbID: 2VPT)[168].

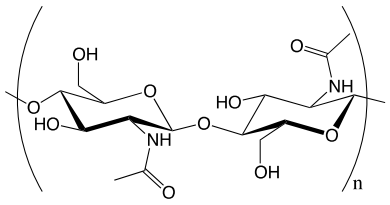
The active site of this enzyme contains the Ser-His-Asp catalytic triad common to enzymes with a serine-protease-like mechanism (Ser44, His208, and Asp205).

3.3.3 Carbohydrate Esterase Family 4

CE4 was selected as the main focus for this project as IcaB is a member of this family. As a result, several specific enzymes were examined to fully define the necessary properties of the CE4 family. Carbohydrate esterase family 4 includes chitin deacetylases, chitooligosaccharides, peptidoglycan deacetylases specific to both N-acetyl glucosamine (NAG) and N-acetylmuramic acid (NAM), and enzymes responsible for deacetylation of polysaccharides necessary for biofilm formation (such as IcaB in *staphylococci* and PgaB in *Escherichia coli*). Carbohydrate esterase family 4 (CE-4) enzymes possess the most structures and sequences available in CAZy compared to other carbohydrate esterases and are classified by the presence of a nodB homology domain (Pfam ID: PF01522). The nodB homology domain is any domain showing high sequence homology to nodB, a chitooligosaccharide deacetylase from rhizobia, which forms a 7-stranded β-barrel structure[169,170]. This nodB homology domain is generally found as part, or all, of the catalytic domain in CE-4 enzymes.

3.3.3.1 Chitin Deacetylase

Chitin deacetylases are involved in the chitin breakdown pathway. Chitin is a N-acetylglucosamine polymer found in the exoskeleton of many arthropods, including crabs and shrimp, as well as the fungal cell wall[171]. The structure of chitin, a polymer of β-1,4-linked N-acetylglucosamine residues, **1**, is shown below.



(2): Chitin[171]

Bacterial chitin deacetylases are found in *Vibrionaceae*, which are marine bacteria responsible for recycling nitrogen present in chitinous debris found in sediment[172]. Chitin deacetylases catalyse the reaction of chitin into chitosan, which has applications in a number of industries including biomedicine, food ingredients, cosmetics and pharmaceuticals[173]. Chitin deacetylases recognize a sequence of four GlcNAc units in

the substrate, one of which undergoes deacetylation, producing chitosan with a regular deacetylation pattern[143]. Chitin deacetylase was found to be active on both xylan and several chitin-like substrates. In addition, the activity was significantly increased in the presence of Co^{2+} addition to these chitin-like substrates. However, the presence of Co^{2+} in the enzymes did not affect the activity when xylan was used as a substrate, which suggests a metal-dependent mechanism for the deacetylation of chitin[139]. There are no reported structures of chitin deacetylases in bacteria in the Protein Data Bank.

3.3.3.2 GlcNAc Peptidoglycan Deacetylase

Streptococcus pneumoniae peptidoglycan GlcNAc deacetylase (SpPgdA, Figure 3.4) deacetylates GlcNAc residues in peptidoglycan in order to protect the Gram-positive bacterial cell wall from host lysozyme. This lysozyme hydrolyzes β-(1,4)-glycosidic linkages between MurNAc and GlcNAc residues in the glycan backbone. Reduced activity of host lysozyme enzyme has been shown to be associated with increased deacetylation of peptidoglycan caused by PgdA in *Enterococcus faecalis*[174]. SpPgdA is able to accept an oligomer of three GlcNAc residues as a substrate, in which the nitrogen-bonded acetyl group of the middle sugar is removed by the enzyme. The structure of SpPgdA shows that it uses a Zn^{2+}-binding triad composed of one aspartate and two histidine residues with a nearby aspartate and histidine acting as the catalytic base and acid, respectively, as shown in Scheme 3.2[175]. This is somewhat similar to the mechanism of histone deacetylases as well as other carbohydrate esterases containing a zinc ion such as LpxC[176]. The efficiency of SpPgdA is approximately 80% deacetylation for the N-acetylglucosamine[146]. The C-terminal domain of this enzyme contains the deacetylase activity while the functions of the mid and N-terminal domain are currently unknown.

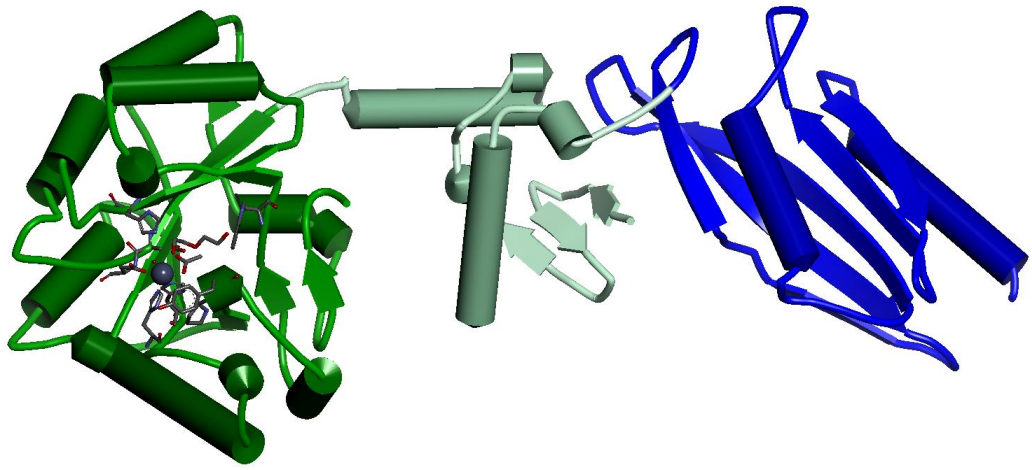
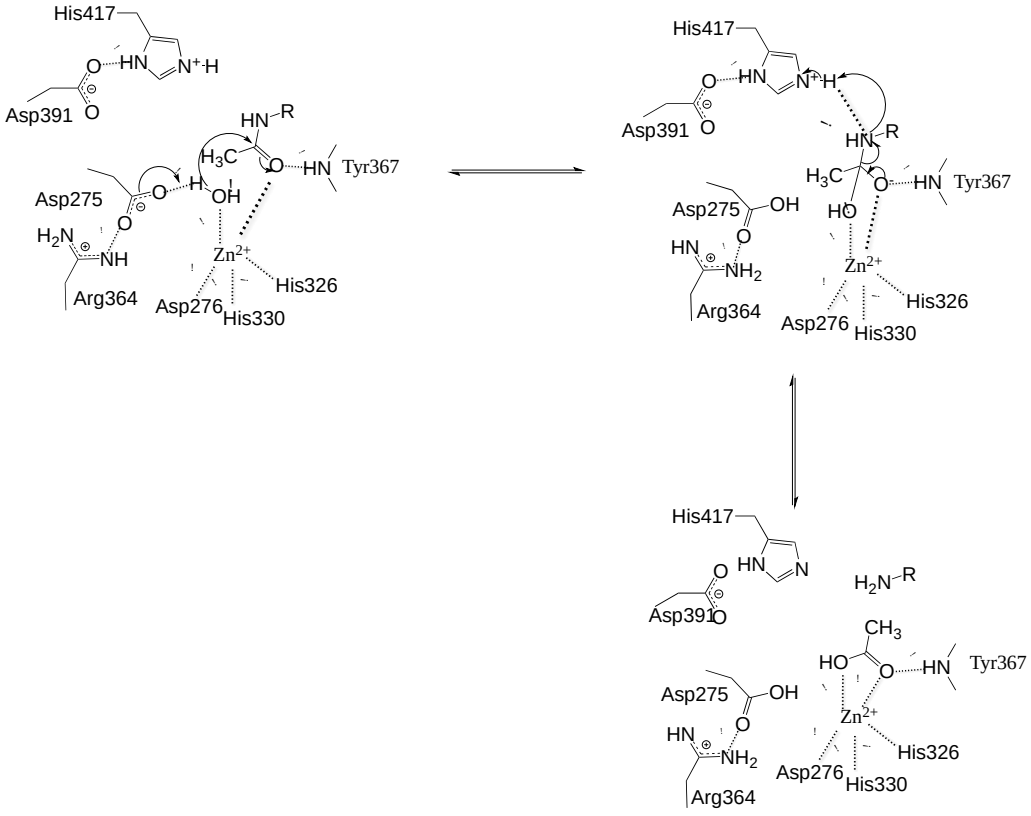


Figure 3.4: Structure of SpPgdA from *S. pneumoniae* at 1.75Å from (pdbID: 2C1G). Blue represents the N-terminal domain, green the C-terminal domain, and pale green the linker domain[175].



Scheme 3.2: Proposed Zn^{2+}-dependent deacetylase mechanism of SpPgdA in *Streptococcus pneumoniae* where Asp275 acts as a catalytic base and His417 acts as the catalytic acid[175].

3.3.3.3 MurNAc Peptidoglycan deacetylase

A peptidoglycan deacetylase from *Bacillus subtilis,* PdaC (originally YjeA), deacetylates N-acetylmuramic acid (MurNAc) as well as GlcNAc. PdaC shares only 25.8% sequence identity with SpPgdA (Genbank IDs: AIX0697.1 and COT19452.1, respectively). Although the biological function of this N-acetylmuramic acid deacetylation is unknown, it has been shown that a mutant deacetylase conferred higher sensitivity to lysozyme, as in PgdA[174,177]. This may suggest a role for PdaC in immune evasion. Unlike the *Streptococcus pneumoniae* PgdA, a highly Zn^{2+}-dependent GlcNAc peptidoglycan deacetylase, there was less dependence on the Zn^{2+} ion for deacetylation of peptidoglycan by PdaC than other divalent metal ions such as Mn^{2+}, Mg^{2+}, and Ca^{2+}. PdaC was also found to possess deacetylase activity against N-acetylglucosamine (GlcNAc) oligomers. Therefore, PdaC possesses GlcNAc deacetylase activity toward GlcNAc oligomers and MurNAc deacetylase activity toward *B. subtilis* peptidoglycan[177]. However, there is no structure available for PdaC in the Protein Data Bank as of July 11, 2015.

3.3.3.4 Chitooligosaccharide Deacetylase

ChbG encodes a monodeacetylase that is essential for growth on the acetylated chitooligosaccharides chitobiose and chitotriose but is not required for growth on cellobiose and chitosan, the deacetylated form of chitobiose[178]. The *chb* operon, originally known as the *cel* operon, of *E. coli* is involved in the utilization of the β-glucoside chitobiose in the *E. coli* chitin breakdown pathway[179]. The gene products of *chbB*, *chbC*, and *chbA* form a permease that transports chitobiose across the inner membrane and phosphorylates it to chitobiose-6-P, while *chbG* codes for a deacetylase and *chbR* codes a repressor/activator protein[180–182]. It is believed that deacetylation of chitobiose-6-P and chitotriose-6-P is necessary for their recognition by ChbR as inducers of function, by changing the conformation of ChbR to an active form[183].

ChbG is active as a homodimer. Putative metal-binding residues that are conserved across the YdjC family of proteins, of which ChbG is a member, are a triad formed by an aspartate and two histidines. A recent report in 2014 by Andrés et al. has shown that the catalytic site of chitooligosaccharide deacetylase from *Vibrio cholerae* resembles those of other

CE4 enzymes, containing both this divalent metal-binding triad and catalytic acid and catalytic base residues. The structure of the active homodimer is shown in Figure 3.5 and the proposed mechanism in Scheme 3.3[178].

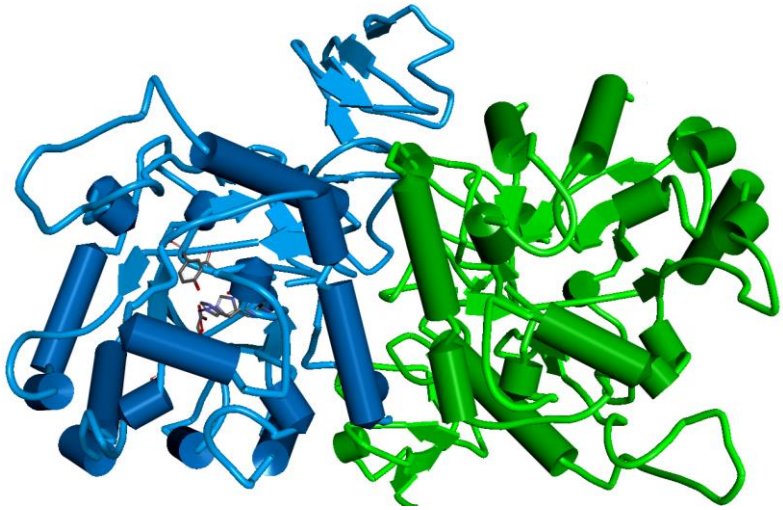
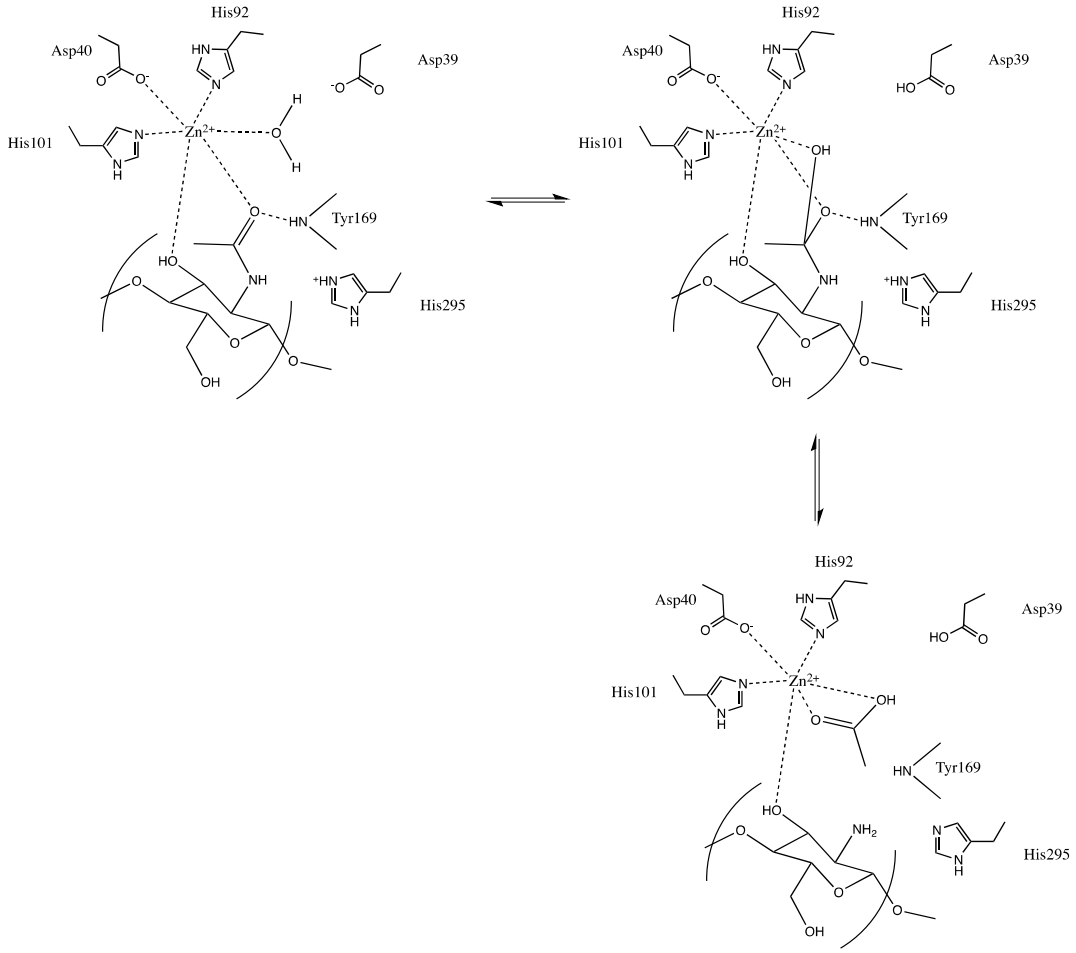


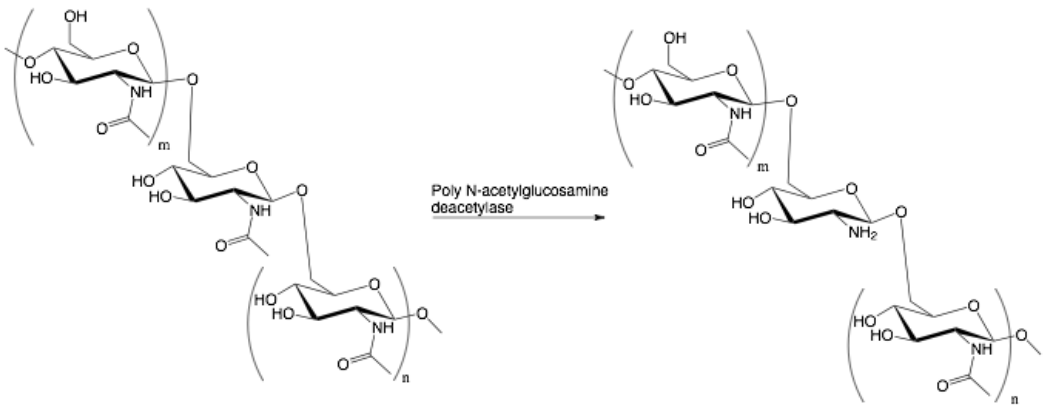
Figure 3.5: Structure of the chitooligosaccharide deacetylase homodimer from *Vibrio cholerae* at 1.88Å (pdbID: 4NY2). Individual monomers are shown in blue and green and the active site residues in stick form[178].



Scheme 3.3: Proposed mechanism of action for chitooligosaccharide deacetylase from *Vibrio cholerae* in which Asp39 acts as the catalytic base and His295 as the catalytic acid[178].

3.3.3.5 Poly-N-acetylglucosamine Deacetylases (PgaB and IcaB)

Although there were a large variety of additional polysaccharide deacetylases examined for features common to deacetylases in general at the beginning of this project, the carbohydrate esterase family 4 (CE4) became the main focus, as IcaB is classified as a CE4 enzyme[184]. IcaB partially deacetylates the growing poly-N-acetylglucosamine chain (PNAG) to yield the deacetylated form, dPNAG (also known as PIA, Polysaccharide Intercellular Adhesin), as shown in Scheme 3.4[144].



Scheme 3.4: Deacetylation of poly-N-acetyl glucosamine (left) yielding randomly deacetylated PNAG (right)[37]

The structure of PgaB is shown in Figure 3.6. The N-terminal domain of PgaB in *Escherichia coli* (residues 1-310) performs the same function as IcaB for *Staphylococcus epidermidis*[186].

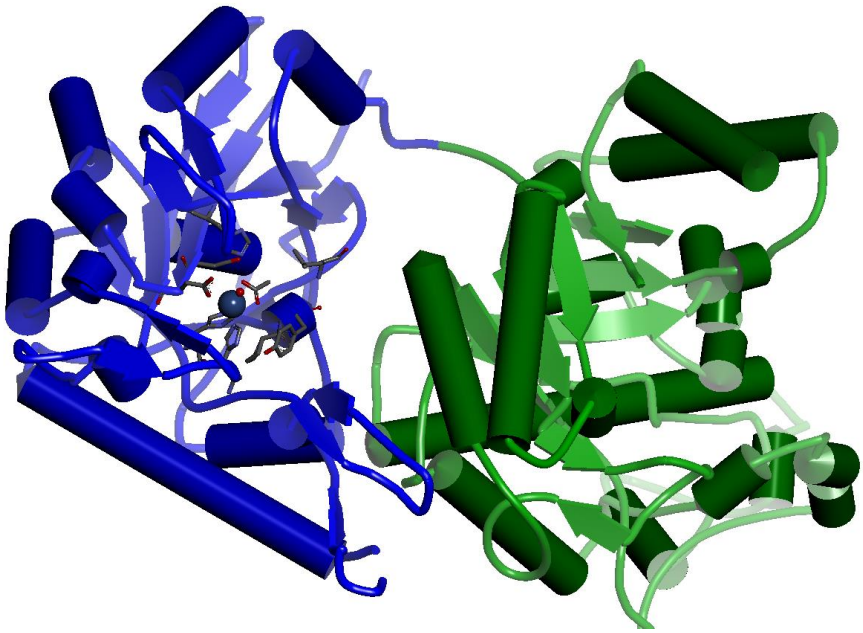
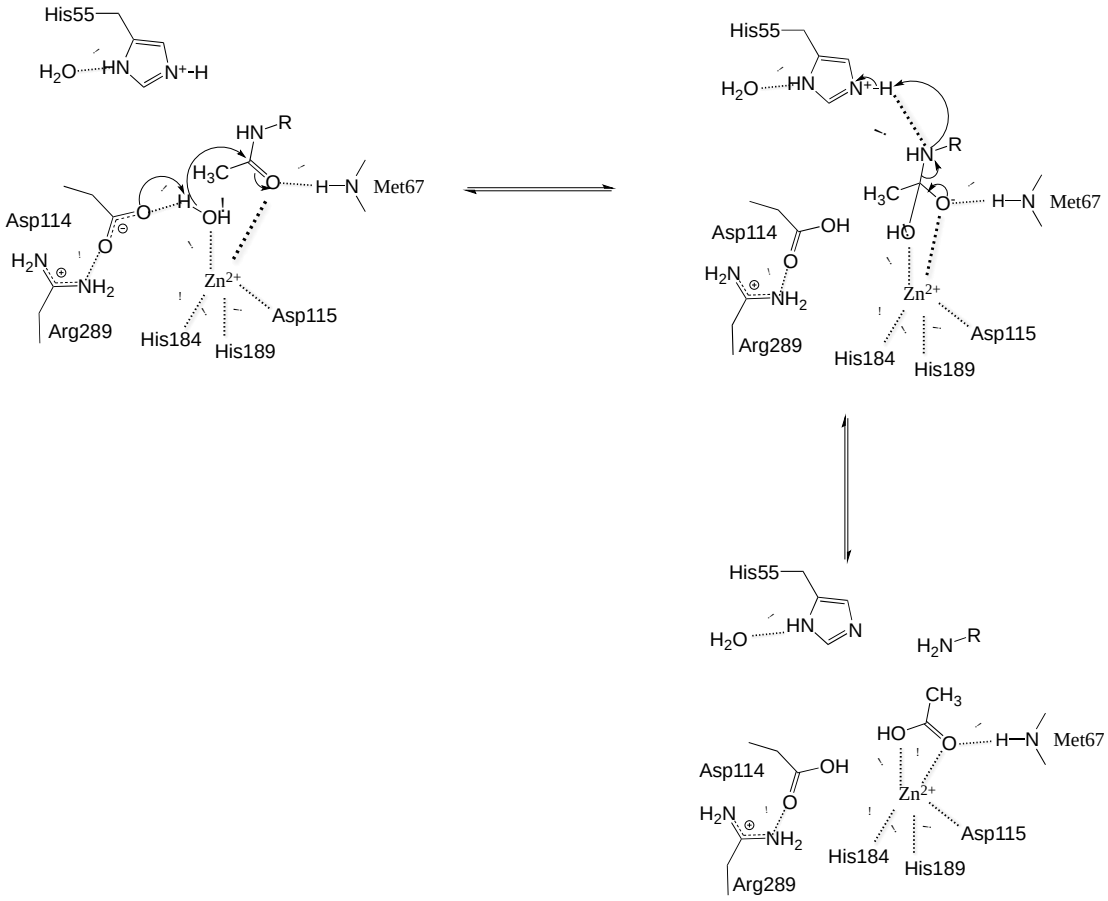


Figure 3.6: Structure of PgaB with Ni^{2+} as the active site metal from *E. coli* from pdbID: 4f9d (blue=N-terminal domain, green=C-terminal domain) containing the active site residues, Ni^{2+} ion and bound acetate in stick form[142].

The active site of PgaB consists of an aspartate, histidine, histidine (His184, His189, Asp115) metal binding triad to bind the active site divalent metal ion as well as an aspartate residue, Asp114, which acts as a catalytic base, and a histidine residue, His55, which acts as a catalytic acid (Scheme 3.5). The catalytic base is held in place in the active site by an arginine residue, Arg289, by hydrogen bonding. The backbone nitrogen of a methionine residue, Met67, helps to hold the substrate carbonyl in place during the reaction by hydrogen bonding to the carbonyl oxygen.



Scheme 3.5: Active site mechanism of deacetylation of PNAG by PgaB in which the catalytic base residue is Asp114 and His55 is the catalytic acid residue[142].

Unlike other carbohydrate esterase family 4 enzymes (11 structures known), in PNAG deacetylases such as PgaB, the catalytic acid histidine residue is not activated by an aspartate residue farther away in the structure, but by a water molecule. This substitution may explain the partial deacetylation of PNAG caused by these enzymes as the polar water molecule does not lower the pKa of the histidine as effectively as a

negatively charged aspartate[143,175]. The optimal substrate length was determined by a length-specific oligomer assay to measure the degree of de-N-acetylation by measuring free amine production using fluorescamine. Fluorescamine reacts with free amines to produce a fluorescent product proportional to the concentration of free amine[187]. The fluorescence is then compared to the length of the deacetylated GlcNAc oligomer. The highest enzymatic activity for PgaB was achieved when the active site contained five monosaccharide units at one time with a k_{cat}/K_m of β-1,6-$(GlcNAc)_5$ = 0.25 $M^{-1}s^{-1}$. This is indicative of much lower activity compared to other CE4 enzymes such as SpPgdA ($0.00015M^{-1}s^{-1}$)[142]. The N-terminal deacetylase domain resembles the C-terminal domain of other CE4 enzymes such as SpPgdA. It is currently unknown whether the position of this domain at the C- or N-terminal affects the deacetylase function[175]**.** A series of hydrophobic residues in PgaB (Phe154, Tyr190, Ile192, Leu200, Tyr209, and Tyr216 from β6–7, β9–10, and β11-12) form a binding groove to hold the substrate PNAG[142]. The surface of this groove is shown in Figure 3.7.

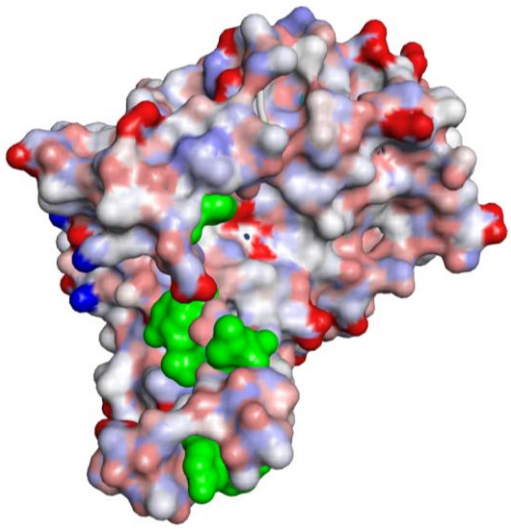


Figure 3.7: Hydrophobic PgaB binding groove surface. Regions of positive charge are shown in blue and negative charge in red, while the green region represents the hydrophobic residues that form the binding groove[142].

3.3.4 CE-7

CE-7 enzymes, such as cephalosporin C deacetylase, display activity towards both acetylated xylooligosaccharides and cephalosporin C, unlike other carbohydrate esterase

enzymes. In addition to the breakdown of acetylated xylan, these enzymes have been investigated for use in the synthesis of novel cephalosporin antibiotics[188]. The structure of cephalosporin C deacetylase is a classical α/β hydrolase fold. Substrates bind predominantly through non-specific contacts with protein hydrophobic residues. Protein residues involved in catalysis are held in place by interactions from the α/β hydrolase fold. This fold also contains interactions between individual monomers to form dimers, three of which in turn form the catalytically competent hexamer containing six active centres aimed towards the central pore. The entrance to the binding channel of cephalosporin C deacetylase may function to block access of sterically hindered substrates to the active site. This would explain the observation that the enzyme is active only on small, acetylated molecules[189]. The structure of a cephalosporin C deacetylase dimer is shown in Figure 3.8.

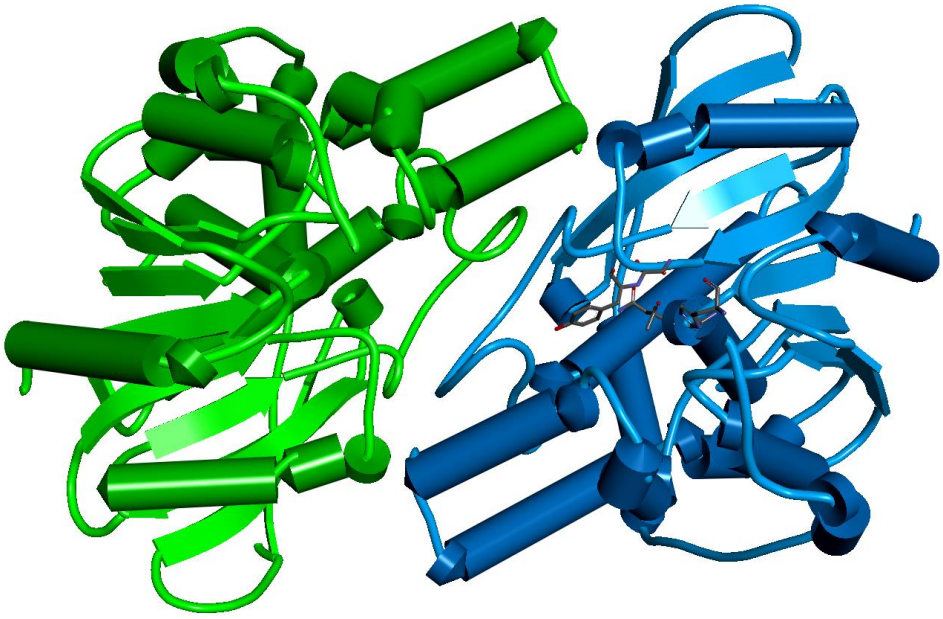
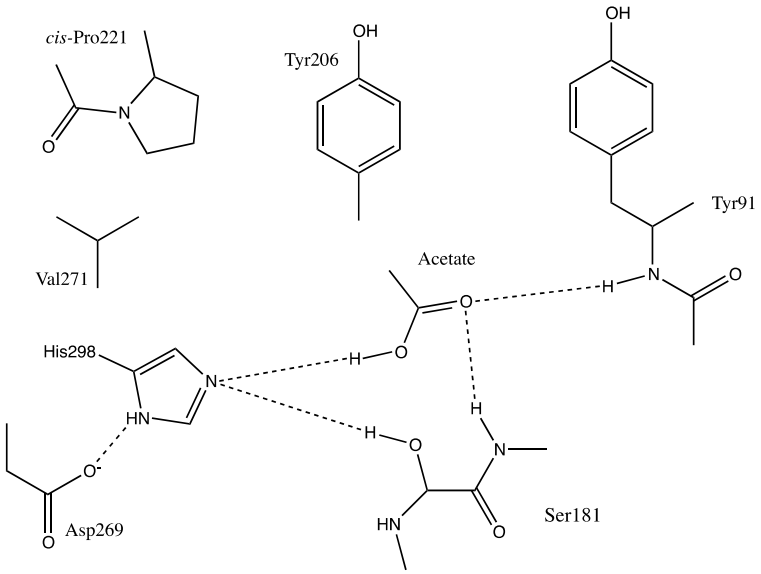


Figure 3.8: Structure of the cephalosporin C deacetylase dimer from *Bacillus subtilis* with acetate in the active site at 1.9Å resolution (pdbID: 1ODT). Blue and green represent individual monomer subunits[189].

The substrate is held in place by residues Ser181, Tyr91, and His298. Tyr206 may be involved in a stacking interaction with the substrate necessary for stability in the active site[189]. The active site of cephalosporin C deacetylase is shown in Scheme 3.6. The

Ser181, His298, Asp269 triad in the active site suggests a serine protease-like mechanism.



Scheme 3.6: Active site residues bound to acetate in an X-ray crystal structure of cephalosporin C deacetylase[189].

3.3.5 CE-8

Pectin methylesterases, enzymes responsible for demethylesterification of pectins from plant cell walls, form family CE-8[190]. In pathogenic bacteria infecting plants, this allows for breakdown of the cell walls allowing the bacteria to invade plant tissues. Pectins are polysaccharides rich in D-galacturonic acid. Pectin methylesterases remove methyl groups in the breakdown of pectin[191,192].

The structure of pectin methylesterase in *Erwinia chrysanthemi* shown in Figure 3.9 was determined to be a β-helix protein characterized by parallel β sheets coiled into a large right-handed barrel connected by surface α helices, similar to other proteins of the parallel β-helix family[193].

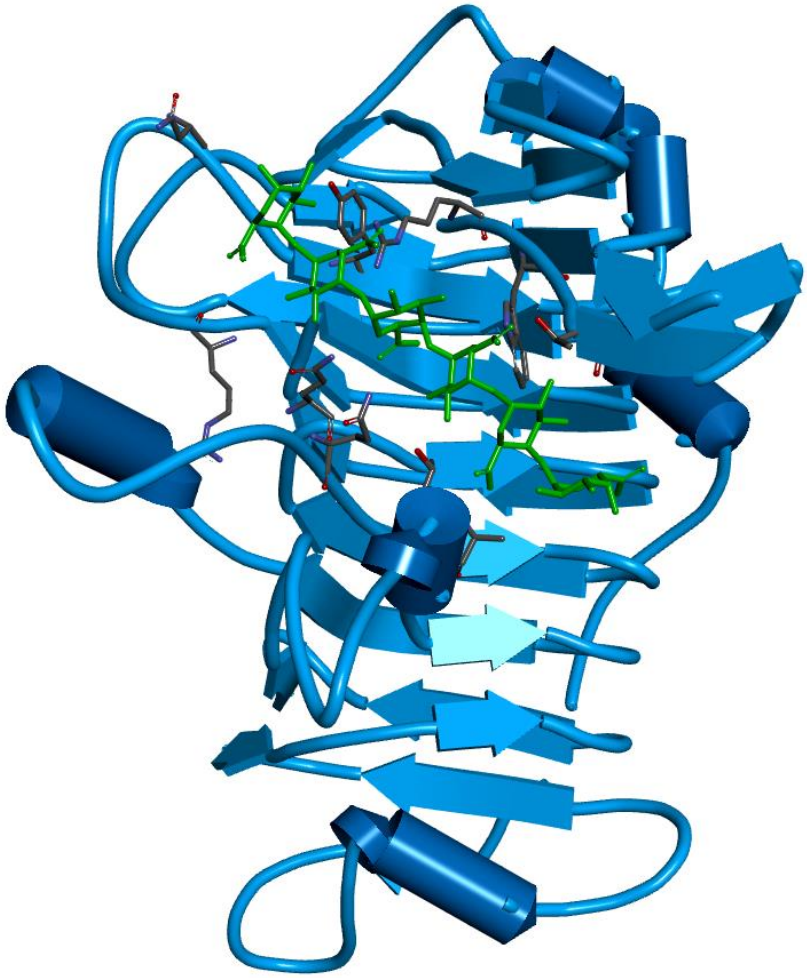


Figure 3.9: Structure of pectin methylesterase from *Erwinia chrysanthemi* containing a hexamer of D-galacturonic acid in stick form (green) and binding site residues (grey)(pdbID: 2NTB)[192]

The organization of the binding site, as well as the substrate binding positions, is shown in Figure 3.10.

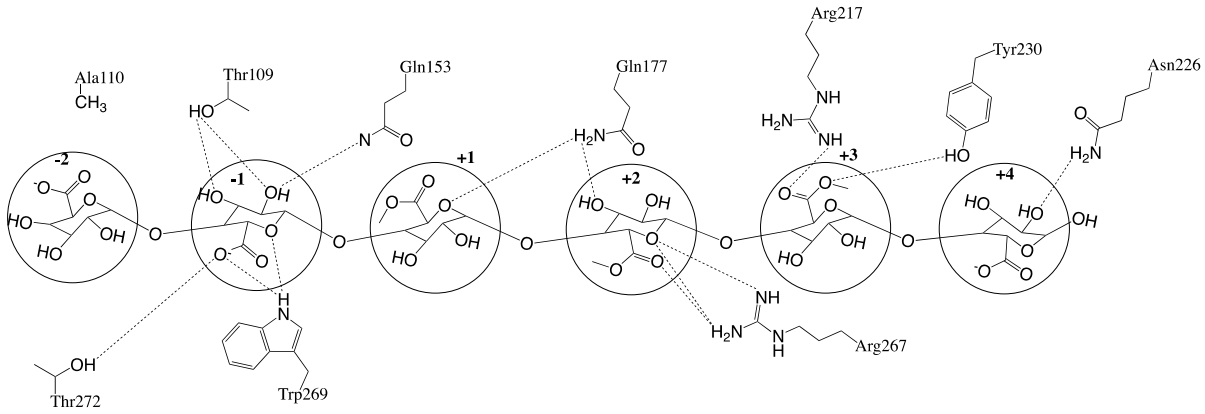
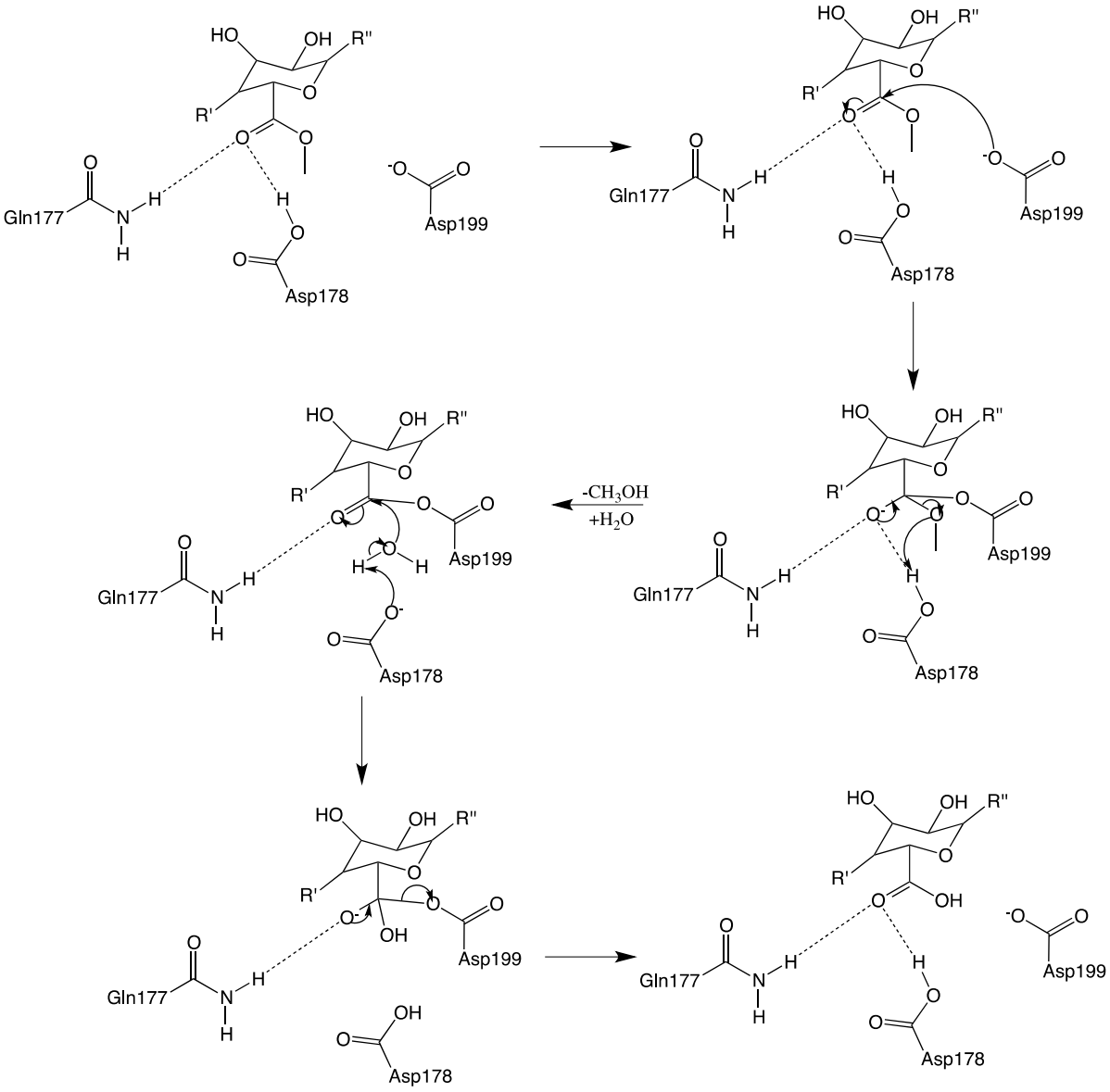


Figure 3.10: Binding cleft residues with labelled D-galacturonic subunit positions set to +1 at the position adjacent to Gln177[192]

It should be noted that within carbohydrate esterases, CE8 is very different mechanistically, relying on the catalytic triad Gln177, Asp178 and Asp199 (kept in place by Arg267) as shown in scheme 3.7.



Scheme 3.7: Mechanism of pectin methylesterase in *Erwinia chrysanthemi*[192]

This mechanism is unique in carbohydrate esterases as it requires neither the Ser-His-Asp triad of serine proteases nor the active divalent metal required in CE4 enzymes.

3.3.6 CE-9

N-Acetylglucosamine-6-phosphate deacetylase catalyses the hydrolysis of the N-acetyl group of GlcNAc-6-P to yield glucosamine 6-phosphate and a free acetate[194]. This is the first step in the synthesis of amino-sugar-nucleotides necessary for biosynthesis of both peptidoglycan and teichoic acids in the Gram-positive bacterial cell wall[195,196]. The structure of this protein consists of a $(\beta/\alpha)_8$ domain containing active residues and a

small β-barrel domain, which contains a carboxylated lysine that bridges the metal ions in the active site. *N*-Acetylglucosamine-6-phosphate deacetylase uses a deacetylation mechanism with a catalytic acid and catalytic base with two Fe^{2+} ions to hold the acetyl group in place, which differs from the general carbohydrate esterase active site by containing two metal ions as opposed to one. The metal ion is held in place by a triad of amino acid residues[194]. The structure is shown in Figure 3.11 and the mechanism in Scheme 3.8.

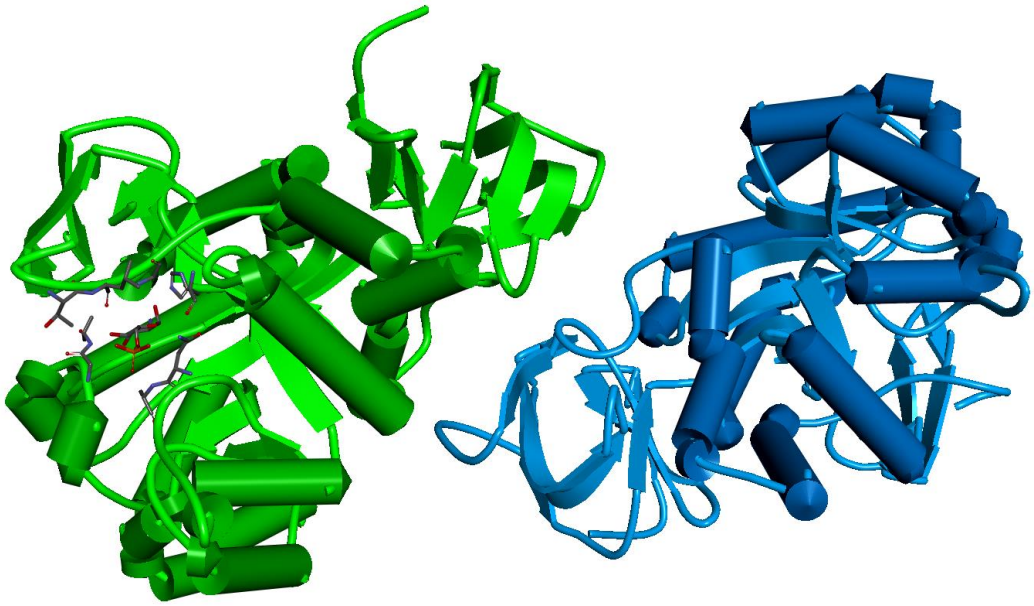
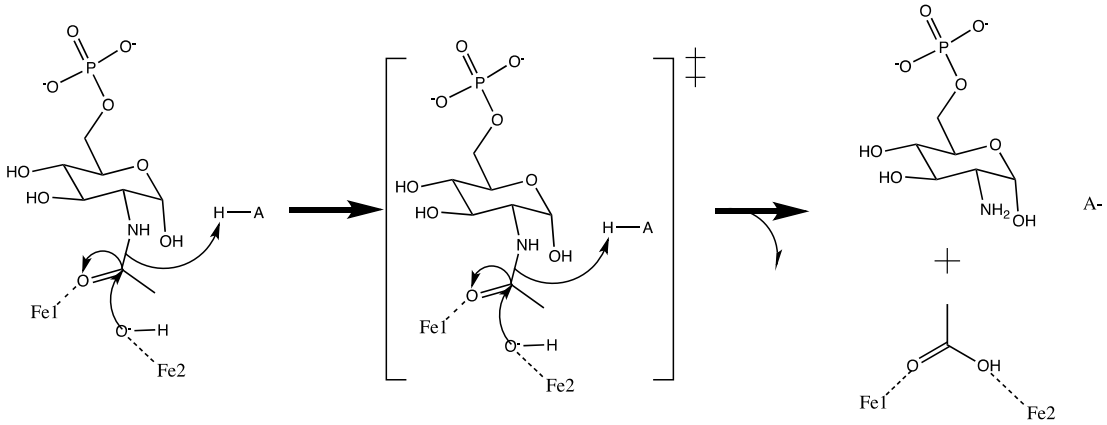


Figure 3.11: Structure of NAG-6-P deacetylase from *Bacillus subtilis* at 2.05Å resolution (pdbID: 1UN7)[194]



Scheme 3.8: Proposed mechanism of deacetylation for NAG-6-P deacetylase[194]

3.3.7 CE-10

Members of this family originally in the CAZy database included arylesterase, sterol esterase, carboxyl esterase and acetylcholinesterase. As the majority of CE-10 esterases act on non-carbohydrate substrates, the information has been withdrawn from the CAZy website[197]. Cholinesterases inhibit nerve transmission in eukaryotes by catalysing the breakdown of the neurotransmitter acetylcholine. These enzymes have similar mechanisms to serine protease enzymes as previously described[198]. Since this review is an overview of bacterial deacetylases, CE10 was not a focus of study within this chapter.

3.3.8 CE-11

LpxC is an N-acetylglucosamine deacetylase that catalyses the hydrolysis of an acetyl group from UDP-3-hydroxymyristoyl-N-acetylglucosamine to UDP-3-O-(R)-3-hydroxymyristoylglucosamine. This hydrolysis is needed for lipid A of lipopolysaccharide (LPS) on the gram-negative cell wall. Lipid A acts as an anchor for LPS in the cell wall and this is most likely why strains that contain defective lipid A have been shown to possess greater susceptibility to antibiotics[199]. The structure of a single LpxC monomer consists of two homologous domains containing two α-helices and a five-stranded β-sheet. LpxC monomers form a homotrimer when complexed with an inhibitor, as shown by the position of the Zn^{2+} ions on the interfaces between monomers. One α-helix and two β-sheet strands from each subunit form a hydrophobic tunnel leading to the active site. The active homotrimer of LpxC from *Pseudomonas aeruginosa* is shown in Figure 3.12.

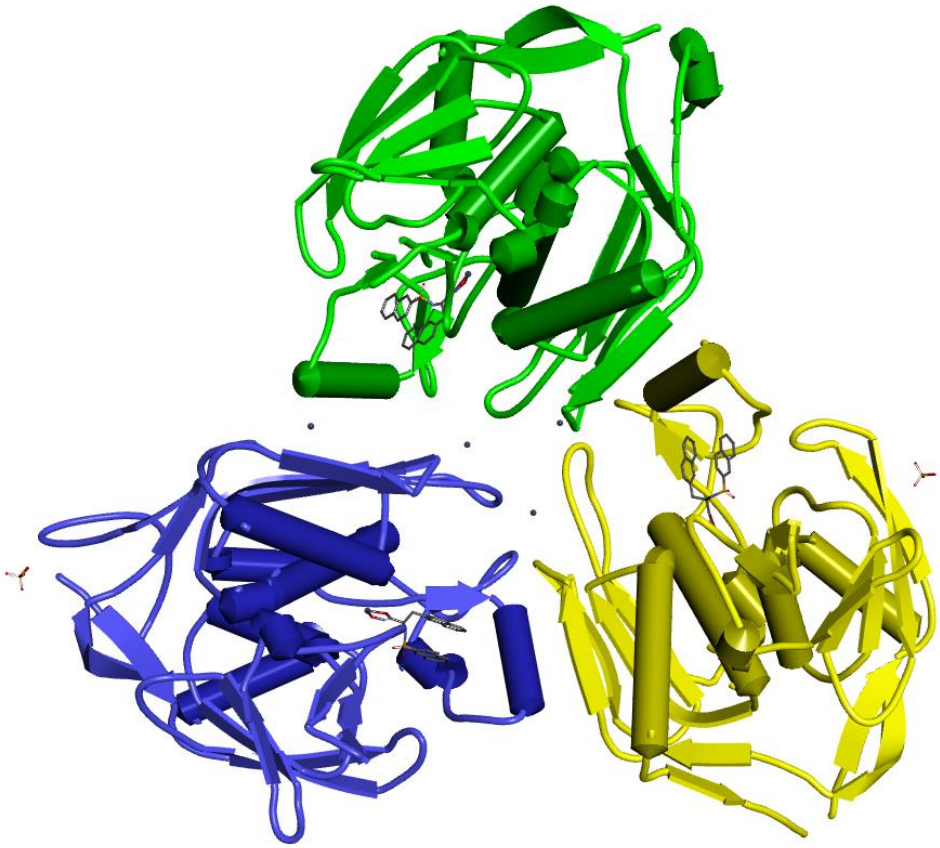
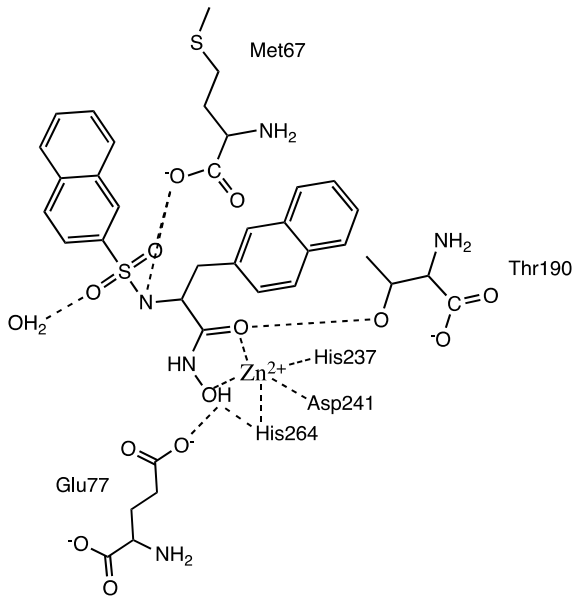


Figure 3.12: Structure of LpxC homotrimer from *Pseudomonas aeruginosa* containing inhibitor BB-78485 (sticks) and zinc ions (balls) in the active site at 1.9Å resolution (pdbID: 2VES)[176].

The active site of LpxC contains a Zn^{2+} coordinated to one aspartate and two histidine residues, as well as a glutamine residue which is within hydrogen bonding distance and may therefore serve as a catalytic acid/base. This active site suggests a metalloprotease mechanism related to CE4 enzymes and is shown in Scheme 3.9 complexed with an inhibitor[176].



Scheme 3.9: A schematic representation of LpxC in complex with an inhibitor BB-78485[176].

3.3.9 CE-12 and CE-13 (Pectin Acetylesterases)

Both families CE-12 and CE-13 consist of pectin acetylesterases. Pectin acetylesterases are involved in the regulation of plant growth and reproduction[200]. These enzymes assist in degradation of pectin in plant cell walls by pathogenic bacteria which infect various plants[201]. A structure of a CE-12 enzyme from *Bacillus subtilis* is in the Protein Data Bank, yet to be published (pdbID: 2O14). This protein (Protein YXIM_BACsu from *Bacillus subtilis*) contains an N-terminal beta barrel and a C-terminal beta sheet sandwiched between two alpha helical regions (Figure 3.13).

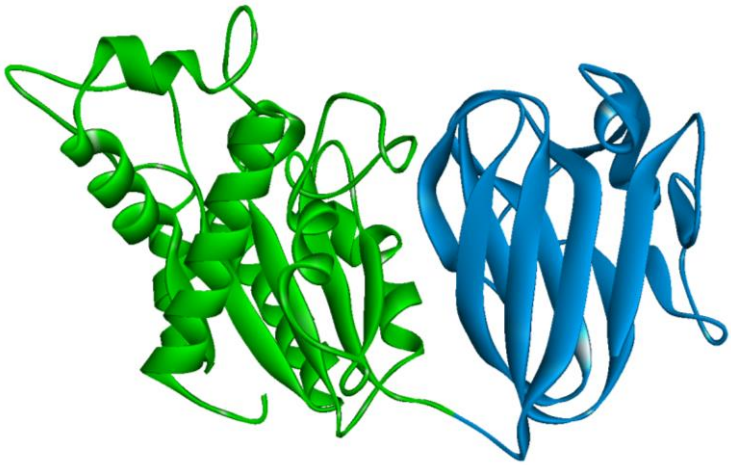


Figure 3.13: Structure of pectin acetylesterase at 2.1Å resolution from *B. subtilis* (pdbID: 2O14). The N-terminal domain is shown in blue and the C-terminal domain in green.

3.3.10 CE-14

The enzyme MshB from *Mycobacterium tuberculosis* (an N-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside deacetylase) is involved in the biosynthesis of mycothiol, which is used for defence against electrophilic toxins and oxidative stress. Specifically, MshB performs the rate-limiting step, which is the removal of the N-acetyl group from N-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside (GlcNAc-Ins) to form the free amine GlcN-Ins. Mycothiol is specific to actinobacteria, which is why MshB has been investigated as a potential drug target for tuberculosis infections[202]. The newest crystal structure of MshB containing glycerol and acetate in the active site for a better active site structure at a resolution of 1.85Å is in the Protein Data Bank (pdbID: 4EWL). The overall structure of the protein is an N-terminal domain containing a lactate dehydrogenase-like Rossman fold and a C-terminal domain consisting of two β-sheets and two α-helices, as shown in Figure 3.14[203]. This structure suggests a mechanism in which a general base (in its deprotonated form) and a general acid (in its protonated form) are involved in the deacetylation mechanism[204].

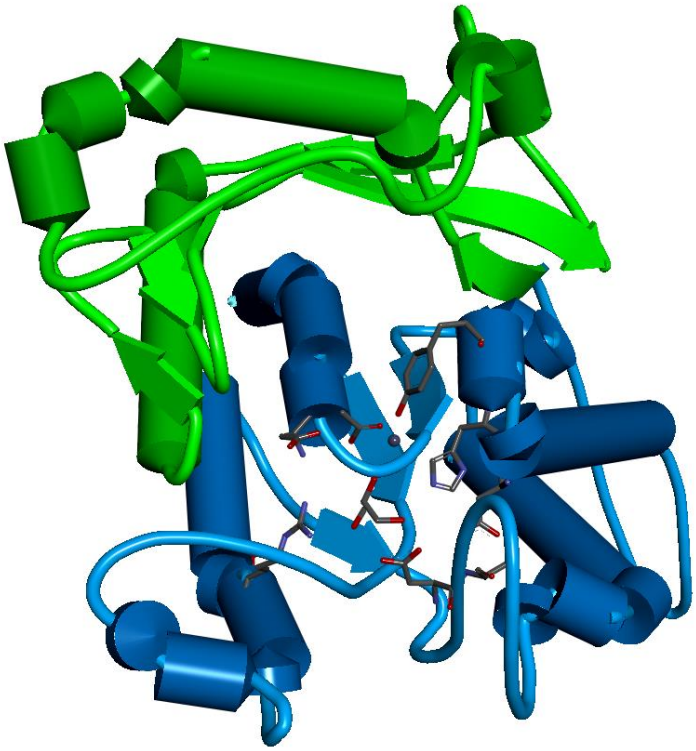


Figure 3.14: Structure of MshB at 1.85Å (pdbID: 4EWL). Blue represents the N-terminal domain and green the C-terminal domain[204].

Acetate, which represents the N-acetyl group of GlcNAc-Ins in the crystal structure, is held in position by interactions with the active site Zn^{2+} as well as the residues Asp15 and Tyr142. The glycerol representing alcohol groups on the lower section of the substrate is positioned by interactions with the side chains of His144, Asp95 and Arg68, while Asp146 stabilizes the position of Asp95 through interactions with Ser96. The active site containing glycerol and acetate is illustrated in Figure 3.15.

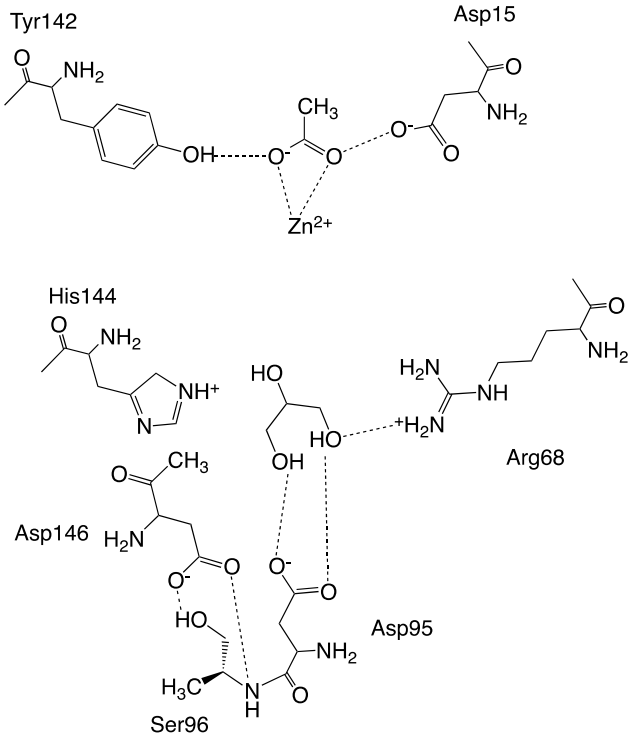
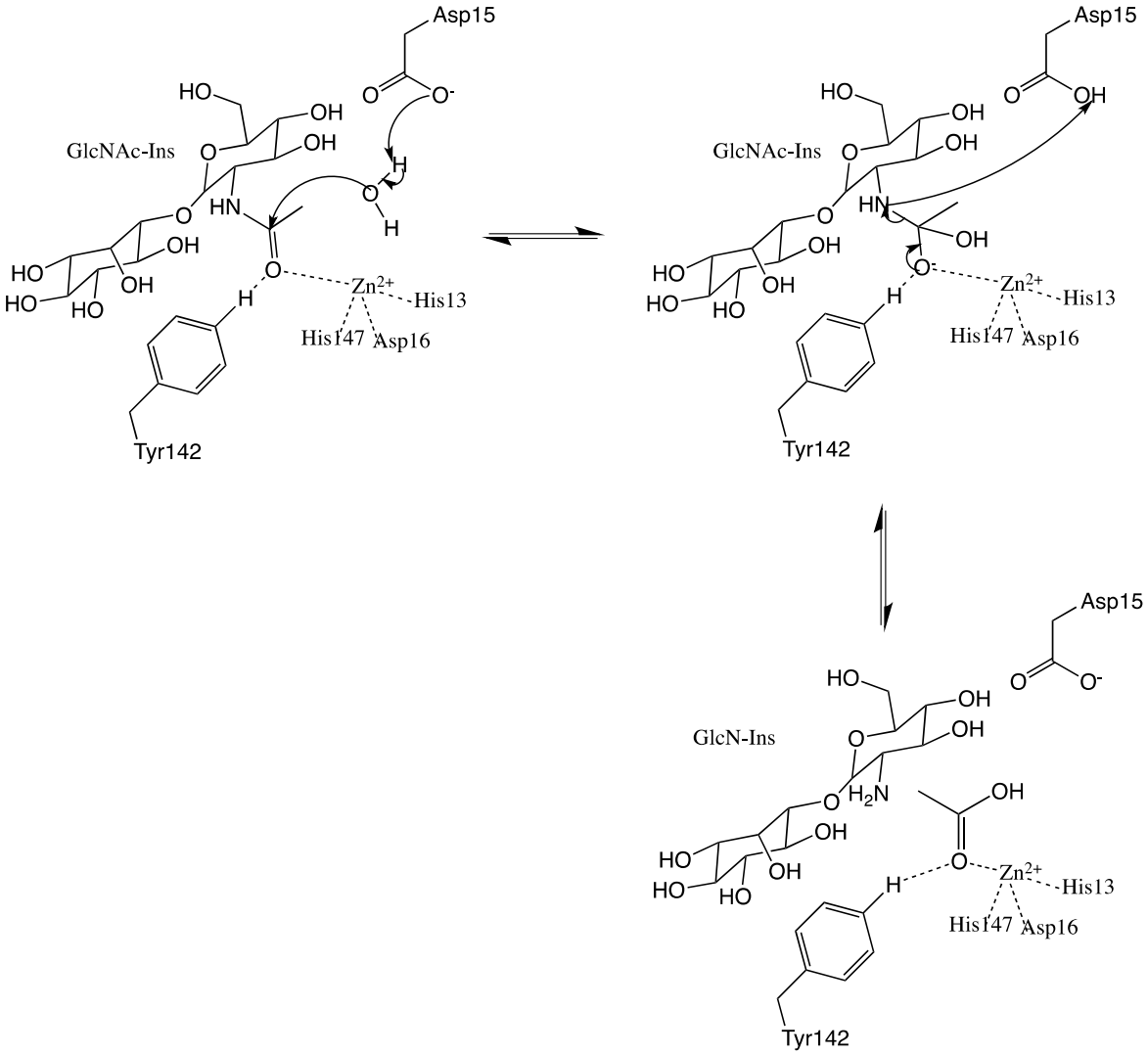


Figure 3.15: Active site of MshB containing glycerol and acetate to simulate mycothiol binding[204]

The deacetylation reaction performed is similar to that of other metal-dependent carbohydrate esterases, with Asp15 acting as a catalytic acid/base residue to activate a water molecule as a nucleophile. The His13, Asp16, His147 triad holds the active site Zn^{2+} in position, while Tyr142 hydrogen bonds to the acetyl oxygen to stabilize the negative charge during the reaction. The mechanism of MshB is shown in Scheme 3.10.



Scheme 3.10: Mechanism of deacetylation of GlcNAc-Ins (native substrate) by MshB[204]

3.3.11 CE-15

Glucuronoyl esterase is an enzyme from the fungus *Hypocrea jecorina*, which has been shown to hydrolyse alkyl and arylalkyl esters of D-glucuronic acid and 4-O-methyl-D-glucuronosyl. This enzyme plays an important role in microbial digestion of plant tissues containing 4-*O*-methyl-d-glucuronoxylan in the xylan metabolism pathway[205]. There are currently no structures of glucuronoyl esterases in bacteria in the Protein Data Bank.

3.3.12 CE-16

Acetyl esterase is an enzyme involved in a step further along the biosynthesis pathway of *Hypocrea jecorina,* mentioned above, as glucoronyl esterases by deacetylation of several positions on 4-O-methyl-d-glucuronoxylan[206]. As with glucoronyl esterase, there are currently no bacterial structures of acetyl esterase available in the Protein Data Bank. The mechanism of these enzymes is currently unknown.

3.4 Citrulline Deacetylase

Citrulline deacetylase is a tetrameric protein responsible for conversion of N-acetyl-citrulline to citrulline in an arginine biosynthetic pathway of *Xanthomonas campestris*. This model is similar to those of serine proteases as well as carbohydrate esterases, involving a metal ion attached to two histidine residues and one aspartate residue. The monomer has a two-domain fold consisting of a catalytic domain (residues 1-166 and 286-365) and a dimerization domain (residues 167-285), as shown in Figure 3.16. The catalytic domain is a six-stranded β sheet between α helices forming the hydrophobic core and a three-stranded β sheet on the solvent-exposed surface, while the dimerization domain is composed of a four-stranded anti-parallel β sheet facing two α helices. The two domains are linked through the H7 helix.

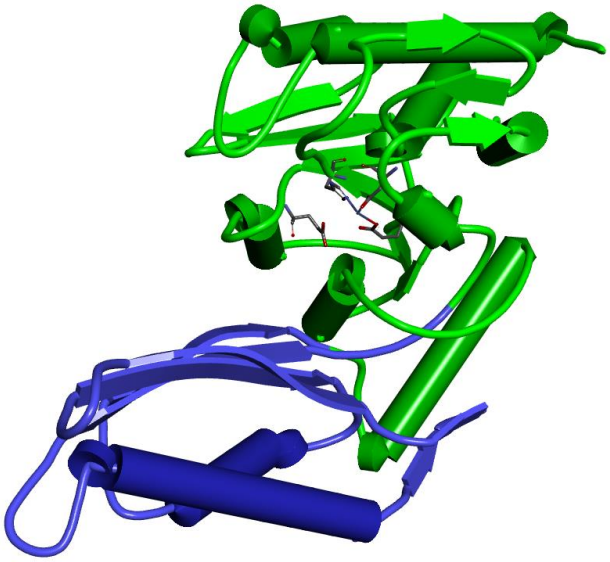
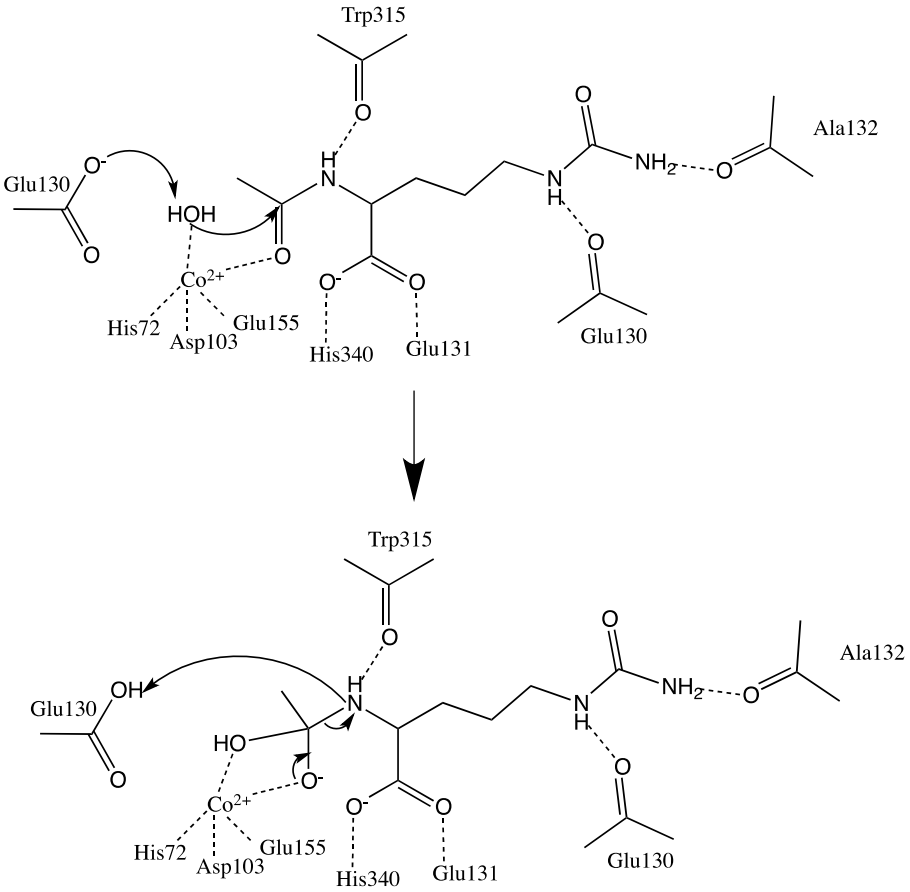


Figure 3.16: Structure of citrulline deacetylase from *Xanthomonas campestris* at 1.75Å (pdbID: 2F7V). The catalytic domain is labelled in green and the dimerization domain in blue[207].

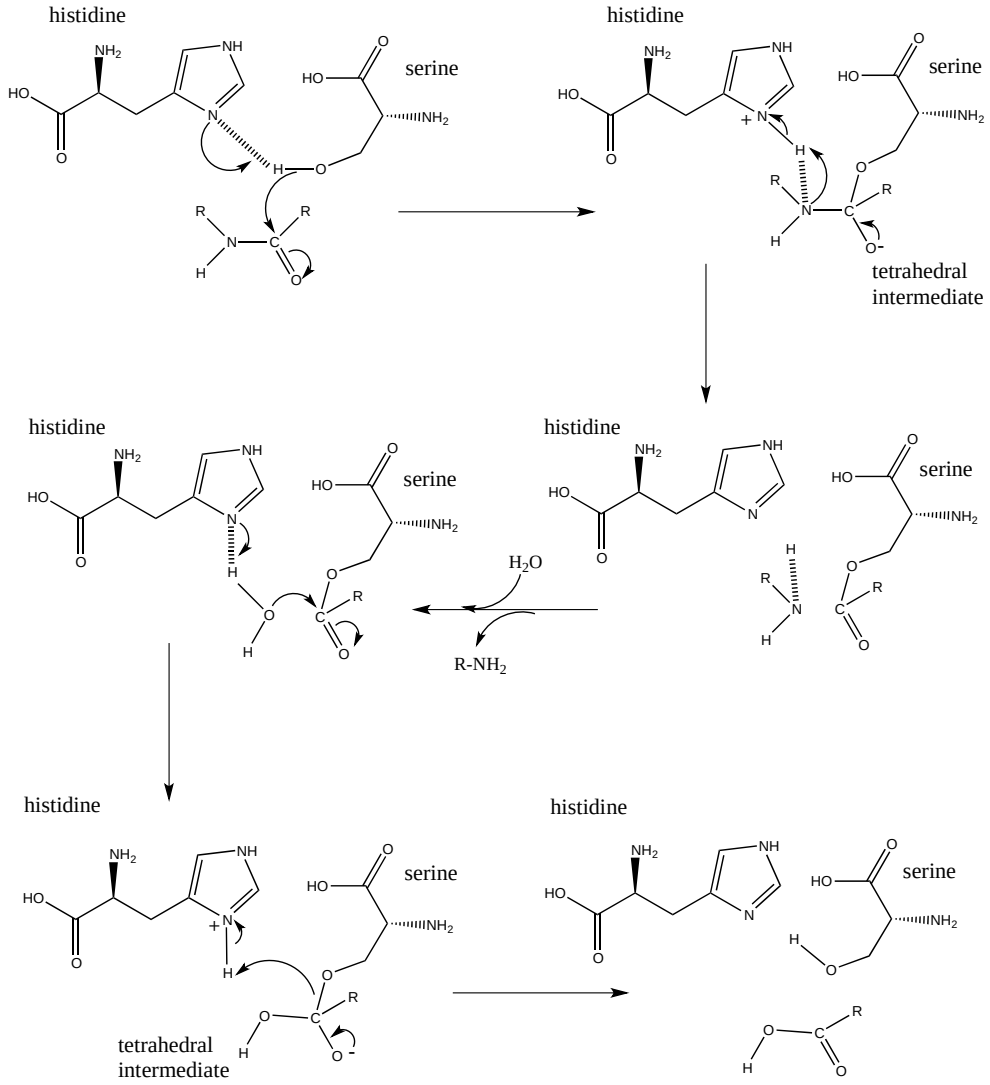
Glu130 functions as a general acid–base catalyst, activating a water molecule coordinated to the Co^{2+} ion, while the ion itself interacts with the acetyl group to polarize the amide bond. The departing amino group is then protonated by Glu130, leading to the breakdown of the tetrahedral intermediate, as shown in Scheme 3.11[207].



Scheme 3.11: Mechanism of deacetylation in citrulline deacetylase[207].

3.5 Common Features of Deacetylase Enzymes

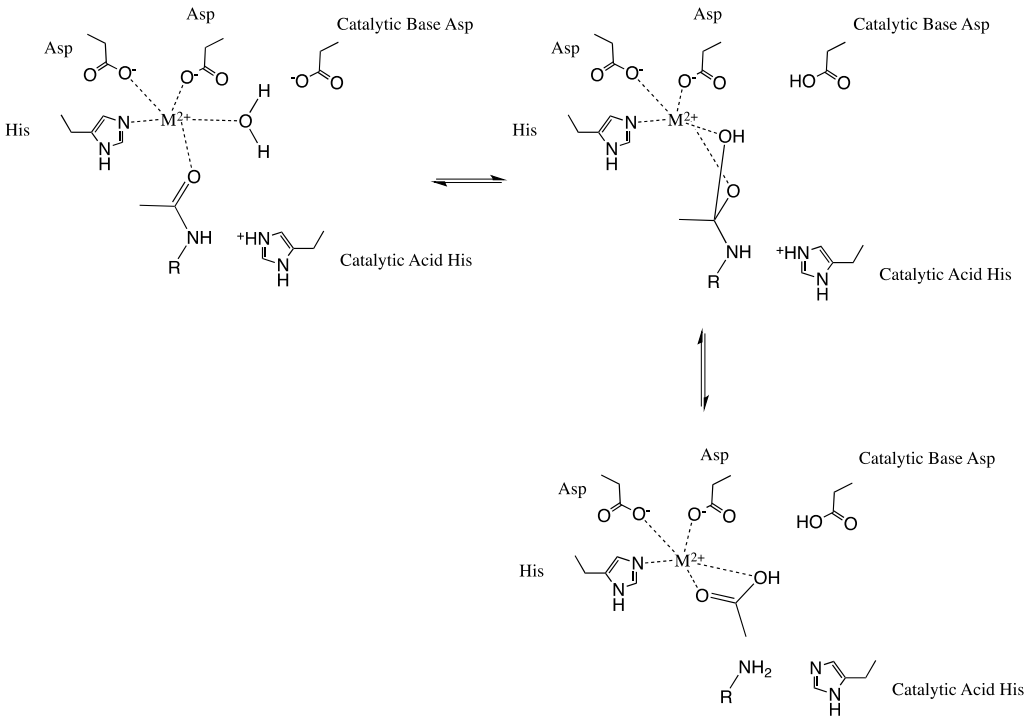
Two common features present in the majority of deacetylase enzymes are a hydrophobic binding pocket and a series of amino acid residues. The majority of deacetylases can be divided into two mechanistic families: serine protease-like and metal-dependent deacetylases. The serine protease mechanism uses a serine residue as a nucleophile, which is activated elsewhere in the protein by an aspartate residue. A histidine residue acts as a catalytic acid/base residue, similar to histidine residues in the histone deacetylase mechanism. A generic version of the serine protease mechanism is shown in Scheme 3.12.



Scheme 3.12: Serine protease mechanism[163]. Histidine is activated by hydrogen bonding to an aspartate residue further in the sequence and carries a H^+ during the mechanism, while serine acts as a nucleophile to the carboxyl carbon of the acetyl group.

The metal-dependent mechanism was used by IcaB and therefore was the focus of this review. In the metal-dependent mechanism, these residues consist of a divalent metal-binding triad, a catalytic acid, and a catalytic base. This metal-dependent mechanism was used by histone deacetylases, citrulline deacetylases, and carbohydrate esterase families 4 and 11. In this mechanism, the triad of negatively charged or polar residues, such as glutamate, serine, aspartate, or histidine, binds a divalent metal ion in the active site of deacetylases to assist in the binding of the substrate. This triad stabilizes the position of the active site metal, allowing for binding of substrate and a water molecule to the metal

ion. Catalytic acid and base residues are involved in the general deacetylase mechanism. In specific enzymes, additional amino acid residues stabilize the transition state molecule. An example of this is the main chain nitrogen atom from Tyr367 in SpPgdA, which holds the carbonyl oxygen in place during the mechanism[175]. In this mechanism, the catalytic base acts as a nucleophile on a water molecule. This in turn attacks the carbonyl carbon of the acetyl group. After the acetate group is released, the acid provides a proton for the nitrogen atom to yield an amine. An example of a general metal-dependent deacetylase mechanism is shown in Scheme 3.13.



Scheme 3.13: A proposed general metal-dependent mechanism of deacetylase enzymes. The triad of amino acid residues that bind the metal ion as well as the divalent ion itself varies between different enzymes[142,175,178,204]

3.6 Discussion of deacetylase review with focus on IcaB

Table 3.2 below shows a summary of which deacetylase families fit into which mechanistic group as well as presence of catalytic metal ions and active site residues.

Table 3.2: Deacetylase enzyme summary

Family	Targets	Mechanism	Divalent Metal Atom	Active site Residues
HDAC	Histone lysines	Serine prot	None	Ser His Asp
Sirtuins	TNF-alpha (possibly)	Unknown	N/A	N/A
CE-1	Acetylxylan	Serine prot	None	Ser His Asp
CE-2	Acetylxylan	Serine prot	None	Ser His Asp
CE3	Acetylxylan	Serine prot	None	Ser His Asp
CE4	Acetylated sugars	Metal-dependent	Varies	His His Asp Triad Catalytic Asp
CE5	Acetylxylan	Serine prot	None	Ser His Asp
CE6	Acetylxylan	Serine prot	None	Ser His Asp
CE7	Acetylated oligosaccharides Cephalosporin C	Serine prot	None	Ser His Asp
CE8	Pectin methylester groups	Gln-Asp-Asp Triad	None	Gln Asp Asp
CE9	N-acetylglucosamine-6-Phosphate	Two metal centres	Two Fe^{2+}	Unknown
CE11	UDP-3-hydroxymyristoyl-N-acetylglucosamine	Metal-dependent	Zn^{2+}	His His Asp Triad Catalytic Glu
CE12	pectin acetylesterases	N/A	N/A	N/A
CE13	pectin acetylesterases	N/A	N/A	N/A
CE14	GlcN-Ins	Metal-dependent	Zn^{2+}	His His Asp Triad Catalytic Asp
CE15	Unknown	N/A	N/A	N/A
CE16	Unknown	N/A	N/A	N/A

While deacetylases in general may be divided into two groups mechanistically, the carbohydrate esterases vary greatly in protein structure. Also, although both mechanisms are found in deacetylases, the group of proteins has such a broad range of targets and activity that CE8 and CE9 enzymes have unique mechanisms not found in any other deacetylase families. Of the carbohydrate esterases, CE4 enzymes contain the most available experimental structures, whereas carbohydrate esterase families 12, 13, 15 and 16 in bacteria have no crystal structures available and therefore little is known about

their mechanisms. Finally, while sirtuins have been historically considered to have histone deacetylase activity, newer research has suggested that they may hydrolyse fatty acid groups from lysine residues of TNF-α to enhance secretion[151].

Searching the Pfam and PROSITE databases for the IcaB sequence (GenBank Acession Number AAQ88122.1) showed the presence of a nodB homology domain specific to the CE4 family (Pfam ID: PF01522, PROSITE: PS51677)[170,208]. Therefore, IcaB is a member of the CE4 family and therefore should share the metal-binding triad of residues composed of His and Asp residues as well as a catalytic Asp residue. PgaB from *E. coli* has the same target, namely deacetylation of PNAG, and therefore was used as the template for modelling in chapter 4.

CE10 enzymes were originally proposed as a family based on sequence homology with other carbohydrate esterases, but were removed from CAZy as they do not possess deacetylase activity or act on carbohydrates. In addition, deacetylases as a family have generally been divided into histone deacetylases and carbohydrate esterases. There has been little discussion of these enzymes as a group, as the focus of most reviews is either on histone deacetylases or carbohydrate esterases.

For both mechanisms, a hydrophobic binding pocket is formed from non-polar residues, such as phenylalanine, tryptophan, or leucine. This allows the polar substrate (carbohydrates for carbohydrate esterases and the lysine residue for HDACs) to travel and bind to the active site without binding to residue sidechains elsewhere in the protein.

The binding pocket and active site residues are important to the activity of deacetylase enzymes and therefore, any deacetylase modelling must maintain these features. Both these features are essential for proper activity of the enzyme and will play a role in guiding the development of the IcaB model.

Chapter 4: Homology Modelling of IcaB

4.1 Biofilm Formation In *staphylococcus* Infections

Staphylococci such as *Staphylococcus capitis, S. aureus,* and *S. epidermidis* form biofilms on indwelling medical devices such as heart valves and catheters, which lead to chronic infections that are responsible for serious costs to the public health system[209,210]. If allowed to breach the epithelial barrier, *staphylococci* can cause chronic opportunistic infections[211]. Staphylococcal infections in the United States result in additional costs up to around $56 000 per person per year, as well as prolonged hospital stays [212]. In the UK, approximately 1 in 3300 individuals will develop an invasive staphylococcal infection per year[213]. In addition, neonatal infections are commonly due to infection with *staphylococci*[214]. Of the coagulase negative *staphylococci*, *S. epidermidis* is of particular interest since it is one of the most common bacteria that cause infections of indwelling medical devices[215]. Biofilms such as those formed by *S. epidermidis* are resistant to almost all antibiotics, which is a hindrance in treating these infections[103,215].

4.2 Structure and Activity of Partially Deacetylated Poly-N-acetylglucosamine

In biofilms of *S. epidermidis*, the extracellular bacterial coat is made of a combination of matrix proteins and non-protein polymers including polysaccharides and teichoic acids[216]. A linear polymer of partially deacetylated N-acetylglucosamine (dPNAG) is found on the bacterial cell surface and makes up the majority of the extracellular polysaccharide coat[144]. This polysaccharide has multiple roles in immune evasion. For example, mutant strains of *S. epidermidis* lacking dPNAG were much more vulnerable to being killed by human polymorphonuclear lymphocytes as well as by antibacterial peptides[184]. Biofilms containing dPNAG have been shown to be physically stronger, as well as less susceptible to antibody binding than those containing only protein factors such as Aap (accumulation associated protein)[217].

4.3 The *ica* Operon of *staphylococcus*

dPNAG is synthesized by the products of the *ica* operon: IcaA, IcaB, IcaC and IcaD. IcaA adds GlcNAc residues to growing poly-N-acetylglucosamine (PNAG) chains, which may be secreted through IcaC. This hypothesis is due to the presence of multiple transmembrane helices in IcaC and the fact that an *icaC* mutant was unable to form long chains of PNAG[218]. However, given that IcaC is classified as an acyltransferase (Pfam ID: PF01757) and there is a lack of evidence for export through IcaC, it is possible that this transport activity may have been misassigned[170]. Recently, it has been proposed that IcaC is responsible for O-succinylation of PNAG, which is necessary for the active form of PNAG specific to staphylococci[219]. IcaD is required for optimal PNAG synthesis activity of IcaA, but its role is currently unknown[216]. A diagram of the proposed organization of the *ica* operon gene products is shown in Figure 4.1.

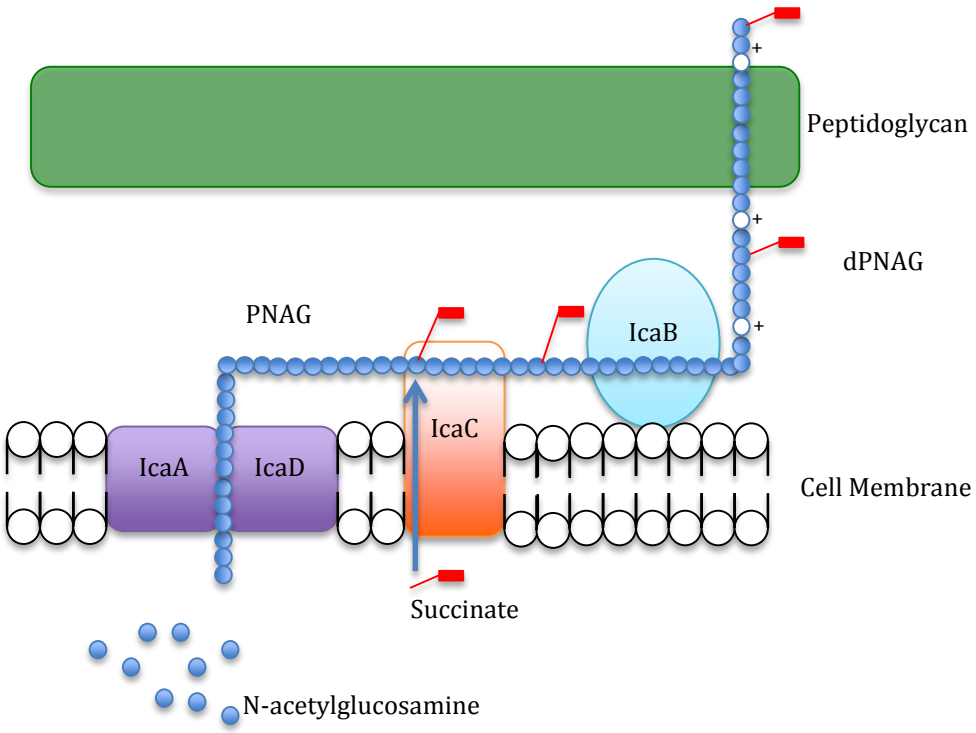


Figure 4.1: Proposed functions of gene products of the *ica* Operon. IcaA and IcaD form a polymer of N-acetylglucosamine, which is then O-succinylated by IcaC. Finally, PNAG is exported from the cell and the N-acetyl group is removed by IcaB (modified from Otto et al., 2012)[216,219].

After export from the cell, 16.5% of GlcNAc residues are then deacetylated by IcaB[216,220]. In 2014, Little et al. have demonstrated that the related protein PgaB, which shares the PNAG deacetylase function of IcaB, remains associated with dPNAG during transport across the bacterial cell membrane[34]. As a result, it is possible that IcaB may be responsible for dPNAG transport instead of IcaC.

4.4 Structure and Function of IcaB and PgaB

IcaB-dependent deacetylation has been implicated in the virulence of *staphylococci*. An IcaB mutant strain that produced non-deacetylated PNAG was unable to attach to the bacterial cell surface, presumably due to a lack of positive charge. Also, no biofilm formation occurred in this strain and there was a lack of resistance to antimicrobial peptides and phagocytosis by neutrophils provided by dPNAG[220]. As a result, this protein has been identified as a potential target for novel therapies. However, the structure of IcaB, as well as a detailed mechanism of deacetylation by IcaB, was unknown at the beginning of this project.

4.4.1 IcaB and PgaB domain structure

The domain structure of IcaB resembles that of PgaB; however, IcaB lacks a C-terminal glycoside hydrolase domain. This supports the similarity between IcaB and the PgaB N-terminal domain despite the low sequence homology between the two proteins. Figure 4.2 shows the domain structures for IcaB and PgaB, both of which are taken from Pfam: *Polysacc_deac_1* (PF01522)**[170]**.

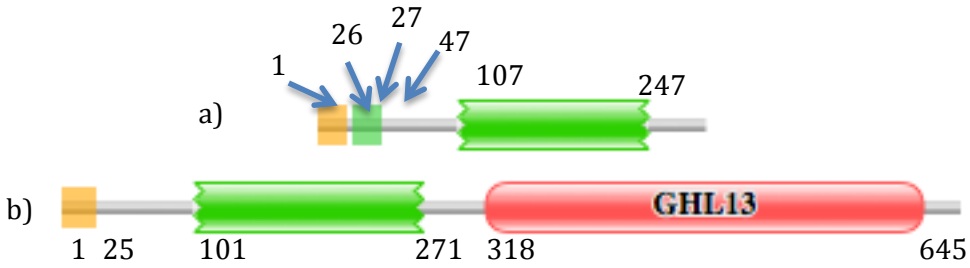


Figure 4.2: a) IcaB domain structure from *Staphylococcus epidermidis*; b) PgaB domain structure from *Escherichia coli* (strain K12). The domains are represented by Yellow = signal sequence, light green = coiled coil, green = deacetylase domain, GHL13 = Glycoside Hydrolase-Like C-terminal domain [170].

Both proteins contain a deacetylase domain; PgaB also contains a C-terminal glycoside hydrolase domain. Due to the role of the C-terminal domain in transport of PNAG in PgaB, this may imply a different export mechanism for PNAG through IcaB in *staphylococci*[34]. The outer membrane of Gram-negative E. coli may require the PgaB C-terminal domain to aid in PNAG transport, while the Gram-positive *S. epidermidis* will require a different mechanism.

4.4.2 Structures of PgaB

There are currently three structures of PgaB in the Protein Database: 3vus, a fragment containing only the N-terminal domain of PgaB with an additional gap at residue 155, as well as 4f9d and 4f9j, which both have a truncated N-terminus and include N-terminal and C-terminal domains (resolutions: 3vus = 1.65Å, 4f9d = 1.90 Å, and 4f9j = 2.10Å.)[142,221]. Other CE4 enzymes have preference for Co^{2+} and Zn^{2+}, but PgaB only shows higher activity in an activity-based metal screen with Co^{2+} and Ni^{2+} and the addition of Zn^{2+} causes a reduction in activity[142,222]. One structure of PgaB contains a Ni^{2+} ion as the active site metal (4f9d). The second structure is similar, but with a Fe^{2+} ion in the active site (4f9j)[142]. The N-terminal domain in all three structures is identical, except for a few gaps caused by missing amino acids from low-resolution structures. An alignment of all three structures' sequences is shown in Figure 4.3.

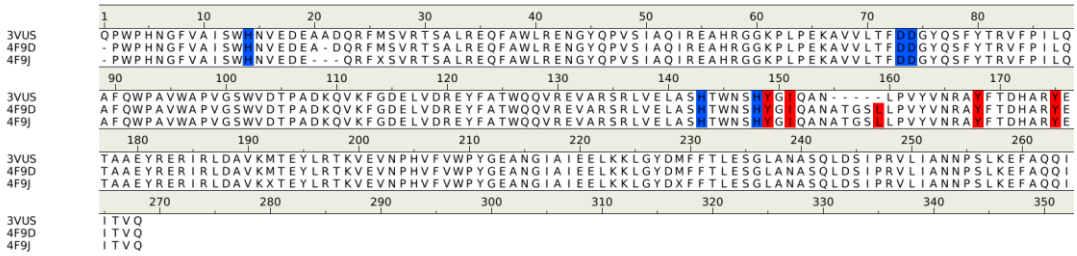


Figure 4.3: Alignment of all sequences of PgaB in Protein Databank (3vus, 4f9d, and 4f9j); active site residues are blue, and binding groove residues are red[142,221]. 4f9j has methionine residues replaced by selenomethionine for crystallization, which are represented by the letter X in the sequence.

The RMSD (Root Mean Square Deviation) of all heavy atoms between the 4f9d and 4f9j structures is 0.674Å and there are no dramatically differing regions between both structures. 4f9d and 4f9j contained crystals with orthorhombic space groups ($P\ 2_1\ 2_1\ 2_1$), while the crystals in 3vus were monoclinic ($P\ 1\ 2_1\ 1$). There are two identical proteins per asymmetric unit (asu) in each structure, which is believed to be an artefact of crystallization[142,221]. The region in the crystal where the helix α26 should be located is instead replaced by the α1 helix from the second protein in the asu. This provides further evidence that the appearance of two molecules in PgaB (residues 43-646 in 4f9d/j and 42-309 in 3vus) seen in the asu is an artefact of crystallization, because the two structures would be unable to pack into a single asymmetric unit if the α26 helix was present. The N-terminal domain of 4f9d was selected for further modelling as it was the most complete structure and the iron-containing structure, 4f9j, was not stable at ambient conditions[142]. Therefore, the activity of PgaB with Fe^{2+} would be difficult to test experimentally[142]. The PDB structures 3vus, 4f9d, and 4f9j are overlapped and shown in Figure 4.4.

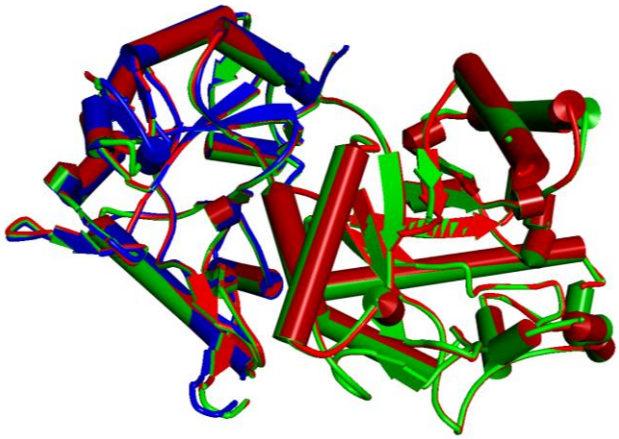


Figure 4.4: Comparison of PgaB structures available in the Protein Data Bank (3vus in blue, 4f9d in green, and 4f9j in red)[142,221]

The active site and N-terminal binding groove of PgaB from 4f9d are shown in Figure 4.5.

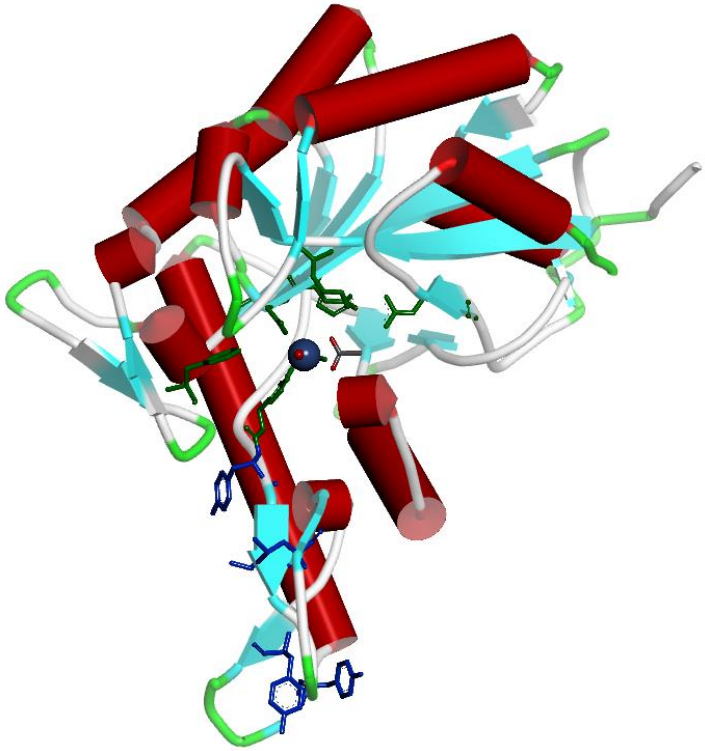


Figure 4.5: Active site and N-terminal binding groove of PgaB from 4f9d. The active site is shown in green (His55, Asp114, Asp115, His184, His189) and the hydrophobic binding groove in blue (Phe-154, Tyr-190, Ile-192, Leu-200, Tyr-209, and Tyr-216)[142].

4f9d and 4f9j contain the C-terminal domain of PgaB, which contains a beta barrel $(\beta/\alpha)_8$ motif. The NodB homology domain is generally present as a unique domain on its own in CE4 enzymes. This is not the case in PgaB, which implies an important role for the C-terminal domain. There is little homology between the complete sequence of IcaB and the C-terminal domain of PgaB, which supports the existence of a different export mechanism in *S. epidermidis*. However, since the C terminus of PgaB was truncated to facilitate crystallization, the eighth helix of the C-terminal domain, α26, is missing in the PDB structures, which may impact the overall structure of this domain. The role of this region is currently unclear, but is important for the ratio of deacetylated to acetylated residues *in vivo*, as removing it from the protein causes a reduction from 22% to 1%[223]. Eight short helices in this domain create a negatively charged surface groove 45Å long and 10Å wide, which may be involved in binding PNAG upstream of the site of deacetylation. Residues Phe-154, Tyr-190, Ile-192, Leu-200, Tyr-209, and Tyr-216 from β6–7, β9–10, and β11-12 provide a hydrophobic environment in the C-terminal domain. This groove is structurally similar to glycoside hydrolases such as β-amylase[224]. This may imply a role for this domain in hydrolysis of the polysaccharide chain or simply to assist in binding of the PNAG substrate. However, this domain has recently been suggested to bind PNAG as part of the export from the bacterial cell[34].

4.5 IcaB Alignment with PgaB

The goal of this research was to produce a homology model of IcaB from *S. epidermidis* using sequence alignment with a related protein, PgaB, from *E. coli*.

The Clustal multiple sequence alignment program was used to align sequences of IcaB and PgaB[130]. Clustal is based on derivation of a phylogenetic tree from a sequence similarity score matrix between individual sequences and the template sequence to be aligned. A multiple alignment may be formed by organizing individual pairwise alignments in the order of most to least homologous sequences compared to the template[225]. The alignment at each amino acid residue is then scored, with the highest score being identical to the template sequence and the lowest score being a residue with a different structure and charge/polarity.

4.5.1 Methods

Several sequence alignments of IcaB from *S. epidermidis* (GenBank Accession number: AAQ88122.1) with the N-terminal domain of PgaB containing a Ni^{2+} ion in the active site (residues 42-310 of pdbID: 4f9d) were produced by the Clustal program[154,225]. The sequence of PgaB was taken from the PDB file 4f9d and individually aligned with the sequence of IcaB. In this alignment, the active site residues of the PgaB sequence were not aligned with the target IcaB sequence. As a result, the PgaB sequence from GenBank (ZP_03070115.1) was aligned with the target sequence and the active site residues of PgaB were aligned with identical residues in the target alignment, which differed due to missing signal sequence residues in the PDB sequence. The GenBank sequence was then truncated to match the 4f9d; the alignment was very similar to the other alignments[226]. This alignment was then manually adjusted to retain the structure of alpha helix 7.

4.5.2 Sequence Alignment

All alignments had low sequence identity: 13.7% for the original alignment, 16.4% for the truncated GenBank sequence, and 15.7% for the manually adjusted alignment. In both the alignment with the truncated GenBank sequence and the manual alignment, the active site residues of the target IcaB sequence were properly aligned with those of PgaB. In addition, the sequence similarity was higher than the identity in each case: 36.2% for the original alignment, 41.5% for the truncated GenBank sequence, and 39.3% for the manual alignment. While the identity may be low, the sequence is fairly similar, which may lead to a sufficiently similar environment for binding. For example, if an electronegative residue such as glutamate were to be replaced by a negatively charged and sterically similar aspartate residue, the binding environment may continue to be negative and in the same position.

There were minor differences between alignments, but only by a few residues, with the most important difference in the large α7 helix. This may be the most problematic region for the model, as the automatic alignment inserted a gap into this region, which caused the long helix structure to be lost. When the gap was moved manually upstream, the helix was retained with little loss of sequence identity (Figure 4.6).

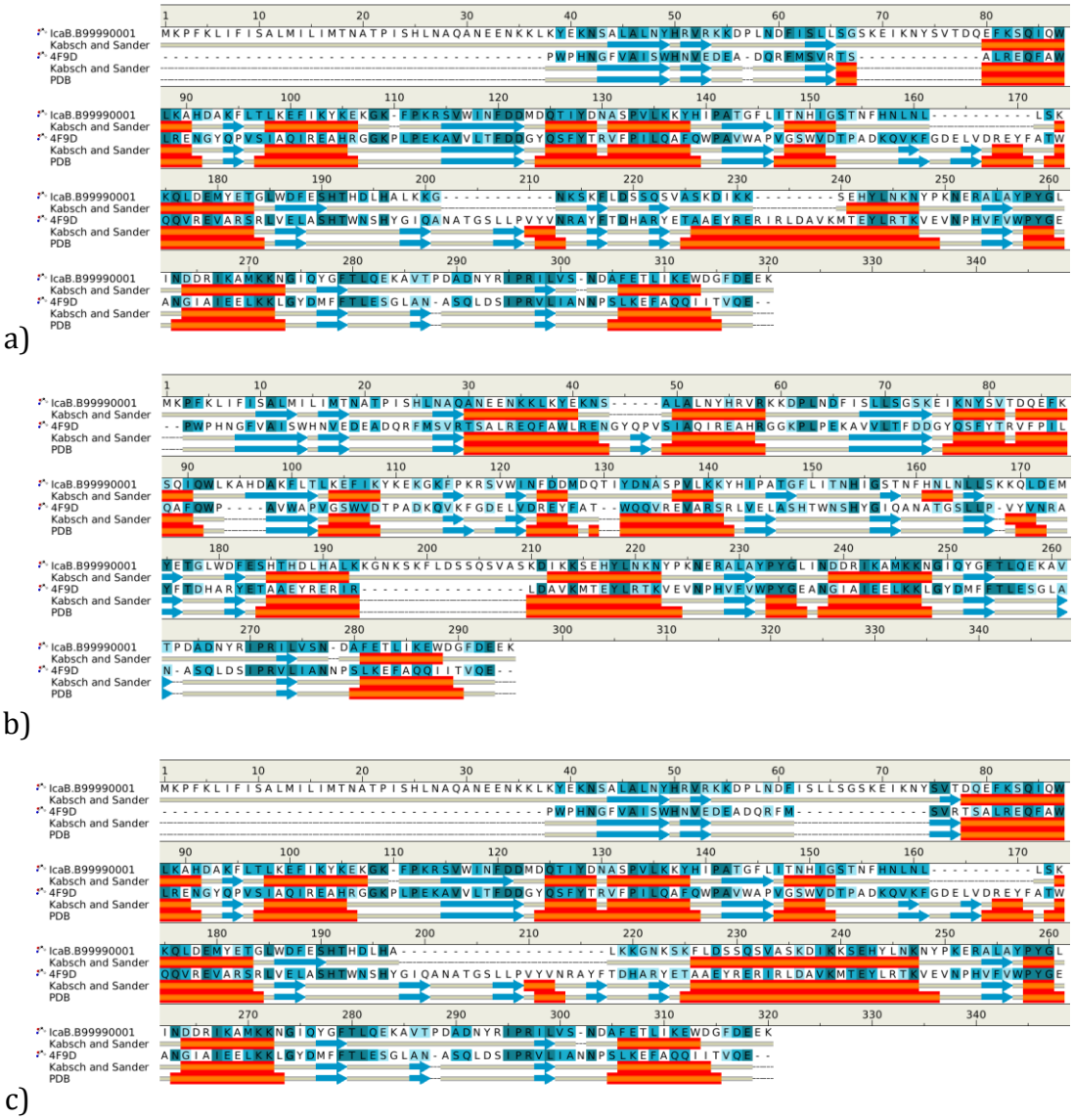


Figure 4.6: a) Sequence alignment of IcaB with 4f9d N-terminal domain; b) Sequence alignment of IcaB aligned with PgaB N-terminal domain NCBI sequence and truncated to match 4f9d (residues 1-42 removed); c) Sequence alignment from b) manually adjusted to preserve α7 helix. Dark blue represents fully conserved residues, light blue strongly conserved residues and pale blue weakly conserved residues. The secondary structure is shown as orange cylinders for α-helices and blue arrows for β-strands.

As the N-terminal domain was truncated by 43 residues to facilitate crystallization, Pro1 of the PDB sequence is equivalent to Pro43 of the GenBank sequence, which was used for

the residue numbering in this project. The N-terminal groove residues were not conserved in IcaB. The only similarities were Phe154, which was replaced by a leucine, and Tyr190 by an alanine, but otherwise these residues were different, several even aligning within the added gap. This may be an issue with the alignment or it may be that the groove is different in IcaB, which may compensate for the lack of C-terminal domain if the groove is larger in IcaB. Unfortunately, without more information about a structure of IcaB or characterization of the export mechanism of PNAG in *staphylococci*, the effect of this groove is still unknown. The active site and N-terminal groove residues from the final alignment are shown in Figure 4.7.

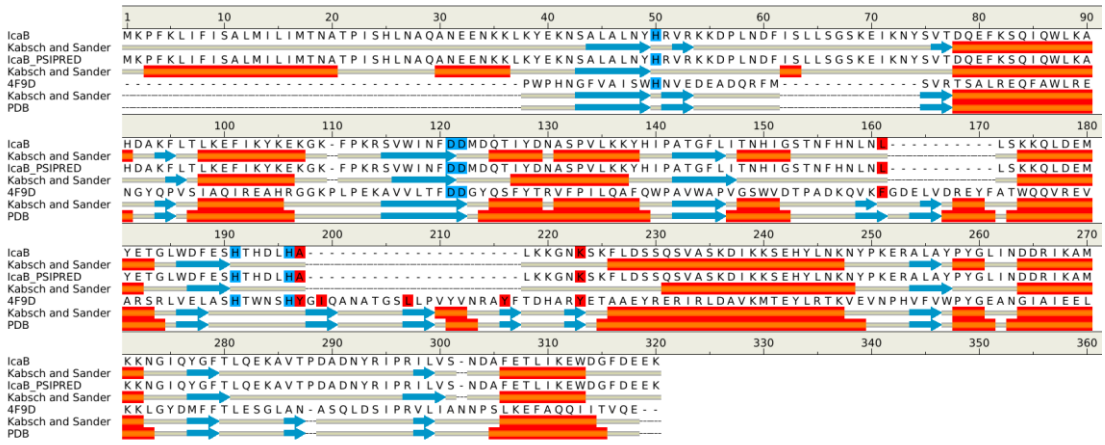


Figure 4.7: Active site and N-terminal groove residues in the final alignment; active site residues are blue and binding groove residues are red. The secondary structure is shown as orange cylinders for α-helices and blue arrows for β-strands.

4.6. Homology Model of IcaB

A homology model of IcaB using the program MODELLER was constructed using the sequence alignment with PgaB as described above.

MODELLER is a homology modelling program that calculates theoretical protein structure based on alignment with a related protein with high sequence identity. MODELLER restraints include C_α-C_α bonds (d), main chain N-O distances (h), stereochemical restraints (e), main chain dihedral angles (φ and ψ), and side chain dihedral angles (χ_1, χ_2, χ_3, and χ_4, referred to in the equations as "c"). These restraints are defined through sequence alignment with structures of related proteins. Therefore, a

sequence identity of >40% is generally needed to produce a model with a high degree of confidence. Restraints are used to make molecular probability density functions (pdfs) for conformations of target sequence. Pdfs contain more information than bounds for bond distances only, including restrictions on bond angles, dihedral angles, stereochemical restraints and solvent accessibility[90]. Descriptions of these functions are shown in Equations 4.1-4.5.

$$p^{D}(d) = p^{v}(d) \sum_{i} \omega_i p_i^{d}\,(d|a, b, \dots)$$

Equation 4.1: C_α-C_α bonds

$$p^{H}(h) = p^{v}(h) \sum_{i} \omega_i p_i^{h}(h|a, b, \dots)$$

Equation 4.2: Main chain N-O bonds

$$p^{E}(e) = p^{e}(e)$$

Equation 4.3: Stereochemical restraints

$$p^{M}(\theta) = \begin{cases} \sum_{i=1}^{n} \omega_i p_i^{m}(\theta|a, b, \dots) & n > 0 \\ p^{M}(\theta|R) & n = 0 \end{cases}$$

Equation 4.4: Main chain dihedral angles

$$p^{S}(c) = \begin{cases} \sum_{i=1}^{n} \omega_i p_i^{s}(c|a, b, \dots) & n > 0 \\ p^{S}(c|R) & n = 0 \end{cases}$$

Equation 4.5: Side chain dihedral angles

Optimization of these functions uses a combination of two methods: conjugate gradients, as previously described, and molecular dynamics with simulated annealing. Molecular dynamics with simulated annealing finds a decreasing probability of high-energy solutions over time, leading to a global maximum at the most stable structure. Simulated

annealing is so named because it simulates the effect of lowering temperature to find the most probable conformation[227].

DOPE (Discrete Optimized Protein Energy) is a statistical potential used by the program MODELLER to calculate the minimized energy of a homology model produced. DOPE is implemented in Python and may be run through the MODELLER environment [228]. Due to the specificity of the parameters used to calculate this energy, DOPE can only be used to compare different iterations of the same model structure.

The secondary structure in the homology model of IcaB was calculated using the Kabsch-Sander method, as implemented in Accelrys Discovery Studio. This method calculates the secondary structure based on recognition of patterns of hydrogen bonding that result in specific secondary structure motifs. Hydrogen bonds are treated as electrostatic interactions between two sets of partial charges, i.e. positive partial charge for the carbonyl carbon and a negative for the oxygen, and a negative partial charge for the amino nitrogen and a positive for the hydrogen. Patterns of hydrogen bonding are divided into "turns" and "bridges". A "turn" is defined as a hydrogen bond between a carbonyl carbon of a given residue and an amino nitrogen 3-5 residues away from it in the backbone, while a "bridge" is a hydrogen bond between two residues that are further apart. Patterns of these hydrogen-bonding interactions are used to determine secondary structure. For example, repeating "turns" that are 4 residues apart are classified as α-helices, while repeating "bridges" form individual β-strands, which if connected, form a β-sheet. The computer program DSSP (Define Secondary Structure of Proteins), included in Discovery Studio, implements these algorithms to predict secondary structure from a given sequence[229].

In addition, the secondary structure was predicted from the sequence using PSIPRED prediction methods. These use a pair of neural networks to analyse PSI-BLAST scoring matrix and to predict the secondary structure based on the most conserved residues[230]. PSI-BLAST (Position-Specific Iterative-Basic Local Alignment Search Tool) performs alignments with the NCBI database to determine the most homologous sequences. Next, a position specific scoring matrix is calculated to store the pattern of

conserved residues. This is then used for additional iterations of BLAST searches until a threshold for convergence is met[231].

The prediction method used to determine the secondary structure by the neural networks in PSIPRED is the Benner and Gerloff method. This method uses patterns of amino acid polarity to determine secondary structure elements such as α-helices and β-strands[232]. For example, to determine whether a protein segment is an α-helix, a helical wheel is set up to identify whether an amphiphilic pattern may be achieved (namely, hydrophobic residues oriented towards the protein core and hydrophilic towards the protein surface). If not, positions are removed from the ends of the chain until a region within the segment showing a 3.6 residue periodicity is achieved. If this periodicity does not appear, a helix is not assigned to any section of this protein segment. For PSIPRED, this method is applied to the conserved residues predicted via the PSI-BLAST scoring matrix.

4.6.1 Methods

The structure of IcaB from *S. epidermidis* (GenBank Accession number: AAQ88122.1) was modelled by the homology modelling program MODELLER, using the manually adjusted alignment with the N-terminal domain of PgaB containing a Ni^{2+} ion in the active site (residues 42-310 of pdbID: 4f9d)[90]. Default settings in MODELLER were used and the Discrete Optimized Protein Energy (DOPE) was calculated for individual iterations of the model. Additional support for the model was provided by comparison with predicted secondary structure by PSIPRED[233].

4.6.2 Results

Homology models were produced for the manually adjusted alignment of IcaB with the N-terminal domain of PgaB (residues 43-310) from 4f9d structure using the program MODELLER[90]. Five models were produced and the model with the lowest DOPE is shown in Figure 4.8[228]. The RMSD between the lowest and second lowest DOPE structures was 3.244Å. The long N-terminal region that was truncated in the PgaB structure was not properly modelled and therefore is not shown.

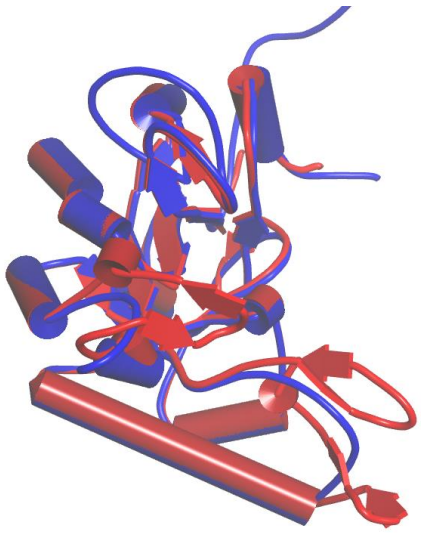


Figure 4.8: Homology model of IcaB using manually adjusted 4f9d NCBI alignment as a template (red = 4f9d template, blue = IcaB model)

4.7 Discussion and Analysis of IcaB Model

The model may have varied between individual runs[234]. To determine whether the model would have varied given different starting positions, five models were produced in MODELLER, and the RMSD (between 4f9d and IcaB) for the main chain and side chain residues of each model was calculated (Table 4.1).

Table 4.1: RMSD comparison between multiple models of IcaB with manually adjusted 4f9d template

Model Number	Main Chain RMSD (Å)	Side Chain RMSD (Å)	DOPE score
1	1.99	2.19	-26400
2	2.19	2.57	-26400
3	2.23	2.26	-26200
4	2.28	2.56	-26300
5	2.08	2.27	-26700

All RMSD are under 3Å and similar enough to be only minor changes to the structure on average. The active site residues in IcaB (His50, Asp121, Asp122, His180 and His185) have a lower RMSD than the average RMSD for all residues, which supports conservation of active site residues in the proper orientation. The charge surface of the IcaB model is shown in Figure 4.9.

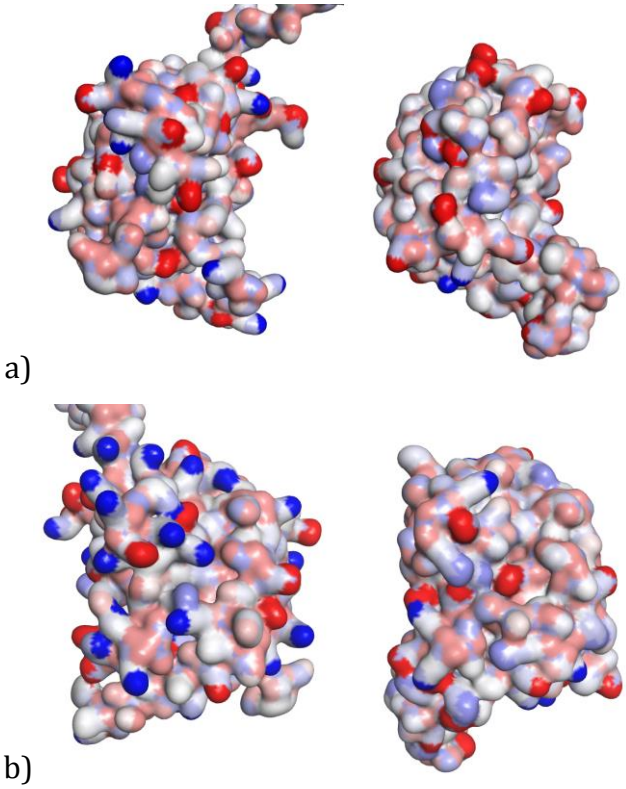
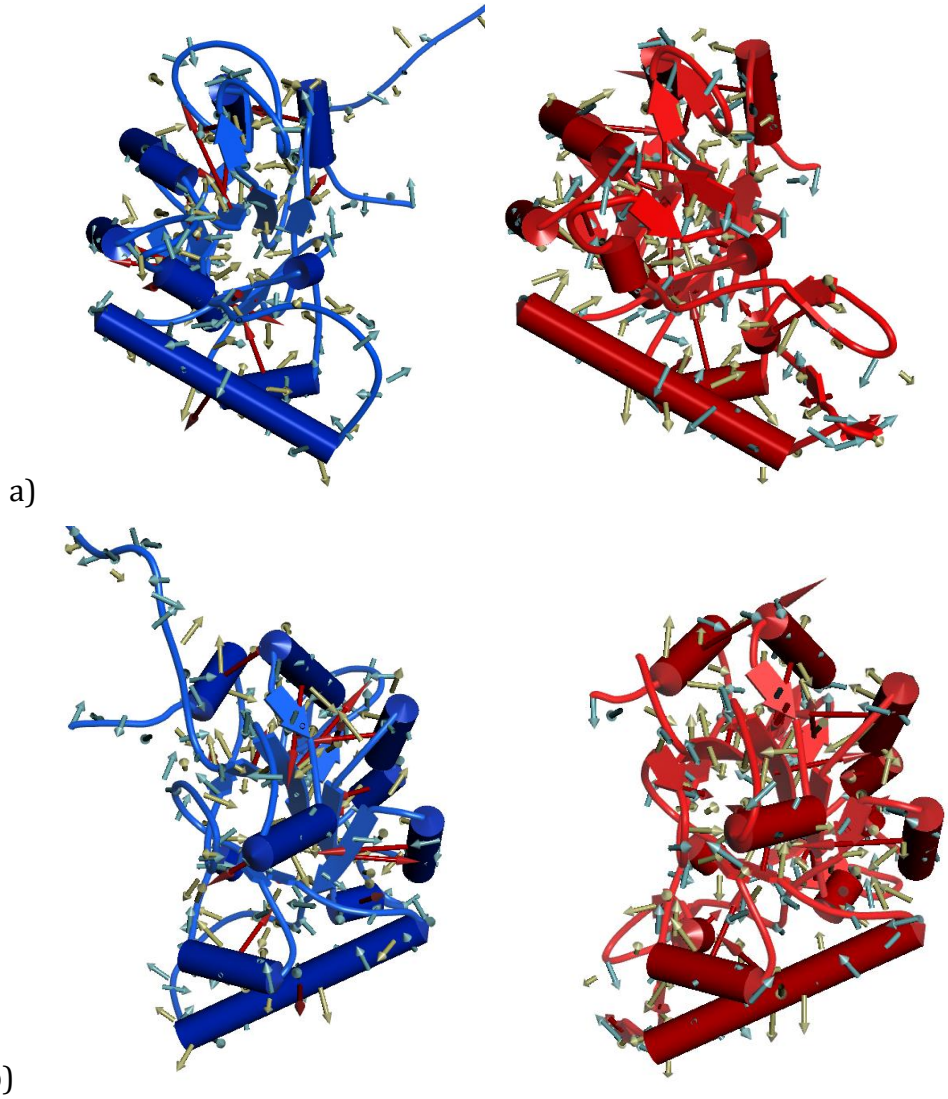


Figure 4.9: a) Charge surface of IcaB homology model (left) and 4f9d template (right) in which red indicates regions of negative charge and blue indicates regions of positive charge b) Charge surface from a) rotated by 180°.

The theoretical pI was calculated by the ExPASy server to be 9.20 for IcaB compared to 5.80 for PgaB, which suggests a more positive surface for IcaB at neutral pH[42]. This is well shown in the surface of the IcaB model when compared to the PgaB surface (Figure 4.9). There are no regions in the hydrophobic core of the protein with large positive/negative charge. Both these points indicate reliability for the structure, as the charge distribution appears normal for a monomeric protein[235].

A hydrophobic moment is a representation of hydrophobicity by searching for recurrence of hydrophobic residues in a one dimensional function[236]. The strongest hydrophobic moments are in the core of the protein and the weakest (most hydrophilic) are on the surface (as shown in Figure 4.10), which agrees with the charge surface in Figure 4.9. The exception is the α7 helix, which has a strongly hydrophobic region in the middle of the helix; this is lacking from the PgaB structure. This may be a problem with manual alteration and may require fine-tuning of the manual alignment or may provide a starting point for the hydrophobic groove of IcaB.



a)

b)

Figure 4.10: a) Hydrophobic moments in IcaB model (left) and PgaB (pdbID: 4f9d, right); b) Hydrophobic moments from a) rotated 180°.

In both properly aligned models, the hydrophobic core of the protein is well conserved, particularly the nodB homology domain containing the active site residues. The α7 helix only maintained its alpha helix character in IcaB when the alignment was manually altered. Most of the structure that differed were the solvent-exposed loops and external helices. Several short β-sheets 2-3, 6-7, 9-10, and 11-12 differ from similar sheets in PgaB (Figure 4.11). However, these sheets are regions unique to the PgaB structure, which implies that, like other CE4 enzymes such as SpPgdA, IcaB has different folds in these regions of the protein[142]. However, as several residues in these folds are important for forming the hydrophobic binding groove, this difference in IcaB may either be a difference caused by the modelling process or it may be due to a different binding groove for the PNAG substrate in IcaB. The reason for this difference is unknown, but may help to explain why IcaB lacks the second C-terminal domain found in PgaB. The active site residues in IcaB (His50, Asp121, Asp122, His180 and His185) were within 2Å of the corresponding PgaB residues and in similar orientations (as shown in Figure 4.12). Although the standard sequence identity required is 40% for a homology model to prevent large deviations away from the template, the important features of PgaB (namely the hydrophobic core and active site residues) were conserved in the IcaB models as seen in Figures 4.11 and 4.12[90].

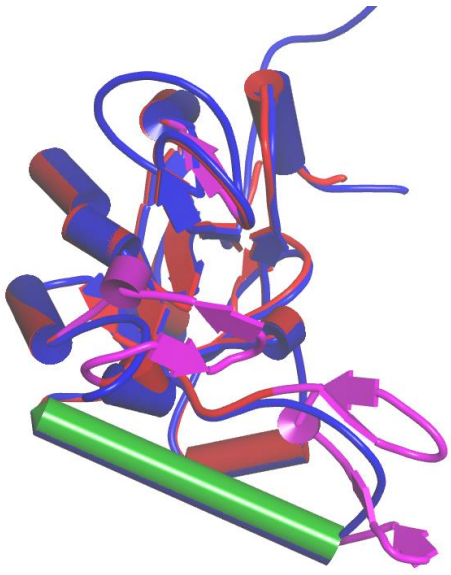


Figure 4.11: Regions of PgaB (red, 4f9d) that differed most from the IcaB model (blue). Pink represents β-sheets 2-3, 6-7, 9-10, and 11-12 and green represents the α7 helix.

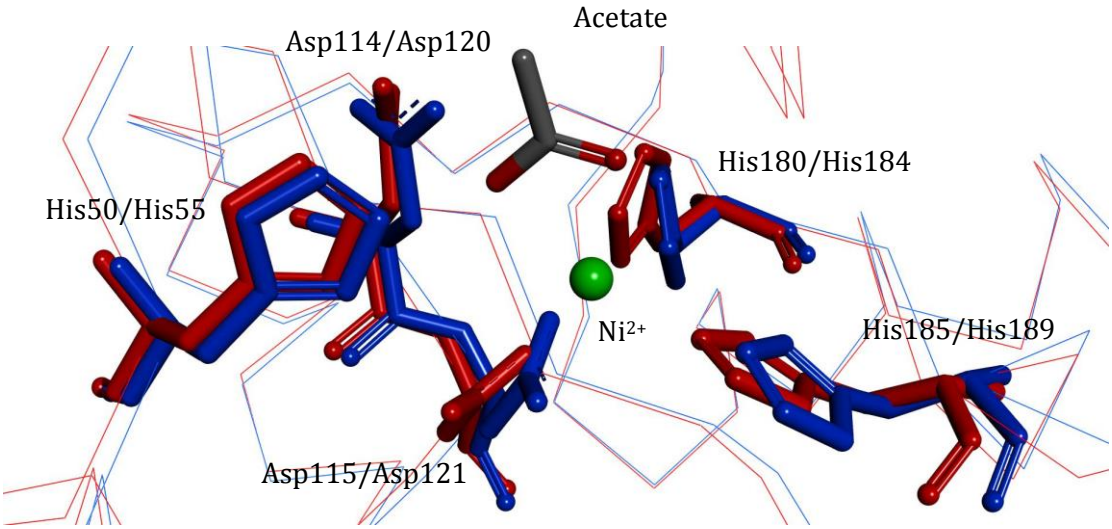


Figure 4.12: Overlap between active site residues of PgaB (red, 4f9d) and the IcaB homology model (blue). Residue labels are shown in the format IcaB/PgaB. Bound Ni^{2+} and acetate structures were taken from the 4f9d structure and both proteins are represented by C_{α} wires.

When compared with the PSIPRED predicted secondary structure (Figure 4.13), the manually adjusted alignment was found to be more similar than the other 4f9d alignments, particularly near the α7 helix[233]. This alignment only differed in several small regions from the PSIPRED structure. While there appears to be a break in a helix at Asn129, this may be due to incorrect secondary structure prediction due to the algorithm used, which may also explain the presence of several small helices, such as one directly after the α7 helix. Comparison with the PSIPRED secondary structure provides further evidence that the original alignment is flawed because the PSIPRED predicted secondary structure possesses more secondary structure elements of the manually adjusted alignment, particularly in the α7 helix region.

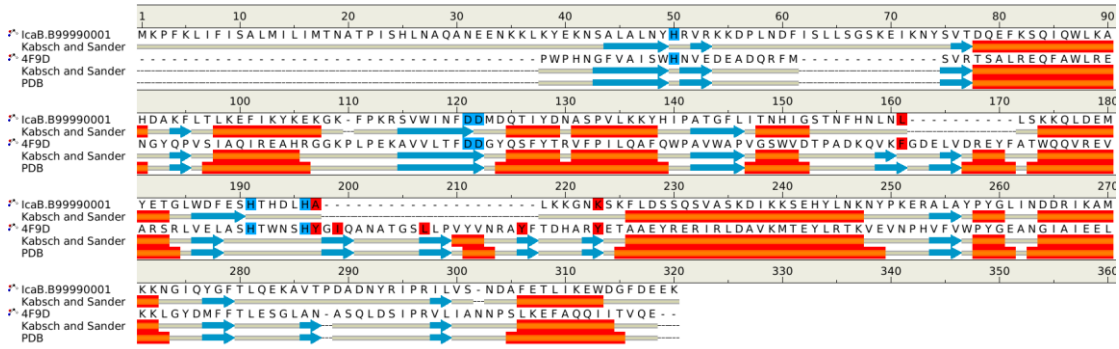


Figure 4.13: Comparison of manually adjusted IcaB alignment with PSIPRED predicted secondary structure; active site residues are blue, and binding groove residues are red

The manually adjusted structure of the IcaB model was best aligned with the PgaB sequence from the NCBI database. IcaB was found to contain the NodB homology domain common to carbohydrate esterase family 4 enzymes. The homology model structure of IcaB possessed a conserved triad of metal binding as well as catalytic acid/base residues necessary for a deacetylase. The pattern of hydrophobicity suggests a stable globular protein. The model still has two main concerns: the location of the α7 helix and the loss of the N-terminal binding groove. This will require additional study (both experimental and computational) to determine whether these are features of the IcaB protein or if the model requires further adjustment. Until such time as an X-ray structure of IcaB from *S. epidermidis* is determined, this model should be refined and used to aid in structure determination, as well as providing a starting point for molecular docking. Protein threading could be used to create a model of IcaB to compare features of the protein with the homology model in order to further refine the model. The main challenge towards refining the homology model would be the lack of highly homologous proteins (>40% sequence identity) to align with the IcaB sequence. A multiple sequence alignment with several CE4 enzymes may provide additional confidence in the resulting IcaB homology model.

After the conclusion of this project, additional research regarding IcaB structure and characterization has been published. The catalytic efficiency and residue specificity of IcaB has been characterized by Pokrovskaya et al., who demonstrated the position-specific deacetylation of N-acetylglucosamine pentamers and hexamers. In addition, the

low rate of turnover for IcaB deacetylation was confirmed, which resembles other related proteins such as PgaB[138]. On October 30, 2014, Nitz and co-workers reported an X-ray crystal structure of IcaB from *Ammonifex degensii* ($IcaB_{AD}$)[237]. From their report, it is clear that the structure of $IcaB_{AD}$ is similar to the homology model's overall structure. While the protein was derived from a different organism and will therefore have a difference in sequence, the overall protein structure should be similar. The metal ion found to associate greatest with this protein was Ni^{2+}. However, $IcaB_{AD}$ was best crystallized with Zn^{2+} and therefore the structure of $IcaB_{AD}$ shown in Figure 4.14 was co-crystallized with a Zn^{2+} ion.

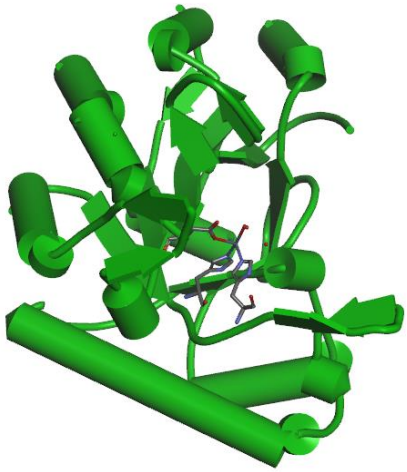


Figure 4.14: Structure of IcaB from *Ammonifex degensii* at 1.70Å reproduced from pdbID: 4WCJ[237]

The larger hydrophobic loop found in both the model, due to positioning of the α7 helix, as well as the $IcaB_{AD}$ structure, has been theorized to attach the protein to the cell membrane. This may help to explain the lack of a C-terminal domain, as this loop may take the place of the C-terminal domain for attachment to and transport across the bacterial cell membrane. Another aspect of the structure that is similar is the organization of the deacetylase active site residues. The triad of metal binding residues and catalytic acid/base residues in the $IcaB_{AD}$ structure are oriented in the same position as the IcaB homology model.

In order to compare the experimental IcaB structure with the homology model produced in this chapter, a simple alignment between the model and the *A. degensii* sequence taken from pdbID: 4WCJ was performed and is shown in Figure 4.15.

```
IcaB      MKPFKLIFISALMILIMTNATPISHLNAQANEENKKLKYEKNSALALNYHRVRKKDPLND
4wcj      -----------------------------------THLQGDGLVVLCYHRVL-------
                                              : : :. :.* ****

IcaB      FISLLSGSKEIKNYSVTDQEFKSQIQWLKAHDAKFLTLKEFIKYK-EKGKFPKRSVWINF
4wcj      ---------PSSRYAISRREFAQQLDYLRQVGVRFVTPQEAEDYLAGRIHLPGKLVLVTF
                    ..*::: :** .*:::*:   .:*:* :*  .*   : ::* : * :.*

IcaB      DDMDQTIYDNASPVLKKYHIPATGFLITNHIGSTNFHNLNLLSKKQLDEMYETGLWDFES
4wcj      DDGDLSVYRHAFPVLKKRKIPFLFFVIAGQVGR-KWEGFSMCSWEQIKEMVASGLCVVGL
          ** * ::* .* ***** :**   *:*: ::*  ::. :.: * :*:.**  :**  .

IcaB      HTHDLHALKKGNKSK-FL-DSSQSVASKDIKKSEHYL-NKNYPKERALAYPYGLINDDRI
4wcj      HTYDLHYWDSQAKKPVFLLPGRERLFAEDTARGTACLKEHLGLKTRYFAYPYGFGTPTTD
          **:***  ..  *.  **  . : : ::*  :.   * ::   * * :*****: .

IcaB      KAMKKNGIQYGFTLQEKAVTPDADNYRIPRILVSNDAFETLIKEWDGFDEEK
4wcj      EILRTQGFSLVFTLRAKVNRPGDAPF-VGRVLVTPDSWPQV-AAWAQA----
          : ::.:*:.  ***: *.  *    : : *:**: *::  :   *
```

Figure 4.15: Alignment of IcaB sequences from *S. epidermidis* (top) and *A. degensii* (bottom) (* = identical, : = conserved, .= semiconserved)[130,237]

All active site residues (*S. epidermidis* His50, Asp120, Asp121, His180 and His185) were identical in both sequences. The sequence identity is still very low at 22.8% even factoring in the unfolded N-terminal in model (removing this region only increases identity to 26.2%). A comparison of the *S. epidermidis* IcaB homology model and the *A. degensii* IcaB crystal structure is shown in Figure 4.16.

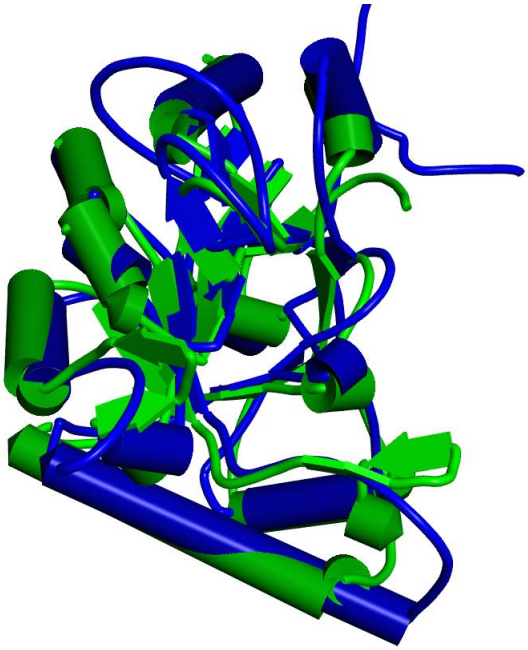


Figure 4.16: Comparison of *S. epidermidis* IcaB homology model (blue) and *A. degensii* IcaB crystal structure from pdbID: 4WCJ (green) [237]

The overall structure is similar, although the loop domains are much more variable and the position of the helices is slightly different. Most importantly, the nodB homology domain containing the catalytic and metal-binding residues is found in both structures. The active site residues are shown in Figure 4.17.

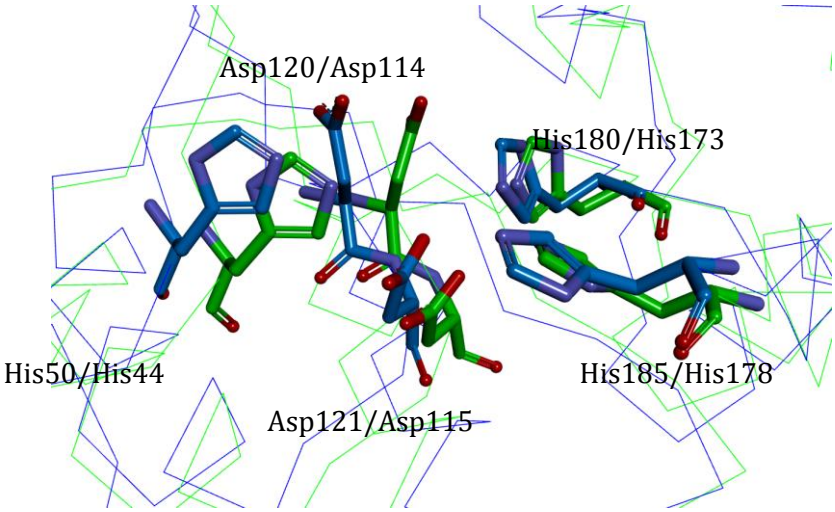


Figure 4.17: IcaB active site residue overlap. Numbering is in the format *S. epidermidis* structure (blue)/*A. degensii* structure (green) [237]

The RMSD between all heavy atoms for the main chain is 4.42Å and 5.49Å for the side chains. However, RMSD for the active site residues of both structures is 1.65Å, which is within the 3Å threshold. As a result, the 4WCJ structure could potentially be used in future research as a template for further homology models of the S. epidermidis structure. Although the low homology means that the homology model is still below the necessary threshold to be accurate, a multiple sequence alignment using the PgaB N-terminal domain and the $IcaB_{AD}$ structure could be used to provide a more accurate alignment on which to base the model.

Chapter 5: Conclusions

Computational modelling of proteins and their interactions with small molecule ligands is a field of increasing importance [92,238,239]. Potential applications of this research includes virtual screening of therapeutic agents to reduce experimental load, *ab initio* design of enzymes and inhibitors, and insight into protein mechanisms[240–242]. This thesis is divided into two main projects: modelling of ciprofloxacin derivatives binding to DNA gyrase and the production of a homology model of the protein IcaB by sequence alignment with PgaB.

Three derivatives of the antibiotic ciprofloxacin using a citrate group were prepared in order to improve the pharmacodynamics and pharmacokinetics of ciprofloxacin while maintaining its antibiotic activity. The activity of these derivatives to inhibit the activity of DNA gyrase was experimentally determined using a gel electrophoresis assay. In addition, the structure of the derivatives bound to DNA gyrase was computationally modelled using molecular mechanics methods with the CHARMm forcefield optimized by molecular dynamics simulations. In both experimental and modelled systems, the lowest energy and highest binding affinity of the ciprofloxacin derivatives belonged to the c-gly-ciprofloxacin. Therefore, the ideal linker group was theorized to be 1-2 atoms long in order to prevent steric clashes and maximize interactions between the DNA gyrase enzyme and the citrate group of the ciprofloxacin derivative. Although the solvent and metal atoms in the system were not fully computationally modelled, the resulting structures were considered to be sufficient for a qualitative comparison between the ciprofloxacin derivatives. A sterically hindered derivative could be also used in order to further test the steric environment of the DNA gyrase binding site.

IcaB, a protein necessary for synthesis of active poly-N-acetylglucosamine (dPNAG) from *Staphylococcus epidermidis*, was chosen as a target for computational modelling as PNAG is an important factor in the formation of staphylococcal biofilms and IcaB had no structures available in the Protein Data Bank. A literature review of deacetylase enzymes was performed to identify key features required for proper modelling. Although deacetylase enzymes are a large superfamily of proteins, the majority were divided into two mechanistic families: a serine protease-like mechanism and a metal-dependent

mechanism using a catalytic acid and a catalytic base residue. The hydrophobic substrate-binding region and active site residues specific to the mechanism were determined to be essential for the activity of deacetylase enzymes.

The IcaB sequence from *S. epidermidis* was aligned with the sequence of PgaB, a protein with similar function from *E. coli* with an X-ray structure available. Using this alignment and the structure of the PgaB N-terminal domain, a homology model was constructed for IcaB. The model resembled PgaB with the exception of several small surface loops and strands. Importantly, the metal-binding and catalytic residues kept their orientation similar to that of PgaB, which would be necessary for proper deacetylase function. However, the binding groove involved in the transport of the PNAG to the active site of PgaB was not replicated in the IcaB model. Without a structure of IcaB from *S. epidermidis*, it is unclear whether this is a modelling error or whether the binding groove consists of differing amino acids.

After completion of this project, an X-ray structure of IcaB from *Ammonifex degensii* was achieved by Little et al. Despite the low sequence identity between the IcaB structure from *A. degensii* and the IcaB from *S. epidermidis* (22.8%), both structures were quite similar and the active site is within 2Å of the IcaB model residues. Therefore, a full comparison or use of the $IcaB_{AD}$ structure to produce a new homology model for IcaB from *S. epidermidis* would be of interest.

References

1. Williamson MP, Williamson M. How Proteins Work. Garland Science; 2012.

2. Stroud RM, Finer-Moore J, Royal Society of Chemistry. Computational and structural approaches to drug discovery [electronic resource] : ligand-protein interactions. Cambridge: Royal Society of Chemistry; 2008.

3. Martin SF, Clements JH. Correlating Structure and Energetics in Protein-Ligand Interactions: Paradigms and Paradoxes. Annu Rev Biochem. 2013;82: 267–293. doi:10.1146/annurev-biochem-060410-105819

4. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational Methods in Drug Discovery. Pharmacol Rev. 2014;66: 334–395. doi:10.1124/pr.112.007336

5. Dror O, Shulman-Peleg A, Nussinov R, Wolfson HJ. Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. Curr Med Chem. 2004;11: 71–90.

6. Schwede T. Protein Modelling: What Happened to the "Protein Structure Gap"? Struct Lond Engl 1993. 2013;21. doi:10.1016/j.str.2013.08.007

7. Lengauer T, Rarey M. Computational methods for biomolecular docking. Curr Opin Struct Biol. 1996;6: 402–406. doi:10.1016/S0959-440X(96)80061-3

8. Baker D, Sali A. Protein structure prediction and structural genomics. Science. 2001;294: 93–96. doi:10.1126/science.1065659

9. Dill KA, Ozkan SB, Shell MS, Weikl TR. The Protein Folding Problem. Annu Rev Biophys. 2008;37: 289–316. doi:10.1146/annurev.biophys.37.092707.153558

10. Dill KA, MacCallum JL. The Protein-Folding Problem, 50 Years On. Science. 2012;338: 1042–1046. doi:10.1126/science.1219021

11. Yan R, Wang X, Huang L, Yan F, Xue X, Cai W. Prediction of structural features and application to outer membrane protein identification. Sci Rep. 2015;5. doi:10.1038/srep11586

12. Semblat J-P, Ghumra A, Czajkowsky DM, Wallis R, Mitchell DA, Raza A, et al. Identification of the minimal binding region of a Plasmodium falciparum IgM binding PfEMP1 domain. Mol Biochem Parasitol. 2015;201: 76–82. doi:10.1016/j.molbiopara.2015.06.001

13. Ghitti M, Musco G, Spitaleri A. NMR and computational methods in the structural and dynamic characterization of ligand-receptor interactions. Adv Exp Med Biol. 2014;805: 271–304. doi:10.1007/978-3-319-02970-2_12

14. Jakubke H-D, Sewald N. Peptides from A to Z: A Concise Encyclopedia. John Wiley & Sons; 2008.

15. Prabakaran S, Lippens G, Steen H, Gunawardena J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. Wiley Interdiscip Rev Syst Biol Med. 2012;4: 565–583. doi:10.1002/wsbm.1185

16. Johnson LN. Glycogen phosphorylase: control by phosphorylation and allosteric effectors. FASEB J. 1992;6: 2274–2282.

17. Sevier CS, Kaiser CA. Formation and transfer of disulphide bonds in living cells. Nat Rev Mol Cell Biol. 2002;3: 836–847. doi:10.1038/nrm954

18. Mizzen CA, Allis CD. Linking histone acetylation to transcriptional regulation. Cell Mol Life Sci CMLS. 1998;54: 6–20. doi:10.1007/s000180050121

19. Chen D, Ma H, Hong H, Koh SS, Huang S-M, Schurter BT, et al. Regulation of Transcription by a Protein Methyltransferase. Science. 1999;284: 2174–2177. doi:10.1126/science.284.5423.2174

20. Helenius A, Aebi M. Roles of N-linked glycans in the endoplasmic reticulum. Annu Rev Biochem. 2004;73: 1019–1049. doi:10.1146/annurev.biochem.73.011303.073752

21. Baker EN, Hubbard RE. Hydrogen bonding in globular proteins. Prog Biophys Mol Biol. 1984;44: 97–179. doi:10.1016/0079-6107(84)90007-5

22. Nič M, Jirát J, Košata B, Jenkins A, McNaught A, editors. IUPAC Compendium of Chemical Terminology: Gold Book [Internet]. 2.1.0 ed. Research Triagle Park, NC: IUPAC; 2009. Available: http://goldbook.iupac.org/H02899.html

23. Pauling L, Corey RB. Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds. Proc Natl Acad Sci U S A. 1951;37: 729–740.

24. Pauling L, Corey RB. The Pleated Sheet, A New Layer Configuration of Polypeptide Chains. Proc Natl Acad Sci. 1951;37: 251–256.

25. Sibanda BL, Blundell TL, Thornton JM. Conformation of β-hairpins in protein structures: A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. J Mol Biol. 1989;206: 759–777. doi:10.1016/0022-2836(89)90583-4

26. Rose GD, Fleming PJ, Banavar JR, Maritan A. A backbone-based theory of protein folding. Proc Natl Acad Sci. 2006;103: 16623–16633. doi:10.1073/pnas.0606843103

27. Pace CN, Shirley BA, McNutt M, Gajiwala K. Forces contributing to the conformational stability of proteins. FASEB J. 1996;10: 75–83.

28. Dunker AK, Bondos SE, Huang F, Oldfield CJ. Intrinsically disordered proteins and multicellular organisms. Semin Cell Dev Biol. doi:10.1016/j.semcdb.2014.09.025

29. Perutz MF. Hemoglobin structure and respiratory transport. Sci Am. 1978;239: 92–125.

30. Komar AA, Kommer A, Krasheninnikov IA, Spirin AS. Cotranslational heme binding to nascent globin chains. FEBS Lett. 1993;326: 261–263. doi:10.1016/0014-5793(93)81803-8

31. Paoli M, Liddington R, Tame J, Wilkinson A, Dodson G. Crystal Structure of T State Haemoglobin with Oxygen Bound At All Four Haems. J Mol Biol. 1996;256: 775–792. doi:10.1006/jmbi.1996.0124

32. Monod J, Wyman J, Changeux JP. ON THE NATURE OF ALLOSTERIC TRANSITIONS: A PLAUSIBLE MODEL. J Mol Biol. 1965;12: 88–118.

33. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. Nature. 2014;508: 331–339. doi:10.1038/nature13001

34. Little DJ, Li G, Ing C, DiFrancesco BR, Bamford NC, Robinson H, et al. Modification and periplasmic translocation of the biofilm exopolysaccharide poly-β-1,6-N-acetyl-d-glucosamine. Proc Natl Acad Sci. 2014;111: 11013–11018. doi:10.1073/pnas.1406388111

35. Ames GFL. Bacterial Periplasmic Transport Systems: Structure, Mechanism, and Evolution. Annu Rev Biochem. 1986;55: 397–425. doi:10.1146/annurev.bi.55.070186.002145

36. Schneider M, Fu X, Keating AE. X-ray vs. NMR structures as templates for computational protein design. Proteins. 2009;77: 97–110. doi:10.1002/prot.22421

37. Bragg WL. The Analysis of Crystals by the X-ray Spectrometer. Proc R Soc Lond Math Phys Eng Sci. 1914;89: 468–489. doi:10.1098/rspa.1914.0015

38. Hammond C. The basics of crystallography and diffraction : International Union of Crystallography texts on crystallography 12 [Internet]. Oxford University Press; Available: https://www.dawsonera.com/readonline/9780191567711

39. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature. 1958;181: 662–666.

40. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28: 235–242. doi:10.1093/nar/28.1.235

41. Nič M, Jirát J, Košata B, Jenkins A, McNaught A, editors. IUPAC Compendium of Chemical Terminology: Gold Book [Internet]. 2.1.0 ed. Research Triagle Park, NC: IUPAC; 2009. Available: http://goldbook.iupac.org/I03275.html

42. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol Clifton NJ. 1999;112: 531–552.

43. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, et al. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis. 1993;14: 1023–1031.

44. Skovbjerg H, Sjöström H, Norén O. Purification and Characterisation of Amphiphilic Lactase/Phlorizin Hydrolase from Human Small Intestine. Eur J Biochem. 1981;114: 653–661. doi:10.1111/j.1432-1033.1981.tb05193.x

45. Guengerich FP. Cytochrome P450 and Chemical Toxicology. Chem Res Toxicol. 2008;21: 70–83. doi:10.1021/tx700079z

46. Yung-Chi C, Prusoff WH. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem Pharmacol. 1973;22: 3099–3108. doi:10.1016/0006-2952(73)90196-2

47. Huang Z, Zhu L, Cao Y, Wu G, Liu X, Chen Y, et al. ASD: a comprehensive database of allosteric proteins and modulators. Nucleic Acids Res. 2011;39: D663–D669. doi:10.1093/nar/gkq1022

48. Watson JD, Crick FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature. 1953;171: 737–738. doi:10.1038/171737a0

49. Helin K, Dhanak D. Chromatin proteins and modifications as drug targets. Nature. 2013;502: 480–488. doi:10.1038/nature12751

50. Kondo Y. Targeting histone methyltransferase EZH2 as cancer treatment. J Biochem (Tokyo). 2014;156: 249–257. doi:10.1093/jb/mvu054

51. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem. 1985;28: 849–857.

52. Zaki MJ, Bystroff C, editors. Protein Structure Prediction Using Threading - Springer. Humana Press; 2008. Available: http://link.springer.com.ezproxy.york.ac.uk/protocol/10.1007/978-1-59745-574-9_4

53. Piana S, Klepeis JL, Shaw DE. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. Curr Opin Struct Biol. 2014;24: 98–105. doi:10.1016/j.sbi.2013.12.006

54. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How Fast-Folding Proteins Fold. Science. 2011;334: 517–520. doi:10.1126/science.1208351

55. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. Science. 2010;330: 341–346. doi:10.1126/science.1187409

56. Bradley P, Misura KMS, Baker D. Toward high-resolution de novo structure prediction for small proteins. Science. 2005;309: 1868–1871. doi:10.1126/science.1113801

57. Mao B, Tejero R, Baker D, Montelione GT. Protein NMR structures refined with Rosetta have higher accuracy relative to corresponding X-ray crystal structures. J Am Chem Soc. 2014;136: 1893–1906. doi:10.1021/ja409845w

58. O'Meara MJ, Leaver-Fay A, Tyka M, Stein A, Houlihan K, DiMaio F, et al. A Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta. J Chem Theory Comput. 2015;11: 609–622. doi:10.1021/ct500864r

59. Moult J, Fidelis K, Kryshtafovych A, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)—round IX. Proteins Struct Funct Bioinforma. 2011;79: 1–5. doi:10.1002/prot.23200

60. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP) — round x. Proteins Struct Funct Bioinforma. 2014;82: 1–6. doi:10.1002/prot.24452

61. Monastyrskyy B, Andrea D D', Fidelis K, Tramontano A, Kryshtafovych A. Evaluation of residue-residue contact prediction in CASP10. Proteins. 2014;82: 138–153. doi:10.1002/prot.24340

62. Hillisch A, Pineda LF, Hilgenfeld R. Utility of homology models in the drug discovery process. Drug Discov Today. 2004;9: 659–669. doi:10.1016/S1359-6446(04)03196-4

63. Biesiada J, Porollo A, Velayutham P, Kouril M, Meller J. Survey of public domain software for docking simulations and virtual screening. Hum Genomics. 2011;5: 497. doi:10.1186/1479-7364-5-5-497

64. van der Kamp MW, Mulholland AJ. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. Biochemistry (Mosc). 2013;52: 2708–2728. doi:10.1021/bi400215w

65. Decherchi S, Masetti M, Vyalov I, Rocchia W. Implicit solvent methods for free energy estimation. Eur J Med Chem. 2015;91: 27–42. doi:10.1016/j.ejmech.2014.08.064

66. Massova I, Kollman PA. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspect Drug Discov Des. 2000;18: 113–135. doi:10.1023/A:1008763014207

67. Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers. 2000;56: 275–291. doi:10.1002/1097-0282(2000)56:4<275::AID-BIP10024>3.0.CO;2-E

68. Böhm HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des. 1994;8: 243–256.

69. Boger J, Lohr NS, Ulm EH, Poe M, Blaine EH, Fanelli GM, et al. Novel renin inhibitors containing the amino acid statine. Nature. 1983;303: 81–84.

70. Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. J Comput Chem. 1992;13: 505–524. doi:10.1002/jcc.540130412

71. Rarey M, Wefing S, Lengauer T. Placement of medium-sized molecular fragments into active sites of proteins. J Comput Aided Mol Des. 1996;10: 41–54.

72. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol. 1997;267: 727–748. doi:10.1006/jmbi.1996.0897

73. Harrison RL. Introduction To Monte Carlo Simulation. AIP Conf Proc. 2010;1204: 17–21. doi:10.1063/1.3295638

74. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, et al. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. PLoS Comput Biol. 2014;10: e1003571. doi:10.1371/journal.pcbi.1003571

75. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31: 455–461. doi:10.1002/jcc.21334

76. Klebe G, Mietzner T. A fast and efficient method to generate biologically relevant conformations. J Comput Aided Mol Des. 1994;8: 583–606.

77. Chen Y, Shoichet BK. Molecular docking and ligand specificity in fragment-based inhibitor discovery. Nat Chem Biol. 2009;5: 358–364. doi:10.1038/nchembio.155

78. Damm-Ganamet KL, Smith RD, Dunbar JB, Stuckey JA, Carlson HA. CSAR Benchmark Exercise 2011–2012: Evaluation of Results from Docking and Relative Ranking of Blinded Congeneric Series. J Chem Inf Model. 2013; doi:10.1021/ci400025f

79. Moal IH, Moretti R, Baker D, Fernández-Recio J. Scoring functions for protein–protein interactions. Curr Opin Struct Biol. 2013;23: 862–867. doi:10.1016/j.sbi.2013.06.017

80. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE. Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol. 2009;19: 120–127. doi:10.1016/j.sbi.2009.03.004

81. Adcock SA, McCammon JA. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. Chem Rev. 2006;106: 1589–1615. doi:10.1021/cr040426m

82. Senn HM, Thiel W. QM/MM Methods for Biological Systems. In: Reiher M, editor. Atomistic Approaches in Modern Biology. Springer Berlin Heidelberg; 2006. pp. 173–290. Available: http://link.springer.com/chapter/10.1007/128_2006_084

83. Eisen MB, Wiley DC, Karplus M, Hubbard RE. HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a

macromolecule binding site. Proteins. 1994;19: 199–221. doi:10.1002/prot.340190305

84. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. Protein Sci Publ Protein Soc. 2000;9: 1753–1773.

85. Rohl CA, Strauss CEM, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with rosetta. Proteins. 2004;55: 656–677. doi:10.1002/prot.10629

86. Rossi KA, Weigelt CA, Nayeem A, Krystek SR. Loopholes and missing links in protein modeling. Protein Sci Publ Protein Soc. 2007;16: 1999–2012. doi:10.1110/ps.072887807

87. Warshel A, Levitt M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol. 1976;103: 227–249. doi:10.1016/0022-2836(76)90311-9

88. Bax BD, Chan PF, Eggleston DS, Fosberry A, Gentry DR, Gorrec F, et al. Type IIA topoisomerase inhibition by a new class of antibacterial agents. Nature. 2010;466: 935–940. doi:10.1038/nature09197

89. Huang N, Kalyanaraman C, Bernacki K, Jacobson MP. Molecular mechanics methods for predicting protein–ligand binding. Phys Chem Chem Phys. 2006;8: 5166–5177. doi:10.1039/B608269F

90. Šali A, Blundell TL. Comparative Protein Modelling by Satisfaction of Spatial Restraints. J Mol Biol. 1993;234: 779–815. doi:10.1006/jmbi.1993.1626

91. Nayeem A, Sitkoff D, Krystek S. A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. Protein Sci Publ Protein Soc. 2006;15: 808–824. doi:10.1110/ps.051892906

92. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014;42: W252–W258. doi:10.1093/nar/gku340

93. Consortium TU. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43: D204–D212. doi:10.1093/nar/gku989

94. Pieper U, Webb BM, Dong GQ, Schneidman-Duhovny D, Fan H, Kim SJ, et al. ModBase, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res. 2014;42: D336–D346. doi:10.1093/nar/gkt1144

95. Schmidt T, Bergner A, Schwede T. Modelling three-dimensional protein structures for applications in drug design. Drug Discov Today. 2014;19: 890–897. doi:10.1016/j.drudis.2013.10.027

96. Høiby N. Recent advances in the treatment of Pseudomonas aeruginosa infections in cystic fibrosis. BMC Med. 2011;9: 32. doi:10.1186/1741-7015-9-32

97. Tümmler B, Kiewitz C. Cystic fibrosis: an inherited susceptibility to bacterial respiratory infections. Mol Med Today. 1999;5: 351–358. doi:10.1016/S1357-4310(99)01506-3

98. Boucher RC. New Concepts of the Pathogenesis of Cystic Fibrosis Lung Disease. Eur Respir J. 2004;23: 146–158. doi:10.1183/09031936.03.00057003

99. Hansen CR, Pressler T, Høiby N. Early aggressive eradication therapy for intermittent Pseudomonas aeruginosa airway colonization in cystic fibrosis patients: 15 years experience. J Cyst Fibros. 2008;7: 523–530. doi:10.1016/j.jcf.2008.06.009

100. Geller DE. Aerosol antibiotics in cystic fibrosis. Respir Care. 2009;54: 658–670.

101. Henwood CJ, Livermore DM, James D, Warner M, Pseudomonas Study Group. Antimicrobial susceptibility of Pseudomonas aeruginosa: results of a UK survey and evaluation of the British Society for Antimicrobial Chemotherapy disc susceptibility test. J Antimicrob Chemother. 2001;47: 789–799.

102. Pitt T, Sparrow M, Warner M, Stefanidou M. Survey of resistance of Pseudomonas aeruginosa from UK patients with cystic fibrosis to six commonly prescribed antimicrobial agents. Thorax. 2003;58: 794–796. doi:10.1136/thorax.58.9.794

103. Stewart PS, William Costerton J. Antibiotic resistance of bacteria in biofilms. The Lancet. 2001;358: 135–138. doi:10.1016/S0140-6736(01)05321-1

104. Costerton JW, Stewart PS, Greenberg EP. Bacterial biofilms: a common cause of persistent infections. Science. 1999;284: 1318–1322.

105. Silva MT. Staphylococcus aureus Biofilms Prevent Macrophage Phagocytosis and Attenuate Inflammation In Vivo. J Leukoc Biol. 2010;89: 675–683. doi:10.1189/jlb.0910536

106. Keren I, Kaldalu N, Spoering A, Wang Y, Lewis K. Persister cells and tolerance to antimicrobials. FEMS Microbiol Lett. 2004;230: 13–18.

107. Shapiro JA, Nguyen VL, Chamberlain NR. Evidence for persisters in Staphylococcus epidermidis RP62a planktonic cultures and biofilms. J Med Microbiol. 2011;60: 950–960. doi:10.1099/jmm.0.026013-0

108. Potera C. Forging a Link Between Biofilms and Disease. Science. 1999;283: 1837–1839. doi:10.1126/science.283.5409.1837

109. Walsh C. Molecular mechanisms that confer antibacterial drug resistance. Nature. 2000;406: 775–781. doi:10.1038/35021219

110. Lambert PA. Mechanisms of antibiotic resistance in Pseudomonas aeruginosa. J R Soc Med. 2002;95: 22–26.

111. Wolfson JS, Hooper DC. Fluoroquinolone antimicrobial agents. Clin Microbiol Rev. 1989;2: 378–424.

112. Hirai K, Aoyama H, Irikura T, Iyobe S, Mitsuhashi S. Differences in susceptibility to quinolones of outer membrane mutants of Salmonella typhimurium and Escherichia coli. Antimicrob Agents Chemother. 1986;29: 535.

113. Dalhoff A. Resistance surveillance studies: a multifaceted problem—the fluoroquinolone example. Infection. 2012;40: 239–262. doi:10.1007/s15010-012-0257-2

114. Emerson J, McNamara S, Buccat AM, Worrell K, Burns JL. Changes in cystic fibrosis sputum microbiology in the United States between 1995 and 2008. Pediatr Pulmonol. 2010;45: 363–370. doi:10.1002/ppul.21198

115. Antonela Antoniu S. Inhaled ciprofloxacin for chronic airways infections caused by Pseudomonas aeruginosa. Expert Rev Anti Infect Ther. 2012;10: 1439–1446. doi:10.1586/eri.12.136

116. Rukhshanda S, Sultana N, Khalid M, Nosheen A, Arayne MS. Effect of Skeletal Modification of Ciprofloxacin on Antibacterial, Antifungal, and Cytotoxic Activities. . J Chin Clin Med. 2007;2: 188–195.

117. Md-Saleh SR, Chilvers EC, Kerr KG, Milner SJ, Snelling AM, Weber JP, et al. Synthesis of citrate–ciprofloxacin conjugates. Bioorg Med Chem Lett. 2009;19: 1496–1498. doi:10.1016/j.bmcl.2009.01.007

118. Zheng T, Bullock JL, Nolan EM. Siderophore-Mediated Cargo Delivery to the Cytoplasm of Escherichia coli and Pseudomonas aeruginosa: Syntheses of Monofunctionalized Enterobactin Scaffolds and Evaluation of Enterobactin–Cargo Conjugate Uptake. J Am

Chem Soc. 2012;134: 18388–18400. doi:10.1021/ja3077268

119. Wencewicz TA, Möllmann U, Long TE, Miller MJ. Is drug release necessary for antimicrobial activity of siderophore-drug conjugates? Syntheses and biological studies of the naturally occurring salmycin "Trojan Horse" antibiotics and synthetic desferridanoxamine-antibiotic conjugates. BioMetals. 2009;22: 633–648. doi:10.1007/s10534-009-9218-3

120. Milner SJ, Carrick CT, Kerr KG, Snelling AM, Thomas GH, Duhme-Klair A-K, et al. Probing Bacterial Uptake of Glycosylated Ciprofloxacin Conjugates. ChemBioChem. 2014;15: 466–471. doi:10.1002/cbic.201300512

121. Zheng H, Taraska J, Merz AJ, Gonen T. The Prototypical H+/Galactose Symporter GalP Assembles into Functional Trimers. J Mol Biol. 2010;396: 593–601. doi:10.1016/j.jmb.2009.12.010

122. Doige CA, Luzzi Ames GF. ATP-Dependent Transport Systems in Bacteria and Humans: Relevance to Cystic Fibrosis and Multidrug Resistance. Annu Rev Microbiol. 1993;47: 291–319. doi:10.1146/annurev.mi.47.100193.001451

123. Shen LL, Mitscher LA, Sharma PN, O'Donnell TJ, Chu DWT, Cooper CS, et al. Mechanism of inhibition of DNA gyrase by quinolone antibacterials: a cooperative drug-DNA binding model. Biochemistry (Mosc). 1989;28: 3886–3894. doi:10.1021/bi00435a039

124. Drlica K, Hiasa H, Kerns R, Malik M, Mustaev A, Zhao X. Quinolones: Action and Resistance Updated. Curr Top Med Chem. 2009;9: 981–998. doi:10.2174/156802609789630947

125. Milner SJ, Snelling AM, Kerr KG, Abd-El-Aziz A, Thomas GH, Hubbard RE, et al. Probing linker design in citric acid–ciprofloxacin conjugates. Bioorg Med Chem. 2014;22: 4499–4505. doi:10.1016/j.bmc.2014.04.009

126. Wagegg W, Braun V. Ferric citrate transport in Escherichia coli requires outer membrane receptor protein fecA. J Bacteriol. 1981;145: 156–163.

127. Yue WW, Grizot S, Buchanan SK. Structural Evidence for Iron-free Citrate and Ferric Citrate Binding to the TonB-dependent Outer Membrane Transporter FecA. J Mol Biol. 2003;332: 353–368. doi:10.1016/S0022-2836(03)00855-6

128. Fernandes F, Neves P, Gameiro P, Loura LMS, Prieto M. Ciprofloxacin interactions with bacterial protein OmpF: Modelling of FRET from a multi-tryptophan protein trimer. Biochim Biophys Acta BBA - Biomembr. 2007;1768: 2822–2830. doi:10.1016/j.bbamem.2007.07.016

129. Ferguson AD, Chakraborty R, Smith BS, Esser L, van der Helm D, Deisenhofer J. Structural Basis of Gating by the Outer Membrane Transporter FecA. Science. 2002;295: 1715–1719. doi:10.1126/science.1067313

130. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23: 2947–2948. doi:10.1093/bioinformatics/btm404

131. Milner SJ, Seve A, Snelling AM, Thomas GH, Kerr KG, Routledge A, et al. Staphyloferrin A as siderophore-component in fluoroquinolone-based Trojan horse antibiotics. Org Biomol Chem. 2013;11: 3461–3468. doi:10.1039/C3OB40162F

132. Maxwell A, Burton NP, O'Hagan N. High-throughput assays for DNA gyrase and other topoisomerases. Nucleic Acids Res. 2006;34: e104. doi:10.1093/nar/gkl504

133. Abramoff M d., Magalhães PJ, Ram SJ. Image processing with ImageJ. Biophotonics Int. 2004;11: 36–42.

134. Brooks BR, Brooks III CL, Mackerell Jr. AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: The biomolecular simulation program. J Comput Chem. 2009;30: 1545–1614. doi:10.1002/jcc.21287

135. Discovery Studio Modeling Environment, Release 3.0. San Diego Accelrys Softw Inc. 2007;

136. Milner SJ, Snelling AM, Kerr KG, Abd-El-Aziz A, Thomas GH, Hubbard RE, et al. Probing linker design in citric acid–ciprofloxacin conjugates. Bioorg Med Chem. doi:10.1016/j.bmc.2014.04.009

137. Duarte F, Bauer P, Barrozo A, Amrein BA, Purg M, Åqvist J, et al. Force Field Independent Metal Parameters Using a Nonbonded Dummy Model. J Phys Chem B. 2014;118: 4351–4362. doi:10.1021/jp501737x

138. Pokrovskaya V, Little DJ, Poloczek J, Griffiths H, Howell PL, Nitz M. Functional Characterization of Staphylococcus epidermidis IcaB, a de-N-acetylase important for biofilm formation. Biochemistry (Mosc). 2013; doi:10.1021/bi400836g

139. Caufrier F, Martinou A, Dupont C, Bouriotis V. Carbohydrate esterase family 4 enzymes: substrate specificity. Carbohydr Res. 2003;338: 687–692. doi:10.1016/S0008-6215(03)00002-8

140. de RUIJTER AJM, van GENNIP AH, Caron HN, Kemp S, van KUILENBURG ABP. Histone deacetylases (HDACs): characterization of the classical HDAC family. Biochem J.

2003;370: 737. doi:10.1042/BJ20021321

141. Guedes-Dias P, Oliveira JMA. Lysine deacetylases and mitochondrial dynamics in neurodegeneration. Biochim Biophys Acta. 2013;1832: 1345–1359. doi:10.1016/j.bbadis.2013.04.005

142. Little DJ, Poloczek J, Whitney JC, Robinson H, Nitz M, Howell PL. The Structure- and Metal-dependent Activity of Escherichia coli PgaB Provides Insight into the Partial De-N-acetylation of Poly-β-1,6-N-acetyl-d-glucosamine. J Biol Chem. 2012;287: 31126–31137. doi:10.1074/jbc.M112.390005

143. Zhao Y, Park R-D, Muzzarelli RAA. Chitin Deacetylases: Properties and Applications. Mar Drugs. 2010;8: 24–46. doi:10.3390/md8010024

144. Mack D, Fischer W, Krokotsch A, Leopold K, Hartmann R, Egge H, et al. The intercellular adhesin involved in biofilm accumulation of Staphylococcus epidermidis is a linear beta-1,6-linked glucosaminoglycan: purification and structural analysis. J Bacteriol. 1996;178: 175–183.

145. Domenech M, García E, Prieto A, Moscoso M. Insight into the composition of the intercellular matrix of Streptococcus pneumoniae biofilms. Environ Microbiol. 2013;15: 502–516. doi:10.1111/j.1462-2920.2012.02853.x

146. Vollmer W, Tomasz A. The pgdA Gene Encodes for a PeptidoglycanN-Acetylglucosamine Deacetylase in Streptococcus pneumoniae. J Biol Chem. 2000;275: 20496–20501. doi:10.1074/jbc.M910189199

147. Vollmer W. Structural variation in the glycan strands of bacterial peptidoglycan. FEMS Microbiol Rev. 2008;32: 287–306. doi:10.1111/j.1574-6976.2007.00088.x

148. Walkinshaw DR, Tahmasebi S, Bertos NR, Yang X-J. Histone deacetylases as transducers and targets of nuclear signaling. J Cell Biochem. 2008;104: 1541–1552. doi:10.1002/jcb.21746

149. Shahbazian MD, Grunstein M. Functions of Site-Specific Histone Acetylation and Deacetylation. Annu Rev Biochem. 2007;76: 75–100. doi:10.1146/annurev.biochem.76.052705.162114

150. Rundlett SE, Carmen AA, Kobayashi R, Bavykin S, Turner BM, Grunstein M. HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. Proc Natl Acad Sci. 1996;93: 14503–14508.

151. Jiang H, Khan S, Wang Y, Charron G, He B, Sebastian C, et al. SIRT6 regulates TNF-α secretion through hydrolysis of long-chain fatty acyl lysine. Nature. 2013;496: 110–113. doi:10.1038/nature12038

152. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009;37: D233–D238. doi:10.1093/nar/gkn663

153. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42: D490–D495. doi:10.1093/nar/gkt1178

154. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res. 2005;33: D34–D38. doi:10.1093/nar/gki063

155. Warrell RP, He L-Z, Richon V, Calleja E, Pandolfi PP. Therapeutic Targeting of Transcription in Acute Promyelocytic Leukemia by Use of an Inhibitor of Histone Deacetylase. J Natl Cancer Inst. 1998;90: 1621–1625. doi:10.1093/jnci/90.21.1621

156. Wade PA. Transcriptional control at regulatory checkpoints by histone deacetylases: molecular connections between cancer and chromatin. Hum Mol Genet. 2001;10: 693–698. doi:10.1093/hmg/10.7.693

157. Wang J, Elahi A, Ajidahun A, Clark W, Hernandez J, Achille A, et al. The interplay between histone deacetylases and c-Myc in the transcriptional suppression of HPP1 in colon cancer. Cancer Biol Ther. 2014;15: 0–1.

158. Bottomley MJ, Surdo P Lo, Di Giovine P, Cirillo A, Scarpelli R, Ferrigno F, et al. Structural and Functional Analysis of the Human HDAC4 Catalytic Domain Reveals a Regulatory Structural Zinc-binding Domain. J Biol Chem. 2008;283: 26694–26704. doi:10.1074/jbc.M803514200

159. Lombardi PM, Cole KE, Dowling DP, Christianson DW. Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes. Curr Opin Struct Biol. 2011;21: 735–743. doi:10.1016/j.sbi.2011.08.004

160. Sanders BD, Jackson B, Marmorstein R. Structural Basis for Sirtuin Function: What We Know and What We Don't. Biochim Biophys Acta. 2010;1804: 1604–1616. doi:10.1016/j.bbapap.2009.09.009

161. Hirsch BM, Burgos ES, Schramm VL. Transition-State Analysis of 2-O-Acetyl-ADP-Ribose Hydrolysis by Human Macrodomain 1. ACS Chem Biol. 2014;9: 2255–2262. doi:10.1021/cb500485w

162. Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B. A hierarchical classification of polysaccharide lyases for glycogenomics. Biochem J. 2010;432: 437–444. doi:10.1042/BJ20101185

163. Kraut J. Serine Proteases: Structure and Mechanism of Catalysis. Annu Rev Biochem. 1977;46: 331–358. doi:10.1146/annurev.bi.46.070177.001555

164. Ronning DR, Vissa V, Besra GS, Belisle JT, Sacchettini JC. Mycobacterium tuberculosis Antigen 85A and 85C Structures Confirm Binding Orientation and Conserved Substrate Specificity. J Biol Chem. 2004;279: 36771–36777. doi:10.1074/jbc.M400811200

165. DeBoy RT, Mongodin EF, Fouts DE, Tailford LE, Khouri H, Emerson JB, et al. Insights into Plant Cell Wall Degradation from the Genome Sequence of the Soil Bacterium Cellvibrio japonicus. J Bacteriol. 2008;190: 5455–5463. doi:10.1128/JB.01701-07

166. Montanier C, Money VA, Pires VMR, Flint JE, Pinheiro BA, Goyal A, et al. The Active Site of a Carbohydrate Esterase Displays Divergent Catalytic and Noncatalytic Binding Functions. PLoS Biol. 2009;7. doi:10.1371/journal.pbio.1000071

167. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, et al. The Path Forward for Biofuels and Biomaterials. Science. 2006;311: 484–489. doi:10.1126/science.1114736

168. Correia MAS, Prates JAM, Brás J, Fontes CMGA, Newman JA, Lewis RJ, et al. Crystal Structure of a Cellulosomal Family 3 Carbohydrate Esterase from Clostridium thermocellum Provides Insights into the Mechanism of Substrate Recognition. J Mol Biol. 2008;379: 64–72. doi:10.1016/j.jmb.2008.03.037

169. John M, Röhrig H, Schmidt J, Wieneke U, Schell J. Rhizobium NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. Proc Natl Acad Sci U S A. 1993;90: 625.

170. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. Nucleic Acids Res. 2011;40: D290–D301. doi:10.1093/nar/gkr1065

171. Ifuku S. Chitin and Chitosan Nanofibers: Preparation and Chemical Modifications. Molecules. 2014;19: 18367–18380. doi:10.3390/molecules191118367

172. Kafetzopoulos D, Martinou A, Bouriotis V. Bioconversion of chitin to chitosan: purification and characterization of chitin deacetylase from Mucor rouxii. Proc Natl Acad Sci U S A. 1993;90: 2564–2568.

173. Tsigos I, Martinou A, Kafetzopoulos D, Bouriotis V. Chitin deacetylases: new, versatile tools in biotechnology. Trends Biotechnol. 2000;18: 305–312. doi:10.1016/S0167-7799(00)01462-1

174. Benachour A, Ladjouzi R, Le Jeune A, Hébert L, Thorpe S, Courtin P, et al. The Lysozyme-Induced Peptidoglycan N-Acetylglucosamine Deacetylase PgdA (EF1843) Is Required for Enterococcus faecalis Virulence. J Bacteriol. 2012;194: 6066–6073. doi:10.1128/JB.00981-12

175. Blair DE, Schüttelkopf AW, MacRae JI, van Aalten DMF. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. Proc Natl Acad Sci U S A. 2005;102: 15429–15434. doi:10.1073/pnas.0504339102

176. Mochalkin I, Knafels JD, Lightle S. Crystal structure of LpxC from Pseudomonas aeruginosa complexed with the potent BB-78485 inhibitor. Protein Sci Publ Protein Soc. 2008;17: 450–457. doi:10.1110/ps.073324108

177. Kobayashi K, Sudiarta IP, Kodama T, Fukushima T, Ara K, Ozaki K, et al. Identification and Characterization of a Novel Polysaccharide Deacetylase C (PdaC) from Bacillus subtilis. J Biol Chem. 2012;287: 9765–9776. doi:10.1074/jbc.M111.329490

178. Andrés E, Albesa-Jové D, Biarnés X, Moerschbacher BM, Guerin ME, Planas A. Structural Basis of Chitin Oligosaccharide Deacetylation. Angew Chem Int Ed. 2014;53: 6882–6887. doi:10.1002/anie.201400220

179. Keyhani NO, Roseman S. Wild-type Escherichia coli grows on the chitin disaccharide, N,N′-diacetylchitobiose, by expressing the cel operon. Proc Natl Acad Sci U S A. 1997;94: 14367–14371.

180. Keyhani NO, Bacia K, Roseman S. The transport/phosphorylation of N,N'-diacetylchitobiose in Escherichia coli. Characterization of phospho-IIB(Chb) and of a potential transition state analogue in the phosphotransfer reaction between the proteins IIA(Chb) AND IIB(Chb). J Biol Chem. 2000;275: 33102–33109. doi:10.1074/jbc.M001045200

181. Plumbridge J, Pellegrini O. Expression of the chitobiose operon of Escherichia coli is regulated by three transcription factors: NagC, ChbR and CAP. Mol Microbiol. 2004;52: 437–449. doi:10.1111/j.1365-2958.2004.03986.x

182. Thompson J, Ruvinov SB, Freedberg DI, Hall BG. Cellobiose-6-Phosphate Hydrolase (CelF) of Escherichia coli: Characterization and Assignment to the Unusual Family 4 of Glycosylhydrolases. J Bacteriol. 1999;181: 7339–7345.

183. Verma SC, Mahadevan S. The chbG Gene of the Chitobiose (chb) Operon of Escherichia coli Encodes a Chitooligosaccharide Deacetylase. J Bacteriol. 2012;194: 4959–4971. doi:10.1128/JB.00533-12

184. Vuong C, Voyich JM, Fischer ER, Braughton KR, Whitney AR, DeLeo FR, et al. Polysaccharide intercellular adhesin (PIA) protects Staphylococcus epidermidis against major components of the human innate immune system. Cell Microbiol. 2004;6: 269–275. doi:10.1046/j.1462-5822.2004.00367.x

185. Chibba A, Poloczek J, Little DJ, Howell PL, Nitz M. Synthesis and evaluation of inhibitors of E. coli PgaB, a polysaccharide de-N-acetylase involved in biofilm formation. Org Biomol Chem. 2012;10: 7103–7107. doi:10.1039/c2ob26105g

186. Wang X, Preston JF, Romeo T. The pgaABCD Locus of Escherichia coli Promotes the Synthesis of a Polysaccharide Adhesin Required for Biofilm Formation. J Bacteriol. 2004;186: 2724–2734. doi:10.1128/JB.186.9.2724-2734.2004

187. Udenfriend S, Stein S, Böhlen P, Dairman W, Leimgruber W, Weigele M. Fluorescamine: a reagent for assay of amino acids, peptides, proteins, and primary amines in the picomole range. Science. 1972;178: 871–872.

188. Mitsushima K, Takimoto A, Sonoyama T, Yagi S. Gene cloning, nucleotide sequence, and expression of a cephalosporin-C deacetylase from Bacillus subtilis. Appl Environ Microbiol. 1995;61: 2224–2229.

189. Vincent F, Charnock SJ, Verschueren KHG, Turkenburg JP, Scott DJ, Offen WA, et al. Multifunctional Xylooligosaccharide/Cephalosporin C Deacetylase Revealed by the Hexameric Structure of the Bacillus subtilis Enzyme at 1.9 Å Resolution. J Mol Biol. 2003;330: 593–606. doi:10.1016/S0022-2836(03)00632-6

190. Micheli F. Pectin methylesterases: cell wall enzymes with important roles in plant physiology. Trends Plant Sci. 2001;6: 414–419. doi:10.1016/S1360-1385(01)02045-3

191. Boraston AB, Abbott DW. Structure of a pectin methylesterase from Yersinia enterocolitica. Acta Crystallograph Sect F Struct Biol Cryst Commun. 2012;68: 129–133. doi:10.1107/S1744309111055400

192. Fries M, Ihrig J, Brocklehurst K, Shevchik VE, Pickersgill RW. Molecular basis of the activity of the phytopathogen pectin methylesterase. EMBO J. 2007;26: 3879–3887. doi:10.1038/sj.emboj.7601816

193. Jenkins J, Pickersgill R. The architecture of parallel beta-helices and related folds. Prog Biophys Mol Biol. 2001;77: 111–175.

194. Vincent F, Yates D, Garman E, Davies GJ, Brannigan JA. The Three-dimensional Structure of the N-Acetylglucosamine-6-phosphate Deacetylase, NagA, from Bacillus subtilis A MEMBER OF THE UREASE SUPERFAMILY. J Biol Chem. 2004;279: 2809–2816. doi:10.1074/jbc.M310165200

195. Yadav V, Panilaitis B, Shi H, Numuta K, Lee K, Kaplan DL. N-acetylglucosamine 6-Phosphate Deacetylase (nagA) Is Required for N-acetyl Glucosamine Assimilation in Gluconacetobacter xylinus. Neves NM, editor. PLoS ONE. 2011;6: e18099. doi:10.1371/journal.pone.0018099

196. Park JT. Identification of a Dedicated Recycling Pathway for Anhydro-N-Acetylmuramic Acid andN-Acetylglucosamine Derived from Escherichia coli Cell Wall Murein. J Bacteriol. 2001;183: 3842–3847. doi:10.1128/JB.183.13.3842-3847.2001

197. CAZy - CE10 [Internet]. [cited 8 Jun 2015]. Available: http://www.cazy.org/CE10_structure.html

198. Aldridge WN, Davison AN. The mechanism of inhibition of cholinesterases by organophosphorus compounds. Biochem J. 1953;55: 763–766.

199. Vaara M. Outer membrane permeability barrier to azithromycin, clarithromycin, and roxithromycin in gram-negative enteric bacteria. Antimicrob Agents Chemother. 1993;37: 354–356.

200. Gou J-Y, Miller LM, Hou G, Yu X-H, Chen X-Y, Liu C-J. Acetylesterase-Mediated Deacetylation of Pectin Impairs Cell Elongation, Pollen Germination, and Plant Reproduction. Plant Cell Online. 2012; doi:10.1105/tpc.111.092411

201. Shevchik VE, Hugouvieux-Cotte-Pattat N. PaeX, a second pectin acetylesterase of Erwinia chrysanthemi 3937. J Bacteriol. 2003;185: 3091–3100.

202. Newton GL, Av-Gay Y, Fahey RC. N-Acetyl-1-D-myo-inosityl-2-amino-2-deoxy-alpha-D-glucopyranoside deacetylase (MshB) is a key enzyme in mycothiol biosynthesis. J Bacteriol. 2000;182: 6958–6963.

203. Maynes JT, Garen C, Cherney MM, Newton G, Arad D, Av-Gay Y, et al. The Crystal Structure of 1-D-myo-Inosityl 2-Acetamido-2-deoxy-α-D-glucopyranoside Deacetylase (MshB) from Mycobacterium tuberculosis Reveals a Zinc Hydrolase with a Lactate Dehydrogenase Fold. J Biol Chem. 2003;278: 47166–47170. doi:10.1074/jbc.M308914200

204. Broadley SG, Gumbart JC, Weber BW, Marakalala MJ, Steenkamp DJ, Sewell BT. A new crystal form of MshB from Mycobacterium tuberculosis with glycerol and acetate in the active site suggests the catalytic mechanism. Acta Crystallogr D Biol Crystallogr. 2012;68:

1450–1459. doi:10.1107/S090744491203449X

205. Li X-L, Špániková S, de Vries RP, Biely P. Identification of genes encoding microbial glucuronoyl esterases. FEBS Lett. 2007;581: 4029–4035. doi:10.1016/j.febslet.2007.07.041

206. Li X-L, Skory CD, Cotta MA, Puchart V, Biely P. Novel Family of Carbohydrate Esterases, Based on Identification of the Hypocrea jecorina Acetyl Esterase Gene. Appl Environ Microbiol. 2008;74: 7482–7489. doi:10.1128/AEM.00807-08

207. Shi D, Yu X, Roth L, Tuchman M, Allewell NM. Structure of a novel N-acetyl-l-citrulline deacetylase from Xanthomonas campestris. Biophys Chem. 2007;126: 86–93. doi:10.1016/j.bpc.2006.05.013

208. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. Nucleic Acids Res. 2013;41: D344–D347. doi:10.1093/nar/gks1067

209. Wojtyczka RD, Dziedzic A, Kępa M, Kubina R, Kabała-Dzik A, Mularz T, et al. Berberine Enhances the Antibacterial Activity of Selected Antibiotics against Coagulase-Negative Staphylococcus Strains in Vitro. Molecules. 2014;19: 6583–6596. doi:10.3390/molecules19056583

210. Otto M. Staphylococcus epidermidis--the “accidental” pathogen. Nat Rev Microbiol. 2009;7: 555–567. doi:10.1038/nrmicro2182

211. Dryden MS. Skin and soft tissue infection: microbiology and epidemiology. Int J Antimicrob Agents. 2009;34, Supplement 1: S2–S7. doi:10.1016/S0924-8579(09)70541-2

212. Raad I, Hanna H, Maki D. Intravascular catheter-related infections: advances in diagnosis, prevention, and management. Lancet Infect Dis. 2007;7: 645–657. doi:10.1016/S1473-3099(07)70235-9

213. Choices NHS. Staphylococcal Infections - NHS Choices [Internet]. 29 Aug 2013 [cited 25 May 2014]. Available: http://www.nhs.uk/Conditions/Staphylococcal-infections/Pages/Introduction.aspx

214. Cheung GY, Otto M. Understanding the significance of Staphylococcus epidermidis bacteremia in babies and children. Curr Opin Infect Dis. 2010;23: 208–216. doi:10.1097/QCO.0b013e328337fecb

215. Uçkay I, Pittet D, Vaudaux P, Sax H, Lew D, Waldvogel F. Foreign body infections due to Staphylococcus epidermidis. Ann Med. 2009;41: 109–119. doi:10.1080/07853890802337045

216. Otto M. Molecular basis of Staphylococcus epidermidis infections. Semin Immunopathol. 2012;34: 201–214. doi:10.1007/s00281-011-0296-2

217. Rohde H, Burandt EC, Siemssen N, Frommelt L, Burdelski C, Wurster S, et al. Polysaccharide intercellular adhesin or protein factors in biofilm accumulation of Staphylococcus epidermidis and Staphylococcus aureus isolated from prosthetic hip and knee joint infections. Biomaterials. 2007;28: 1711–1720. doi:10.1016/j.biomaterials.2006.11.046

218. Gerke C, Kraft A, Süßmuth R, Schweitzer O, Götz F. Characterization of theN-Acetylglucosaminyltransferase Activity Involved in the Biosynthesis of the StaphylococcusepidermidisPolysaccharide Intercellular Adhesin. J Biol Chem. 1998;273: 18586–18593. doi:10.1074/jbc.273.29.18586

219. Atkin KE, MacDonald SJ, Brentnall AS, Potts JR, Thomas GH. A different path: Revealing the function of staphylococcal proteins in biofilm formation. FEBS Lett. 2014;588: 1869–1872. doi:10.1016/j.febslet.2014.04.002

220. Vuong C, Kocianova S, Voyich JM, Yao Y, Fischer ER, DeLeo FR, et al. A Crucial Role for Exopolysaccharide Modification in Bacterial Biofilm Formation, Immune Evasion, and Virulence. J Biol Chem. 2004;279: 54881–54886. doi:10.1074/jbc.M411374200

221. Nishiyama T, Noguchi H, Yoshida H, Park S-Y, Tame JRH. The structure of the deacetylase domain of Escherichia coli PgaB, an enzyme required for biofilm formation: a circularly permuted member of the carbohydrate esterase 4 family. Acta Crystallogr D Biol Crystallogr. 2012;69: 44–51. doi:10.1107/S0907444912042059

222. Deng DM, Urch JE, Cate JM ten, Rao VA, van Aalten DMF, Crielaard W. Streptococcus mutans SMU.623c Codes for a Functional, Metal-Dependent Polysaccharide Deacetylase That Modulates Interactions with Salivary Agglutinin. J Bacteriol. 2009;191: 394–402. doi:10.1128/JB.00838-08

223. Itoh Y, Rice JD, Goller C, Pannuri A, Taylor J, Meisner J, et al. Roles of pgaABCD Genes in Synthesis, Modification, and Export of the Escherichia coli Biofilm Adhesin Poly-β-1,6-N-Acetyl-d-Glucosamine. J Bacteriol. 2008;190: 3670–3680. doi:10.1128/JB.01920-07

224. Oyama T, Miyake H, Kusunoki M, Nitta Y. Crystal Structures of β-Amylase from *Bacillus cereus* var. *mycoides* in Complexes with Substrate Analogs and Affinity-Labeling Reagents. J Biochem (Tokyo). 2003;133: 467–474.

225. Higgins DG, Sharp PM. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene. 1988;73: 237–244. doi:10.1016/0378-1119(88)90330-7

226. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2009;37: D5–15. doi:10.1093/nar/gkn741

227. Bassolino-Klimas D, Tejero R, Krystek SR, Metzler WJ, Montelione GT, Bruccoleri RE. Simulated annealing with restrained molecular dynamics using a flexible restraint potential: theory and evaluation with simulated NMR constraints. Protein Sci Publ Protein Soc. 1996;5: 593–603.

228. Shen M, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci Publ Protein Soc. 2006;15: 2507–2524. doi:10.1110/ps.062416606

229. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22: 2577–2637. doi:10.1002/bip.360221211

230. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999;292: 195–202. doi:10.1006/jmbi.1999.3091

231. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25: 3389–3402.

232. Benner SA, Gerloff D. Patterns of divergence in homologous proteins as indicators of secondary and tertiary structure: A prediction of the structure of the catalytic domain of protein kinases. Adv Enzyme Regul. 1991;31: 121–181. doi:10.1016/0065-2571(91)90012-B

233. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000;16: 404–405. doi:10.1093/bioinformatics/16.4.404

234. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A. Comparative Protein Structure Modeling of Genes and Genomes. Annu Rev Biophys Biomol Struct. 2000;29: 291–325. doi:10.1146/annurev.biophys.29.1.291

235. Wada A, Nakamura H. Nature of the charge distribution in proteins. Nature. 1981;293: 757–758. doi:10.1038/293757a0

236. Eisenberg D, Weiss RM, Terwilliger TC. The hydrophobic moment detects periodicity in protein hydrophobicity. Proc Natl Acad Sci U S A. 1984;81: 140–144.

237. Little DJ, Bamford NC, Podrovskaya V, Robinson H, Nitz M, Howell PL. Structural basis for the de-N-acetylation of poly-β-1,6-N-acetyl-D-glucosamine in Gram-positive bacteria. J Biol Chem. 2014; jbc.M114.611400. doi:10.1074/jbc.M114.611400

238. Barril X, Brough P, Drysdale M, Hubbard RE, Massey A, Surgenor A, et al. Structure-based discovery of a new class of Hsp90 inhibitors. Bioorg Med Chem Lett. 2005;15: 5187–5191. doi:10.1016/j.bmcl.2005.08.092

239. Konc J, Lešnik S, Janežič D. Modeling enzyme-ligand binding in drug discovery. J Cheminformatics. 2015;7: 48. doi:10.1186/s13321-015-0096-0

240. Forli S. Charting a Path to Success in Virtual Screening. Mol Basel Switz. 2015;20: 18732–18758. doi:10.3390/molecules201018732

241. Wasko MJ, Pellegrene KA, Madura JD, Surratt CK. A Role for Fragment-Based Drug Design in Developing Novel Lead Compounds for Central Nervous System Targets. Front Neurol. 2015;6. doi:10.3389/fneur.2015.00197

242. Muratcioglu S, Guven-Maiorov E, Keskin O, Gursoy A. Advances in template-based protein docking by utilizing interfaces towards completing structural interactome. Curr Opin Struct Biol. 2015;35: 87–92. doi:10.1016/j.sbi.2015.10.001