Structural Studies of Lumazine Synthases – Thermostability, Catalytic Mechanism and Molecular Assembly

張肖鋒
Xiaofeng Zhang

Stockholm 2005
Cover Illustration:
Electron density of the active site of lumazine synthase from the hyperthermophilic bacterium *Aquifex aeolicus*.
To My Parents
献给我的父亲母亲
ABSTRACT

Riboflavin, also known as vitamin B₂, is biosynthesized in plants, bacteria, archaea and fungi. The primary biological function of riboflavin is related to its existence as a component of the two coenzymes, flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD), which play an important role for electron transfer in energy metabolism.

This project is mainly focused on structural studies of lumazine synthase (LS) from the hyperthermophilic bacterium *Aquifex aeolicus* (LSAQ). The enzyme is involved in the penultimate step of biosynthesis of riboflavin. The aim of this study is to gain insights into the structural basis of thermostability, catalytic mechanism as well as the molecular assembly of the enzyme. Methods used for these studies include X-ray crystallography, electron microscopy (EM), small angle X-ray scattering (SAXS) and differential scanning calorimetry (DSC).

Lumazine synthase from the hyperthermophile *A. aeolicus* displays dramatic stability against high temperature. The calorimetric melting profile indicates an apparent melting temperature (T_m) of 120°C. The factors that determine the thermostability of *A. aeolicus* LS were revealed by structural comparisons (Paper I, 2001).

In the second last step of riboflavin biosynthesis, lumazine synthase catalyzes the formation of 6-7-dimethyl-8-ribityllumazine, which is subsequently converted to riboflavin. In light of the structural studies of the enzyme in complexes with inhibitors (four complex structures were studied in this work), which were designed to mimic the substrates, reaction intermediate and the product at different stages of the reaction, a structural model of the catalytic process, which illustrates binding of substrates, enantiomer specificity, proton abstraction/donation, inorganic phosphate elimination, formation of the Schiff base and cyclisation, was proposed (Paper II, 2003).

Lumazine synthase assumes at least four assembly forms, namely the virus-like icosahedral capsid with a diameter of about 160 Å, the pentameric form, the stacking pentamers and larger capsids with a diameter of about 300 Å (metamorphosis of the enzyme is reviewed in Appendix A). The pH and/or buffer dependence of the assembly states of LS from *B. subtilis*, *A. aeolicus* and a designed mutant LS from *A. aeolicus* (structure determined by cryo-EM in manuscript IV) were studied using small angle x-ray scattering (SAXS) and cryo-EM. The results indicate that multiple assembly states are a general feature of lumazine synthases. Furthermore, the catalytic function of the enzyme is closely correlated with the assembly state (Manuscript, III).

Sequence alignment revealed that an insertion of 1–4 residues after Gly138 is unique for the pentameric lumazine synthases. Structural comparisons and modeling studies suggested that this insertion may inhibit the formation of icosahedral capsids. The structure of lumazine synthase from *A. aeolicus* with a four-residue’s insertion (IDEA) is studied by cryo-EM. It is shown that the mutant forms large capsids with a diameter of 292 Å. The analysis of the subunit interactions indicated that the assembly of the mutant does not follow the theory of “quasi-equivalence”, because the contact surfaces are non-equivalent. Compared to that of the wild type enzyme, the pentamer of the mutant is widened. The expanded pentameric structure provides a model for an alternative conformation of the LS pentamer as it could also be formed during the catalytic reaction in the T=1 capsid (Manuscript IV)
List of Publications


CONTENTS

1 Introduction .................................................................................................. 1
  1.1 Structure of Proteins........................................................................... 1
  1.2 X-ray Scattering ................................................................................. 1
    1.2.1 Solution Scattering (Small angle X-ray Scattering) .................. 2
    1.2.2 Crystal Scattering ....................................................................... 4
    1.2.3 The Electron Density and the Phase Problem ........................... 8
  1.3 Protein X-ray Crystallography ........................................................... 8
    1.3.1 Protein Crystallization and Data Collection .............................. 8
    1.3.2 Symmetry and Symmetry Operations ..................................... 11
    1.3.3 Phasing Methods ...................................................................... 12
    1.3.4 Patterson Function and Patterson Map .................................... 13
    1.3.5 Rotation, Translation and Molecular Replacement ................. 14
    1.3.6 Density Modification ............................................................... 15
    1.3.7 Model Building and Crystallographic Refinement .................... 16
  1.4 Hyperthermophiles and Proteins from Hyperthermophiles............. 18
    1.4.1 Hyperthermophiles................................................................... 18
    1.4.2 Aquifex aeolicus ...................................................................... 19
    1.4.3 Protein Stability and Folding ................................................... 19
    1.4.4 Dominant Forces of Protein Stability ...................................... 21
  1.5 Riboflavin and Enzymes Involved in Riboflavin Biosynthesis....... 23
    1.5.1 Riboflavin................................................................................. 23
    1.5.2 Enzymes involved in Riboflavin Biosynthesis........................ 24
    1.5.3 Crystal Structures of Lumazine Synthase............................... 25
    1.5.4 Catalytic Mechanism of Lumazine Synthase ......................... 28
    1.5.5 Molecular Assembly of Lumazine Synthase........................... 31

2 Crystal Structure of Lumazine Synthase from Aquifex aeolicus – Determinants of Thermostability (Paper I) ................................................................. 35
  2.1 Results and Discussion..................................................................... 35
  2.2 Conclusions ..................................................................................... 39

3 Structures of Aquifex aeolicus Lumazine Synthase in Complex with Inhibitors – A Model of the Reaction Process (Paper II) ........................................ 40
  3.1 Results and Discussion..................................................................... 40
  3.2 Conclusions ..................................................................................... 45

4 Multiple Assembly states of Lumazine Synthases – A Model Relating Catalytic Function and Molecular Assembly (Paper III)................................................. 46
  4.1 Results and Discussion..................................................................... 46
  4.2 Conclusions ..................................................................................... 51

5 A 180 subunit complex of a lumazine synthase mutant violates quasi-equivalence in capsid assembly (Paper IV) ................................................................. 52
  5.1 Results and Discussion..................................................................... 52
  5.2 Conclusions ..................................................................................... 55

6 Acknowledgements .................................................................................... 57

7 References .................................................................................................. 59
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>Three-dimensional</td>
</tr>
<tr>
<td>ASA</td>
<td>Solvent accessible surface area</td>
</tr>
<tr>
<td>AU</td>
<td>Asymmetric unit</td>
</tr>
<tr>
<td>B-factor</td>
<td>Temperature factor ($B = 8\pi^2 &lt;\mathbf{x}^2&gt;$, where $&lt;\mathbf{x}^2&gt;$ is the mean square displacement of the atom from its equilibrium position)</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DM</td>
<td>Density modification</td>
</tr>
<tr>
<td>DSC</td>
<td>Differential scanning calorimetry</td>
</tr>
<tr>
<td>EM</td>
<td>Electron microscopy</td>
</tr>
<tr>
<td>Fc</td>
<td>Calculated structure factor</td>
</tr>
<tr>
<td>Fo</td>
<td>Observed structure factor</td>
</tr>
<tr>
<td>FAD</td>
<td>Flavin-adenine dinucleotide</td>
</tr>
<tr>
<td>FMN</td>
<td>Riboflavin mononucleotide</td>
</tr>
<tr>
<td>LS</td>
<td>Lumazine synthase</td>
</tr>
<tr>
<td>LSAQ</td>
<td>Lumazine synthase from <em>Aquifex aeolicus</em></td>
</tr>
<tr>
<td>LSAQ-IDEA</td>
<td>Mutant of lumazine synthase from <em>Aquifex aeolicus</em> with the IDEA insertion</td>
</tr>
<tr>
<td>LSBS</td>
<td>Lumazine synthase from <em>Bacillus subtilis</em></td>
</tr>
<tr>
<td>MAD</td>
<td>Multiple-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>MIR</td>
<td>Multiple isomorphous replacement</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum likelihood</td>
</tr>
<tr>
<td>MR</td>
<td>Molecular replacement</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance spectroscopy</td>
</tr>
<tr>
<td>NCS</td>
<td>Non-crystallographic symmetries</td>
</tr>
<tr>
<td>ORF</td>
<td>Open-reading frame</td>
</tr>
<tr>
<td>RDL</td>
<td>6,7-dioxo-5H-8-ribitylaminolumazine</td>
</tr>
<tr>
<td>RNOP</td>
<td>6-ribitylamino-5-nitroso-2,4(1H,3H)-pyrimidine-dione</td>
</tr>
<tr>
<td>RNO2P</td>
<td>6-ribitylamino-5-nitro-2,4(1H,3H)-pyrimidine-dione</td>
</tr>
<tr>
<td>RPL</td>
<td>3-(7-hydroxy-8-ribityllumazine-6-yl)propionic acid</td>
</tr>
<tr>
<td>RPP</td>
<td>5-(6-D-ribitylamino-2,4(1H,3H)pyrimidine-dione-5-yl)pentyl-1-phosphonic acid</td>
</tr>
<tr>
<td>RS</td>
<td>Riboflavin synthase</td>
</tr>
<tr>
<td>SAD</td>
<td>Single-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small angle X-ray scattering</td>
</tr>
<tr>
<td>SIR</td>
<td>Single isomorphous replacement</td>
</tr>
<tr>
<td>T</td>
<td>Triangulation number</td>
</tr>
<tr>
<td>Tm</td>
<td>Melting temperature</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

1.1 STRUCTURE OF PROTEINS

Proteins are essential for all organisms. A special class of proteins, namely the enzymes, is involved in catalyzing almost all biochemical reactions in living cells. The function and activity of proteins depend on their three-dimensional (3D) structures. Therefore the knowledge of protein structures is a key for the understanding of many fundamental biological problems. After the discovery of the double helix structure of the DNA molecule and the determination of the first 3D protein structure of myoglobin by X-ray crystallography in the middle of last century, solving 3D structures of biological macromolecules has become one of the most demanding tasks in the field of biosciences. Thanks to the development and application of many new techniques such as gene sequencing and cloning, protein expression, purification and high-throughput crystallization, multi-dimensional nuclear magnetic resonance spectroscopy (NMR), cryogenic technique, the increasing accessibility of synchrotron X-ray beams as well as the development of computer programs, the determination of protein structures is becoming a standard technique. To date, more than 30,000 protein structures have been determined and all these structures together provide us with quite a comprehensive source of information for insights into many essential biological processes of the living systems.

Electron microscopy (EM), NMR spectroscopy and X-ray crystallography are the most important methods for protein structure determination. Using techniques such as 2D crystallization, sample freezing and image reconstruction, EM is now widely applied to study, at low resolution (usually at nm level), the structure of biological macromolecules including viral proteins, membrane proteins and complexes, which can virtually have a molecular weight up to millions of Daltons. NMR is a versatile technique used not only to determine protein structures in solution but also to study enzyme kinetics, ligand binding and molecular dynamics. However, the application of NMR for high-resolution structural analysis is limited to proteins with a molecular weight of less than 25–30 kDa. X-ray diffraction is the first and, until now, the most powerful technique to solve even very large structures of biological macromolecules. With diffraction by crystals, X-ray crystallography is able to provide structural details of the protein at atomic resolution.

1.2 X-RAY SCATTERING

X-rays are electromagnetic waves with a wavelength of 1000-0.1 Å. When electrons with high velocity are decelerated due to collision with metal atoms (conventional X-ray generators) or influenced by a magnetic field, which bends the electron beam (synchrotron radiation source), a part of the kinetic energy is emitted as X-rays.
The scattering by an electron is shown in Figure 1-1, where \( s_0 \) and \( s_1 \) represent the unit vectors of the incident and scattered waves, respectively.

![Figure 1-1 X-ray scattering by one electron \( e_1 \). The incident and scattered waves are represented using the unit vectors \( s_0 \) and \( s_1 \) with a modulus of \( 1/\lambda \). The “reflecting plane” is perpendicular to the scattering vector \( S \).](image)

The scattering vector \( S \), which illustrates the difference between the incident and scattered waves, is given by:

\[
S = (s_0 - s_1), \quad |S| = \frac{2\sin \theta}{\lambda}
\]  

(1)

The angle between the vectors \( s_0 \) and \( s_1 \) (20) is the scattering angle. Scattering by an electron can also be regarded as being “reflected” by a plane, which is perpendicular to the scattering vector \( S \) (Figure 1-1).

X-ray scattering by different objects, e.g. solutions and crystals of proteins, are based on the same electromagnetic theory; however they differ in their character.

**1.2.1 Solution Scattering (Small angle X-ray Scattering)**

Small angle X-ray scattering (SAXS), usually observed within a rather small angular range, e.g. 2-5°, occurs in sample solutions containing particles with a colloidal size (tens to several thousand Å). Due to the fact that the particles in solution are randomly orientated, the resulting scattering pattern is centro-symmetric (Figure 1-2a). X-ray scattering of protein solutions is a monotone function of the scattering angle (Figure 1-2b).

Suppose that the solution is considerably diluted, i.e. particles in solution do not interact with one another, it can be assumed that all particles make independent contributions to the intensity, which is given by:

\[
I(S) = 4\pi \int_0^\infty p(r) \frac{\sin(2\pi rS)}{2\pi rS} dr
\]

and \( p(r) = r^2 \gamma(r) \)

(2)

where \( p(r) \) is a distance distribution function describing the geometrical character of the particles in solution and \( \gamma(r) \) gives the probability of finding a volume element
Figure 1-2. Small angle scattering experiments (a) a schematic 3D view of solution scattering, colors from blue to white indicate the intensity from low to high. Due to the random orientation of the particles in solution, the pattern of solution scattering is centro-symmetric; (b) The scattering curve (solid) of lumazine synthase from *B. subtilis* overlaid with the calculated scattering function (dashed) for a hollow sphere, where $I(h) = 4\pi\sin\theta/\lambda$ and $\lambda$ is the X-ray wavelength; (c) the distance distribution function of the same sample indicates the size of the largest particle in solution.

(e.g. an electron) at a given position represented by the vector $r$.\textsuperscript{11, 12} The maximum distance within the particles ($D_{\text{max}}$) can be directly obtained from the distance distribution function $p(r)$ (Figure 1-2c).

In the vicinity of the origin (*i.e.* small $S$ or small $2\theta$, Figure 1-1), the scattered intensities can be approximated according to the Guinier law$^{13}$:
where the radius of gyration \( R_g \) gives a measure of the distance between the electrons and the center of gravity of the molecule.

In most cases, the application of small angle X-ray scattering can provide structural details such as size, shape and molecular weight indirectly; some of these parameters are calculated by fitting of the experimental data to theoretical scattering curves (Figure 1-2a).

### 1.2.2 Crystal Scattering

In the case of crystal scattering, an atom can be represented by the electron cloud \( \rho(r) \), the total scattering of all electrons of an atom is:

\[
\hat{f} = \int \rho(r) \exp[2\pi \mathbf{r} \cdot \mathbf{S}] d\mathbf{r}
\]

where the integration is over the entire space represented by the vector \( \mathbf{r} \). The function \( \hat{f} \) in (4) is called the atomic scattering factor, which represents the scattering power of an atom.

In common diffraction experiments, scattering (\( \hat{f} = f_0 \)) is usually assumed, i.e. the incident X-rays are mainly scattered by electrons in outer shells, which are considered as free electrons. The scattered beam has no phase change with respect to the incident beam and the atomic scattering factor will have a real value. The diffraction pattern of scattering is centrosymmetric, i.e. \( I(hk\ell) = I(h\bar{k}\ell) \). This is the so-called “Friedel’s law”.

However, when the incident X-ray beam has a certain wavelength, close to the absorption edge of the atom (element), the energy of the incident X-rays is strong enough to excite an electron from the inner shell to a higher orbital or just eject it. The resulting scattering will have a phase change with respect to the incident beam. This phenomenon is called anomalous scattering. The atomic scattering factor is then:

\[
f = f_0 + f^* + if^*
\]

---

\* A vector \( Acos(\alpha) + iAsin(\alpha) \) in the complex plane can also be written in the exponential way: \( A\exp[i\alpha] \), where \( A \) is the amplitude and \( \alpha \) is the angle between the vector and the real axis.

Properties of exponential terms:

\[ \exp(a) \times \exp(b) = \exp(a + b) ; \quad \exp(a)/\exp(b) = \exp(a - b) ; \quad \exp(k \cdot a) = [\exp(a)]^k \]
where \( f' \) and \( f'' \) are the real and imaginary components of the anomalous scattering factor. Anomalous scattering varies with change of the wavelength. The presence of anomalous scattering results in the inequivalent changes on phase of the reflections \( (hkl) \) and \( (\overline{hkl}) \), which will eventually lead to the breakdown of the Friedel’s law, i.e. \( I(hkl) \neq I(\overline{hkl}) \). The reflections \( (hkl) \) and \( (\overline{hkl}) \) are designated as the Bijvoet pairs. Anomalous scattering provides phase information, which can be used for structure determination (See 1.3.3).

Suppose a unit cell has \( n \) atoms at positions \( \mathbf{r}_j \) \((j = 1, 2, \ldots, n)\) with respect to the origin of the unit cell, the scattering of the unit cell is then the sum of scattering factors of all atoms in it (Figure 1-3):

\[
F(S) = \sum_{j=1}^{n} f_j \exp[2\pi i \mathbf{r}_j \cdot \mathbf{S}]
\]

\( F(S) \) is called the structure factor because it depends on the arrangement of atoms \((\mathbf{r}_j, \text{i.e. the structure})\) in the unit cell.

![Figure 1-3. The structure factor is the sum of the scattering by all atoms in the unit cell.](image)

A crystal contains a large number of unit cells arranged in three dimensional space defined by the unit vectors \( \mathbf{a}, \mathbf{b} \) and \( \mathbf{c} \). The position of any unit cell in the crystal can be represented by \( t \mathbf{a} + u \mathbf{b} + v \mathbf{c} \), in which \( t, u \) and \( v \) are whole numbers. The total scattering of the crystal is then obtained by a summation over all unit cells in it:

\[
K(S) = F(S) \times \sum_{j=0}^{n_1} \exp[2\pi ita \cdot S] \times \sum_{u=0}^{n_2} \exp[2\pi iub \cdot S] \times \sum_{v=0}^{n_3} \exp[2\pi ivc \cdot S]
\]

(6)

where \( n_1, n_2 \) and \( n_3 \) are the number of unit cells in the directions of \( \mathbf{a}, \mathbf{b} \) and \( \mathbf{c} \), respectively.
Figure 1-4. Each vector represents the scattering of one unit cell in the crystal. The directions of these vectors are different depending on the phase $t \cdot 2\pi a \cdot S$. Because $t$ is a large number, the resulting summation over the total scattering by the unit cells with $t = 0$ to $n1$ is almost always equal to zero.

Regarding the term $\exp[2\pi i a \cdot S]$ in equation (6) as a vector with a phase of $2\pi a \cdot S$ (Figure 1-4), the summation over a very large number (e.g. $n1$) is almost always equal to zero. The same conclusion applies on the other two terms $\exp[2\pi i b \cdot S]$ and $\exp[2\pi i c \cdot S]$ as well. However, when $a \cdot S$, $b \cdot S$ and $c \cdot S$ are all integers:

$$
\begin{align*}
a \cdot S &= h \\
b \cdot S &= k \\
c \cdot S &= l
\end{align*}
$$

namely $h$, $k$ and $l$ are whole numbers, all vectors have the phases of $2\pi$-integer, i.e. the direction of all vectors points to the right, thus the scattering by the crystal can be observed due to the amplification effect. Therefore, a crystal does not scatter X-rays unless equations (7) are fulfilled. They are the well-known “Laue conditions”. The numbers $h$, $k$ and $l$ in (7) are called the indices of reflections and the vector $S$ can now be written as $S(hk l)$.

Rearranging the Laue conditions (7) as:

$$
\begin{align*}
\frac{a}{h} \cdot S &= 1; \\
\frac{b}{k} \cdot S &= 1; \\
\frac{c}{l} \cdot S &= 1
\end{align*}
$$

the projections of the vectors $\frac{a}{h}$, $\frac{b}{k}$ and $\frac{c}{l}$ on the scattering vector $S$ all have the same length of $1/|S|$ (Figure 1-5a) *.

---

* Let vectors $a$ and $b$, with lengths $|a|$ and $|b|$, be inclined at an angle $\theta$. The scalar product is the number:

$$a \cdot b = ab \cos(\theta) \text{ and } a \cdot b = b \cdot a.$$  

The vector product is a vector $c$ with a length $|c| = absin(\theta)$ and points to the direction perpendicular to both $a$ and $b$. $a \times b = -b \times a$. 

---
Figure 1-5. (a) A 2D unit cell is drawn for simplicity. The endpoints of the vectors \( \mathbf{a}/h \), \( \mathbf{b}/k \) (and \( \mathbf{c}/l \)) form a lattice plane perpendicular to vector \( \mathbf{S} \), \( d \) is the distance between these lattice planes. (b) A graphic representation of Bragg’s law. Two lattice planes are separated by a distance \( d \). The condition for constructive interference is that the path difference between the two scattered beams is \( 2d \sin \theta = n\lambda \).

Letting this length \( 1/|\mathbf{S}| = d \) and from equation (1) where \( |\mathbf{S}| = \frac{2\sin \theta}{\lambda} \), Bragg’s law emerges:

\[
2d \sin \theta / \lambda = 1 
\]  

(8)

where \( d \) is the distance between two successive lattice planes defined by \((h \ k \ l)\). Bragg’s law (Figure 1-5b) describes the condition of diffraction as: the path difference between the beams scattered by two lattice planes is equal to \( n\lambda \) (\( n \) is an integer). It can also be explained as: the phase difference between the beams scattered by two lattice planes is equal to \( \pi \times 2n \) (\( n \) is an integer).

Considering a special vector \( \mathbf{S}(100) \), it is perpendicular to the plane (100) and has a length \( 1/d(100) \). If we call this vector \( \mathbf{a}^* \), in the same way \( \mathbf{S}(010) \perp \) plane (010) and \( \mathbf{S}(001) \perp \) plane (001), we can get the vectors \( \mathbf{b}^* \) (with a length of \( 1/d(010) \)) and \( \mathbf{c}^* \) (with a length of \( 1/d(001) \)), respectively. The vectors \( \mathbf{a}^*, \mathbf{b}^* \) and \( \mathbf{c}^* \) are related to the vectors \( \mathbf{a}, \mathbf{b} \) and \( \mathbf{c} \) as:

\[
\begin{align*}
\mathbf{a} \cdot \mathbf{a}^* &= h \\
\mathbf{b} \cdot \mathbf{b}^* &= k \\
\mathbf{c} \cdot \mathbf{c}^* &= l \\
\mathbf{a} \cdot \mathbf{b}^* &= \mathbf{a} \cdot \mathbf{c}^* = \mathbf{b} \cdot \mathbf{c}^* = 0
\end{align*}
\]  

(9)

Similar to the vectors \( \mathbf{a}, \mathbf{b} \) and \( \mathbf{c} \), which define the real space, the vectors \( \mathbf{a}^*, \mathbf{b}^* \) and \( \mathbf{c}^* \) define another space called the “reciprocal space”. Applying (9) to the structure factor (5) where

\[
\mathbf{r}_j \cdot \mathbf{S} = (x_j \mathbf{a} + y_j \mathbf{b} + z_j \mathbf{c}) \cdot (ha^* + kb^* + lc^*)
\]  

\[= hx_j + ky_j + lz_j \]
The structure factor $F(S)$ can then be written as $F(hkl)$ or $F(h)$:

$$F(hkl) = \sum_{j=1}^{n} f_j \exp[2\pi i (hx_j + ky_j + lz_j)]$$

**1.2.3 The Electron Density and the Phase Problem**

In equation (10) the summation is over all atoms $j$ in the unit cell. It can also be calculated by the integration of the electron density over the unit cell:

$$F(hkl) = \sum_{j=1}^{n} f_j \exp[2\pi i (hx_j + ky_j + lz_j)] = V \int_{cell} \rho(xyz) \exp[2\pi i (hx + ky + lz)] dxdydz$$

The Fourier transformation of the structure factor (11) gives the electron density function $\rho(xyz)$ (The mathematical principles of the Fourier transformation are given in Drenth, 2002):

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) \exp[-2\pi i (hx + ky + lz)]$$

Given the structure factors, equation (12) can be used to calculate the electron density of the crystal structure. However in X-ray diffraction experiments, the complete structure factors can not be obtained directly. Instead, only the intensities are accessible, the phase information is lost:

$$I(hkl) = F(hkl) \cdot F^*(hkl) = |F(hkl)|^2 \exp[-2\pi i \alpha] \cdot \exp[2\pi i \alpha] = c |F(hkl)|^2$$

and $|F(hkl)| = c \sqrt{I(hkl)}$

where $c/c'$ are parameters related to absorption and temperature. Shown in (13) only the amplitude of the structure factor $|F(hkl)|$ is obtained from the intensity; however the phase is not accessible, because it cancels out in the exponential terms. This is the so-called “phase problem” of X-ray crystallography.

**1.3 PROTEIN X-RAY CRYSTALLOGRAPHY**

**1.3.1 Protein Crystallization and Data Collection**

Protein crystallization is a process, which slowly and evenly decreases the protein solubility until the system reaches a supersaturated state. The thermodynamic force can then drive the system to a new equilibrium state with the formation of crystalline or amorphous precipitate. The crystallization process may be affected by many factors including the purity and homogeneity of samples, temperature, pressure, pH and the presence of crystallization nuclei. As a consequence, the results of crystallization are to a large degree unpredictable.
In practice, pure and homogenous protein with a concentration higher than 5 mg/ml is normally needed. A number of crystallization conditions may be screened. Crystallization buffers containing one or more precipitating agents such as salts, polyethylene glycol (PEG) and/or other additives like organic compounds, are prepared at different pHs. Protein samples are mixed with buffer solutions and the concentrations of precipitants are slowly increased by vapor diffusion, dialysis or other methods. Crystals that are larger than 200μm in all dimensions are usually required in order to obtain diffraction data at high resolution with good quality.

Diffraction data can be collected at either room temperature or cryogenic conditions, i.e. around 100 Kelvin. For data collection at room temperature, the crystal is transferred into a glass capillary. For data collection at low temperatures, a cryoprotectant, typically a water soluble organic material (e.g. 20% glycerol), is added to the crystallization buffer. The cryoprotectant slows down the rate of formation of the crystalline ice, which causes the damage of the crystal. Instead, a glass-like solid state is obtained by shock freezing. Collection at low temperature also reduces the radiation damage of the crystals.

Diffraction data are collected by different kinds of detectors, among which the image plate and area detectors are most often used nowadays. The image plate works in a similar way as the conventional film. It “records” the energy of the reflections by an inorganic storage phosphor and the reflections can be read out (scanned) and erased by light with different wavelengths. The original design of the area detector contains a chamber, which is filled with gas e.g. xenon. The gas atoms can be ionized by the diffracted X-rays and the ionization of the gas causes the formation of ions and electrons, which can be amplified and recorded electronically. At synchrotron X-ray beam stations, a new kind of area detector, the charge coupled device (CCD) is widely used to replace the gas chamber of the area detector.

A schematic drawing of the X-ray diffraction by a crystal is shown in Figure 1-6. As the orientation of the crystal is fixed, still exposure of a stationary crystal only contains a small number of reflections (Figure 1-6a). However when the crystal is rotated by a small angle (the oscillation angle Δφ) around an axis, which is perpendicular to the beam, more reflections are collected on one image (Figure 1-6b). In practice, many images, all collected with a small oscillation angle of the crystal, result in a full data set. This procedure is called “rotation method”. Parameters for data collection including the incident wavelength, starting angle, oscillation angle and total rotation range and the distance from crystal to the detector can be optimized according to the absorption, space group, cell dimensions, shape and orientation of the crystal as well as the quality of the crystal (e.g. the mosaicity). A diffraction image from the data collection of lumazine synthase from *Aquifex aeolicus* is shown in Figure 1-6c.
Figure 1-6. (a) A still exposure with a stationary crystal results in only a small number of reflections arranged in a set of narrow ellipses; (b) When the crystal is rotated, reflections from the same reciprocal lattice plane form a lune, limited by two ellipses corresponding to the start and end positions; (c) X-ray diffraction pattern of lumazine synthase from *Aquifex aeolicus*. The data were collected at the synchrotron beam line X11 at HASYLAB (EMBL Outstation, DESY, Hamburg). The wavelength of the incident X-ray beam was 0.909 Å and the highest resolution of this image is 2.0 Å.

After a full dataset is collected, the images need to be processed. This procedure is started by finding the spots on one (or several) chosen images. It is usually done by comparing the intensity of the spots and the local background on the image. Once the locations of these spots are determined, the coordinates of the spots (i.e. their positions on the image) are converted and mapped onto the surface of an imaginary Ewald sphere, from where the indices (hkl) of these spots are retrieved. Using the indices, the approximate unit cell parameters \(a, b, c, \alpha, \beta, \text{ and } \gamma\) and the relative orientation of the crystal with respect to the detector can be calculated. These parameters are then used to predict reflections on the image. The positions of the predicted reflections and the observed reflections are compared and the differences between the positions are minimized resulting in a set of refined parameters. The refined parameters, including the unit cell parameters, crystal orientation, beam center, crystal to detector distance, are then used to process the whole dataset. After all reflections are indexed, the data are processed by integrating intensities merging partial reflections and reducing the reflections to a unique set (applying the
crystallographic symmetries). Data processing can be done using automated computer programs like XDS and HKL.\textsuperscript{21, 22}

### 1.3.2 Symmetry and Symmetry Operations

As a consequence of energy optimization, molecular assembly and regular packing, molecules in crystal are usually arranged in a symmetric way.\textsuperscript{15} A symmetry operation is an action that can be performed on an object without changing the appearance of it. The symmetry operations include: rotation, mirror planes, translation, inversion and certain combinations of them. Mathematically, there are 230 different ways to combine these operations in the crystals leading to 230 possible space groups.\textsuperscript{23} However not all of them are allowed in protein crystals. The reason is, for instance, a mirror plane would change the chirality of an L-amino acid to a D-amino acid, which does not naturally occur in any protein.

The asymmetric unit of the crystal is the smallest unit required to generate a whole unit cell by the crystallographic symmetry operations (space group symmetry operations). However, the asymmetric unit can also contain more than one molecule, which are often related by symmetries. The symmetry relations between the molecules within an asymmetric unit are called the “non-crystallographic symmetries” (NCS) or the “local symmetries”. The NCS are sometimes not strictly followed due to the flexibility of the molecules. A good example for explaining both the crystallographic symmetries and the non-crystallographic symmetries is the icosahedral capsid in space group I\textsubscript{23} (\textbf{Figure 1-7}). The crystallographic symmetries of space group I\textsubscript{23} include the 2-fold and 3-fold rotation axes. However, the 5-fold rotation axes are local symmetries.

![Figure 1-7. The symmetries of an icosahedral capsid in space group I\textsubscript{23}: the crystallographic symmetries of the space group I\textsubscript{23} including the 2-folds and 3-folds are shown with filled-symbols. The non-crystallographic 5-folds are shown with hollow-symbols.](image)

A translation operation can be mathematically described as:

\[
A(x' y' z') = A(xyz) + T(t)
\]
where \((x,y,z)\) and \((x',y',z')\) are the coordinates before and after the operation and \(t\) is the translation vector. In the same way a rotation operation can be described as:

\[
A(x'y'z') = A(xyz)R(C)
\]  

where \(R(C)\) is a rotation operator and \(C\) is a 3x3 rotation matrix defining the rotation operation. A rotation operation can be graphically represented in 3D space in different ways, \(e.g.\) in the Eulerian system, the three angles \(\theta_1, \theta_2\) and \(\theta_3\) define a rotation (Figure 1-8). The Eulerian angle system can be converted to the polar angle system for the application by different programs, \(e.g.\) for construction of 2D stereogram plots.

![Figure 1-8](image)

Figure 1-8. A rotation operation can be represented using Eulerian angles. The rotation can be regarded as rotating the Cartesian system \((x, y, z)\) to \((x', y', z')\), the rotation is defined by the three angles \(\theta_1, \theta_2\) and \(\theta_3\).

### 1.3.3 Phasing Methods

During the past century, a number of methods have been developed for solving the phase problem in X-ray crystallography:

1. **Direct methods**, based on statistical and probability analysis, derive phase values \(ab\ initial\) from the observed structure amplitudes. Direct methods are normally used to determine structures of small molecules. It is also, very often in combination with other methods, used to determine heavy atom positions in macromolecules.\(^{24-28}\)

2. **Single/multiple isomorphous replacement** (SIR/MIR) are used to calculate initial phases by studying the differences in diffraction intensities between the native protein crystals and isomorphous crystals containing heavy atoms like Hg, Ag, Au, Pt … as reference scatterers. Isomorphous replacement was one of the fundamental methods to determine structures of new proteins (new folds).\(^{29-32}\)
(3) **Single-/Multiple-wavelength anomalous dispersion** (SAD/MAD) In the presence of anomalous scattering, Friedel’s law breaks down, i.e. \( I(hkl) \neq I(\bar{h}\bar{k}\bar{l}) \). SAD and MAD extract the information from the differences between the Bijvoet pairs to determine the positions of the anomalous scatterers, from which the phase values can be derived. Therefore both methods require the presence of anomalous scatterers, which are either naturally occurring (Fe, Mn, Cu) or specially introduced into the molecule by recombinant techniques (seleno-methionine) or by quick-soaking (bromine). SAD may also use weak anomalous scatterers such as sulfur, which exists in almost all proteins. SAD and MAD are regarded as the “high-throughput” phasing methods and are today widely applied for structure determination of new proteins.33-37

(4) Homologous proteins often have similar structures. **Molecular replacement** (MR) is a method, by which the initial phases are obtained using the structure of a homologous protein as the template. With the increasing number of determined protein structures, MR has become the most commonly used technique in the practice of solving the phase problem.38-40

Molecular replacement, as was used to determine all the structures in this work, will be described briefly in the following sections. For detailed discussions of the phasing methods mentioned above, the reader is referred to relevant literature of crystallography.14, 15

### 1.3.4 Patterson Function and Patterson Map

The Patterson function plays an essential role for solving the phase problem. It is combined with almost all other phasing methods in the process of phase determination. The Patterson function can be obtained by Fourier transformation \((F^*)\) of the intensities \( I(h) \) as the Fourier coefficients: \(^{41}\)

\[
\hat{P}(u) = \mathbb{F}[I(h)] = \mathbb{F}[F(h) \cdot F^*(h)] \\
= \mathbb{F}[F(h)] \cdot \mathbb{F}[F^*(h)] \\
= \rho(r) \cdot \rho(-r) \tag{14}
\]

By the convolution integral \(C(u) = \int f(r) \cdot g(u-r)dr \) we obtain:

\[
P(u) = \int \rho(r) \cdot \rho(r+u)dr \tag{15}
\]

By insertion of \( \rho(r) \) and \( \rho(r+u) \), it can be shown, that

\[
P(u) = V^{-1} \sum_{l=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |F(h)|^2 \exp[-2\pi i h \cdot u] \tag{16}
\]

Mathematically, the Patterson function is the convolution of the electron densities at positions \((x, y, z)\) and \((x+u, y+v, z+w)\). The coordinate system \((u)\)
defines a space known as the Patterson vector space or the Patterson space. In the Patterson space each peak indicates a vector between two atoms in a crystal unit cell. When positions \((x, y, z)\) and \((x+u, y+v, z+w)\) are both occupied by atoms, the Patterson Function \(P(u)\) will have a positive value, which is shown as a peak on the so-called “Patterson map”.

The Patterson map is centrosymmetric and the height of each peak is roughly proportional to the product of the atomic numbers of the two atoms that contribute to the vector. As a result, heavy atoms produce much higher peaks in the Patterson map with respect to those of the other peptide atoms. Peaks in a Patterson map are grouped into two categories: (1) those arising from vectors between atoms within the same molecule (self-vectors); (2) those arising from vectors of atoms from different symmetry-equivalent molecules (cross-vectors). In principle, a Patterson map can be directly used to calculate atom positions in a crystal of simple molecules. However the Patterson maps of macromolecules, such as proteins, contain a huge number of peaks (theoretically, the number of peaks is \(N^2 - N\), where \(N\) is the number of atoms in protein) and many of them may overlap making the interpretation of the map very complex. Therefore the analysis of the Patterson map is often combined with correlation methods in the procedure of phase determination.

1.3.5 Rotation, Translation and Molecular Replacement

Proteins belonging to a same family usually have identical or similar fold. In molecular replacement (MR), a homologous structural template is used to replace the unknown structure in the crystal and provides the information needed for the calculation of initial phases. The phases calculated from the structural template and the diffraction intensities of unknown protein can then be used for Fourier synthesis to compute the electron density map. In order to locate the exact position of the template in the crystal, rotation- and translation-searches are performed in Patterson space.

The rotation search is performed by calculating the rotation function \(R(C)\), in which \(P_{\text{Crys}}\) is the Patterson function of the crystal calculated using the observed intensities and \(P_{\text{Mod}}\) is the self-Patterson function of a molecule of the structural model (containing only the self-vectors). The integration in (17) is over the volume of the Patterson map \(u\), where the self-Patterson peaks are located. Applying the rotation operation \(C\) on \(P_{\text{Mod}}\), \(R(C)\) will give a maximum when the two functions, \(P_{\text{Crys}}\) and \(P_{\text{Mod}}\), superimpose.

\[
R(C) = \int_{\text{volume}} P_{\text{Crys}}(u)P_{\text{Mod}}(Cu)\,du
\]

The resulting rotation matrix \(C\) (derived from the Eulerian angle triplet, \(01, 02, 03\)) defines the relative orientation of template molecule with respect to the unknown
structure in its unit cell. After the rotation operation \( C \) is applied to the model, translation searches are performed in order to find the absolute position of the model in the crystal:\(^\text{44, 45}\)

\[
T(t) = \int P_{\text{Cryp}}(u)P_{\text{mod}}(u,t)du 
\] (18)

The translation search is performed in a similar way as the rotation search. However, only cross-vectors are used for the calculations, as self-vectors are not sensitive to translational operations. When cross-vectors of the two Patterson functions superimpose, \( T(t) \) has a maximum value with a corresponding translational vector \( t \). Thus the absolute position of the template structure in crystal of the unknown protein is obtained.\(^\text{39, 40, 46}\)

### 1.3.6 Density Modification

Initial phases obtained by the phasing methods discussed above contain errors, which may often result in a rather poor agreement between the density map and the model structure. The calculated phases need to be improved so that the quality of the density map can also be improved. This procedure is called “density modification” (DM).\(^\text{47, 48}\) Density modification combines different methods to improve the quality of both the phase and the electron density map.

**Solvent flattening:** Protein crystals contain about 30% to 70% solvent,\(^\text{49}\) most of which is disordered. These solvent molecules often confound the real density with background noise at the initial stage of model building and refinement. Therefore improvement of the electron densities can be obtained by flattening the residual density. The solvent flattening is done by estimating the solvent content of the crystal and defining the protein region using an “envelope” that covers the protein molecule.\(^\text{50-52}\)

**Averaging:** Most proteins form oligomers. In many cases the asymmetric unit of the protein crystal contains multiple copies of the subunits, which are related by the so called “local symmetries” or non-crystallographic symmetries (NCS). The information on NCS may be used to average the density map.\(^\text{50, 53}\) In this thesis, the quality of the electron density maps was improved dramatically after averaging against the non-crystallographic 5-fold (icosahedral 5-fold).

Other techniques, such as **histogram matching, solvent flipping/correction** are also used in DM, for detailed discussions, the readers are referred to the relevant literature.\(^\text{54-58}\)

Solvent flattening is normally performed before crystallographic refinement and model building (1.3.7), however density averaging may be applied throughout the whole refinement procedure.
1.3.7 Model Building and Crystallographic Refinement

Model building is a computer aided procedure, in which the model structure is fitted into the density map.\textsuperscript{59} However, errors introduced by model building may result in deviations between the calculated structure factors ($F_{\text{calc}}$) and the observed structure factors ($F_{\text{obs}}$). Crystallographic refinement is then performed in order to minimize the deviations between $F_{\text{obs}}$ and $F_{\text{calc}}$.

$$Q = \sum_{h} W_{h}(F_{\text{c}}(h) - F_{\text{v}}(h))^2 = \min$$ \hspace{1cm} (19)

where $W_{h}$ is a weight factor.

**Least squares refinement:** It was shown that structure refinement of macromolecules could be performed in reciprocal space using a least squares method on the individual atomic parameters. As the high resolution 3D structures of all the 20 amino acids are available, the structural information including stereo chemistry, bond length and bond angle can be parameterized and used in the target function for minimization\textsuperscript{60, 61}:

$$F_{\text{DT}}Q = w_{1}DF + w_{2}DD + w_{3}DT$$ \hspace{1cm} (20)

where $DF$ represents the differences between the observed and calculated structure factors; $DD$ restrains the stereochemistry and $DT$ is the deviation of coordinates between the target molecule and the model atoms.\textsuperscript{61} Least squares refinement requires a high ratio of observations ($F_{\text{obs}}$) / the number of parameters to refine ($N_{P}$) for convergence the refinement, therefore it is usually applied for the refinement of small molecules, which can often diffract to very high resolution.

**Maximum likelihood refinement:** assuming that the best structure model would be most consistent to observations (data), maximum likelihood (ML) can be used for crystallography refinement.\textsuperscript{62-64}

$$P(F_{\text{calc}}(\mathbf{r}) + \varepsilon(\mathbf{h})) = F_{\text{obs}}(\mathbf{h}))$$

$$= \prod_{h} \frac{1}{\sigma(h)\sqrt{2\pi}} \exp \left[ -\frac{(F_{\text{calc}}(\mathbf{r}) - F_{\text{obs}}(\mathbf{h}))^2}{2\sigma^2(h)} \right] = \max$$ \hspace{1cm} (21)

The assumption of using (ML) for crystallographic refinement is that the errors ($\varepsilon(\mathbf{h})$) in the magnitudes ($F_{\text{obs}}(\mathbf{h})$) are random variables distributed in accordance to a Gaussian law and the ultimate goal of ML is to find the coordinates ($\mathbf{r}$), which maximise the probability $P$, provided the experimental errors are $\varepsilon(\mathbf{h})$.

The basic concepts of using ML in crystallographic refinement are (a) given the current model, consistency is measured statistically by the probability that the
reflections would be observed; (b) if changes of the model make the observations more probable, the model gets better and the likelihood increases; (c) the probabilities include the effects of all sources of error. As the model gets better, the errors get smaller and the probabilities get sharper, which also would increase the likelihood.

Given a target function, many different mathematical methods can be used for the minimization. The most popular algorithms including the conjugate gradient method or simulated-annealing have been integrated in many crystallographic refinement software packages such as REFMAC (CCP4), XPLOR/CNS etc.\textsuperscript{65-67}

The whole refinement procedure is monitored by the so-called crystallographic R-factor:

\begin{equation}
R = \frac{\sum_{\text{all}} |F_{\text{obs}}| - k |F_{\text{calc}}|}{\sum_{\text{all}} |F_{\text{obs}}|}
\end{equation}

where \(k\) is a scale factor. It has been noticed that the crystallographic R-factor can be refined to a very low value, which, however does not necessarily stand for a good result. This is due to the bias introduced by model building and refinement. In order to overcome the disadvantage of the crystallographic R-factor, a cross-validation scheme was developed by Brünger\textsuperscript{68, 69}

\begin{equation}
R_{\text{free}} = \frac{\sum_{\text{all} \in \text{Test}} |F_{\text{obs}}| - k |F_{\text{calc}}|}{\sum_{\text{all} \in \text{Test}} |F_{\text{obs}}|}
\end{equation}

The free R-factor or \(R_{\text{free}}\) is calculated using a test set of the reflections (usually 5-10\% of the observed reflections), which are set aside from refinement and are therefore unbiased.

The quality of the refined structures can be verified using Ramachandran plot,\textsuperscript{70, 71} which examines the stereochemistry of the main-chain dihedral angles. The temperature factors \(B = 8\pi^2 \langle \overline{x}^2 \rangle\) (\(\langle \overline{x}^2 \rangle\) is the mean square displacement of the atom from its equilibrium position), as one of the refined parameters, can also be used to check the quality of the refined structure. Because the information on mobility is included in the temperature factor, residues that are flexible or poorly defined normally have higher B values. A properly refined structure would have an averaged B factor closed to the statistical value directly obtained from the reflection data from the Wilson plot.\textsuperscript{72}
1.4 HYPERTHERMOPHILES AND PROTEINS FROM HYPERTHERMOPHILES

1.4.1 Hyperthermophiles

Hyperthermophiles are microorganisms, which grow optimally at temperatures above 80 °C and are usually unable to reproduce at temperatures below 60 °C.73-75 Hyperthermophilic microorganisms live and thrive at geothermal and volcanic environments like solfataras, hot springs, geysers and deep-sea vents.76, 77

To adapt to the extreme environment of their biotopes, hyperthermophiles have developed some unique strategies. Most hyperthermophiles are chemolithoautotrophic, i.e. producing energy via inorganic redox reactions (chemolithotrophic) and utilizing only inorganic carbon sources, like CO₂, to build up the cell material (autotrophic).78 A novel type of glycerol ether membrane lipids, which dramatically increases the resistance against hydrolysis at high temperature and low pH, has been found in the hyperthermophilic bacterium *T. maritima*.79-81 Recent studies have revealed that reverse gyrase is the only hyperthermophile-specific protein indicating the important contribution of the enzyme to the stability of DNA molecules in hot environment.82, 83

![Figure 1-9. Phylogenetic tree based on 16S rRNA sequences.](image)

The branches of the hyperthermophiles are highlighted with bold lines.

From a phylogenetic point of view, hyperthermophiles appear in the deepest branches of the genealogy tree (Figure 1-9), studies on these organisms have a great impact on our understanding of the evolution of life on earth.85, 86
1.4.2 Aquifex aeolicus

Most hyperthermophiles belong to the domain of archaea, whereas only two families, Thermotogales and Aquificales belong to the domain of bacteria. *Aquifex aeolicus*, originally found in Sicily, is one of the most thermophilic organisms known to date. Successful isolation from hot springs at 95°C in the Yellowstone National Park has also been reported. A. *aeolicus* is chemolithoautotrophic and can only grow in a medium without organic source such as sugars, amino acid. *A. aeolicus* was cultured at 85°C in a H₂/CO₂/O₂ (79.5/19.5/1.0) atmosphere. The complete genome of *A. aeolicus* has been sequenced. Among 1,512 classified open-reading frames (ORFs), 1,105 have been identified for coding of proteins.

1.4.3 Protein Stability and Folding

Proteins isolated from hyperthermophiles usually display abnormal tolerance against heat, extreme pHs and high salinity. Enzymes from hyperthermophilic organisms often function optimally at temperatures above 80°C. Knowledge on the structural basis of protein stability is essential for the understanding of some fundamental problems in biology. The functional and structural features as well as the catalytic mechanisms of thermostable enzymes, once understood, may be used for modification and engineering of enzymes for industrial applications.

Protein stability and folding are strictly related. The stability of a protein is determined by the difference in the free energy between the *Native* (folded) state and the *Unfolded* (denatured) state:

\[
\Delta U G = G^U - G^N
\]

Suppose that the native (N) and the unfolded states (U) are in equilibrium described by the equilibrium constant:

\[
K = \frac{[U]}{[N]} \tag{25}
\]

a reversible unfolding reaction *Native* $\xrightleftharpoons{K}$ *Unfolded* can be represented in terms of thermodynamics. The difference in free energy between the unfolded and native states is then:

\[
\Delta _U G = G^U - G^N = -RT \ln K \tag{24'}
\]

where $R$ is the universal gas constant and $T$ is the absolute temperature in Kelvin.

There are two important forms of enthalpy as far as protein thermal denaturation is concerned: the Van't Hoff enthalpy ($\Delta _H ^{vH} (T)$), determined from the temperature dependence of the equilibrium constant and the enthalpy measured experimentally by a calorimeter ($\Delta _H ^{cal} (T)$, the area under the peak, see below). If
these enthalpies are equal, *i.e.* \( \Delta H^\text{V} = \Delta H^\text{C} \), the system is considered as a two-state system. Taking into account both the enthalpic and the entropic contributions to the free energy, a two-state reversible unfolding reaction can be described using the thermodynamic model:

\[
\Delta G = G^\text{V} - G^\text{S} = -RT \ln K = \Delta H - T \Delta S
\]

(24’’)

Equation (24’’) allows the calculation of the change of the free energy upon unfolding \( \Delta G \) as a function of the temperature, when the temperature dependence of \( \Delta H \) and \( \Delta S \) are both known.

![Figure 1-10](image)

Figure 1-10. The stability curve of a protein given by \( \Delta G(T) \).

A schematic stability curve of a protein is shown in Figure 1-10. The temperature of maximum stability is indicated by \( T_s \), where \( \Delta S \) is equal to zero. At any higher or lower temperature, the protein is less stable. The unfolding/denaturation of the protein occurs at the temperatures where \( \Delta G = 0 \). Theoretically, two such temperatures exist: the melting temperature \( (T_m) \), which indicates the temperature of thermal unfolding, and \( T_m' \), which indicates the temperature of “cold denaturation”.

The unfolding of proteins can be induced by temperature, pH and pressure changes or by adding of chemical denaturants, such as guanidine hydrochloride (GuHCl) or urea.

Differential scanning calorimetry (DSC) is widely used for studies of protein thermostability. In a DSC unit a sample cell and a reference cell are separately heated with a constant power input. The reference cell obtains slightly more power input from an offset heater, which thus creates a small temperature difference between the two cells. In order to keep temperature difference close to zero, the sample cell also receives more power input form a second heater. As the two cells have the same mass and volume, the difference of the energy input is a direct measure of the difference of
heat capacity. A number of thermodynamic data can be obtained by DSC experiments. The thermal unfolding profile of a DNA binding protein Sso7d measured by differential scanning calorimetry is shown in Figure 1-11 as an example. The melting temperature $T_m$ is indicated by the peak, the change of enthalpy ($\Delta H_m$) is calculated as the area under the curve and change of heat capacity ($\Delta C_p$) is measured by the difference between the pre- and post- transition baselines as shown in the picture. The change of free energy $\Delta G(T)$ upon thermal unfolding can be calculated.

![Figure 1-11. Differential scanning calorimetric measurement of the thermal unfolding of the DNA binding protein Sso7d from Sulfolobus solfataricus.](image)

In this thesis, the stability of lumazine synthase is studied by DSC. However the unfolding of this enzyme is irreversible, therefore only apparent melting temperatures were obtained and used for an approximate comparison of thermostability.

### 1.4.4 Dominant Forces of Protein Stability

It has been shown that proteins isolated from hyperthermophiles display dramatic stability against heat. Considerable effort has been made during the past decades to investigate the dominant forces for protein stability and folding. Structural comparisons, mutagenic studies as well as theoretical calculations have revealed a number of the most important forces responsible for the stability of proteins: the hydrophobic effect, hydrogen bonding, electrostatic interactions, the formation of disulfide-bridges, aromatic interactions, helix dipoles, binding of small molecules or ions, packing and oligomerization. The environmental factors, such as temperature, pH and salts, affect the stability by changing the contributions of the hydrophobic effect, hydrogen bonding and configurational entropy or other forces listed above.97, 98
Spassov et al pointed out that the adaptation of proteins from hyperthermophiles to their extreme environment requires optimization of these factors.99

The hydrophobic effect generally refers to the low solubility of hydrophobic (i.e., apolar) compounds in water. The hydrophobic effect in protein can also be defined as the energy associated with the transfer of hydrophobic surface from the protein interior to water.100, 101 In proteins, side chains of the apolar amino acid residues have a tendency to evade water. Applying the hydrophobic concept by Frank & Evans to proteins, Kauzmann suggested that hydrophobic effect is the dominating force for protein folding and thermostability.102-104 This theory has been supported by accumulating experimental evidences.105 Theoretical studies by Privalov, et. al. suggested that the energetically most favorable structure would correspond to those with minimized hydration of apolar atoms.106, 107

Hydrogen bonding refers to partial sharing of a hydrogen atom between a donor atom, to which it is covalently bound, and an acceptor atom which has a lone pair of electrons. In protein solutions, hydrogen bonds and hydrogen bond networks are formed between the main-chain atoms and side-chain atoms of the amino acids. Water molecules or other compounds from the solvent are also involved in formation of hydrogen bonds. Studies of model compounds and protein mutants suggested that the formation of hydrogen bonds contribute to stability of proteins.108

Earlier work by Perutz and Raidt109, 110 showed that in thermostable proteins the number of surface ion-pairs is increased. This observation was further confirmed by statistical analysis99 and a large number of studies based on structure comparisons.111-116 With increasing melting temperature, ion-pairs in protein show a tendency of being organized into large networks. A largest increase of the ion-pair content and the largest ion-pair network in hyperthermostable proteins reported to date have been observed in a comparison of glutamate dehydrogenase from Pyrococcus furiosus117 and the enzyme from the mesophilic bacterium Clostridium symbiosum.118 In this thesis, structural studies showed that the number of ion pairs and ionic networks is more than doubled in lumazine synthases form the hyperthermophilic bacterium A. aeolicus with respect to that of the enzyme from the mesophilic B. subtilis (Paper I).

The thermodynamic contribution to stability resulting from an increased number of surface ion pairs, albeit being found in many of the existing structures of hyperthermostable proteins, is still a matter of lively debate. However, in a study of salt-bridge stability at high temperatures, model calculations have suggested that a considerable energy barrier exists for the solvation (breaking) of a salt-bridge and that the height of this barrier increases with temperature.119 A similar barrier is not seen with isosteric hydrophobic groups. The presence of this energy barrier suggests an apparent role of salt-bridges in increasing the kinetic barrier towards thermal inactivation or unfolding. It has also been shown that the desolvation penalty is reduced because of a lower $\varepsilon_{\text{solv}}$ at high temperature.120
A number of approaches by site-directed mutagenesis of ion-pair interactions and ionic networks, performed in several laboratories, has still not given the clear answers that were expected. However, Vetriani et al. recently observed that the melting temperature and the half-life of glutamate dehydrogenase from Thermococcus littoralis are increased over the values of the native enzyme after introduction of a double mutant. The construction of a 16-residue ion-pair network in the less thermostable Thermotoga maritima glutamate dehydrogenase resulted in an enzyme with a half-life of 240 minutes at 85 °C, compared to the wild-type protein with a half-life of 210 minutes, suggesting increased kinetic stability of the mutant protein.

Other important forces and their contributions to protein stability are discussed in detail in Merphy (2001).

1.5 RIBOFLAVIN AND ENZYMES INVOLVED IN RIBOFLAVIN BIOSYNTHESIS

1.5.1 Riboflavin

Riboflavin (also known as vitamin B₂, Figure 1-12a) is the precursor of two coenzymes, namely riboflavin mononucleotide (FMN, Figure 1-12b) and flavin-adenine dinucleotide (FAD, Figure 1-12c), which are fundamental for the metabolism of carbohydrates, fats, and proteins into energy. FAD and FMN serve as electron carriers in oxidation-reduction reactions (Figure 1-13) catalyzed by flavoenzymes (Table 1-1).

![Figure 1-12. The chemical formulae of: (a) riboflavin; (b) riboflavin mononucleotide (FMN) and (c) flavin adenine dinucleotide (FAD).]
Figure 1-13. FAD and FMN transfer electrons in oxidation-reduction reactions. R represents the ribityl side chain of the compounds.

Table 1-1. Some flavoenzymes

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>Flavin nucleotide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty acyl-CoA dehydrogenase</td>
<td>FAD</td>
</tr>
<tr>
<td>Dihydrolipoyl dehydrogenase</td>
<td>FAD</td>
</tr>
<tr>
<td>Succinate dehydrogenase</td>
<td>FAD</td>
</tr>
<tr>
<td>α-Glycerophosphate dehydrogenase</td>
<td>FAD</td>
</tr>
<tr>
<td>NADH dehydrogenase</td>
<td>FMN</td>
</tr>
<tr>
<td>Glycolate dehydrogenase</td>
<td>FMN</td>
</tr>
</tbody>
</table>

It has been shown that flavoproteins play important roles in many biological processes, such as fetus development in birds and mammals,\textsuperscript{125,126} signal transduction in apoptosis,\textsuperscript{127} regulation of biological clocks\textsuperscript{128}. They are also involved in the biodegradation of aromatic pollutants,\textsuperscript{129} in photosynthesis as well as in the light dependent DNA reparation.\textsuperscript{130}

Bacteria, archaea, certain yeasts and plants produce riboflavin by endogenous synthesis. However mammals must take up riboflavin through their diet. The fact that riboflavin is indispensable for living cells makes the enzymes involved in riboflavin biosynthesis ideal targets for the development of new antimicrobial drugs,\textsuperscript{131-133} as highly specific inhibitors are unlikely to interfere with human metabolism.\textsuperscript{134,135}

1.5.2 Enzymes involved in Riboflavin Biosynthesis

Enzymes involved in the last two steps of riboflavin biosynthesis were first isolated and purified in the form of a large complex from the cell contents of \textit{Bacillus subtilis} in 1980 at the Technical University of Munich.\textsuperscript{136} The complex, initially designated as “heavy riboflavin synthase”, has a total molecular weight of more than 1 MDa. Further studies characterized two types of functional subunits in the complex, namely the α subunits (MW = 23500 Da) and the β subunits (MW = 16000 Da).\textsuperscript{10,137}

The α subunit of heavy riboflavin synthase catalyzes the dismutation of 6,7-dimethyl-8-ribityllumazine (Figure 1-14, 3) resulting in the formation of riboflavin (Figure 1-14, 4) and was therefore named “riboflavin synthase”.\textsuperscript{138} The mechanism of the dismutation of 6,7-dimethyl-8-ribityllumazine (Figure 1-14, 3) has been investigated in a number of studies.\textsuperscript{139-142} Recent structural studies of the riboflavin synthase revealed more details of the catalytic mechanism.\textsuperscript{143-148} Further studies on structure and function of riboflavin synthase are underway.\textsuperscript{149-152}
Figure 1-14. The biosynthesis of riboflavin.

Earlier ligand binding experiments indicated that the β subunit plays a role in catalyzing the biosynthesis of 6,7-dimethyl-8-ribityllumazine (Figure 1-14, 3), the precursor of riboflavin. However, the exact function of the β subunit remained obscure for almost 40 years until the 4-carbon substrate was identified as (3S)-3,4-dihydroxy-2-butanone-4-phosphate (Figure 1-14, 2). It was then clear that the β subunit catalyze the formation of 6,7-dimethyl-8-ribityllumazine (Figure 1-14, 3) from 5-amino-6-ribitylamino-2,4-(1H,3H)pyrimidine-dione (Figure 1-14, 1) and (3S)-3,4-dihydroxy-2-butanone-4-phosphate (Figure 1-14, 2). The β subunit of heavy riboflavin synthase was thus named as lumazine synthase (LS).

1.5.3 Crystal Structures of Lumazine Synthase

Structural studies of lumazine synthase were pioneered by R. Ladenstein and A. Bacher et al. The first crystal structure of the enzyme from *Bacillus subtilis* (LSBS) was determined in the late 80’s of last century. It was shown that the enzyme forms a virus-like icosahedral capsid with a molecular weight of about 1MDa. Sixty subunits are arranged in twelve pentameric building blocks on the basis of icosahedral symmetries giving rise to a hollow capsid with inner and outer diameters of 80 Å and 156 Å, respectively.

In the subsequent 15 years, structures of lumazine synthase from a number of other organisms, including *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Magnaportha grisea*, *Mycobacterium tuberculosis*, *Escherichia coli*, *Spinacia oleracea*, *Brucella abortus* and *Aquifex aelutus* (this work) have been investigated. It has been shown that the enzyme exists in three association forms in crystals: icosahedral capsids in *B. subtilis*, *E. coli* and *S. oleracea*; pentamers in *S. pombe* and *S. cerevisiae* as well as stacking pentamers in *Brucella sp* and *M. tuberculosis*. (Earlier studies on the multiple assembly states of the enzyme in solution are summarized in section 1.5.5.)
Figure 1-15. Crystal structures of lumazine synthase: (a) a subunit of the enzyme from *B. subtilis*; (b) a pentamer of the enzyme from *B. subtilis*; (c) a capsid of the enzyme from *B. subtilis*; (d) a pentamer of the enzyme from *S. cerevisiae*. The arrows indicate the positions of the N-termini.

Despite the differences in assembly states, the subunit structures are nearly identical in different species. A subunit of lumazine synthase is formed by a four-stranded $\beta$-sheet flanked by two $\alpha$-helices on one side and three $\alpha$-helices on the other side (Figure 1-15a). In the icosahedral enzymes, the first seven N-terminal residues form an inter-subunit $\beta$-sheet with the neighboring monomer and this pattern is repeated five times leading to the formation of a pentamer (Figure 1-15b). Twelve pentamers are arranged in an icosahedral capsid (Figure 1-15c). The inner void of the capsid is connected to the solvent space through channels along the non-crystallographic 5-folds at the center of every pentamer.

In the pentameric enzymes from *S. pombe* and *S. cerevisiae*, the N-termini do not form an inter-subunit $\beta$-sheet with the adjacent monomer as observed in the icosahedral enzymes.\textsuperscript{166, 167, 170} As a result, the N-termini are more flexible and exposed to the solvent. A pentamer from *S. cerevisiae* is shown in Figure 1-15d.
Figure 1-16. Amino acid sequence alignment of lumazine synthase from 8 species. Icosahedral species are highlighted. Fully conserved residues are highlighted. Residues forming the active site pocket are marked with "+". Position of the N-terminal proline, which is unique for pentameric LS, is marked with "v". Positions of the single-site mutations, which lead to the formation of large capsids, are marked with "v". Positions of the 2-4-residue insertion, which is unique for the pentameric enzymes, is marked with "v". Secondary structure elements are labeled according to the crystal structure of lumazine synthase from *A. aeolicus*.

The substrate binding sites are located at the interface between two adjacent subunits within a pentamer. The secondary structural segments involved in ligand binding are β-turns (21-24, 54-58, 81-92) from one subunit and α-helices (127'-142') as well as a β-strand (113'-116') from the neighboring subunit (Figure 1-16). The structures of the active site in enzymes from different species are highly identical. In the structural studies of LS in complex with a number of designed inhibitors, the
detailed substrate binding schemes were revealed. The results are briefly summarized in the next section.

1.5.4 Catalytic Mechanism of Lumazine Synthase

The mechanism of this enzymatic reaction was investigated in a number of structural and kinetic studies after the function of the β-subunit of heavy riboflavin synthase was identified as catalyzing the formation of 6,7-dimethyl-8-ribityllumazine.155, 156

Substrate binding and the conformation of the active sites were revealed by structural analysis of the enzymes from Bacillus subtilis, Saccharomyces cerevisiae, Spinacea oleracea and Magnaporthe grisea in complex with different inhibitor compounds. The substrate binding pocket of LS from B. subtilis is shown in Figure 1-17a.164, 165 Arg127 forms a salt bridge with the phosphate ion. It was thus proposed that Arg127 is mainly responsible for the binding of substrate 2 via an ionic contact.165 His88 is in close proximity to the substrate analogue indicating the possibility of serving as a proton donor / acceptor during catalysis.165 The pyrimidine ring of the inhibitor was found in a parallel conformation with respect to the aromatic ring of residue 22 (Phe22/Trp22). Theoretical studies suggested that offset stacking aromatic interactions (π-π / σ-π) in protein energetically stabilize the structure.175-180 Thus it was suggested that Phe22 facilitates the binding of substrate 1 and the reaction intermediate in the active site. The structure of the empty active site of LS has been studied and described in LS from Brucella abortus173 and Aquifex aeolicus (paper I). The observed conformational changes, revealed by comparisons with the occupied active sites of the enzymes in complex with inhibitors, have suggested an induced-fit mechanism of substrate binding in LS. The aromatic ring of Phe22 adapts to the pyrimidine system by a swing movement, which results in a stacking aromatic interaction and proper orientation of the pyrimidine.
Hydrogen bond networks are formed between the inhibitor and protein. As shown in the schematic drawing of the inhibitor binding interactions (Figure 1-17b), the hydroxyl groups of the ribityl moiety of the inhibitor compound form hydrogen bonds with the residues Gly55, Ser56, Phe57, Glu58, Phe113' and Lys135'. The keto groups of the pyrimidine ring form hydrogen bonds with Asn23, Ala56, Thr80 and Ile82. The active site of the pentameric enzymes is similar to that of the LSBS, detailed comparison and analysis have been presented in earlier work.166-170

On the basis of crystal structures of LS, mutagenic studies on LSBS have confirmed a list of residues that are essential for enzyme activity (Table 1-2).181 In addition, in vitro experiments on the wild-type and the mutants of LS have revealed considerable insights into the catalytic mechanism of the enzyme.149, 182-189

<table>
<thead>
<tr>
<th>Table 1-2. Approximate reaction rates of lumazine synthase mutants (% of wild-type enzyme activity).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Phe22</td>
</tr>
<tr>
<td>Ala56</td>
</tr>
<tr>
<td>Phe57</td>
</tr>
<tr>
<td>Phe61</td>
</tr>
<tr>
<td>Val80</td>
</tr>
<tr>
<td>Ile82</td>
</tr>
<tr>
<td>Phe113</td>
</tr>
<tr>
<td>Asn23</td>
</tr>
<tr>
<td>Thr90</td>
</tr>
<tr>
<td>Ser142</td>
</tr>
<tr>
<td>Glu58</td>
</tr>
<tr>
<td>Glu126</td>
</tr>
<tr>
<td>Asp138</td>
</tr>
<tr>
<td>His88</td>
</tr>
<tr>
<td>Arg127</td>
</tr>
<tr>
<td>Lys131</td>
</tr>
<tr>
<td>Lys135</td>
</tr>
</tbody>
</table>

It was shown that the K_m value for substrate 2 (130 μM) exceeds that of the pyrimidine substrate 1 (5 μM) by more than 20 times.184 The kinetic data have also revealed that in absence of the pyrimidine substrate 1, lumazine synthase does not catalyze the proton exchange on the chiral C-3 atom of 2 with solvent water, nor does the enzyme function as a racemase. The spontaneous proton exchange at the methyl group of 2 is not accelerated by the enzyme. It was also shown that the enzyme does not catalyze the formation of diacetyl from 2. It was therefore suggested that the initial step of the enzyme-catalyzed reaction requires the presence of the second substrate (the pyrimidine 1) at the active site. These results indicate an ordered bi-bi mechanism of the reaction, however no conclusive evidence supporting this hypothesis has been obtained.184

Earlier in vitro experiments suggested that the regio-specificity of the reaction is about 90%.190-192 This result was later confirmed by experiment using the 13C labeled substrate 2, which was prepared in situ. The 13C atoms were found in lumazine and riboflavin at positions shown in Figure 1-18. Taking into account the
experimental sensitivity, Kis et al concluded that the regio-specificity of the enzymatic reaction is at least 98%.

![Figure 1-18](image)

Figure 1-18. The regio-specificity of the lumazine synthase catalyzed reaction. The carbon 1 atom of substrate 2 is labeled with an asterisk.

It has also been shown that both enantiomers of the four carbon compound 3,4-dihydroxy-2-butanone-4-phosphate, namely the 3S-2 and 3R-2 enantiomers (Figure 1-19), can serve as substrates for lumazine synthase, nonetheless stereo-specificity of the enzyme was detected. The reaction rate of the natural S-enantiomer is about 5-6 times higher than that of the R-enantiomer.

![Figure 1-19](image)

Figure 1-19. The 3S (left) and 3R (right) enantiomers of the substrate 3,4-dihydroxy-2-butanone-4-phosphate. The chiral carbon atom is labeled with an asterisk.

Based on the results of both structural and kinetic studies, Kis et al postulated a reaction mechanism (Figure 1-20) initiated by the formation of a Schiff base (6) upon the reaction of the 5-amino group of the pyrimidine 1 with the carbonyl group of 2. This step could be followed by proton abstraction and phosphate elimination. The resulting intermediate (7) contains a double bond, which is in favorable conjugation with the pyrimidine system. This enolate intermediate could tautomerize under formation of a carbonyl group, which would then be attacked by the 6-amino group, leading to the closure of the second ring. In the final step of the reaction, a water molecule would be released and the heterocyclic double ring system of the product 3 would thus be in an energetically favorable conjugation.
1.5.5 Molecular Assembly of Lumazine Synthase

The function of protein oligomers is often related to their assembly state. A paradigm for studying molecular assembly represent the viral proteins, which very often form icosahedral capsids enclosing DNA or RNA molecules.

A simple icosahedral capsid consists of 60 copies of the same subunit, which is the largest possible number of subunits that can be assembled with each of them in an identical (equivalent) environment. However, most viral proteins display icosahedral symmetry but contain more than 60 subunits. In the light of the pioneering studies of virus structures including the tobacco mosaic virus (TMV), tomato bushy stunt virus (TBSV) and turnip yellow mosaic virus (TYMV), Caspar and Klug developed the theory of quasi-equivalence of viral protein assembly. Caspar and Klug demonstrated that an icosahedral capsid can be constructed by inserting pentamers in a hexagonal lattice at certain positions described by the triangulation number \( T = h^2 + hk + k^2 \) (Figure 1-21).

With \( T > 1 \) the resulting capsid consists of more than 60 subunits arranged in hexamers and pentamers, which would show similar conformations and inter-subunit contact environments. Thus the pentameric and hexameric building blocks are regarded as "quasi-equivalent". Pentamers and hexamers are generally considered as the basic building blocks of a quasi-equivalent icosahedral capsid. (For a detailed
discussion of the principles of quasi-equivalence, please refer to Johnson et al. 1997.)

Figure 1-21. The construction of icosahedral capsids: (a) inserting pentamers into the hexagonal lattices at positions defined by the triangulation number $T = h^2 + 3h + k^2$ results in a closed icosahedral capsid. Replacing hexamers with pentamers at positions (0, 1) and (1, 0), indicated by the red triangle, results in a closed $T = 1$ capsid (b); similarly, a $T = 3$ capsid (c) can be constructed by replacing the hexamers at positions indicated by the blue triangle in (a). Inserting pentamers at positions indicated by the green and yellow triangles in (a) results in icosahedral capsids with $T = 7$ and $T = 13$, respectively.

A large body of structural evidence supports the quasi-equivalence theory, however, examples which may disobey part of these rules were also found. It was shown that the same polypeptide chains could assume different conformations in a virus shell. The in vitro assembly experiments of the bacterial virus P22 indicated that the formation of pentamers and hexamers did not seem to be obligatory for the association of the icosahedral capsid. It was shown by this experiment that although the matured capsid contains quasi-equivalent pentamers and hexamers, the capsid could be assembled from a solution containing only the monomers but without the presence of pentamers and/or hexamers, i.e. the monomers are the basic building blocks of the capsid.

A theory based on the so-called “local rules”, which illustrates virus shell assembly from a geometrical point of view, was introduced by Berger et al. It was proposed that the self-assembly of icosahedral capsids may be regarded as only relying on the “lower-level” interactions of each polypeptide chain with its neighbors, rather than the interactions of large structural building blocks, e.g. pentamers or hexamers. According to this theory, the “lower-level” interactions simply represent all non-covalent inter-molecular interactions. Computer based simulations followed
by energy minimization have successfully modeled the assembly of several viral shell proteins.211-213

Together with the pyruvate dehydrogenase (PDH) complex,214-217 the tricorn protease218-221 and the light-harvesting complex of the photosystem in pea chloroplast,222, 223 lumazine synthase is among the four enzymes that are able to form icosahedral capsids. Lumazine synthase was in fact the first known enzyme complex with icosahedral symmetry. As mentioned above, besides the existence of LS pentamers and T = 1 icosahedral capsids, earlier pH-induced dissociation experiments have revealed another association form. LSBS was observed to reassemble to large hollow capsids in Tris hydrochloride buffer at pH > 7.5. Small angle X-ray scattering and negative staining EM analysis showed that the large capsids were about 300 Å in diameter. Ultracentrifugation showed that these aggregates had a molecular mass of about 3 MDa indicating the existence of a T = 3 icosahedral assembly. It was further observed that the substrate analogue, 5-nitroso-6-ribitylamino-2,4(1H,3H)-pyrimidine-dione, could trigger the reassociation of large particles to empty T = 1 capsids.10, 163, 224

Structural determinants that may lead to the difference in assembly forms of lumazine syntheses have been discussed extensively. Braden et al. proposed that the enzyme with a five residue-kink, i.e. GxKxG, in the C-terminal helix would form an icosahedral capsid.173 Structure comparisons showed that the N-termini of all the pentameric enzymes contain a proline residue (Figure 1-16). The unique backbone torsion angle of proline may prevent the formation of the inter-subunit β-sheet, which is important for capsid assembly and stability. It was therefore proposed that the enzyme with a proline in the N-terminus would favor the pentameric assembly.168 In icosahedral LSs the loop linking helix α4 and helix α5 is positioned at the interface between two pentamers. Sequence alignment showed that an insertion of one to four residues after Gly129 (Figure 1-16) is unique for all pentameric LS. Computer modeling indicated that these extra residues, if inserted in an icosahedral capsid, would lead to clashes at the subunit interface between adjacent pentamers of icosahedral capsid (Figure 1-22).167 It was then proposed that lumazine syntheses with the four-residue insertion would rather assemble to pentamers.167 Recent sequence analysis and quaternary structure comparisons of lumazine synthase from different species have identified in total eight regions that appeared to be important for the assembly state.225

As the sixty active sites of icosahedral LS are located close to the inner surface of the capsid, substrates and products need to be transported to and from the active site through a rather densely packed capsid wall. The largest openings in the icosahedral shell are channels with a diameter of 5-12 Å located along the local 5-fold axes. The attempt to block the channels of “heavy riboflavin synthase” (complex formed of 60 LS subunits and 3 RS subunits inside of the LS capsid) from B. subtilis with a five-fold-symmetric tungsten-compound ([NaP5W30O110]14-) resulted in
unperturbed enzymatic activity.\textsuperscript{161} The results implied that the transport of substrates and products as well as the catalysis seem to be related to the assembly state of LS or to local perturbations causing widening of the icosahedral capsid (see section 5).

\textbf{Figure 1-22.} The subunit interface between adjacent pentamers of the icosahedral LS from \textit{B. subtilis} (gray); a subunit of the pentameric \textit{S. cerevisiae} LS (blue) is superimposed on one of the LSBS subunits. Note: the clash of the loop is indicated by an arrow.\textsuperscript{167}
2 CRYSTAL STRUCTURE OF LUMAZINE SYNTHASE FROM *AQUIFEX AEOLICUS* – DETERMINANTS OF THERMOSTABILITY (PAPER I)

To obtain sufficient protein material for structure analysis, an open reading frame (ORF) optimized for the expression of 6,7-dimethyl-8-ribityllumazine synthase from the hyperthermophilic bacterium *Aquifex aeolicus* (LSAQ) was synthesized and expressed in an *Escherichia coli* strain to a level of around 15%. The protein was purified by heat-treatment and gel-filtration. The protein was crystallized in the cubic space group I23 with cell dimensions a = b = c = 180.8 Å and diffraction data were collected to 1.6 Å resolution at the synchrotron beam line at DESY (Hamburg, Germany). The structure was solved by molecular replacement using lumazine synthase from *Bacillus subtilis* as the search model. The structure of the *A. aeolicus* enzyme was refined to a resolution of 1.6 Å. The empty protein capsid consists of 60 identical subunits with strict icosahedral symmetries. The apparent melting temperature measured by a differential scanning calorimetry (DSC) is 120°C. The structure was compared with other icosahedral and pentameric lumazine synthases and the possible determinants of thermostability were discussed.

2.1 RESULTS AND DISCUSSION

The recombinant lumazine synthase from *A. aeolicus* was crystallized in a solution containing 12% (W/V) PEG 4000, 0.1 M lithium sulfate and 0.1 M sodium citrate at pH = 5.6 using the sitting drop vapor diffusion method. Cubic crystals were obtained within two weeks and the diffraction data were collected at the synchrotron beam line X11 at HASYLAB (EMBL Outstation, DESY, Hamburg). The parameters for data collection and the results of data evaluation are shown in Table 2-1.

<table>
<thead>
<tr>
<th>Table 2-1. Data collection and evaluation of LSAQ.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>Observations/Unique (I &gt; 0)</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
</tr>
</tbody>
</table>

The structure of LSAQ was determined by molecular replacement using the structure of lumazine synthase from *B. subtilis* as the search model. The structure was refined to a resolution of 1.6 Å using the programs Refmac and CNS. During refinement, electron density maps were averaged against the non-crystallographic 5-fold axis. The stereochemistry of the refined structure was examined using the
Ramachandran plot.\textsuperscript{70} Results of the crystallographic refinement are shown in Table 2-2.

<table>
<thead>
<tr>
<th>Table 2-2. Results of crystallographic refinement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of atoms</strong>\textsuperscript{a}</td>
</tr>
<tr>
<td>Protein</td>
</tr>
<tr>
<td>Ligand</td>
</tr>
<tr>
<td>Solvent</td>
</tr>
<tr>
<td><strong>The refinement</strong></td>
</tr>
<tr>
<td>Resolution range for refinement (Å)</td>
</tr>
<tr>
<td>R-factor (%)</td>
</tr>
<tr>
<td>Free R-factor (%)\textsuperscript{b}</td>
</tr>
<tr>
<td><strong>Ramachandran plot</strong></td>
</tr>
<tr>
<td>Most favored regions (%)</td>
</tr>
<tr>
<td>Allowed regions (%)</td>
</tr>
<tr>
<td>Disallowed regions (%)</td>
</tr>
<tr>
<td><strong>B-factors (Å\textsuperscript{2})</strong></td>
</tr>
<tr>
<td>Wilson plot</td>
</tr>
<tr>
<td>All atoms</td>
</tr>
<tr>
<td>Protein atoms</td>
</tr>
<tr>
<td>Solvent atoms</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Numbers were counted per asymmetric unit.

\textsuperscript{b}5 % of the unique reflections were set aside for the calculation of the free R-factor.

The overall structure of lumazine synthase from *A. aeolicus* is similar to that of the enzyme from *B. subtilis* (Figure 2-1). Each asymmetric unit of the crystal cell contains a pentamer. Twelve pentamers are arranged in a $T = 1$ icosahedral capsid. Apart from the N-terminal and C-terminal ends, the largest difference of main chain $C_\alpha$ atoms of LSAQ and LSBS are in the loop region between helix $\alpha_2$ and the strand $\beta_3$ with a maximum deviation of 1.8 Å at residue Lys69. The region is situated at the outer surface of the capsid and is fully exposed to the solvent. The averaged r.m.s deviation of the atom positions between the LSAQ subunit and LSBS subunit is 0.8 Å.

Thermal unfolding experiments of lumazine synthase from *B. subtilis*, *A. aeolicus* and *S. cerevisiae* were performed using differential scanning calorimetry (DSC). Due to the irreversibility of the thermal unfolding process, only the apparent melting temperatures ($T_m$) could be obtained. The apparent $T_m$ of LSAQ reached 120°C, which is one of the highest melting temperatures for a protein reported to date. Nevertheless, the apparent $T_m$ of LSBS is 93°C, which is unusually high among those of mesophilic proteins. The $T_m$ of the pentameric lumazine synthase from *S. cerevisiae* was 74°C (Figure 2-2).
Figure 2-1. Structure alignment of lumazine synthase from *A. aeolicus* (black) and *B. subtilis* (gray). The loop region between the helix $\alpha_2$ and the strand $\beta_3$ with the maximum backbone deviation is denoted with an arrow.

Figure 2-2. Melting profiles of lumazine synthase from *S. cerevisiae*, *B. subtilis* and *A. aeolicus*.

On binding of the inhibitor compounds 5-(6-D-ribitylamino-2,4(1H,3H)pyrimidine-dione-5-yl)pentyl-1-phosphonic acid (RPP) and 6-ribitylamino-5-nitroso-2,4(1H,3H)pyrimidine-dione (RNOP), the apparent $T_m$ of LSBS increased by about 8°C and 6°C, respectively indicating stabilization of the structure by inhibitor binding (Melting curves not shown).
In order to reveal the structural determinants responsible for thermostability, the structure of LSAQ was compared with those of the other icosahedral and pentameric enzymes. The solvent accessible surface areas (ASA) were calculated using the program CNS (Figure 2-3).

Figure 2-3. Accessible surface area calculations on lumazine synthase from *A. aeolicus* (LSAQ), *B. subtilis* (LSBS), *S. oleracea* (LSSO), *S. cerevisiae* (LSSC), *M. grisea* (LSMG) and *B. abortus* (LSBA).

Generally speaking, the icosahedral enzymes have a smaller hydrophobic ASA with respect to the pentameric species. This is due to the formation of capsids. Among all the compared species, LSAQ has the smallest hydrophobic ASA (27%) indicating the predominate role of the reduced hydrophobic ASA for the stability of this enzyme.

Among the icosahedral enzymes, LSAQ has the largest charged surface (14.4%) indicating a stabilizing effect of the increased surface ion pairs and ionic networks. A comparison of the capsid surfaces of LSBS and LSAQ are shown in Figure 2-4.

The statistics of surface ion-pair and ionic networks are summarized in Table 5 and 6 of paper I. The results clearly show that lumazine synthase from *A. aeolicus* has the largest number of surface ion-pairs (0.11 / residue) and ionic networks (0.065 / residue) compared to those of the enzyme from *B. subtilis* (0.052, 0.032), *S. cerevisiae* (0.083, 0.043) and *M. grisea* (0.057, 0.027). Structural comparisons of the icosahedral LS’s revealed several highly conserved ion pairs at the subunit interface. Interestingly, arginine pairs in stacking conformation were involved in most of the conserved ion pair networks (Figure 5 of paper I). Theoretical studies by Karshikoff et al. have demonstrated that arginine pairs can behave as single titratable sites and that the titration curves may have a well-defined sigmoidal shape with an apparent pK
value of about 7. It was therefore suggested that the earlier reported instability of *B. subtilis* LS in 0.1 M Tris-HCl at pH > 7 might be due to the disruption of the ionic interactions at the interface between two pentamer blocks, initiated by decharging of the central arginine pair. 163

![Figure 2-4](image.png)

**Figure 2-4.** Accessible surface areas of lumazine synthase from (a) *A. aeolicus* and (b) *B. subtilis*. Atom-wise color codes are: red for negatively charged surface; blue for positively charged surface; green for polar surface and white for hydrophobic surface.

### 2.2 CONCLUSIONS

As the apparent $T_m$ of the pentameric LS from *S. cerevisiae* is much lower than that of the icosahedral LSBS and LSAQ, it is suggested that capsid assembly has a great impact on thermostability of the icosahedral enzymes. Moreover, binding of the inhibitor compounds stabilizes the enzyme.

Compared to that of the pentameric enzymes, the hydrophobic ASA of the icosahedral species are generally smaller, indicating the contribution of the hydrophobic effect to thermostability.

Among all the studied structures, LSAQ has the largest number of surface ion-pairs and networks. Thus the increased ion pairs and ionic networks at the subunit interface are suggested to play a role for capsid assembly and contribute to the overall stability. It is further proposed that the arginine pairs at the subunit interface may be responsible for the pH dependence of the capsid assembly.
3 STRUCTURES OF *AQUIFEX AEOLICUS* LUMAZINE SYNTHASE IN COMPLEX WITH INHIBITORS – A MODEL OF THE REACTION PROCESS (PAPER II)

The enzymes involved in lumazine biosynthesis have been studied in considerable detail. However the mechanism of the enzymatic reaction has remained obscure. In this study, four crystal structures of the enzyme from the hyperthermophilic bacterium *Aquifex aeolicus* in complex with different inhibitor compounds, which were designed to mimic the substrate, the putative reaction intermediates and the final product, were solved. The structures were refined to resolutions of 1.72 Å, 1.85 Å, 2.05 Å and 2.2 Å, respectively. Earlier kinetic studies of the enzyme from *B. subtilis* and structural comparisons of the native enzyme from *A. aeolicus* and the complexes (this work) showed that several highly conserved residues at the active site, namely Phe22, His88, Arg127, Lys135 and Glu138 are most likely involved in catalysis. A structural model of the catalytic process, which illustrates binding of substrates, enantiomer specificity, proton abstraction/donation, inorganic phosphate elimination, formation of the Schiff base and double ring cyclization is proposed.

3.1 RESULTS AND DISCUSSION

The crystal structures of lumazine synthase in complex with four inhibitor compounds\[140, 227, 228\] (Figure 3-1) were determined in this work.

![Figure 3-1](image)

Figure 3-1. The structure formulae of the inhibitors: (a) 6,7-dioxo-5H-8-ribitylaminolumazine (RDL); (b) 3-(7-hydroxy-8-ribityllumazine-6-yl)propionic acid (RPL); (c) 6-ribitylamino-5-nitroso-2,4(1H,3H)pyrimidine-dione (RNOP); (d) 5-(6-ribitylamino-2,4(1H,3H)pyrimidine-dione-5-yl)pentyl-l-1-phosphonic acid (RPP).

Crystals of the complexes were prepared using the sitting drop vapor diffusion method. Crystals of the native protein, used for soaking of the RPL and RPP complexes, were obtained in a solution containing 4% PEG 400 (W/V), 0.3 M lithium sulphate and 0.1 M MOPS (pH 6.5). Protein in complex with RDL and RNOP were co-crystallized by adding 1 µl of saturated inhibitor-water solution (ca. 1 µM) to 2 µl of protein solution (ca. 15 mg/ml) and 3 µl of a solution, which contained 0.9 M sodium-potassium tartrate and 0.1 M HEPES (pH 7.5). Large crystals were obtained within two weeks.
Data collection was performed at HASYLAB, Beam Line X11 (DESY, EMBL Hamburg Outstation). The reflection data were evaluated, merged and scaled using the program package HKL.\textsuperscript{22} The results of data collection and evaluation are shown in Table 3-1.

<table>
<thead>
<tr>
<th>Table 3-1. X-ray diffraction data evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS-RDL</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
</tr>
<tr>
<td>Observations/Unique (I &gt; 0)</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>Overall $R_{merge}$ (%)</td>
</tr>
</tbody>
</table>

All the structures were solved by molecular replacement using the structure of native LSAQ as the search model.\textsuperscript{174} Results of the crystallographic refinement are summarized in Table 3-2.

<table>
<thead>
<tr>
<th>Table 3-2. Refinement of $A. aeolicus$ lumazine synthase in complex with inhibitors</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDL</td>
</tr>
<tr>
<td>Number of atoms $^a$</td>
</tr>
<tr>
<td>Protein (atoms with 0 occupancy)</td>
</tr>
<tr>
<td>Ligand</td>
</tr>
<tr>
<td>Solvent</td>
</tr>
<tr>
<td>Refinement</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
</tr>
<tr>
<td>$R$-factor (%)</td>
</tr>
<tr>
<td>Free $R$-factor (%) $^b$</td>
</tr>
<tr>
<td>Ramachandran diagram</td>
</tr>
<tr>
<td>Most favored regions (%)</td>
</tr>
<tr>
<td>Allowed regions (%)</td>
</tr>
<tr>
<td>Additionally allowed (%)</td>
</tr>
<tr>
<td>$B$ factors (Å$^2$)</td>
</tr>
<tr>
<td>Wilson Plot</td>
</tr>
<tr>
<td>All atoms</td>
</tr>
<tr>
<td>Protein</td>
</tr>
<tr>
<td>Ligand</td>
</tr>
<tr>
<td>Solvent</td>
</tr>
</tbody>
</table>

$^a$Numbers were counted per asymmetric unit.

$^b$5% of the unique reflections were set aside for calculations of the free R-factor.

The active sites of the enzyme are located at the interface between every two neighboring subunits within a pentamer. The secondary structural elements
constructing the active sites are $\beta$-turns (residues 21-24, 54-58 and 81-92) from one subunit and a $\beta$-strand (residue 127'-142') as well as an $\alpha$-helix (residue 113'-116') from the adjacent subunit. The substrate binding pocket of the enzyme-RPL complex is shown as an example in Figure 3-2a. The superposition of substrate binding sites in the native protein and the four complexes are shown in Figure 3-2b.

**Figure 3-2.** (a) The substrate binding pocket is located at the interface of two adjacent subunits; (b) Comparison of the active site structures of lumazine synthase in complex with 6,7-dioxo-5H-8-ribitylaminolumazine (RDL, wheat), 3-(7-hydroxy-8-ribityllumazine-6-yl)propionic acid (RPL, green), 6-ribitylamino-5-nitroso-2,4(1H,3H)pyrimidine-dione (RNOP, blue), and 5-(6-D-ribitylamino-2,4(1H,3H)pyrimidine-dione-5-yl)pentyl-1-phosphonic acid (RPP, yellow) as well as the native enzyme (red) with an empty active site. Note the alternate side-chain locations of Lys135' and Glu138' (labels are colored blue) in the complex structures, the adaptation movements of His88 and also the tilted phenyl ring of Phe22 in the structure of the native enzyme.

Shown in Figure 3-2b, the phenyl ring of Phe22 of the complex structures is in an offset parallel conformation with respect to the heteroaromatic ring system of the inhibitor compounds, whereas in the native enzyme it is rotated away by more than 30°. Theoretical studies have shown that the offset-stacking aromatic interactions may stabilize the protein structure. It is proposed that the phenyl ring of Phe22 acts like a “gate” controlling the pathway between the active site and the solvent environment. The aromatic interaction stabilizes the bound substrate and the reaction intermediates. After conclusion of the reaction, the phenyl ring rotates back to release the product.

Kinetic studies indicated that the reaction does not commence without binding of substrate 2. Shown in earlier studies and structures in this work, the phosphate group of substrate 2 is in ionic contact with Arg127'. And it also forms hydrogen bonds with Gly64, Ala85 and Thr86. Thus the orientation of substrate 2 is determined mainly by the interactions shown in the schematic drawing (Figure 3-3). It is therefore suggested that Arg127' and other residues that bind to the phosphate group of substrate 2 are responsible for the regio-specificity of the reaction. Sequence alignments (Figure 1-9) showed that Arg127' is highly conserved. Replacing Arg127'
with a hydrophobic, polar or negatively charged residue led to more than 90% loss of the enzyme activity (Table 1-2). However, replacing Arg127' with a histidine resulted in a reduction of the activity to only 62% with respective to the wild-type enzyme. It indicates that Arg127' (or a positively charged residue), which may form a salt-bridge with the phosphate group, is crucial for catalysis.

Figure 3-3. A schematic drawing of the interactions between the phosphate ion and the enzyme residues (contacts via side-chain atoms are marked by frames).

Arg127’, Glu126’ Lys131’ and His132’ form a charged tetrad at the subunit interface. This tetrad constructs a pocket with its counterpart from the neighboring pentamer, which is related by the crystallographic 2-fold symmetry (Figure 3-4). The substrate binding site opens towards the 2-fold pocket, which may serve as an alternate channel for substrate entry. As the charged tetrad is located at the subunit interface, it is proposed that binding of a phosphate ion or an inhibitor / substrate molecule that contains a phosphate group is strictly related to the stability and assembly of the capsid.

Figure 3-4. Solvent-accessible surface representation of the opening at the crystallographic 2-fold. The product analogue 6,7-dioxo-5H-8-ribitylaminolumazine (RDL) is shown by space-filling models. The color codes are: wheat and blue for subunits from one pentamer, salmon and green for subunits from the neighboring pentamer.

According to the reaction pathway proposed by Kis et al (1995), the N=C double bond of the Schiff base intermediate must be formed in cis-configuration in order to close the second ring of lumazine (Figure 3-5).
Figure 3-5. The cis- configuration of the hypothetical Schiff base intermediate (left) allows nucleophilic attack, which leads to the ring closure; the trans- configuration of the intermediate is shown on the right.

Lys135’ and Glu138’ form an ion pair in the active site. Both residues were observed to adopt either one of the two alternative configurations in all the structures (Figure 3-2b). Lys135’ is able to form a salt bridge with the phosphate group of substrate 2 in one of the conformations. It is therefore suggested that the movement of the side chains of Lys135’ and Glu138’ may facilitate the reorientation of the phosphate moiety of the reaction intermediate leading to a cis- configuration. Following this assumption, the phosphate group of the intermediate may leave the original binding site before the cleavage. Whether the resulting phosphate ion is removed from the active site or bound back to the original site is still unclear. However, it unlikely that the active site would be completely “empty” without either a phosphate ion or a substrate molecule bound to it, because phosphate binding is needed for capsid stabilization. It is therefore proposed that the incoming phosphate-containing substrate (2) would replace the phosphate ion at the active site.

The imidazole group of His 88 is flexible. It is twisted away from the inhibitor by about 0.7 Å in all the complex structures except LSAQ–RNOP, in which the atom N61 on the imidazole ring forms a hydrogen bond with the oxygen atom of the nitroso group of the inhibitor. Earlier studies suggested that the reaction involves several proton transfer steps. Seen from the structures, His88 is in an appropriate position and is therefore a strong candidate for the involvement in the proton transfer steps (Figure 3-6). In the ring closure step of the reaction, the imidazole ring is directly accessible to the oxygen atom of the intermediate, if the S-2 enantiomer is used as the substrate. On the other hand, with the R-2 enantiomer, the distance between His88 and the oxygen atom would not allow a direct proton transfer. It is therefore suggested that His88 is related to the substrate stereo-specificity (note, as mentioned earlier, that the reaction rate of the natural S-2 substrate 2 is about sixfold higher than that of the R-2 enantiomer).
Proton donation from His88: the His88-Ne atom serves as a proton donor, which protonates the 15-hydroxyl oxygen of the intermediate (indicated by the arrow). The chiral C15 atom (originally from the S-2 enantiomer of substrate 2) is labeled with an asterisk.

3.2 CONCLUSIONS

Based on kinetic studies and structural comparisons, a reaction mechanism which illustrates key steps of the reaction including substrate binding, nucleophilic attack, formation of the Schiff base, proton transfer, elimination of the phosphate ion, ring closure and the release of the product, is proposed. In this mechanism, the structural features and residues, which may be responsible for the regio-specificity and stereo-specificity of the reaction, were revealed. A hypothetical dual function of the phosphate binding site at Arg127′ is proposed, which is on one hand responsible for capsid stabilization and on the other hand for binding of the phosphate-containing substrate 2.
4 MULTIPLE ASSEMBLY STATES OF LUMAZINE SYNTHASES – A MODEL RELATING CATALYTIC FUNCTION AND MOLECULAR ASSEMBLY (MANUSCRIPT III)

Lumazine synthases appear in nature in form of pentamers, dimers of pentamers, icosahedral capsids consisting of 60 subunits and larger capsids with unknown molecular structure. In this paper we describe the analysis of the assembly of wild type and mutant forms of lumazine synthases from *Bacillus subtilis* and *Aquifex aeolicus* at various pH values and in presence of different buffers using small angle X-ray scattering and electron microscopy. The data show that multiple assembly states are a general feature of lumazine synthases and there are indications that catalysis is correlated with quaternary structure.

4.1 RESULTS AND DISCUSSION

The pH/buffer dependence of the assembly states of LS from *B. subtilis* and *A. aeolicus* as well as a number of mutants was studied using SAXS.

The scattering profiles of LSBS in phosphate (pH = 6.0 ~ 8.0) and Tris hydrochloride (pH = 7.0 ~ 9.0) buffers are shown in Figure 4-1A and 4-1B, respectively. The distance distribution function, \( p(r) \), indicates that the assembly of LSBS in phosphate buffers is largely invariant to pH change (Figure 4-1F, curve 1). Capsids with an outer diameter of about 160 Å (which is in line with the size of the icosahedral \( T = 1 \) capsid observed in crystal structures) and a small amount of larger particles (yielding a small contribution of distances from 160 to 330 Å, Figure 4-1F curve 1) are formed. In contrast, the distance distribution curves of LSBS in Tris hydrochloride (Figure 4-1F, curves 2-5) point out to large capsids with the outer diameter of about 320 Å. Interestingly, the curves at pH 7.0 and 9.0 (Figure 4-1F, curve 2, 5) display pronounced maxima suggesting rather isometric particle shapes whereas those at pH 7.6 and 8.4 the maxima appear smeared (Figure 4-1F, curve 3, 4), indicating possible polydispersity or deformations of the particles. One can conclude that a structural transition is taking place in Tris at around pH 8.0. In borate buffers (pH = 7.0 ~ 10.0, Figure 4-1C) LSBS displays a similar but even more pronounced dependence of the capsid formation on pH variation than in Tris hydrochloride buffers. At pH 10.0, mostly large particles are formed (Figure 4-1G, curve 4), whereas at other pH 7.0, 8.0 and 9.0 mainly small particles are present (Figure 4-1G, curves 1-3).
Figure 4-1. Experimental scattering curves of wild type LSBS and R127 mutant (circles) and fits obtained from the MIXTURE-M program (solid lines). (A) curves (1-5): wild type LSBS in phosphate buffer at pH = 6.0, 6.5, 7.0, 7.5, 8.0, respectively; (B) curves (1-4): wild type LSBS in Tris.HCl buffer at pH = 7.0, 7.6, 8.6 and 9.0, respectively; (C) curves (1-4): wild type LSBS in borate buffer at pH = 7.0, 8.0, 9.0 and 10.0 respectively; (D) curves (1-5): the R127T mutant in phosphate buffer at pH = 6.0, 6.5, 7.0, 7.5 and 8.0, respectively; (E) curves (1-5): the R127T mutant in Tris.HCl buffer at pH = 7.0, 7.5, 8.0, 8.5 and 9.0, respectively. (F): Distance distribution functions p(r), curve (1): wild type LSBS in phosphate buffer pH 6.0; curves (2-5): wild type LSBS in Tris.HCl buffer at pH 7.0, 7.6, 8.4 and 9.0, respectively; (G) curves (1-4): LSBS in borate at pH 7.0, 8.0, 9.0, and 10.0, respectively; (H) curve (1): the R127T mutant in phosphate buffer pH = 6.0; curves (2-6): the mutant in Tris.HCl buffer at pH 7.0, 7.5, 8.0, 8.5 and 9.0, respectively.
The scattering curves of the **LSBS R127T mutant in phosphate and Tris buffers** are shown in Figure 4-1D and 4-1E, respectively. The distance distribution function, \( p(r) \), of the mutant in phosphate buffers (Fig. 4-1H, curve 1) indicates the presence of capsids with an outer diameter of about 300 Å. However, in Tris hydrochloride buffers, the distance distribution function (Figure 4-1H, curves 2-6) varies considerably with change of pH. At pH 6.0, the distance distribution function is similar to that of the mutant in phosphate buffer (corresponding to large particles with a diameter of 300 Å), but at pH 7.5 and 8.0 yet larger particles with a diameter of about 320-330 Å appear.

The scattering curves (Figure 4-1D, B) and the distance distribution functions of **LSAQ in phosphate and Tris hydrochloride buffers** (Figure 4-1E, curve 1, 2) are very similar to each other and they display practically no pH dependence. Here, the
profile of p(r) corresponds to the small particles (160Å in diameter) with the presence of a minor amount of larger particles (diameter > 300 Å). The **LSAQ mutant with an IDEA insertion in phosphate and Tris buffers** are similar and show very little pH dependence (Figure 4-2C, D). The p(r) function (Figure 4-2E, curve 3, 4) indicates hollow capsids with an outer diameter of about 275Å. Structural studies of the LSAQ-IDEA mutant are presented in section V (manuscript IV). Detailed results of the SAXS experiments and particle fraction size analyses are summarized in Tables 2 and 3 in manuscript III. The results of the particle size and population analyses obtained using CryoEM are comparable to those of the SAXS experiments (Table 1 in manuscript III).

Mutagenic studies on LSBS have shown that a number of residues at the substrate binding site or the subunit interface are important for both assembly and enzyme activity.

The replacement of Arg127, which is essential for substrate binding, by a polar, hydrophobic or acidic residue results in more than 95% reduction of catalytic activity with respect to the wild type enzyme. The Arg127Thr mutant is not catalytically active, whereas the activity of the Arg127His mutant is reduced to 62%. Both mutants Arg127His (Figure 4-3b) and Arg127Thr (Figure 4-3c) form capsids of at least two different sizes, one resembling the known wild-type T = 1 capsid (Figure 4-3a) and the others with a diameter of at least 280 Å.

![Figure 4-3. Negative staining electron micrographs of recombinant lumazine synthase from B. subtilis (A) and the mutants R127T (B) and R127H (C). The scale bar corresponds to 100 nm.](image)

The replacement of either Phe57 or Phe113, involved in substrate binding and inter-subunit interactions, by a serine causes reduction of the reaction rate to 36% and 5%, respectively. These two mutants assemble to both T = 1 icosahedral capsids and large capsids.

Arg21 is involved in a highly conserved ionic network connecting two neighboring pentamer subunits in the capsid. Earlier structural studies suggested that
this ionic network is important for capsid assembly. The mutant Arg21Ala has 43% residual catalytic activity with respect to the wild type LSBS. It assembles to both T = 1 capsids and large capsids.

Based on observations presented earlier and in this work, a model relating the capsid assembly and the catalytic function is postulated.

The active site of icosahedral lumazine synthases is located close to the inner wall of the capsids, which implies that the free passage to the solvent is obstructed by the capsid wall. From a structural point of view, the 5-fold channels of the T = 1 capsids could potentially serve as the substrate / product diffusion pathway. However, the available high-resolution structures indicate that the 5-fold channels do not seem wide enough to allow riboflavin, the product of the reaction catalyzed by heavy riboflavin synthase from *B. subtilis* (the complex [(LS)$_{60}$(RS)$_{3}$] described earlier), to pass through. Furthermore, incubation of the enzyme with a tungsten-compound [NaP$_5$W$_{30}$O$_{110}$]$^{14-}$, which was shown to block the 5-fold channels of heavy riboflavin synthase, did not result in a decreased enzymatic activity. Alternative channels close to the 2-fold and 3-fold axis might allow the diffusion of LS substrates and products, but might also be too narrow for riboflavin. This indicates that the transfer of substrates and product could follow another mechanism. An icosahedral LS with a larger number of subunits will most likely be less densely packed and allow for easier access to the binding sites.

As shown in an earlier comparative study, the reaction is unlikely to proceed with rearrangement of the binding site. There is no experimental evidence whether the eliminated phosphate ion would keep being attached to Arg127 during the reaction or at least transiently be removed from the phosphate-binding site. However it was proposed in paper II, that the substrate 2 is moved away from the phosphate-binding site, whereupon phosphate is eliminated from substrate 2. In the subsequent step it might either bind back to the phosphate-binding site or leave the active site. The results and considerations presented in this paper indicate that the formation of proper subunit contacts critically depend on the presence of phosphate or a ligand in the active site and the correct alignment of non-covalent contacts spanning the interface. The absence of these stabilizing contacts, for instance the release of the inorganic phosphate ion from the binding site, would thus destabilize the T = 1 capsids.

Upon conclusion of the reaction, the substrate-binding site could open up involving a widened state of the pentamers (see structural evidence in paper IV). The formation of a large capsid requires a substantial extent of subunit rearrangements, which probably do not proceed on the same time scale as the reaction. It is therefore unlikely that the enzyme would form larger capsids during the catalytic cycle. However, the opening of the active site might be driven by similar forces as the formation of large capsids and eventually lead to a similar local conformation of the
pentamers. When the binding site is opened and the interactions in the active site are partially disrupted, the enzyme is not capable of catalysis anymore. This new state might also provide a local widening of the pentamer channel and therefore allow for easier passage of substrates and products through the capsid wall.

As observed by Bacher et al., large capsids are readily converted to \( T = 1 \) capsids after incubation with a substrate analogue, 5-nitroso-6-ribitylamino-2,4(1H,3H)-pyrimidine-dione. It is therefore concluded that in the process of substrate binding, the opened and unfunctional binding site rearranges back to the functional state and is thus available for a new catalytic cycle. The rearrangement of LS during the reaction cycle might have a biological and physiological meaning: it has evolved to slow down and control a chemical reaction rather than to accelerate it, in this way keeping the concentration levels of lumazine and riboflavin low in the cell.

4.2 CONCLUSIONS

It has been shown that the wild type lumazine synthase from both \( B. subtilis \) and \( A. aeolicus \) assembles in multiple states in different buffers. Different assembly forms are in equilibrium, which can be affected by changing of pH. It is suggested that the multiple assembly states are a general feature of icosahedral lumazine synthases.

Studies on the single site mutants of LS from \( B. subtilis \) indicate that the capsid assembly is correlated with binding of the substrates (Arg127, Phe57, Phe113) and particular contacts between subunits (e.g. ionic interaction: Arg21, aromatic interaction: Phe57, Phe113). A detailed discussion of structure and assembly of the LSAQ-IDEA mutant is presented in the next section (paper IV).

It is postulated that the substrate binding site in large capsids may have a different conformation with respect to that in the wild type \( T = 1 \) capsid. This conformation would allow easier substrate access to and product release from the active site. The local conformational change of the active site may be triggered by binding of the substrate and releasing of the product.
5 \hspace{1cm} A 180 SUBUNIT COMPLEX OF A LUMAZINE SYNTHASE MUTANT VIOLATES QUASI-EQUIVALENCE IN CAPSID ASSEMBLY (MANUSCRIPT IV)

As shown above, multiple assembly states are a general feature of lumazine synthases. The sequence alignment indicates that an insertion of one to four residues after Gly138 (\textit{S. cerevisiae} numbering) in helix $\alpha_4$ is unique for pentameric lumazine synthases. In order to test the hypothesis that introducing this insertion into an icosahedral enzyme could lead to the formation of pentamers, a mutant of \textit{Aquifex aeolicus} lumazine synthase, in which four amino acids (Ile-Asp-Glu-Ala, IDEA) were inserted after Gly129, was prepared. The structure was determined using cryo-electron microscopy and image reconstruction. Surprisingly, the mutant does not form pentamers, nor does it assemble to the $T = 1$ capsid like the wild type crystal structure. Instead, it forms large capsids with an outer diameter of 292 Å. The capsid follows icosahedral symmetry with three subunits within each icosahedral asymmetric unit resulting in a complex of 180 subunits. However, the analysis of subunit interactions revealed the lack of quasi-equivalent contacts between subunits indicating that the icosahedral assembly of the LSAQ-IDEA mutant violates the principles of quasi-equivalence. The expanded pentameric structure provides a model for an alternative conformation of the pentamer in the wild-type enzyme as it could be formed during catalysis.

5.1 RESULTS AND DISCUSSION

The LSAQ-IDEA mutant particles were observed on low-dose micrographs as spherical projections with less density in the center (Figure 2 in paper IV).

The refined particle has a diameter of 292 Å, which is 138 Å larger than that of the wild type enzyme, however the thickness of the capsid shell (38 Å) is similar (Figure 5-1 A). The particle shows well-defined 532 symmetry and each icosahedral asymmetric unit contains 3 subunits, i.e. A, B and C (Figure 5-1B), resulting in a capsid with 180 subunits.
Figure 5-1. 3D cryo-EM reconstruction of the mutant of lumazine synthase from *A. aeolicus* with the IDEA insertion. (a) The LSAQ-IDEA mutant follows icosahedral symmetry; the outer diameter is 292 Å. (b) Three unique subunits can be identified in one icosahedral asymmetric unit (the yellow triangle). They are named A, B, and C. Filled ovals, triangles and pentagons (green) indicate the location of the icosahedral 2-, 3- and 5-folds, respectively. Open ovals, triangles and hexagons (red) indicate the locations where the quasi 2-, 3- and 6-folds are expected according to the theory of quasi-equivalent assembly.

Figure 5-2. The movement needed to transform wildtype LSAQ (T = 1) subunit from its original location (blue) into the position of the LSAQ-IDEA subunit (yellow). The T = 1 subunit must undergo a small translation by 3 Å (along the green line) away from the fivefold axis (grey line) and a rotation by 36° (about the red line).

Earlier modeling studies suggested that introducing the IDEA insertion into an icosahedral enzyme, e.g. LSBS or LSAQ, would lead to clashes at the interface between neighboring pentamers. However, the EM structure shows that pentamers in the mutant are expanded. The diameter of the 5-fold channel is increased is by 8 Å compared to that of the wild-type enzyme. To adopt the opened conformation, each subunit within the pentamer, is to be rotated (as a rigid body) by 36° and translated by 3 Å with respect to the position of the subunits in the native T = 1 capsid (Figure 5-2).
As a result, the IDEA loop, which is inserted after amino acid Ala129 between helices α₄ and α₅, is located towards the inner surface of the capsid. Therefore the clashes are avoided.

In the wild-type enzyme, the active site is located at the interface between two adjacent subunits within the pentamer and both subunits are involved in substrate binding. Therefore a proper conformation of the pentamer is essential for catalytic competence of the enzyme. It has been proven that the large LSAQ-IDEA is enzymatically inactive (Fischer et al unpublished result), thus it is suggested that the widening of the pentamer may be one of the reasons for losing the catalytic activity.

The analysis of subunit interactions shows that the contact surfaces between subunits of LSAQ-IDEA are generally smaller with respect to that of the wild-type enzyme. The averaged contact surface area of subunit A, B and C in the large capsid are reduced from 48.9% (the native T = 1 capsid) to 20.2% with respect to the total accessible surface area of an isolated monomer (Table 5-1). The contact surface areas are mainly formed by hydrophobic residues indicating the predominant role of the hydrophobic effect for the assembly of LSAQ-IDEA capsid. The results of surface area calculations are summarized in Table (5-1).

<table>
<thead>
<tr>
<th></th>
<th>Total ASA</th>
<th>Total CSA (a)</th>
<th>Charged CSA</th>
<th>Polar CSA</th>
<th>Hydrophobic CSA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(% of a)</td>
<td>(% of b)</td>
<td>(% of b)</td>
<td>(% of b)</td>
<td>(% of b)</td>
</tr>
<tr>
<td>LSAQ-IDEA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subunit A (in pentamer)</td>
<td>6721 (79.5)</td>
<td>1734 (20.5)</td>
<td>204 (11.8)</td>
<td>271 (15.6)</td>
<td>1259 (72.6)</td>
</tr>
<tr>
<td>Subunit B (in capsid)</td>
<td>6971 (82.5)</td>
<td>1483 (17.5)</td>
<td>325 (21.9)</td>
<td>197 (13.3)</td>
<td>962 (64.8)</td>
</tr>
<tr>
<td>Subunit C (in capsid)</td>
<td>7087 (83.8)</td>
<td>1367 (16.2)</td>
<td>342 (25.0)</td>
<td>279 (20.4)</td>
<td>747 (54.6)</td>
</tr>
<tr>
<td>Averaged (in capsid)</td>
<td>6790 (80.3)</td>
<td>1664 (19.7)</td>
<td>351 (21.1)</td>
<td>253 (16.5)</td>
<td>1061 (63.7)</td>
</tr>
<tr>
<td>LSAQ-WT Subunit in capsid</td>
<td>4231 (51.1)</td>
<td>4053 (48.9)</td>
<td>632 (15.6)</td>
<td>696 (17.2)</td>
<td>2725 (67.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Total ASA of an isolated subunit of LSAQ-IDEA and LSAQ-WT are 8454 Å² and 8284 Å², respectively.

For a quasi-equivalent assembly, interactions located around quasi-m-fold symmetry axes at positions Qₘ would be expected to be similar to each other (Figure 5-1B). Furthermore, they would also be similar to the interactions between subunits around related icosahedral n-fold symmetry axes Iₙ. That means that the intersubunit interactions at Q₂ would be similar to the intersubunit interactions at I₂ and the subunit contacts at Q₆ (which is at the same position as I₃) would be similar to each other and to those at I₃ (Figure 5-1B). The contacts between individual subunits within the asymmetric unit at Q₃ would also be similar. However, contact surface comparisons have revealed no similarities between any subunit interfaces at the icosahedral n-fold symmetry axes and the quasi-m-fold axes. On the other hand, there
are no similarities between the subunit interfaces among the 3 subunits within the icosahedral asymmetric unit (Figure 8 in paper V).

Modeling studies showed that by a slight expansion of the capsid, the widened LSAQ-IDEA pentamers can be fitted into a wild type T = 1 icosahedral capsid without serious clashes. Recent EM studies have confirmed that the wild-type enzyme from *B. subtilis* forms an enlarged T = 1 icosahedral capsid with an outer diameter of 186 Å instead of 156 Å (L. Xing, unpublished results). Pentamers in this enlarged T = 1 capsid are also widened and their conformation is similar to that of the pentamers in the LSAQ-IDEA mutant. It is therefore suggested, that the expanded pentameric structure, observed in this study, may serve as a model for an alternative conformation of the wild type LS pentamer as it could also be formed during the catalytic process in the form of local conformational fluctuations (Figure 5-3). Adopting the opened conformation, the product could be released. Furthermore, there would be no steric hindrance for a substrate or product analog, e.g. 5-nitroso-6-ribitylamo-2,4(1H,3H)-pyrimidine-dione, to bind to the more exposed pyrimidine binding site. By this process, sufficient binding energy could be gained, which could trigger the conformational rearrangement leading in a ligand-driven reaction to the reconstitution of the closed pentamer structure, as it is seen in the native T = 1 capsids. 10, 163

**Figure 5-3.** A hypothetical model of the catalytic cycle, in which the widening of the pentamer may play a role in binding of the substrates (S1 and S2) and in releasing of the products (lumazine and/or the phosphate ion). Note that the phosphate ion could either bind back to the original binding site or be released (paper II).

### 5.2 CONCLUSIONS

The LSAQ-IDEA mutant does not form either T = 1 capsids or pentamers, it forms large capsids with 180 subunits instead.
In the large capsid, the pentameric building blocks are widened compared to that of the wild-type enzyme. The widened pentamers may serve as a model for an alternative conformation of the wild-type enzyme playing a role in catalysis. It is shown that the opened form of the pentamer would exert no steric hindrance to substrate binding or product release. The energy gained by binding of the substrate could trigger the conformational rearrangement from the open enzymatically inactive form of the pentamer structure to the closed enzymatically active form, in accordance with ligand-driven reconstitution of \( T = 1 \) capsids from 180 subunit capsids.\(^{10,163}\)

Although 180 LSAQ-IDEA subunits are assembled into a large icosahedral capsid and the number of subunits fits to the theory of quasi-equivalence assembly with \( T = 3 \), contact surface analysis has revealed no similar interactions at the interface between subunits within the icosahedral asymmetric unit or subunits around the icosahedral 2-fold, 3-fold, 5-fold and the quasi 2-fold, 3-fold and 6-fold. It is therefore concluded that the LSAQ-IDEA mutant doesn’t follow the principles of quasi-equivalence in capsid assembly.
6 ACKNOWLEDGEMENTS

The entire work and every single piece of it are largely due to the expert help of my supervisors. I would like to thank:

Prof. Rudolf Ladenstein, for offering me the opportunity to work in the X-ray group, for guiding me through the entire project, for his endless support, constant encouragement and kindest understanding.

Dr. Winfried Meining, for those valuable advices and fruitful discussions on both theoretical and practical issues, for sharing great ideas on the project and many other scientific topics also for the great company during the past years.

Thanks to all my collaborators:
Prof. Adelbert Bacher, Dr. Markus Fischer, Dr. Ilka Haase, Prof. Mark Cushman, Dr. Dmitri I. Svergun, Dr. Petr V. Konarev, Prof. Holland Cheng, Dr. Li Xing, Josefina Nilsson and Wit Suphamungmee for their contributions, which were indispensable for the project.

Thanks to:
Dr. Assen Koumanov and Mikael Karlström for their enormous help on studying, working and living in Sweden and for their great friendship, which I have always been enjoying very much.

Current and former colleagues in X-ray group: Dr. Andrey Karshikoff, Dr. Bin Ren, Dr. Jordi Benach, Dr. Wei Liu, Dr. Ekaterina Morgunova, Gudrun Tibbelin, Dr. Petras Kundratas, Linda Arnfors, Dr. Heiko Bönisch, Dr. Joyce Lebbink, Niklas Hellgren, Dr. Stefan Knapp, Dr. Thijs Kaper and Dr. Ida Helene Steen for their help and pleasant company.

Colleagues in CSB: Prof. Kurt Berndt, Prof. Lenart Nilsson, Prof. Hans Hebert, Ana Caballero-Herrera, Dr. Jan Norberg, Dr. Jianxin Duan, Cesar Santiago, Tobias Elgán, Johan Sagemark, Dr. Philip Koeck, Dr. Yijing Xian, Lars Haag, Leif Bergman and other colleagues, whose help and company have been very important for me during that time.

I greatly appreciate:
The IT and service group for their great support, especially Erik Lundgren, for his expert help with computing at CSB. The administration board, especially Kristina Bergholm and Ingvar Lennerfors for their help throughout these years. My dearest friends in Sweden:
Jiang Wu, Kejun Li, Shujing Zhang, Li Lan, Roger Eriksson, Xin Ma, Yuli Cao, Yu Shi, Zhong He, Yaofeng Zhao, Jinjing Pei, Wei Jia, Yintong Xue, Yongtao Xue and Hui Zhang for friendship and supports.
I shall also thank Prof. Yonggeng Hou and Prof. Lipu Li who introduced me to the field of X-ray crystallography and have been encouraging me all the time.

I am deeply grateful to my wife, Hailin Wang, for her understanding, encouragement and love.
I appreciate my parents and brother, who have given me so much.

This work was supported by Karolinska Institutet, Södertörns Högskola and the Swedish Natural Science Research Council.
7 REFERENCES


63


