Master of Science thesis

Structural studies of *Erwinia carotovora* L-Asparaginase by X-ray crystallography

Charlotta S. Andersson

LiTH - IFM - EX - - 06/1580 - - SE
Structural studies of *Erwinia carotovora* L-Asparaginase

X-ray crystallography

Physics and Measurement Technology, Linköping University

Charlotta S. Andersson

LiTH - IFM - EX - - 06/1580 - - SE

Master of Science thesis: 20 p

Level: D

Supervisor: Tassos Papageorgiou,
Turku Centre for Biotechnology, Turku

Examiner: Lars-Göran Mårtensson,
Physics and Measurement Technology, Linköping University

Linköping: March 2006
Bacterial L-asparaginases (E.C.3.5.1.1) are enzymes that catalyze the hydrolysis of L-asparagine to aspartic acid. For the past 30 years these enzymes have been used as therapeutic agents in the treatment of acute childhood lymphoblastic leukemia. The presence of a low rate glutaminase activity however causes serious side-effects to patients in treatment, as glutamine depletion give rise to neurotoxicity, anaphylaxis, and other hypersensitivity reactions. The interest in the enzyme from *Erwinia carotovora* originates from the fact that it shows a decreased glutaminase activity, and therefore the enzyme is expected to exhibit fewer side-effects when used in therapy.

The main focus of this thesis is the crystal structure determination of L-asparaginase from *Erwinia carotovora* in the presence of aspartic acid at 2.5 Å resolution. The structure was refined to an R/Rfree factor of 19.9/28.6 with good stereochemistry.

L-Asparaginases are homotetrameric enzymes with a known 222 symmetry and an identical fold. The *Erwinia carotovora* asparaginase consists of eight monomers of 330 amino acid residues each. In this case the enzyme is active as a dimer of tetramers. The two tetramers have an inner twofold non-crystallographic symmetry. Each monomer forms two identifiable domains a large N-domain and a small C-domain. The active sites are found at a topological switch-point between those domains.
Abstract

Bacterial L-asparaginases (E.C.3.5.1.1) are enzymes that catalyze the hydrolysis of L-asparagine to aspartic acid. For the past 30 years these enzymes have been used as therapeutic agents in the treatment of acute childhood lymphoblastic leukemia. The presence of a low rate glutaminase activity however causes serious side-effects to patients in treatment, as glutamine depletion give rise to neurotoxicity, anaphylaxis, and other hypersensitivity reactions. The interest in the enzyme from *Erwinia carotovora* originates from the fact that it shows a decreased glutaminase activity, and therefore the enzyme is expected to exhibit fewer side effects when used in therapy.

The main focus of this thesis is the crystal structure determination of L-asparaginase from *Erwinia carotovora* in the presence of aspartic acid at 2.5 Å resolution. The structure was refined to an R/Rfree factor of 19.9/28.6 with good stereochemistry.

L-Asparaginases are homotetrameric enzymes with a known 222 symmetry and an identical fold. The *Erwinia carotovora* asparaginase consists of eight monomers of 330 amino acid residues each. In this case the enzyme is active as a dimer of tetramers. The two tetramers have an inner twofold non-crystallographic symmetry. Each monomer forms two identifiable domains a large N-domain and a small C-domain. The active sites are found at a topological switch-point between those domains.

**Keywords:** Protein crystallography, enzyme, crystal structure determination, asparaginase, leukemia treatment.
Acknowledgements

This thesis was performed in the Protein Crystallography group at the Turku Centre for Biotechnology (Turun Biotekniikkakeskus, BTK) under the supervision of Tassos Papageorgiou, PhD. It has been a dream come true to finally step into the world of structural biology and protein structure determination.

Firstly I would like to thank my supervisor Tassos Papageorgiou, head of the Protein Crystallography group in BTK. I would like to thank him for his support but by far most for letting me explore the world of structural Biology. In the group of Protein Crystallography I would also like to express my gratitude to Anni Kauko for being a true educationalist in the area of crystallography, and also in the area of understanding the standards in Finnish society. I must thank Susanna Saarinen whom has enlightened my otherwise dark and cold hours in the graphics room.

I would like to express my sincere gratitude to my supervisor Lars-Göran Mårtensson and his co-worker Per Hammarström for enabling a profiling-course in Protein chemistry and protein engineering. I came to Linköping solely to participate in this course, and I made the right choice. Also I must pay my regards to Marianne Kratz or else none of this would have been possible. Furthermore my opponent Katarina Norén also deserves my thanks, not just for reading and questioning my work but for being a part of my time in Linköping.

I would also like to thank the people closest to me, whose support have helped me to finish this graduate work. I thank my family for their encouragement and patience. Also I pay my deepest regards to my friends Maria Finne and Patrik Ståhl whom have taken me under their wings in Turku, and whom have been not just great friends but also good tutors of life. Thank you all for a great opportunity!
Nomenclature

Most of the recurring abbreviations and symbols are described here.

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>Amino Acid</td>
</tr>
<tr>
<td>ALL</td>
<td>Acute Lymphoblastic Lymphoma</td>
</tr>
<tr>
<td>CCP4</td>
<td>CCP4-program suite [17]</td>
</tr>
<tr>
<td>CNS</td>
<td>Crystallography and NMR System [4]</td>
</tr>
<tr>
<td>ECOLI</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>ErCAR</td>
<td><em>Erwinia carotovora</em></td>
</tr>
<tr>
<td>ErCHR</td>
<td><em>Erwinia chrysanthemi</em></td>
</tr>
<tr>
<td>MR</td>
<td>Molecular Replacement</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>rmsd</td>
<td>root mean square deviation</td>
</tr>
</tbody>
</table>

Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Da</td>
<td>dalton</td>
</tr>
<tr>
<td>$F_o$</td>
<td>observed structure factor</td>
</tr>
<tr>
<td>$F_c$</td>
<td>calculated structure factor (Equation 2.3)</td>
</tr>
<tr>
<td>$k_{cat}$</td>
<td>catalytic constant</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>$V_{max}$</td>
<td>maximum enzyme velocity</td>
</tr>
<tr>
<td>Vm</td>
<td>Matthews volume</td>
</tr>
</tbody>
</table>
Contents

1 Introduction 1
  1.1 Background 1
    1.1.1 Therapeutic enzymes 2
    1.1.2 L-Asparaginase 3
  1.2 Structure determination methods 4
    1.2.1 X-ray crystallography 4
    1.2.2 Nuclear magnetic resonance 5

2 Experimental approach 7
  2.1 Crystal characterization 7
    2.1.1 Theorem: Bragg's Law 8
    2.1.2 Matthews coefficient 9
  2.2 Diffraction pattern 10
    2.2.1 Indexing and scaling 11
  2.3 Phase problem and determination 12
    2.3.1 Isomorphous replacement 12
    2.3.2 Molecular Replacement 13
    2.3.3 Multiple-wavelength Anomalous Diffraction 13
  2.4 Structure visualization 13
    2.4.1 The R-factor 14
    2.4.2 Difference map 14
    2.4.3 Electron density map 15
    2.4.4 Omit map 15

3 Experimental details 17
  3.1 Crystallization 17
  3.2 Molecular Replacement for ErCAR 19
  3.3 Refinement 19
    3.3.1 The B-factor 20
  3.4 Rebuilding 20
    3.4.1 Water molecules 21

4 Results 23

Andersson, 2006. xiii
4.1 Quality of the structure ........................................ 23
4.2 Data collection and refinement statistics .................. 26
4.3 Structural features of L-asparaginase ....................... 27
  4.3.1 The monomer structure .................................. 27
  4.3.2 Description of the active site .......................... 29
  4.3.3 The tetramer structure .................................. 30
  4.3.4 The octamer structure .................................. 32

5 Discussion ................................................................ 33

6 Conclusion .............................................................. 35

A Protein Structure ..................................................... 41
  A.1 Primary Structure .............................................. 41
  A.2 Secondary Structure .......................................... 43

B Algorithms and mathematical functions ..................... 45
  B.1 The Patterson function ...................................... 45
  B.2 The intensity function ....................................... 45
  B.3 The least square method ..................................... 46

C External color print
Chapter 1

Introduction

This text is written as a master of science final thesis at Linköping University by Charlotta Andersson with Tassos Papageorgiou as supervisor and Lars-Göran Mårtensson as examiner, with Emacs in \LaTeX{} in 2005/2006. The project was supported by the Sigrid Jusélius Foundation.

This first chapter introduces L-asparaginases in clinical use. It will give some background on the studied enzyme and its importance in the human biological system.

1.1 Background

L-Asparaginases are enzymes that primarily catalyze the conversion of L-asparagine to L-aspartic acid and ammonia, they are also able to hydrolyze L-glutamine but at a lower rate (further information can be found in section 1.1.2). Asparaginases are expressed in many bacterial organisms, but only L-asparaginases from \textit{Escherichia coli} (ECOLI) and \textit{Erwinia chrysantemi} (ErCHR) have been used as chemotherapeutics in Acute Lymphoblastic Lymphoma (ALL) for the last three decades [1, 6]. Although there are therapeutic asparaginases present on the market, recent discoveries have indicated that the L-asparaginase from \textit{Erwinia carotovora} (ErCAR) might be more efficient and also to exhibit fewer side-effects [7, 8]. The need for new therapeutic enzymes is of great interest in both biotechnology and medicine. The aim of this study was to determine and analyze the three-dimensional structure of L-asparaginase from ErCAR.

By structure determination of biological molecules theories of chemical bonding and properties can be laid out and tested. This is possible
due to a close connection between the three-dimensional structure and the properties of the biological macromolecules. Knowledge of a protein structure offers a clue to what role a protein plays in the body and what modifications are possible to make. Also the three-dimensional structures hold a key towards the development of new drugs/medicines and provide a good starting point of protein-engineering studies. Prospective studies of structural features eventually will contribute to the optimization of a protein's therapeutic effect and minimization of its toxicity.

1.1.1 Therapeutic enzymes

The term 'therapeutic enzyme' has been known for at least 40 years [20]. What distinguishes therapeutic enzymes from other drugs are two main features; firstly that the enzymes act on their target with a great specificity and with high affinity, secondly they are catalytic and able to convert a substrate into a desired product. These features render possible the production of potent drugs, that could carry out therapeutic biochemistry \textit{in vivo}. Biotechnological advancements have enabled for enhanced potency and specificity among enzymes with a production at a lower cost.

Therapeutic enzymes have a broad variety of specific uses as oncolytic, anticoagulants or thrombolytic, and as replacements for metabolic deficiencies (Figure 1.1). The favored kinetic properties of these enzymes are low $K_m$ and high $V_{max}$ in order to get maximal efficiency even at very low enzyme and substrate concentrations. It is of great importance to fully understand the enzyme properties and catalytic activity, in order to optimize its use and limit potential side effects. Within the area of cancer treatment one exploits the knowledge of differences between normal- and malignant cells, i.e. malignant cells lack of certain functions. Also one should choose with care the sources of such enzymes to avoid any contamination or structural changes.

Insulin was the first genetically engineered biotechnology drug, introduced in 1982 [20]. For a period of 60 years the sources of insulin had been cattle and pigs. Although these products were highly effective the growing diabetic population arose some concerns about the long-term supply and potential allergic reactions. The ability of expressing recombinant human insulin (humulin) in bacteria gave rise to a whole new industry.
Figure 1.1: Therapeutic enzymes are used in the treatment of various disorders and diseases. Abbreviations of the genetic diseases are as follows: cystic fibrosis (CF), mucopolysaccharide (MPS), severe combined immunodeficiency disease (SCID), and phenylketonuria (PKU) [20].

1.1.2 L-Asparaginase

L-asparaginase has been used as a chemotherapeutic agent for over 30 years, mainly from the bacterial strains of ECOLI and ErCHR [1, 6, 18]. The asparaginase is used in the treatment of lymphoblastic malignancies\(^1\) in children. The enzyme catalyzes the deamidation of L-asparagine to produce L-aspartic acid and ammonia, but is also able to hydrolyze L-glutamine (Figure 1.2).

The antileukemic effect is believed to result from the depletion of circulating asparagine. Certain tumors have decreased or absent activity of asparagine synthase, and hence are dependent on externally supplied L-asparagine for growth [1, 6, 10, 7, 20, 22]. By administration of L-asparaginase the blood levels of asparagine are reduced and this leads to a selectively induced inhibition of malignant growth. In other words with a degrading component, as L-asparaginase, the cancer cells will not be able to survive.

\(^{1}\)Leukemia cancers: ALL, AML, CLL, Hodgkin, Non-Hodgkin and melanoma.

Andersson, 2006.
There is always a back-side with administrating drugs to a biological system. The side-effects in asparaginase therapy are mainly dependent on the glutaminase activity. The main side-effects are hypersensitivity reactions, anaphylaxis, hepatotoxicity, diabetes, and coagulation abnormalities [6, 7]. The elimination of glutaminase activity could result in a safer and more reliable treatment of leukemia. It has been shown that L-asparaginase from ErCAR has decreased glutaminase activity about 1.5%, and also different immunological specificity compared to ECOLI [6], therefore one would expect to exhibit fewer side-effects [8]. Other studies indicate a difference in the catalytical ability, where ErCAR has an approximately 200-fold higher $k_{cat}$ than ECOLI [7].

According to previous reports, slight differences between ECOLI and ErCHR strains in respect to toxicity and efficacy, have been found [6]. ECOLI is more toxic, since it has shown more coagulation abnormalities, although it keeps a higher clinical efficacy than ErCHR. The Erwinia strain has mostly been used as an alternative in cases where allergic reactions force the discontinuation of the ECOLI treatment [20].

### 1.2 Structure determination methods

There are a few methods used for the visualization of the complex arrangement of atoms within molecules. At present the only two techniques that can elucidate proteins to atomic resolution are X-ray diffraction and Nuclear magnetic resonance (NMR) analysis [5].

#### 1.2.1 X-ray crystallography

To be able to perform X-ray crystallography, it is necessary to grow crystals, since they have a repeated unit cell within them. The X-ray diffraction from one unit cell would not be significant but needs to be amplified. This is achieved by the repetition of unit cells within
a crystal. Electrons of structure atoms will scatter the incoming X-rays and cause a diffraction pattern specific for an ordered array of molecules. Using the mathematical Fourier transform these patterns can be converted into maps of electron density showing the position of atoms.

Solving the structure with X-ray crystallography demands an independent knowledge of the primary structure for interpretation of the density map. A more thorough explanation is presented in Chapter 2.

1.2.2 Nuclear magnetic resonance

From NMR the obtained structure will not be as detailed and accurate as with X-ray crystallography but rather a general topology for the polypeptide chain. The advantage here is that the protein used is in solution rather than in a crystal lattice. NMR spectra are generated by a magnetic field with applied radio-frequency pulses, where the absorption of energy by a nucleus gives a change in its orientation in the magnetic field [11]. Transient time domain signals are detected as the system returns to equilibrium [5]. Spin-spin interactions, interatomic shielding and ring currents are properties that make each atom identifiable. NMR gives much more detailed information about the flexibility of the protein structure in solution but is limited to small molecules (up to 20 kDa).
Chapter 2: Explanation of the ground theory of x-ray crystallography. This chapter introduces standard procedures for structure determination on protein molecules, briefly describing the theoretical properties of a crystal, interpretation of diffraction patterns to visualization of atom positioning.

Chapter 3: Experimental details, about the procedures used for structure determination. Initial modeling and improvement of the atomic model, called refinement, together with rebuilding will be explained under this part.

Chapter 4: Brings on the results, describing the structural parts of L-asparaginase along with data collection and refinement statistics.

Chapter 5+6: Discussion and conclusion, questioning the results and choice of methods as well as suggestions for further improvements.

Appendix A: Description of the protein structure.

Appendix B: Describes some of the algorithms and mathematical functions used in the creation of the model and refinement.

Appendix C: External color print of some pictures from chapter 4, reference to this will be done in the Thesis.
Chapter 2

Experimental approach

The most commonly used experimental technique for obtaining a detailed picture of a protein molecule allowing the resolution of individual atoms is X-ray diffraction. Obtaining a diffraction pattern is achieved by placing a crystal in a narrow, focused beam of monochromatic X-rays. These beams can’t be focused by lenses and therefore measurements are based on directions and intensity of the diffracted X-rays. Scattering of the X-ray beams by the electron clouds of the atoms provide the basis for an X-ray experiment.

A single molecule is a very weak scatterer of X-rays, hence most of the incoming X-rays will just pass through the single molecule without being diffracted. Use of crystals will give a positive interference and reinforce a detectable diffraction pattern. A data collection from the diffraction pattern of protein crystals is recorded and then processed to acquire a comprehensive electron density map. The crystal is rotated around an axis perpendicular to the beam and in each rotation phase the image of the diffraction spots is recorded by a detector. The resolution at which the diffraction is recorded will be crucial for the structure determination as it relates with the level of accuracy for the observed molecule.

2.1 Crystal characterization

The basic building block of a crystal is the unit cell, in theory infinitely repeated in three dimensions. A unit cell is characterized by three vectors, denoted a, b, and c starting with the shortest one, corresponding to the edges of a parallelepiped. The cell is also defined by three angles

Andersson, 2006.
between these vectors ($\alpha, \beta, \gamma$ see Figure 2.1(a)). Any crystal can belong to one of seven symmetries $^1$.

The asymmetric unit contain one or more molecules. The number of molecules is determined by the Matthews coefficient (Section 2.1.2). In biological systems the unit cell may possess an internal symmetry containing more than one biological molecule, related to others via axes or planes of symmetry $^2$. The crystal is ordered in three dimensions, where the individual and identical unit cells are arranged in a way that the points of its corners makes an array called a lattice (Figure 2.1(b)).

Generation of coordinates for atoms in a series of unit cells is made from symmetry operations, such as translation, rotation, reflection, and inversion. The crystallographic arrangement from a collection of these symmetry operations define particular space groups $^2$. The crystal structure consists of a basic motif that is repeated in three dimensional space by the symmetry operators of the crystallographic space group.

A crystallographer determines the coordinates of the atoms in this basic motif, called the asymmetric unit. It is the smallest part of a crystal structure from which the complete structure can be built using space group symmetry.

\begin{figure}[h]
\centering
\includegraphics[width=0.7\textwidth]{figure2.1.png}
\caption{The angles and vectors defining a unit cell (a), many unit cells form a lattice represented by the blue dots in the Figure (b).}
\end{figure}

\subsection{2.1.1 Theorem: Bragg’s Law}

The rules for diffraction are given by Bragg’s law, which correctly describes the conditions of constructive interference. It also shows that incident radiation on sets of parallel lattice planes selects those wavelengths corresponding to integral multiples of this wavelength. Peaks

$^1$The seven possible system design: triclinic, monoclinic, orthorhombic, tetragonal, rhombohedral, hexagonal or cubic

$^2$There are 230 different space groups [9]
and intensity will be observed when the angle of incident X-rays is equal to the angle of scattering and the path length difference is equal to an integer multiple number of the wavelength.

\[ n\lambda = 2d\sin\theta \]  

This equality gives information about the structure of the crystal and allows for the structure determination, since the wavelengths of X-rays are closely controlled. To find all of the planes in a crystal one must rotate the crystal or the X-ray beam. Only certain orientations which satisfies the Bragg’s conditions will give rise to spots, in the diffraction pattern.

### 2.1.2 Matthews coefficient

This coefficient allows for an estimation of the total number of molecules, in the asymmetric unit. Based on the following equation Matthews (1968) observed that the acceptable solutions lie between 2.0 to 5.0Å³/Da.

\[ V_m = \frac{\text{Vol.of unit cell}}{M_w \times Z \times X} \]  

where Z is the number of asymmetric units in the unit cell, i.e. the number of symmetry operators in your space group. The unknown variable, X, is the number of molecules in the asymmetric unit [14].

---

Andersson, 2006.
2.2 Diffraction pattern

A diffraction pattern depends on the crystal symmetry and the intensity of each spot is modulated by the protein structure (Figure 2.3). From equation 2.1 one can see that the spacings are inversely proportional to the lengths of the crystal unit cell, and $\theta_n$ is the angle of diffraction for the $n$:th diffraction order [5, 11]. The spacings are not affected by the molecules within a unit cell, although the symmetry of these molecules will explicitly define the symmetry of the diffraction pattern. Therefore these spacings can be used for determining the dimensions, angles, and space group of the unit cell, regardless of its molecules. Further more the orientation of the reflections in the reflection sphere mirrors the orientation of the primary axes from the unit cell [11]. This describes the inverse relationship between the spacing of unit cells in the crystalline lattice (real lattice) and the spacing of reflections of the recorded diffraction pattern (reciprocal lattice). The diffraction pattern relates to the diffracted waves from the object through a mathematical operation, the Fourier Transform.

Figure 2.3: Diffraction photograph of L-asparaginase obtained by Tassos Papageorgiou, June 2004. Data were collected at the EMBL X11 beamline at DORIS storage ring, DESY, Hamburg.
The structure factor is used to see how well the final model of the molecular structure in the crystal fits the observations from the X-ray diffraction pattern. The space of a crystal has a probability for containing electrons, described by the structure factor over a volume, $V$, (Equation 2.3).

$$F_{hkl} = F_{re} + iF_{im} = V \int \int \int \rho(x, y, z)e^{2\pi i(hx + ky + lz)}dxdydz \quad (2.3)$$

The molecular structure factor, $F_{hkl}$, has two components the real and imaginary components. The net values for the real and imaginary components are illustrated in the Argand diagram (Figure 2.4). This diagram shows how the phase is an angle between $F_{im}$ and $F_{re}$. The structure factor is the volume of the unit cell times the integration over the electron density per unit volume, with respect to the phase factor.

![Argand diagram for the structure factor, $F_{hkl}$](image)

**Figure 2.4:** Argand diagram for the structure factor, $F_{hkl}$

### 2.2.1 Indexing and scaling

Some reflections appear as bright intense spots whereas others are weak or missing (Section B.2 describes how to measure the intensity, $I$, of the spots) in the otherwise so evenly spaced pattern. The indexing of the reflection spots are made in two steps. First one uses a primitive, generic lattice to give each reflection an index, to identify a reflection in reciprocal space. Then the reflections related due to the space group symmetry are reduced and collected together.

Indexing data can be accomplished in several ways, using different pro-
grams, for example HKL\textsuperscript{3}. This has automated a procedure that earlier was made manually. An important factor for indexing data is the R-sym factor, which gives a hint of the amount of errors in the data set. Existing reflections are compared to symmetry-related ones thus the lower the R-sym factor, the better.

\[
R_{\text{symm}} = \frac{\sum_{hkl} \sum_i |I_i(h, k, l) - I(h, k, l)|}{\sum_{hkl} \sum_i I_i(h, k, l)}
\]  

(2.4)

Another factor to be taken into consideration while evaluating the quality of the data is the completeness, i.e. the ratio of the number of measured reflections to the number of all reflections possible. A rule of thumb is not to let this factor go beneath 80% \cite{2}.

The concept upon scaling the data is utilized in order to get a reliable data set, free from disturbances caused from external factors. This follows the same concept as in indexing only with a scale factor.

### 2.3 Phase problem and determination

To obtain the relation between the protein and its diffraction pattern one needs to know the amplitudes and phases. The first can be directly measured but the second not. With known phases a picture of the molecule could easily be computed but that information is lost in the experiment. This is the phase problem and a large part of crystallography is devoted to solving it. In crystallography the aim is to determine the positions of atoms (x, y, and z).

There are a few methods available for deriving the phase, the most commonly used are: isomorphous replacement (IR), molecular replacement (MR), and Multiple-wavelength Anomalous Diffraction (MAD).

#### 2.3.1 Isomorphous replacement

This method involves incorporation of heavy atoms into a protein crystal, giving changes in the X-ray intensities. This demands keeping an isomorphous crystal towards the heavy atoms, like none significant distortion to the structure, none altering of it’s space groups, and leave the unit cell parameters intact. There is a very large repertoire of compounds known to produce this sorts of derivatives \cite{11}. The intensity difference is used to deduce the positions of the heavy metals

\textsuperscript{3}The HKL have been written by Dr. Zbyszek Otwinowski (Southwestern Medical Center, University of Texas) and Dr. Wladek Minor (University of Virginia).
within the crystal unit cell. Fourier summations of these differences give vector maps between the different heavy atoms, called Patterson maps. These vector maps simplify the solution for the atomic arrangement. With a knowledge of the atomic position it is now possible to use the R-symm factor (Equation 2.4) for calculating the amplitudes and phases.

2.3.2 Molecular Replacement

This is a conceptually straightforward technique, using co-ordinates of a well defined structure as a "search structure". Common elements between the previously known structure and the one being solved will generate the desired phase data. It is therefore of great importance to use structures that are very similar to each other, due to the correctness of the phasing. Computer applications for carrying out MR lies within the knowledge of modern crystallography and computing, for example CNS [4].

2.3.3 Multiple-wavelength Anomalous Diffraction

Certain X-ray wavelengths cause the electrons to absorb energy which, in turn, causes a change in the scattering, called anomalous scattering. The size of this change of energy is negligible for light atoms but measurable for heavy atoms, such as iron, zinc and mercury. This produces a measurable difference of intensity in the diffraction pattern like the one in IR. This method is especially useful for metallo-proteins and can be applied directly on the native protein, however it is not as strong as with two derivatives with different atomic coordinates.

2.4 Structure visualization

For visualization of the electron density one needs to solve the Fourier Transform for the diffraction pattern, since electron density is considered a function [4]. The problem with Fourier maps is that they require phases in order to be calculated and the only data available is the set of amplitudes from the diffraction images. Calculation of the electron density map is made from recombining mathematically the individual reflections of the diffraction pattern [5, 11]. The electron density \( \rho(x, y, z) \) in the unit cell is given by:

\[
\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F(h, k, l) e^{i\alpha(h,k,l)} e^{-2\pi i(hx+ky+lz)}
\]  

(2.5)

Andersson, 2006.
Consequently this shows that every reflection contains information about all parts in the unit cell, like every atom contributes to each reflection. The quality of the map is dependent on the resolution of the diffraction data, and the resolution is highly influenced from the quality of the crystal.

2.4.1 The R-factor

The model will strongly affect the calculated electron density map, due to the phase problem. To be able to evaluate how well the model fits the map one can use the R-factor (Equation 2.6). This is the average fractional error in the calculated amplitude compared to the observed amplitude [3, 11, 21]. A structure is judged by the crystallographic factor R, defined as the average fractional error in the sum of the differences between calculated structure factor $F_c$ and observed structure factor $F_o$.

$$R = \sum (|F_{obs}(h,k,l)| - k|F_{calc}(h,k,l)|)^2$$ (2.6)

The following statistical concept of cross-validation is based on the partition of the observed reflections into a test set and a working set. The test set is omitted from a small portion of the data and taken after refinement, which in other words will give the correlation coefficient between the refined model against the complete data set.

$$R_{free} = \frac{\sum \|F_{obs}(h,k,l)| - |F_{calc}(h,k,l)\|}{\sum |F_{obs}(h,k,l)|}$$ (2.7)

A rule of thumb for a good structure will be between $R_{free}$ 15-20%, whilst a random structure keeps an $R_{free}$ around 60% [4, 11].

2.4.2 Difference map

This map visualizes the difference between the observed and calculated amplitudes within the model map. Since the Fourier Transform is additive, this is achieved through subtraction of the structure factors: $F_o - F_c$. This is a convenient way for finding solvent molecules, locating missing atoms or residues, as well as finding mistakes in the spatial alignment. This will show the difference of electron density between the original model and the one obtained after phasing.

The $F_o - F_c$ map will have negative features representing the densities created by the model which are not seen in the structure. Its positive
features comes from densities present in the crystal which are not in the model.

2.4.3 Electron density map

The $2F_o - F_c$ show the current best estimate of the electron density for the structure. This is the map in which the model is ought to fit, and therefore it is particularly useful for rebuilding. It does not show the exact atom positions, but rather suggests placements, and it is altered from maximum likelihood refinement. One should remember that the quality of this map relies highly on the quality of the phases, i.e. a map with high R-factor is rather doubtful since it then is badly correlated to the model.

2.4.4 Omit map

Due to incorrect spatial ordering of the model, bias might be introduced into the density map. A way to investigate these more doubtful regions is to calculate an omit map. An estimate of phase angles will be made for a small volume at a time, where parts of the model are left out, and the calculation is made from the remainder of the structure. This method is certainly less accurate but will give an unbiased estimate in troubled regions of the model. The main problem in this case is to choose the proper scaling factors for $F_o$ and $F_c$. As part of the model is left out the sum of $F_c$ will be smaller than $F_o$. 

Andersson, 2006.
Chapter 3

Experimental details

Interpretation of the electron density maps requires knowledge of the primary sequence. Building the initial model is a trial and error process, since the initial models often contain a lot of errors and poor phases. Starting with a matching between the known polypeptide sequence and its density, is followed by finding the best fit of atoms in the density. Although only density maps with atomic resolution resolve individual atoms the maps enable the identification of the side-chains. The atomic model is never perfect but it can be improved by a process called refinement, where the model is adjusted to improve the agreement to its measured diffraction data.

The B-factor reflects the spreading or blurring of electron density and represents the mean square displacement (Subsection 3.3.1). Also the success of the atomic model is measured, this time through the standard crystallographic R-factor (Subsection 2.4.1).

3.1 Crystallization

In order to perform an X-ray experiment one needs to grow large and stable crystals with a sufficient long-range order. Two experimental methods used to form crystals are: vapor diffusion and equilibrium dialysis. Usually equilibrium dialysis is used for crystallization at low or high salt concentrations, whilst the more common vapor diffusion is used at small volumes [21]. Crystals are grown under slow, controlled precipitation from aqueous solutions under non-denaturing conditions. Precipitation is caused by ionic compounds, organic solvents or most commonly with PEG. Whether the protein will be able to form crystals
or not depends on many properties of the solution, like protein concentration, pH, temperature, and ionic strength.

Figure 3.1: Flowchart of structure determination.
3.2 Molecular Replacement for ErCAR

For finding the phases in the diffraction pattern the method used was MR, the phases were found by using PHASER [15]. By the use of a previously known structure it is quite easy to define another molecule, where the search-model was a poly-alanine asparaginase from ErCHR.

An important starting point is to withdraw information about symmetry and cell dimensions from the crystal. This information offers an estimation of the molecular content in the asymmetric unit (Subsection 2.1.2). The internal symmetry of an oligomer does not show up in the crystal symmetry, and the first step will be to find out how these are arranged within the asymmetric unit [2]. It is possible to find out the arrangement for different subunits. In rigid objects one subunit is related to another by an operation of rotation and translation. Rotation is made through the Patterson function (Equation B.1). A rotation of the search-molecule is made around a reference point until it is parallel to the unknown structure in the unit cell. Translation is a movement in the three dimensional space and is made after the rotation to get the search-model "on top" of the displaced molecule. Translation also uses the Patterson function, but slightly differently. This time the model molecule is placed at all positions of the unit cell, and for each position the vectors are calculated and compared to the actual function of the unknown molecule.

Using a model-structure to generate the phase angles will create some bias in the electron density map. Some features represented in the model are not correctly represented by the unknown structure and the other way around, but they will be visible on the map. Careful examination of the electron density difference map and rebuilding is needed to remove the bias.

3.3 Refinement

Refinement is an iterative process in which the atomic model is modified, structure factor amplitudes are calculated from the modified model, and the agreement between these calculated structure factor amplitudes and the observed ones is determined. The goal is to find the model that produces the best agreement between the experimental and the calculated factor amplitudes, by the Least square method (Appendix B.3).

All refinement as well as the electron density map calculations were
done with the program CNS \(^1\) and Refmac \(^2\). CNS seems superior at early refinement stages (R worse than 30%), since it uses a tight stereochemistry, simulated annealing, and is keeping a good radius of convergence. Refmac uses a more aggressive minimization algorithm and is therefore more suitable for later stages of the refinement.

The rigid body protocol was applied in an early stage of the refinement, at a resolution of 3 Å. Explicit refinement was done on water molecules and with the bound L-aspartate in a later stage of the procedure. In rigid body refinement large sections of the protein, such as subunits, move as rigid bodies. In the simplest case the entire protein is treated as one rigid body, which results in 6 degrees of freedom. L-Asparaginase is a dimer of tetramers, so a natural rigid body scheme would be to treat each subunit as a separate body. Rigid body refinement is useful in the early stages of structure determination and it is usually done with low resolution data determination (15-3Å).

### 3.3.1 The B-factor

The B-factor is also called temperature factor or the Debye-Waller factor. Originally the B-factor was introduced as a measure of the thermal motion of the atom \([4, 11]\). In other terms this reflects the extent of disorder of each atom to the diffraction pattern. Since it’s only assigned a single parameter, one can assume this to be an isotropic thermal motion, and the best model should contain thermal motion for all three directions in space \((B_x, B_y, \text{ and } B_z)\). The B-factor will also reflect how often an atom is positioned in a particular spot in space, note that proteins show different flexibility in various regions. This is the partial occupancy of an atom that is within the crystal a specified spatial area.

\[
B_{factor} = e^{-\frac{B_x \sin^2 \theta}{\lambda^2}}
\]  

(3.1)

In a practical meaning the B-factor is a good measure for the overall disorder of the atom. Accordingly this will affect the observed intensity by the B-factor (Equation 3.1) from the original intensity, \(I_0\).

### 3.4 Rebuilding

Manual rebuilding of a structure is required because the use of efficient algorithms the human eye contributes to the greatest process in pattern recognition and interpretation. The main idea for a functional program

---

\(^1\) Crystallography and NMR System, [4].

\(^2\) Supported program within the CCP4-suite, [17].
for rebuilding is to display electron density map and the actual model in three-dimensional space. The programs enable for introducing new features into the model, like adding residues, positional and orientational change, and also changing the torsional angles of particular residues. For the rebuilding and also for visual inspection the program 'O' was used in this project [12].

3.4.1 Water molecules

An important part of the structure are water molecules since a well ordered water molecule may even contribute more to the scattering than the poorly ordered parts of the protein. The waters are clearly visible through experimental maps and in difference maps. It is important to make inspections of the added waters, so that waters will not put in features that are representative of other things. Whilst some waters are added manually others are automatically added by ARP/WARP, from the CCP4 suite. One should keep in mind that double conformations, ligands etc might be occupied by this automated procedure, thus a visual inspection is always required. Waters with a B-factor higher than 50 were all excluded from the structure.

---

3 Used with Refmac, supported by the CCP4-suite [17]
Andersson, 2006.
Chapter 4

Results

After conducting iterative refinement, with additional steps the decreasing R-factor finally reached an acceptable level below 0.20. This Chapter presents the results and gives a view of the three-dimensional properties in the enzyme. Some Figures in this chapter appear better in color and therefore these black and white figures will be referring to separate color prints.

4.1 Quality of the structure

The structure was refined to an R = 19.9% ($R_{\text{free}} = 28.6\%$). For eight molecules in the asymmetric unit the Matthews coefficient is $V_M = 2.4 \, \text{Å}^3/\text{Da}$, from Equation 2.2. This indicates close packing of eight L-asparaginase molecules in a unit cell, each molecule contains 327 residues, with a total amount of 453 water molecules. The space group of L-asparaginase from *Erwinia carotovora* was determined to be $P2_12_12_1$ based on the systematic absences [2, 9] (for data collection and refinement statistics see Table 4.1).

The refinement started off with CNS and continued in CCP4, after a few cycles of rebuilding the density for the loop between residues 20 and 35 was still unclear. Some attempts were made to rebuild this part, but failed due to lack of sufficient density. The final statistics from the refinement are presented in Table 4.1.

The Ramachandran plot in Figure 4.1 shows the distribution of $(\phi, \psi)$ conformational angles along the polypeptide backbone of the protein. This is a very good indicator of the quality in the modeled protein. The distribution gives a picture of the secondary structure as some angles

Andersson, 2006. 23
Figure 4.1: Ramachandran plot of the final L-asparaginase model. Shown in red are those combinations of phi and psi that are "allowed" i.e. that do not result in steric hindrance. The dark yellow and light yellow areas are allowed if some steric hindrance is permitted. The abbreviations correspond accordingly: A/a to alpha, B/b to beta, L/l to lefthanded alpha, and p to epsilon. For color figure see separate pages.
are preferred in helices and others in sheets. Most of the main-chain
torsion angles are found in the most favored region of the Ramachan-
dran plot (86.9%) or in the additionally allowed regions (12.2%). Few
conformational angles are in the "generously allowed" and "forbidden"
regions (0.5 and 0.4% respectively) the last one mainly due to Thr204
from all the independent subunits.

A study of the B-factors shows how conformations near the surface,
expectedly mobile due to solvent exposure, will have low contributions
to scattering. This appears especially in the loop, around 20 residues
long, that is sited near the active site. High B-factors indicates that
coordinates cannot be entirely trusted, this goes for both side- and
main-chains.
## 4.2 Data collection and refinement statistics

<table>
<thead>
<tr>
<th><strong>Data collection</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>$P2_12_12_1$</td>
</tr>
</tbody>
</table>
| Cell dimensions (Å) | \[ \begin{align*}
a &= 73.65 \\
b &= 135.65 \\
c &= 250.10 \\
\alpha &= \beta = \gamma = 90^\circ \\
\end{align*} \] |
| Resolution range (Å) | 20.0 - 2.50 (2.56 - 2.50) |
| Data collection temperature (K) | 100K |
| Wilson B-factor (Å²) | 45.9 |
| No. of observations | 368 260 |
| No. of unique reflections | 83 599¹ |
| Completeness (%) | 95.8 |
| Rsym (%) | 7.8 (43.8) |
| $I/\sigma(I)$ | 13.9 (3.2) |
| No. of molecules | 8 |

<table>
<thead>
<tr>
<th><strong>Refinement statistics</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>20.0 - 2.5</td>
</tr>
<tr>
<td>No. of reflections in working set</td>
<td>4 666</td>
</tr>
<tr>
<td>No. of reflections in test set</td>
<td>79 386</td>
</tr>
<tr>
<td>Protein atoms</td>
<td>18 744</td>
</tr>
<tr>
<td>Water molecules</td>
<td>452</td>
</tr>
<tr>
<td>$R_{cryst}/R_{free}$ (%)</td>
<td>19.9/28.6</td>
</tr>
<tr>
<td>RMSD</td>
<td>0.011</td>
</tr>
<tr>
<td>bond lengths (Å)</td>
<td>0.011</td>
</tr>
<tr>
<td>bond angles (°)</td>
<td>1.434</td>
</tr>
<tr>
<td>Average B factors (Å²)</td>
<td>46.2</td>
</tr>
<tr>
<td>Main chain</td>
<td>46.2</td>
</tr>
<tr>
<td>Side chain</td>
<td>46.1</td>
</tr>
<tr>
<td>Waters</td>
<td>43.6</td>
</tr>
<tr>
<td>Ligand</td>
<td>50.7</td>
</tr>
</tbody>
</table>

Table 4.1: Data collection and refinement statistics for native *Erwinia carotovora* L-asparaginase.

¹ Note: Some values are rounded for simplicity.
4.3 Structural features of L-asparaginase

4.3.1 The monomer structure

Figure 4.2 shows the amino acid sequence for ErCAR. Each monomer consists of about 330 amino-acid residues and it forms 14 β-strands and 13 α-helices.

The independent subunits can easily be divided into two subunits, connected by a 20 residue long linker. The larger N-terminal domain is built up from an 8 stranded antiparallel mixed β-sheet, and the smaller C-terminal subunit of a parallel β-sheet (Figure 4.3).

Figure 4.2: Alignment between L-asparaginase from: ErCAR, ErCHR (PDB-code: 1HG1), and ECOLI (PDB-code: 4ECA). Secondary structure visualized for ErCAR in blue and active-site residues are marked with the black arrow. For color figure see separate print.

The first two residues in N-terminal could not be modeled in any of the eight individual subunits. Also there was some residues in the linking loop that could not be modeled. These residues include 25-29 in chain A, 23-33 in chain B, 23-31 in chain C, 27-30 in chain D, 23-33 in chain E, 25-32 in chain F, 19-32 in chain G, and 21-30 in chain H.
H. No obvious double conformations were found at this resolution (2.5 Å). Some residual disorders appear due to the lack of interactions, especially to residues located near the surface. For illustration of the monomer subunit see Figure 4.4.

**Figure 4.3:** Topology of L-asparaginase from ErCAR. Arrows are representing β-sheets and rectangles α-helices, darkly colored helices are placed on the opposite side of the sheet. The dashed box is representing the N-domain.

**Figure 4.4:** Illustration of subunit 'A' from the L-asparaginase. The two domains (N and C) are marked in the figure. For colors see separate print.

Andersson, 2006.
4.3.2 Description of the active site

The active site is located between the N- and C-terminal domains of the two adjacent monomers, consistent with previously reported results [1, 13, 18]. Interaction between two of the subunits make up the active sites in the created intimate dimer (A-C, B-D, E-G, and F-H). No active site is found in any of the distant dimer parts (Figure 4.5). The residues in each monomer that are involved in the active site are Thr15, Ser62, Thr96, and Asp120 (Figure 4.6), and the catalytic residues are Thr15 and Thr95 [1, 16]. Evaluating contacts within the structure shows that also the surrounding residues have some contacts to the ligand, but mainly responsible for binding are the previous mentioned residues. The four individual active sites do not seem to be cooperative to one another [13]. The crossover between the fourth and fifth $\beta$-strands of the N-terminal domain is left-handed, something that is usually seen of great relevance to the activity [13, 16].

![Figure 4.5](image)

Figure 4.5: This cartoon illustrates the dimer formation between the monomers A and C, the ligands labeled by red spheres. For color figure see separate print.
4.3.3 The tetramer structure

There are two tetramers in the asymmetric unit, the designated monomers are A-H, where ABCD make up one tetramer and EFGH creates the second. The two tetramers are extremely similar, in every aspect. Each subunit of the tetramer makes two types of contact with the neighboring subunits. Intimate contact, forming the intimate dimer, and distant contact forming the tetramer [13, 18]. The intimate dimer accommodates two separate active sites in its interface, see previous section. The assembly of the two dimers A-B and C-D respectively E-F and G-H forms the globular shaped tetramer (Figure 4.7 and 4.5).

Figure 4.6: The active site residues and aspartic acid (ligand).
Figure 4.7: The interactions shown in a and b gives a view of the close contacts within the dimer. Evaluation of intermolecular contacts also show that the dimer to dimer interactions are fewer than the A-C, B-D, E-G, and F-H contacts, indicating a looser fit.
4.3.4 The octamer structure

The active site formation indicates a dimer interaction for the activity of the asparaginase assembler into a tetramer form, and further into the octamer, illustrated in Figure 4.8. Each tetramer has a 222 symmetry and the octamer is characterized by a 2-fold symmetry between tetramers. Calculation of the accessible area\(^2\) indicated a stabilizing interaction between the tetramers, as the accessible area is similar to an expanded van der Waals surface.

Observed interactions were for example: C148 Arg to F215 Asp, creating a salt bridge (2.62 Å) between the two tetramers. These types of contacts are electrostatic and in some degree also hydrogen bonding, and they participate with 2-3 kcal in the molecule [5].

![Ribbon diagram of L-asparaginase showing the two tetramers. Each color represents a monomer. Figure generated in PyMOL. For colors see external print.](image)

**Figure 4.8:** Ribbon diagram of L-asparaginase showing the two tetramers. Each color represents a monomer. Figure generated in PyMOL. For colors see external print.

\(^2\)The difference between tetramer surfaces and octamer surface was: 1806.26. 

---

Andersson, 2006.
Chapter 5

Discussion

Figure 4.2 shows the amino acid sequence alignment resulting from ErCAR, ErCHR, and ECOLI. After superposition of $C_\alpha$ ErCAR shows a 78% sequence identity with the asparaginase from ErCHR (rmsd at 0.42). The sequence identity to ECOLI was 49% (rmsd at 0.88), and it shows even lower identity to the other bacterial L-asparaginases. A comparison shows that the active site residues in ErCAR are likely to be the same as those observed for the structure of ErCHR L-asparaginase. Also the topological structure (Figure 4.3) resembles the previous studied asparaginases [18, 19]. But there seems to be a difference between the ErCAR and ECOLI, since ECOLI has a single disulfide bond in each subunit [18], which is not present in the ErCAR. This disulfide bond is placed near the surface near the substrate 'canal', giving extra stabilization to the ECOLI enzyme.

All known asparaginase are homo-tetramers with 222 symmetry, with a molecular mass of 140-150kDa [1, 7, 16, 18, 19]. These have four identical subunits, each monomer consisting of 330 amino acids. The symmetry found in ErCAR matches previous determined structures, although in the aspects of the total amount of subunits as well as in certain sequential parts there are some visible differences between those features. L-Asparaginase from ErCAR appear as a dimer of tetramers, showing a close connection between the two tetramers.

The first three residues in the N-terminal end cannot be seen in the structure, due to high mobility in the surface residues. Also the electron density for the the active site flexible loop, residue 15-30, was absent for the eight monomers thus indicating a highly mobile region.

The Ramachandran plot revealed that Thr204 was found in the disallowed conformation, from all eight monomers in the enzyme. This has
also been observed in previous studies on ErCHR [13]. The explanation has been that Thr204 is part of the inter-domain linker, and this fragment is characterized by an increased flexibility. This phenomena has also been recognized in ECOLI, but this time in Thr198 and Ser199 [18].
Chapter 6

Conclusion

Despite a low resolution (2.5 Å) the final modeling has been successful. Further improvements on the missing residues, from the gap, could be made collecting new data at a higher resolution. To be able to do this a new crystal batch is needed. Continuing with some engineering studies on the protein could improve the stability of the protein and its features, to achieve a lower glutaminase activity with fewer side-effects.

The accuracy of the model is presented in the results, the modeling was somewhat troublesome due to its size and the protein flexibility. The crystal structure will be a good starting point for further studies and evaluation of the precise role of specific residues in the activities of asparaginase.

Andersson, 2006.
Bibliography


Appendix A

Protein Structure

Description of a the protein structure involves four different levels of structure: primary, secondary, tertiary, and quaternary structure. The primary structure refers to the exact sequence of amino acids present in the protein (for further information see A.1). The secondary structure refers to regular structures of linear segments of the polypeptide chains, i.e. $\alpha$-helices and $\beta$-strands. The tertiary structure is the overall structure of the folded chains, and quaternary structure arises when a protein contains more than one separate polypeptide chain.

Since proteins are not rigid, static objects, but dynamic rapidly changing molecules, that move, bend, expand and contract, these features are causing some obstacles when solving the protein structure.

Detailed information about the structure of ErCAR will be found in Chapter 4.

A.1 Primary Structure

Proteins are built from amino acids, that can be divided into four different groups: hydrophobic, charged, polar, and glycine. The amino acids are divided into these groups accordingly:

2. Charged amino acids: D, E, K, and R.

Andersson, 2006.
<table>
<thead>
<tr>
<th>Amino acids</th>
<th>Abbreviation</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>ALA</td>
<td>CH₃−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Arginine</td>
<td>ARG</td>
<td>HN=CH−NH−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Asparagine</td>
<td>ASN</td>
<td>H₂N−CO−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>ASP</td>
<td>HOOC−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Cysteine</td>
<td>CYS</td>
<td>HS−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Glutamine</td>
<td>GLU</td>
<td>HN−CO−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>GLN</td>
<td>HOOC−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Glycine</td>
<td>GLY</td>
<td>NH₂−CH₂−COOH</td>
</tr>
<tr>
<td>Histidine</td>
<td>HIS</td>
<td>NH₂−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>ILE</td>
<td>CH₃−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Leucine</td>
<td>LEU</td>
<td>(CH₃)₂−CH−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Lysine</td>
<td>LYS</td>
<td>H₂N−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Methionine</td>
<td>MET</td>
<td>CH₃−S−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>PHE</td>
<td>Phenyl−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Proline</td>
<td>PRO</td>
<td>NH−(CH₂)₃−COOH</td>
</tr>
<tr>
<td>Serine</td>
<td>SER</td>
<td>HO−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Threonine</td>
<td>THR</td>
<td>CH₃−CH(OH)−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Tryptophane</td>
<td>TRP</td>
<td>Phenyl−NH−CHC−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>TYR</td>
<td>HO−Phenyl−CH₂−CH(NH₂)−COOH</td>
</tr>
<tr>
<td>Valine</td>
<td>VAL</td>
<td>VAL = HO−CH(NH₂)−COOH</td>
</tr>
</tbody>
</table>

Table A.1: Amino acids, the building blocks of a protein.
A.2 Secondary Structure

The Ramachandran plot shows the allowed combinations of the torsion angles phi and psi of the peptide backbone, since the structure is not steric certain combinations are preferably from others but not all fixed. Common secondary protein structure elements are marked at the positions of their average phi and psi values.
Appendix B

Algorithms and mathematical functions

B.1 The Patterson function

Patterson function for self-rotation [2]:

\[ R(C) = \int_V P_1(u)P_2(Cu)dV \]  \hspace{1cm} (B.1)

The Patterson function is also called the self convolution of a structure, or in other words the Fourier Transform of the intensities. For all the possible rotations the function should find the rotation operation that align the search molecule with the model.

B.2 The intensity function

To calculate the intensity of the scattering one the product of structure factor F (Section 2.3) and its complex conjugate [11].

\[ I(s) = |F(hkl)|^2 = |F(s)|^2 \]  \hspace{1cm} (B.2)
B.3 The least square method

The method generally used in X-ray crystallographic refinements is the method of least squares. This finds model parameters that minimize the sum of the square difference between the observed quantities and the suggested theoretical model [2].

\[ L_2 = \| \omega_i [y_i - f_i(x)] \| = \sum_i \omega_i [y_i - f_i(x)]^2 \]  

(B.3)
Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a period of 25 years from the date of publication barring exceptional circumstances. The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for your own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility. According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement. For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its WWW home page: http://www.ep.liu.se/

Upphovsrätt

Detta dokument hålls tillgängligt på Internet - eller dess framtidare ersättare - under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår. Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns det lösningar av teknisk och administrativ art. Upphovsmannens ideella rätt innefattar rätt att bli nämnt som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart. För ytterligare information om Linköping University Electronic Press se förlagets hemsida http://www.ep.liu.se/

© 2006, Charlotta S. Andersson

Andersson, 2006.