

Structural Analysis of the Enzyme
N-Formylmethanofuran:Tetrahydromethanopterin
Formyltransferase

Dissertation zur Erlangung
des Doktorgrades der Naturwissenschaften

vorgelegt beim Fachbereich 14
Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Michael C. Merckel
aus Cornwall, NY, USA

Frankfurt am Main 2008
(D30)

vom Fachbereich 14: Biochemie, Chemie und Pharmazie der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Herr Professor Dr. Dieter Steinhilber

1. Gutachter: Herr Professor Dr. Hugo Fasold
2. Gutachter: Herr Professor Dr. Ernst Bamberg

Datum der Disputation: 10.Juli.2009

Die Arbeiten zur vorliegenden Dissertation wurden im Zeitraum vom Januar 1992 bis zum Dezember 1997 am Max-Planck-Institut für Biophysik (Frankfurt am Main) in der Abteilung von Prof. Dr. Hartmut Michel unter der Leitung von Dr. Ulrich Ermler Max-Planck-Institut für Biophysik (Frankfurt am Main) durchgeführt.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig angefertigt habe und keine weiteren Hilfsmittel und Quellen als die hier aufgeführten verwendet habe.

Michael Christopher Merckel
Dezember 2008

Dedicated to Mrs. Ursula Hildegard Merckel (geb. Gremme), Offenbach.

In memory of Dr. Günther Christopher Merckel M.D., Frankfurt am Main.

Table of Contents

Zusammenfassung	1
Abstract	6
1. Introduction	7
2. Macromolecular Crystallography	17
2.1 Preparing Protein Single Crystals	19
2.2 X-ray Production	21
2.3 X-ray Diffraction by Single Crystals	23
2.4 Data Collection and Processing	26
2.5 The Phase Problem	28
2.5.1 Isomorphous Replacement	29
2.5.2 Multiple Anomalous Dispersion	32
2.5.3 Molecular Replacement	33
2.6 Electron Density Modification	35
2.7 Map Interpretation	37
2.8 Model Refinement	39
2.9 Crystallographic Terminology	41
3. Results and Discussion	43
3.1 MkFTR from Crystal Form P.	43
3.1.1. MIR, Model Building, Refinement and Model Quality	43
3.1.2. Monomer Structure	47
3.1.3. Quaternary Structure	52
3.1.4. Surface Topography and Putative Active Site	59
3.1.5. Thermal Stability Factors	60
3.1.6. High Salt Solubility Factors	64
3.2. MkFTR from Crystal Form S	65
3.2.1. Phasing, Refinement and Model Quality	65
3.2.2. Comparison to PEG Structure	66
3.2.3. Solubility at high Salt Concentration	67
4. Materials and Methods	75
4.1. Cloning, expression and purification	75
4.2. Crystallization and data collection	75
4.3. Phase determination	76
4.3.1. P form phasing	76

Table of Contents

4.3.2. S form phasing	77
4.4. Substrate co-crystallization	77
4.5. Model building and refinement	77
4.6. Analysis of the structure	78
5. References	79
6. Appendices	88
Curriculum Vitae	105
Publications	106
Acknowledgements	108

Zusammenfassung

Molekularbiologische Methoden wie die Sequenzanalyse der 16S rRNA und die Bestimmung phylogenetischer Distanzen anhand der gewonnenen Sequenzdaten führten zu einer Neuorganisation des biologischen Bereiches auf drei Domänen, Archaea, Eukarya und Bacteria. Neben den genetischen Befunden zeigen sich auch morphologische und molekularbiologische Eigenschaften, wie Ribosomstruktur, Lipidstruktur und Membranaufbau, für die Berechtigung der neuen taxonomischen Verteilung. Die Archaea kommen häufig in widrigen Umgebungen vor, die durch Extreme in Temperatur, Salzkonzentration und Druck gekennzeichnet sind.

Eine kleine Gruppe von Organismen, die Methanogene, sind fähig, Methan zu erzeugen. Diese Gruppe von Organismen gehört ausschließlich der phylogenetischen Domäne der Archaea an. Die Methanogene sind von zunehmendem Interesse, weil Methan ein industriell nutzbares Gas ist und sowohl ein Energieträger als auch ein Treibhausgas. Weiterhin sind die spezifischen Eigenschaften methanogener Organismen vom Standpunkt der Evolutionsbiologie, Biochemie und Molekularbiologie von großem Interesse. Fast alle anaeroben Standorte, wie zum Beispiel Kläranlagen, Salzseen, heiße Quellen auf dem Meeresgrund und Verdauungstrakte von Wiederkäuern und Termiten, weisen auf Methanogene hin.

Methanogene sind unter mesophilen, thermophilen, hyperthermophilen und halophilen Archaea vertreten. Die letzteren Drei sind die so genannten Extremophile. Thermostabile und salztolerante Enzyme aus Extremophilen sind aus biotechnologischen und kommerziellen Gründen von großem Interesse. Enzyme aus Extremophilen funktionieren unter extremen Bedingungen effizienter, ergiebiger oder funktionieren eigentlich überhaupt, wobei normale, mesophile Enzyme nicht mehr funktionieren. Das Hauptproblem bei hoher Salzkonzentration besteht darin, die Löslichkeit des Proteins noch zu erhalten. Daher ist die strukturelle Basis dieser Eigenschaften äußerst wichtig.

Das methanogene Archaeon *Methanopyrus kandleri* wurde in Sedimenten in der Umgebung einer hydrothermalen Tiefseequelle entdeckt und wächst optimal bei Temperaturen zwischen 84°C und 110°C und bei pH-Werten von 5,5 bis 7,0 und NaCl-Konzentrationen zwischen 0,2% und 4%. Bis auf wenige Ausnahmen sind Methanogene in der Lage, auf Wasserstoff und Kohlendioxyd als einzigen Energiesubstraten zu wachsen und die Reaktion $4\text{H}_2 + \text{CO}_2 > \text{CH}_4 + 2\text{H}_2\text{O}$ zu katalysieren. Andere mögliche Substrate sind zum Beispiel Acetat und Methanol. Die Hauptreaktion gliedert sich in sieben Enzymschritte. Schritt 1 ist die Reduktion von Kohlendioxyd zu N-Formylmethanofuran (FMF) und wird durch N-

Zusammenfassung

Formylmethanofuran-Dehydrogenase katalysiert. Schritt 2: der Transfer einer Formylgruppe von FMF auf den Folate-ähnlichen Cofaktor Tetrahydromethanopterin (H_4MPT) wird von der N-Formylmethanofuran:Tetrahydromethanopterin-Formyltransferase (FTR) durchgeführt. Die Struktur des FTR-Enzyms wird in dieser Arbeit vorgelegt. Schritt 3 ist mittels der N5,N10-Methenyltetrahydromethanopterin-Cyclohydrolase die Abspaltung von Wasser aus dem C1- H_4MPT -Intermediat. Schritt 4 ist der erste Reduktionsschritt des C1- H_4MPT -Intermediats durch das Enzym N5,N10-Methylenetetrahydromethanopterin-Dehydrogenase. Die nötige Redoxäquivalente wird durch den Cofactor F_{420} übertragen. Schritt 5 ist der zweite Reduktionsschritt des C1- H_4MPT -Intermediats durch die N5,N10-Methylenetetrahydromethanopterin-Reduktase. Schritt 6: der Transfer der Methylgruppe auf Coenzym-M wird durch die N5-Methyltetrahydromethanopterin-Coenzym-M-Methyltransferase katalysiert. Als Cofaktor verwendet die Methyltransferase einen Ni-tetrapyrrol - Faktor F_{430} . Der letzte Schritt 7 ist die Freisetzung von Methan durch die Methyl-Coenzym-M-Reduktase.

Um den Reaktionsmechanismus und die Funktionsweise eines Biomoleküls zu erklären, muss die Struktur sehr genau untersucht werden, um möglichst detaillierte Daten zu gewinnen. Neben der Kernspinresonanz (NMR) ist die Röntgenkristallographie die Hauptmethode, um Daten atomarer Strukturen in hoher Auflösung zu ermitteln. Kristallographie benötigt als Ausgangspunkt einzelne Kristalle, deren Qualität ausreichend ist, um Röntgenstrahlen zu hoher Auflösung zu streuen. Die Züchtung solcher Kristalle wird durch Änderungen der Löslichkeit des Biomoleküls erreicht. Bei ausreichend reinen und konzentrierten Lösungen kann die Keimung und das Wachstum eines Kristalls oft durch Änderungen der Salzkonzentration erreicht werden, den so genannten "salting-in"- oder "salting-out"-Effekten. Genau dies ist das Problem bei Proteinen, die bei hoher Salzkonzentration funktionieren müssen. Bei der Proteinlöslichkeit besteht ein komplizierter Zusammenhang zwischen physicochemischen Eigenschaften des Proteins und pH, Temperatur und Salzkonzentration. Die Hofmeister-Serie ("lyotropic series"), citrate > phosphate > sulphate > acetate~chloride > nitrate > thiocyanate zeigt, wie effektiv verschiedene Anionen als Fällungsmitteln sind. Eine ähnliche Serie besteht für Kationen. Der Zusatz von einem Precipitanten, wie dem Polyethylenglycol, kann durch den "crowding"-Effekt auch zur Kristallbildung führen.

Ziel der vorliegenden Arbeit war die kristallografische Aufklärung der Proteinstruktur der FTR aus dem extrem thermophilen und salztoleranten Archaeon *Methanopyrus kandleri* (MkFTR). Ausgangsbasis daher waren Diffraktionsdaten, Phasen und ein zum Teil gebautes

Proteinmodell. Die Bestimmung und die Analyse der MkFTR-Struktur könnte die strukturelle Basis der enzymatischen Aktivität und die Basis der extremen Thermostabilität und Salztoleranz der Enzyme erklären. Weiterhin könnte die strukturelle Homologie zu anderen Enzymstrukturen auf evolutionäre Verwandtschaften hinweisen. Zu dem Zeitpunkt dieser Arbeit gab es nur wenige Proteinstrukturen von thermophilen Enzymen. Der Vergleich zu anderen Enzymstrukturen aus Extremophilen könnte zeigen, ob eine Strategie oder mehrere zur Thermostabilität und Salztoleranz verwendet werden.

Die molekulare Basis der Salztoleranz von MkFTR könnte durch den Vergleich von Strukturen erklärt werden, die unter verschiedenen Bedingungen, wie zum Beispiel Unterschiede in Salzkonzentrationen, gelöst werden. Die Aktivität von MkFTR weist auf eine Abhängigkeit vom oligomeren Zustand hin. Dieser Zustand hängt von der Salzkonzentration ab, wobei die Aktivform ein Tetramer ist. Bei niedrigen Salzkonzentrationen kommt MkFTR hauptsächlich als Monomer vor und bei steigender Salzkonzentration bilden sich Dimere und zuletzt hauptsächlich aktive Tetramere. Die Struktur MkFTR könnte diese oligomere Abhängigkeit erklären.

Welche Aminosäuren einer Proteinstruktur zu dem enzymatischen Mechanismus beitragen, lässt sich oft beweisen, wenn die Struktur in Zusammenhang mit Substraten gelöst wird. Daraufhin ergab sich die Kristallisation des MkFTR-FMF-Komplexes als ein weiteres Ziel. Zu dem Zeitpunkt dieser Arbeit stand kein H₄MPT zur Verfügung, um die Kristallisation des MkFTR-FMF-H₄MPT-Komplexes zu versuchen.

Ein wichtiges Ergebnis dieser Arbeit war zunächst der fertige Ausbau der MkFTR-Struktur aus Kristallform P, also PEG8000, als Fällungsmittel. Die Struktur wurde in Elektronendichte-Karten eingebaut, die anhand von 1,73Å Streuungsdaten und von durch MIR (Multipler Isomorpher Ersatz) ermittelte Phasen berechnet wurden. Die Phasen wurden durch Mersalyl und (CH₃)₃-Pb-Acetat Derivate bestimmt. Die komplette Struktur ist schrittweise in der Auflösung gegen Nativdaten verfeinert worden. Das Modell besteht aus vier identischen Ketten pro asymmetrischer Einheit. Die Ketten bilden ein Tetramer mit D₂-Symmetrie. Anscheinend waren die PEG8000-Bedingungen trotz niedriger Salzkonzentration ausreichend, um die Oligomerisierung des Komplexes zu fördern. Die unterschiedliche Größe der Kontaktfläche zwischen verschiedenen Monomeren lässt das MkFTR-Tetramer eher als ein Dimer von zwei Dimeren bezeichnen.

Eine genauere Betrachtung des Monomers führt zu dem Schluss, dass das MkFTR-Monomer trotz fehlender Sequenzhomologie aus zwei Domänen besteht, und dass die beiden

Zusammenfassung

Domänen eine ähnliche Faltung haben. Die Domänen konnten mit dem DALI-Server überlagert werden. Dieses Ergebnis deutet eine evolutionäre Genverdoppelung an.

Oft sind es die größten Merkmale, zum Beispiel Klüfte oder Mulden auf einer Proteinoberfläche, die mit Substratwechselwirkung, Bindung und katalytischer Aktivität zusammenhängen. Die Oberfläche des MkFTR-Tetramers zeigt vier große, symmetrisch verteilte Spalten. Diese Spalten haben jeweils drei Abzweigungen und liegen zum Teil direkt auf der Grenzfläche zwischen MkFTR-Monomeren. Der Zusammenhang dieser Stellen mit der Grenzfläche zwischen Monomeren verweist auf eine mögliche strukturelle Erklärung des Zusammenhangs zwischen der Aktivität und dem Oligomerezustand des MkFTR-Enzyms.

Ein weiteres Indiz dafür, dass diese Stellen Substrate-bindend sind, liegt daran, dass eine zusätzliche Elektronendichte in diesen Stellen zu finden war. Die Elektronendichte lässt sich nicht als Wassermoleküle modellieren. Eine mögliche Alternative wäre Folat, vielleicht in Form von Tetrahydro- oder Dihydrofolat. MkFTR könnte dies als Substratanalog zu H₄MPT nach der heterologen Proteinproduktion in *E. coli* binden. Wegen zu niedriger Besetzung wurde die Elektronendichte nicht als Folate modelliert. Dieser Schluss hätte mittels mass-spektroskopischer (MS) Methoden ermittelt werden können. Es wäre vielleicht möglich gewesen, die Besetzung durch Zugabe von zusätzlichem Folat in den Kristallisationsansatz zu erhöhen. Eine enzymatische Analyse hätte auch prüfen können, ob eine Anbindung von Folaten stattfindet und dieser Cofaktor dem Substrat H₄MPT gegenüber als Inhibitor fungiert. Die zu einem späteren Zeitpunkt veröffentlichten FTR Strukturen aus *Archeoglobus fulgidus* und *Methanosarcina barkeri* wiesen ebenso eine zusätzliche Elektronendichte in diesen Stellen auf. Obwohl nicht in der Veröffentlichung erwähnt, konnte die Dichte aus den deponierten Datensätzen berechnet werden. (Mamat B., *et al.* 2002)

Ob diese Stellen richtige "Active-Sites" sind, könnte nur dann ermittelt werden, wenn MkFTR in Zusammenhang mit Substraten kristallisiert würde. Gezielte Mutagenese von einzelnen Aminosäuren hätte auch wichtige Reste identifizieren können, wurde aber hier nicht angewendet. Weiterhin würde die Kristallisation in Zusammenhang mit Substraten Licht auf Aminosäuren werfen, die wichtig für Substratwechselwirkungen und für die katalytische Aktivität der MkFTR sind. Trotz mehrfacher Versuche, MkFTR mit dem Substrat FMF zusammen zu kristallisieren, entstanden keine nutzbaren Kristalle. Letztendlich schien es erst möglich, MkFTR bei Verwendung beider Cofaktoren, sowohl FMF als auch H₄MPT, mit Substraten zu kristallisieren. Weiterhin musste der Kristallisationsansatz unter verdunkelten und anaerobischen Bedingungen stattfinden, wie vor eineinhalb Jahren veröffentlicht (Archaya P., *et al.* 2007). Zum Zeitpunkt dieser Arbeit stand nur das Substrat FMF zur

Verfügung. Die Veröffentlichung zeigt, dass die erwähnten Stellen tatsächlich mit Substratbindung und Aktivität zusammenhängen.

Anhand der Struktur konnten die thermophilen Eigenschaften von MkFTR erklärt werden. Die höhere thermische Stabilität von MkFTR liegt wahrscheinlich an einer höheren Zahl von Wasserstoff- und Ionenbrücken, sowohl innerhalb eines Monomers als auch zwischen den Monomeren. Die Zahl der Ionenbrücken pro Aminosäurerest liegt deutlich höher gegenüber dem Durchschnittswert für mesophile Proteine.

Die MkFTR-Struktur zeigt eine große Anzahl an negativ geladenen Resten auf der Moleküloberfläche, die für die höhere Löslichkeit und salzabhängige Thermostabilität der MkFTR verantwortlich sind. Diese Eigenschaft ist auch für die salzabhängige Oligomerisierung der MkFTR verantwortlich, weil die elektrostatische Abstoßung der Monomere erst bei erhöhter Salzkonzentration kompensiert wird, indem sich also an der Oberfläche Kationen anbinden.

Die molekulare Basis der Salztoleranz von MkFTR könnte durch den Vergleich von Strukturen erklärt werden, die unter verschiedenen Salzkonzentrationen gelöst werden. Daraufhin wurde die MkFTR bei hoher Salzkonzentration (1% PEG 8000, 1.2M $(\text{Na}_4)_2\text{SO}_4$) kristallisiert, Kristallform S und die gelöste Struktur mit der Struktur aus niedriger Salzkonzentration, Kristallform P (22% PEG 8000, 0.3M $(\text{Na}_4)_2\text{SO}_4$), verglichen. Die Struktur der MkFTR Kristallform S wurde mittels Molekular-Ersatz (MR) gelöst und verfeinert. Die asymmetrische Einheit enthielt zwei MkFTR Monomere.

Weiterhin zeigt der Vergleich der MkFTR Kristallform P- und S-Strukturen, dass Glutamatreste in Kristallform S eine geringfügig höhere Anzahl von Wassermolekülen gegenüber Kristallform P binden, circa 1,5 mal so viele Wassermoleküle pro Glutamat. Die größere Anzahl von gebundenen Wassermolekülen könnte zu der höheren Löslichkeit der MkFTR bei erhöhter Salzkonzentration beitragen. Die zusätzlichen gebundenen Wassermoleküle könnten die Hydratationsschicht des Proteins erweitern und dadurch die hydrophobe Oberfläche effektiv minimieren. Dies würde dazu führen, dass die hydrophobischen Wechselwirkungen, also der "salting-out"-Effekt, reduziert, und die MkFTR-Löslichkeit bei hoher Salzkonzentration erhöht würde.

Dieses Schreiben beinhaltet Anhänge, die weitere Ergebnisse zu Nebenprojekten im Bereich struktureller Biologie beschreiben. Der Autor hegt die Hoffnung, dass diese Anhänge in Verbindung mit der MkFTR-Arbeit als ausreichend für eine Doktorarbeit begutachtet werden.

Abstract

Archaea represent a third domain of life and some archaea exhibit a high degree of tolerance to extreme environmental conditions. Several members are methanogens and present in many anaerobic environments. Most methanogens are able to maintain growth simply on H_2 and CO_2 via the enzymatically catalyzed reaction $4H_2 + CO_2 \rightarrow CH_4 + 2 H_2O$.

The archaeon *Methanopyrus kandleri* grows optimally at temperatures of 84°C to 110°C, pH values of 5.5 to 7.0 and NaCl concentrations 0.2% to 4%. The enzyme N-formylmethanofuran tetrahydromethanopterin formyltransferase (MkFTR) catalyzes the transfer of a formyl group from the cofactor N-formylmethanofuran (FMF) to the cofactor tetrahydromethanopterin (H_4MPT), the second step of the above reaction.

X-ray crystallographic analysis yielded insights into the structure and function of MkFTR, (1) the MkFTR monomer exhibits a pseudo-two fold structure suggestive of an evolutionary gene duplication. (2) The structure is a D_2 homo-tetramer with prominent cleft-like surface features. Analysis of the interface contacts showed that the tetramer is best described as a dimer of dimers. The clefts were associated with the monomer:monomer interface and were weakly occupied by extra electron density which might be attributed to the H_4MPT analog folate. (3) This suggested that the clefts are active sites and their association with oligomer interfaces suggested a basis for the dependence of activity on oligomerization. (4) The thermal stability of MkFTR most likely arises from the greater number of H- and ionic-bonds within the monomer and between monomers with respect to mesophilic protein structures. (5) The structure showed a large number of surface exposed negatively charged, glutamate and aspartate residues. These residues explain the salt dependent oligomerization, as only at high enough salt concentration is the electrostatic charge compensated by cation binding and neutralized allowing oligomerization. (6) These residues also improve the solubility of MkFTR at high salt concentration by increased charge repulsion. (7) Comparison of MkFTR structures from low and high salt conditions showed that surface glutamate residues bind slightly more water molecules at high salt conditions further contributing to MkFTR solubility at high salt concentration.

1. Introduction

1. Introduction

Based on 16S and 18S ribosomal RNA (rRNA) sequence data, Woese proposed a third kingdom of life (Woese, 1977). Further analysis of 16S rRNA and 18S rRNA sequences forced Woese to conclude that *Halobacterium*, regarded previously as a halophilic pseudomonad and *Sulfolobus* regarded as a gram-positive bacterium, were members of domain *Archaea*. Woese proposed that life on earth is made of three primary lineages which he referred to as *Eubacteria* (Eu = true/good), *Archaea* (Archaea = ancient) and *Eukarya* (Karya = kernel) (Woese, 1990) (Figure 1.1).

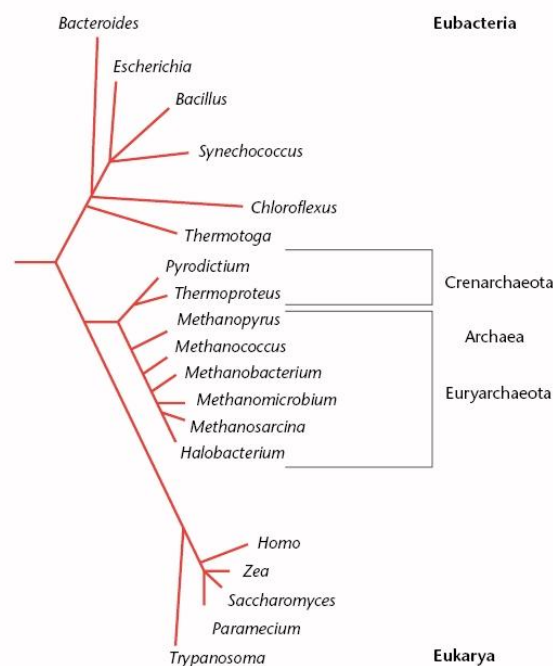


Figure 1.1. Three domain Tree of Life. The division of domains is based on the analysis of 16S and 18S ribosomal RNA sequence data by Woese.

Along with these genetic results, further morphological and molecular biological traits such as ribosome structure, lipid structure and membrane composition strengthen the argument for this new taxonomic delineation. Members of domain *Archaea* are structurally diverse and include spheres, spirals, rods and may be lobed, plate-shaped or irregular-shaped. They may also exist as single cells, aggregates or filaments. The cell diameter may range from 0.1 to 15 μm with lengths up to 200 μm .

The *Archaea* are in some ways a missing link between *Bacteria* and *Eukarya*. Archaeal features often resemble both *Eucaryal* and *Bacterial* features more than they do either one alone. This hints at their primitiveness. With three evolutionary groups, it becomes possible to identify primitive traits of a common ancestor. With only two

groups, it's impossible to tell which version of a trait, if either, is primitive. It appears that the *Archaea* share a common ancestry with *Eucarya* and not *Bacteria*. *Archaea*, then, are primitive relatives of *Eucarya*, and are ideal organisms to shed light on the complexity of eukaryotic organisms.

Archaeal microorganisms are often found in environments characterized by extreme temperature, pressure, salinity and pH. These organisms are often referred to as "extremophiles" and are of considerable biochemical and biotechnological interest. The ability of these organisms and their enzymes to function under extreme conditions where mesophilic organisms and enzymes cannot makes them targets for a variety of potential industrial applications. The presence of archaeal species thriving under extreme conditions has also led to new theories on the evolution of life on the primordial earth.

Members of the domain *Archaea* are phylogenetically divided into three kingdoms, *Euryarchaeota*, *Crenarchaeota* and *Korarchaeota*. Methanoarchaea, the largest and most phylogenetically diverse group in the kingdom *Euryarchaea* are methanogens, but two other phenotypes are found in this group sulphur metabolizing thermophiles and extreme halophiles. Methanogens produce methane and were originally regarded as bacteria but the reanalysis of Woese led to the delineation entirely to the domain of the *Archaea*. Mesophilic, thermophilic and halophilic representatives of the methanogens have been identified.

Methanogens are attracting interest because methane is an industrially useful and energy rich gas. Furthermore methane has come into focus as a greenhouse gas of far more potency, over 20 times than that of CO₂. Nearly all anaerobic environments studied, such as salt lakes, rice fields, waste water treatment facilities, the ruminant and termite gut, deep sea hydrothermal vents and landfills have shown evidence for the presence of methanogens. In anaerobic habitats methane is the product of the degradation of organic material and roughly ~10⁹ tons are generated yearly. Approximately two-thirds diffuses into aerobic habitats and is oxidized by methanotrophic bacteria to CO₂. The remainder is transported to the atmosphere to be photochemically converted to CO₂. (Thauer R.K., 1998)

Bacteria in anaerobic environments ferment and convert biopolymers and organic material into products such as CO₂, H₂, acetate and formate. These are then utilized as substrates by methanogenic archaea specialized in metabolising C1-

1. Introduction

compounds. Methanoarchaea have evolved energy-yielding pathways marked by C1-biochemistry featuring novel cofactors and enzymes. These substrates suffice as the only carbon and energy source for methanogens. (Thauer R.K., 1998), (Wolfe 1996) Recent progress on the enzymology of one-carbon reactions in these pathways has raised the level of understanding with regard to the physiology and molecular biology of methanogenesis. These advances have also provided a foundation for further studies on the structure and function of these novel enzymes and exploitation of the recently completed sequences for the genomes from methanoarchaea.

With few exceptions, methanogens are able to maintain growth simply on H₂ and CO₂ via the reaction $4\text{H}_2 + \text{CO}_2 \rightarrow \text{CH}_4 + 2\text{H}_2\text{O}$ with CO₂ as the terminal electron acceptor and H₂ as primary electron donor. The reduction of CO₂ to CH₄ is enzymatically catalyzed and includes a series of specialized bound cofactors. The reaction proceeds with an overall free energy change of $\Delta G_0' = -131\text{kJ/mol}$.

The reaction can be split into 7 enzymatic steps (Figure 1.2). Step 1 is the reduction of CO₂ to N-Formylmethanofuran and H₂O and is catalyzed by N-formylmethanofuran-dehydrogenase. The reaction is coupled to the translocation of Na⁺ across the cell membrane providing the driving force for the reaction. Step 2 is the transfer of the formyl group from N-formylmethanofuran (FMF) to tetrahydromethanopterin (H₄MPT) and is catalyzed by the enzyme N-formylmethanofuran:tetrahydromethanopterin formyltransferase (FTR). The reaction proceeds by a nucleophilic attack of the N5 atom of H₄MPT onto the formyl carbon atom of formyl-MFR. Step 3 involves formation of N10,N5-methenyl-H₄MPT by removal of H₂O from C1-intermediate N5-formyl-H₄MPT. This step is catalyzed by the enzyme N10,N5-methenyl-H₄MPT-cyclohydrolase. Step 4 is the first reduction step of the C1-intermediate N10,N5-methenyl-H₄MPT. The redox reaction is carried out by N5-N10-methylene-H₄MPT. Redox equivalents for this step are provided by a flavin analog F₄₂₀ cofactor. Step 5 is the second reduction step of the C1-H₄MPT-intermediate and is catalyzed by N5,N10-Methylenetetrahydromethanopterin Reductase. This also involves the F₄₂₀ cofactor. Step 6 is catalyzed by N5-Methyltetrahydromethanopterin:coenzyme M methyltransferase and involves transfer of the methyl group to coenzyme-M. The methyltransferase is an integral membrane protein which requires Na⁺ ions for activity and, in addition to methyl transfer, generates a Na⁺ ion gradient across the membrane contributing to the cellular

electromotive force utilized for instance for ATP synthesis and active transport. The enzyme utilizes a corrinoid cofactor (5'-hydroxybenzimidazolyl cobamide). The Co^{1+} atom functions as a super-reduced nucleophile accepting the methyl group from $\text{CH}_3\text{-H}_4\text{MPT}$ in the first of two half-reactions catalyzed. The second half-reaction involves transfer of the methyl group from $\text{CH}_3\text{-Co}^{3+}$ to coenzyme M (HS-CoM) producing $\text{CH}_3\text{-S-CoM}$ and regenerating the activated Co^{1+} form of the corrinoid. Step 7 is the release of CH_4 and is catalyzed by Methyl-coenzyme M reductase and utilizes the Ni-porphinoid cofactor F_{430} .

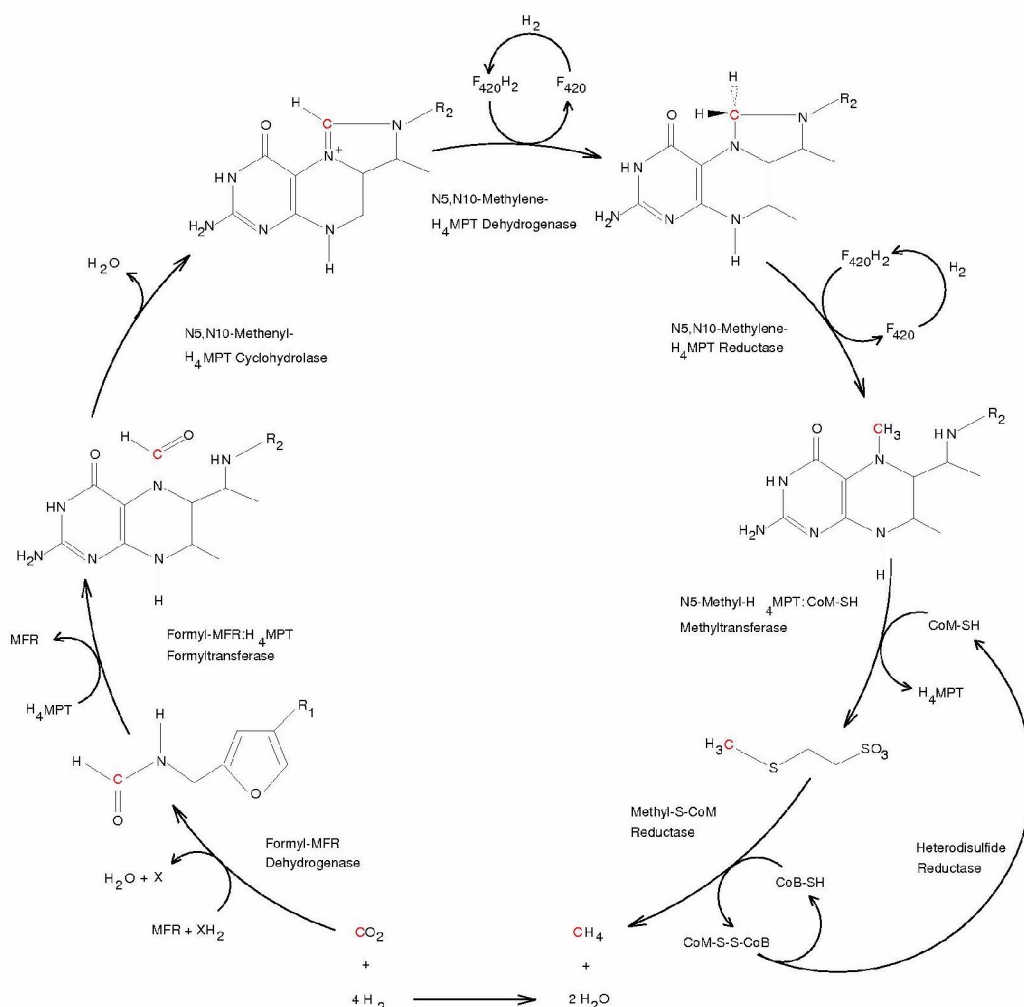


Figure 1.2. Biological Methanogenesis. The pathway catalyzing the overall reaction, $4\text{H}_2 + \text{CO}_2 \rightarrow \text{CH}_4 + 2\text{H}_2\text{O}$. The carbon transferred is highlighted in red. The enzyme N-formylmethanofuran:tetrahydromethanopterin formyltransferase catalyzes step 2, the transfer of a formyl group from formylmethanofuran to tetrahydromethanopterin via a ternary complex mechanism.

The biomolecules of extremophilic microorganisms must evolve to be stable and function properly in extreme conditions of temperature, salt concentration, pH

1. Introduction

and pressure. Strategies and mechanisms must be adopted which allow the enzymes to maintain proper folding and solubility under these conditions. Much effort has gone into understanding the structural basis of enzyme stability of extremozymes in terms of their three-dimensional structure. This route requires high-resolution structural data for homologous enzymes from both mesophiles and extremophiles. With such data in hand differences between homologous enzymes, which might result in enhanced stability, can be identified by structural comparison

Structural features related to thermostability have been revealed by studies of a homologous series of citrate synthase (Russell RJM, et al. 1997) and glutamate dehydrogenase (Yip KSP, et al. 1998) structures. Analysis of these structures indicated the importance played by ionic interactions. The hyperthermophilic members of both these families of enzymes show an increase in the number and extent of ion-pair networks. In the case of the hyperthermophilic citrate synthase, these networks involve intersubunit interactions. In the case of citrate synthase there is an increase in compactness associated with increasing thermostability, along with additional intersubunit interactions involving amino acid residues near the carboxyl terminus. (Russell RJM, et al. 1997)

The structure of the thermostable β -glycosidase from *Sulfolobus solfataricus* revealed a novel feature, the existence of solvent-filled hydrophilic cavities in the core of the protein along with surface ion-pair networks (Aguilar CF, et al. 1997). Surface ion-pair and buried solvent networks have been implicated in providing a degree of “resilience” and resistance to thermal denaturation. This is in contrast to the belief that increasing thermostability is associated with an increase in structural rigidity.

Several thermophilic proteins show a greater proline content as well as an increase in the ratio of arginine to total arginine and lysine content. Proline has lower conformational freedom relative to other residues. This could indicate a stabilizing effect of proline by effectively reducing the entropic contribution to the free energy of the denatured protein. The protein will effectively become more rigid. The increase in rigidity afforded by proline would however seem to run counter to the resilience factors cited above. The increase in arginine preference could be due to arginines ability to hydrogen bond to more solvent molecules than lysine (Delboni, L et al. 1995).

These studies have identified a number of features that are possible contributors to increased thermostability. There are however significant differences

between the conclusions reached with different proteins and there appears to be no general rule for the structural basis of thermostability. Comparisons are qualitative in nature and there is little quantitative information on the significance of various structural features relative to thermal stability.

High ionic strength: Enzymes from halophilic organisms have evolved and adapted to be active and stable under conditions of high ionic strength. This is essential since the intracellular milieu of extreme halophiles is isotonic with the growth environment (Danson MJ, et al 1997). Amino acid sequence comparisons show that halophilic proteins are highly acidic with respect to their mesophilic counterparts, and three-dimensional structures obtained by X-ray crystallography (Dym O, et al. 1995, Frolov F, et al. 1996) or homology modelling (Jolley KA, et al. 1997) show that these enzymes have an excess of negatively charged amino acids on their surfaces. The function of the negative charge on the surfaces of halophilic enzymes is to bind hydrated ions. These bound ions retain a surface hydration layer which tends to reduce the surface hydrophobicity of the protein and leads to a reduction in the tendency for the protein to aggregate at high salt concentration.

Osmolytes: Microorganisms which have adapted to growth in environments of extreme temperature or salt concentration have also evolved to accumulate low molecular weight organic compounds which protect the cells from external stress. (Ignatova Z and Gierasch LM 2006), (Lentzen G, Schwarz T, 2006). These compounds or osmolytes function as physiologically compatible cosolutes that reduce denaturation of macromolecules and function to counteract osmotic stress at high salt concentrations. Some examples of identified osmolytes are polyol phosphates such as di-myoinositol-1,1'-phosphate, cyclic 2,3-diphosphoglycerate, and alpha-diglycerol phosphate. Osmolytes from extremophilic microorganisms are of growing biotechnological interest for application as stabilizers of biopolymers in a variety of processes and applications. These compounds could extend the viability of enzymes during storage or dehydration as well as extend the catalytic regime in which enzymes are active, allowing use under conditions of higher temperature or salt concentration. Osmolyte stabilization may also benefit cell viability and prove useful in medical applications.

Methanopyrus kandleri is a methanoarchaeon isolated from sediment in the vicinity of a deep sea hydrothermal vent off of the coast of California, USA at a depth of 2000m. (Figure 1.3) The organism exhibits optimal growth at temperatures

1. Introduction

between 84°C and 110°C, pH values from 5.5 to 7.0 and NaCl concentrations between 0.2% and 4.0%.

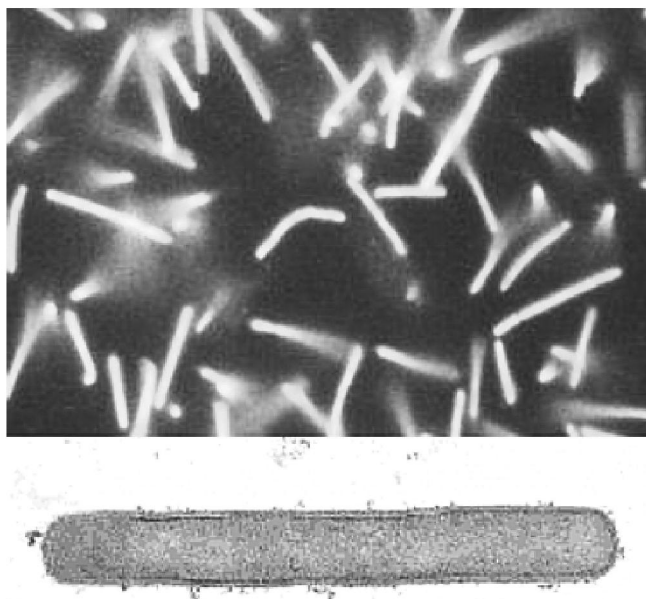


Figure 1.3. Methanopyrus kandleri cells. Micrographs of *M. kandleri* cells viewed in conventional light microscopy above and by transmission electron microscopy (TEM) below.

M. Kandleri like other extremophilic microorganisms has developed a variety of strategies for survival under conditions of high or low temperature, extreme pressure, and high salt concentrations. Cyclic 2,3-diphosphoglycerate has been identified in *M. kandleri* as an osmolyte effective in the activation and thermostabilization of FTR (Shima, S. et al. 1998).

FTR has been purified and characterized from various methanoarchaea, such as the mesophilic *M. barkeri*, the thermophilic *M. thermoautotrophicum*, and the hyperthermophilic *M. kandleri* and *M. fervidus*. (Kunow, J. et al. 1996) The purified enzymes contain one type of subunit with an apparent molecular mass of approximately 32 kDa and exhibit a sequential kinetic mechanism consistent with formation of a ternary complex with FMF and H₄MPT. Some properties of the various FTR enzymes are given in Table 1. The *M. kandleri* FTR (MkFTR) sequence codes for a protein which is the most thermophilic, salt stable and has the lowest isoelectric point and hydrophobicity (Table 1.1). This reflects the special environmental conditions in which *M. kandleri* evolved and grows. The MkFTR exhibits a salt concentration dependant mono- \rightleftharpoons di- \rightleftharpoons tetramer equilibrium with the tetramer as the active and thermostable form (Shima S. et al. 1998).

Organism	<i>M.kandleri</i>	<i>M.fervidus</i>	<i>M.thermoautotrophicum</i>	<i>M.barkeri</i>
Formyltransferase				
Molecular mass (Da)	31,664	31,835	31,401	31,701
Isoelectric point	4.2	5.2	4.5	4.9
Sequence identity (%)	100	64	61	59
Hydrophobicity	-31.4	-20.8	+9.9	+21.6
Temperature optimum (°C)	90	70	65	65
Stimulation of activity at 1.5M K ₂ HPO ₄ (-fold)	>1,000	3	80	4
Thermostability in 1.5M K ₂ HPO ₄ (°C)	>90	-	80	70
Proline content (%)	5.1	4.0	4.0	4.4
Ratio Arg/(Arg+Lys) (%)	29	17	19	24
Property	<i>M.kandleri</i>	<i>M.fervidus</i>	<i>M.thermoautotrophicum</i>	<i>M.barkeri</i>
G+C content (mol %)	60	33	48	42
Growth temperature optimum (°C)	98	83	65	37
Intracellular salt concentration (M)	>1.1	1.0	0.7	0.3

Table 1.1. Comparison of some properties of FTR from a representative group of methanogens. Additionally some physiological properties of these organisms are given.

Sequence comparison. Amino acid sequence data for FTR from several organisms is available. Sequence alignment of the top ten highest scoring BLAST results from known organisms when queried with the *M.kandleri* sequence is given in Figure 1.4. The FTR sequences indicate a high degree of conservation and implies related three-dimensional structures.

The amino acid sequence composition for MkFTR exhibits features consistent with those identified from other thermophilic enzyme structures. The proline content for MkFTR is at 5.1% the highest of those given in Table 1.1. However the range of values from 4.0%-5.1% is rather narrow. The ratio of arginine to total basic residue content is also higher than the other FTR enzymes. An additional characteristic of MkFTR relative to the others is the low isoelectric point of the enzyme. This is a result of an excess of acidic residues with respect to the other enzymes. The excess negative charge imparted by the higher number of acidic residues might function to reduce particle-particle interaction and the “salting-out” effect at high physiological salt concentration.

1. Introduction

```

                *           20           *           40           *
M_kandle : ---MEINGVEIEDTFAEAFEAQMARVLITAASHKWMATAVKEATGFGTSSVIM : 49
M_fervid : ---MKVNGVEIEDTFAEAFD IKVSRILVTAASKRLAKIAAREATGYGTSVIG : 49
A_fulgid : ---MKVNGVEVEEETFAEAFD IKIARVLITGYDYYAWVAANEATGFGTSSVIM : 49
M_thermo : ---MEINGVEIDDTFAEAFPIKIGRVLITAISERWALEAAREATGFGTSSVIM : 49
M_therma : MIVMEINGVEIEDTFAEAFGIKVSRLVTAATKRLAKIAATEATGYGTSVIG : 52
M_barker : ---MEINGVEIEDTYAEAFPIKIARVLITAATKRWALVAATEATGFATSSVIM : 49
M_jannas : ---MEINGVYIEDTFAEAFPIWVSRVLITAATKRWAKIAATEATGFGC SVIM : 49
M_acetiv : ---MEINGVEIEDTYAEAFPIKIARVLITAATKRWAVAAATEATGFATSSVIM : 49
M_burton : ---MELNGVEIEDTFAEAFPIKISRILITAATKRWATVAAQAEATGFGTSVIG : 49
M_aeolic : ---MEINGVEIEDTFAEAFPIWVSRVLITAATEKRLAKIAATEATGFGC SVIM : 49
                Me6NGve6edT5AEAF ik6 R6L6Taa A Aa EATG5gtSVI

                60           *           80           *           100
M_kandle : CPAEAGTDCGYVPPEETPDGRPGVTIMIGHNDEDELKEQLLDRIGOCVMTAP : 101
M_fervid : CPAEAGTDT-YIPPKNTPDQRPGFTIIICNPSKKKLDHELLERVGMTILTAP : 100
A_fulgid : CPAEAGTEI-KAKPSETPDGRPGYITQICHMSKKGLEEQLLARLIGOCVLTAP : 100
M_thermo : CPAEAGTEC-IVPSTETPDGRPGVYITQICNMSWKSLETSLARIGOCVLTAP : 100
M_therma : CPAEAGTDC-YVPPEETPDGRPGYIIMICNPSKKSLDHELLERIGMGILTAP : 103
M_barker : CPAEAGTER-LASPSETPDGRPGVYITQCTFKYEALLEEQLLERIGOCVLTAP : 100
M_jannas : CPAEAGTEK-YVPPSKTPDGRPGFIITQICHPKKSELEHOMLERIGOCVLTCP : 100
M_acetiv : CPAEAGTEK-FASPSETPDGRPGVYITQCTFKYEALLEEQLLERIGOCVLTAP : 100
M_burton : CPAEAGTEK-YADASETPDGRPGVYITQCTFGFKSLEEQLLERVGCILTAP : 100
M_aeolic : CPAEAGTEK-YVPATETPDGRPGYIIEICHPKKSALEHOLLERLIGOCVLTAP : 100
                CPAEAGT eTPDgRPG 6 ic L 6L R6Gqc66TAp

                *           120           *           140           *
M_kandle : TAAAFDAMPFAEKEDEDRVGYKLSFFGDGYQEEDEL DGRKVVKIPVVEGEFI : 153
M_fervid : TAAAFDAL-ENAEF-KVNTGFKLKFFGDGYEKEGEGVQGRKVHIIPIMSGEFI : 150
A_fulgid : TTAVFNGLPDAEEK--DDTGFKLKFFADGYQKEVEVGGRKCVAVPMMEGDFI : 150
M_thermo : TTAVFNGLPFAEKQ--FDTGKKLGYFGDGYQCEIKYCNRNFCFKIPIMEGDFL : 150
M_therma : TTAVFDAL-DDEDE-KLNI GFKLKFFGDGYEKELELDGRKTHSIPIMSGDFL : 153
M_barker : TTAVFNGLPFAEKQ--DNVGFKLKFFADGMESETQIAGRKVYKVPIMEGDFL : 150
M_jannas : TTAIFDAMGDMADE-QLKVGFKLKFFGDGYEKKDEL YGRKVYKIPIMGGEFI : 151
M_acetiv : TTAVFNGLPDAEKQ--FNI GFKLKFFGDGMESEAQVAGRKVYKVPIMEGDFV : 150
M_burton : TTAVFNGLPDAEKQ--FDTGRKLYEADGTESETEVGGRKMHVIPMMEGDFL : 150
M_aeolic : TTAVFDAMGEDATE-HLKVGFKLKFFGDGYETKEKLNDKVMVSVIPIMGGDFK : 151
                Tta F1 6 G KLk5F DG 2 e g4k 6P66 G F

                160           *           180           *           200
M_kandle : VEDSFGITTVAGGNFY IMAESQPAGLQAAEA AVDA IKGVEGAYAPFPGGIV : 205
M_fervid : VEEKFGIKSGVAGGNFF IMAETQGSALLAAEMAVDA IKSVEGVMTPFPGGIV : 202
A_fulgid : IENDIGYNTGIAGGNFF IMAETQPSALAAAKAAVDA ISDVEGVI TFPFGGIV : 202
M_thermo : VEETIGAVDGIAGGNFY I LGQNQPAALMAEA AVDA I SKLRGTI TFPFGGVV : 202
M_therma : IESEFGIKDGVAGGNFF I MGDSQASALLAAQA AVDA IAAVEGTV TFPFGGVV : 205
M_barker : AEENIGATAGIAGGNFF I FGDSQMTALTAEEA AVDT IAELEGTI TFPFGGIV : 202
M_jannas : TEAKFGIKKGVAGGNFF I MADINASALIAEEA AVNA IASVDGVI TFPFGGVV : 203
M_acetiv : TEENIGATAGIAGGNFF I FGDSQMSALTAEEA AVDA IEELEGTI TFPFGGIV : 202
M_burton : VEDTLGAVTATAGGNFF I FGDTQMTLTAEEA AVDA I GAVDGTI TFPFGGIV : 202
M_aeolic : IESKLGIKKGVAGGNFF I MADSNSSALVAEEA AIDA I NGVEKTI TFPFGGIV : 203
                E G g6AGGNF5I q aL AA aA61aI 6 g tFPFGG6V

```

Figure 1.4. Alignment of FTR sequences. The top ten highest scoring BLAST results from known organisms queried with the M.kandleri FTR sequence. Sequence abbreviation; M_kandle - Methanopyrus kandleri, M_fervid - Methanothermus fervidus, A_fulgid - Archaeoglobus Fulgidus, M_thermo - Methanosaeta thermophila, M_therma - Methanothermobacter thermautotrophicus, M_barker - Methanosarcina Barkeri, M_jannas - Methanocaldococcus jannaschii, M_acetiv - Methanosarcina acetivorans, M_burton - Methanococcoides burtonii and M_aeolic - Methanococcus aeolicus. Alignment to M. kandleri FTR performed with CLUSTALW and highlighted by conservation with GeneDoc; black – 100%, dark grey – >80% and light grey - >60% . Entries with known structure are M. kandleri, A. fulgidus and M. barkeri.

The goal of this work was to carry out the structural analysis of MkFTR. In order to study the structure and function relationships of a biological macromolecule in the greatest detail the techniques of X-ray crystallography or NMR are the primary approaches for obtaining data of necessary resolution. In the case of MkFTR phased diffraction data along with a partial model were available. With a structure of the enzyme in hand it would be possible to identify features relevant to its thermal stability and tolerance to high salt concentrations. The MkFTR enzyme is of particular interest due to its dual features of thermal stability and salt tolerance. The structure would be expected to show features of both halophilic adaptation and features that might relate to its thermostability. The structure could be analysed in terms of the enzymes combined thermophilic and halophilic properties and reveal if strategies adopted are similar to those already seen or are unique. Furthermore the structure could explain the basis of the salt dependant mono- \rightleftharpoons di- \rightleftharpoons tetramer equilibrium of MkFTR. Finally the analysis could lead to insights into possible active sites and substrate binding.

2. Macromolecular Crystallography

Conventional and near field light microscopy does not allow visualization of matter with atomic resolution. It would in principle be possible to imagine an X-ray microscope, however it has proven until now to be difficult to prepare lenses with the ability to efficiently focus X-rays of energies useful for atomic resolution studies. Crystallographic diffraction methods provide an indirect way to bypass the need for lenses thus achieving atomic resolution with electromagnetic radiation.

The foundation of X-ray crystallography was laid with the discovery of X-rays by Wilhelm Conrad Röntgen. The X-rays or cathode rays which he described also bear his name, Röntgenstrahlen. (Röntgen, 1895). The diffraction of X-rays was first shown by Friedrich, Knipping and von Laue (Friedrich *et al.* 1912) and followed shortly by the first structural elucidation of Sodium and Potassium Chloride crystals by Bragg (Bragg, 1913). The theoretical interpretation of the diffraction of X-rays by crystals was developed by many researchers. An important step was the establishment of the reciprocal lattice construction by Ewald as an important aid in the construction of the observed reflections (Ewald 1921).

The first crystals of a protein were described in 1840 (Hünefeld, 1840). The first diffraction patterns from protein crystals required nearly a century however and were first reported by Bernal and Crawford (Bernal and Crawford, 1934). Their work showed that macromolecules could organize themselves into crystal lattices with periodic ordering. The breakthrough achieved by Bernal and Crawford was to maintain the protein crystal in a hydrated condition during the diffraction experiment by mounting in a glass capillary along with the crystallization medium. Earlier attempts to obtain diffraction from protein crystals involved freely mounting the crystals. Protein crystals consist on average of 50% solvent and free mounting in air, without maintaining a stream of moist air leads to loss of water and concomitant loss of crystalline order and diffraction. The next important step in the development of macromolecular crystallography was the demonstration that the method of multiple isomorphous replacement could also be applied to large molecules such as proteins (Green *et al.*, 1954). With this the structures of myoglobin (Kendrew *et al.*, 1960) and hemoglobin (Perutz *et al.*, 1960) were the first protein structures to be elucidated.

Even though in the early years of protein crystallography it was only possible in a few cases to determine a structure, the method has developed into established itself as the standard technique to obtain high resolution structural detail of

2. Macromolecular Crystallography

biologically interesting macromolecules. To underline the dramatic increase in the application of crystallography, the exponential growth in deposited structures to the PDB database is shown in Figure 2.1.

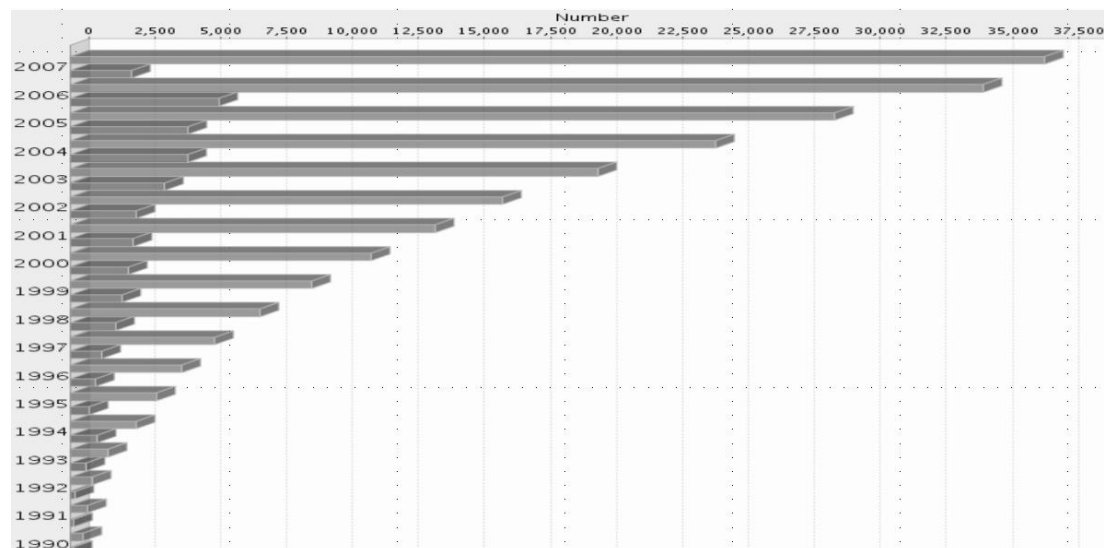


Figure 2.1 Protein database growth. The number of deposited protein structures determined by X-ray crystallography. Deposits for year, dark grey, total deposits, light grey. Statistics shown only since 1990.

The necessary steps in obtaining a protein crystal structure are in principle; (1) the preparation of a pure protein sample, (2) the determination of suitable crystallization conditions, (3) the collection and processing of diffraction data, (4) obtaining phases, (5) the interpretation of electron density maps, specifically the building an atomic model of the protein into the maps and (6) the refinement of the model against the observed scattering data. Considerable developments in all aspects of the above process, in the form of new methods and empirical experience along with steady increases in computational power have contributed to the enormous growth, success and applicability of protein crystallography since the first structure determinations in 1960.

Various aspects of the steps and methods involved are briefly described below. Further extensive details of protein crystallography can be found in, for example the following; (Blundell and Johnson, 1976; McPherson, 1982; Stout and Jensen, 1989; Drenth, 1999; McRee, 1993; and *Methods In Enzymology Volumes*, 114, 115 276, 277, 368 and 374).

2.1 Preparing Protein Single Crystals

The successful crystallization of the protein of interest forms the basic prerequisite for every crystallographic structure determination and stands at the beginning of every macromolecular crystallography project. Unfortunately this step is still rather poorly understood thus trial and error approaches are still necessary. In analogy to small molecule crystallization it is assumed that the protein, in particular solution conditions must be brought into the supersaturated region of the solubility phase space for the protein of interest Figure 2.2. Once in this saturated region it is assumed that aggregates and crystalline nuclei can form. The nuclei can then grow into larger crystals as the system relaxes from the supersaturated state. If conditions of slight supersaturation can be found, then fewer nuclei will form and these will grow into fewer larger crystals rather than when a higher degree of supersaturation is achieved and more nuclei form leading to more and smaller crystals. If the system is brought too far or rapidly into a supersaturated state then amorphous non-crystalline precipitate may be the result.

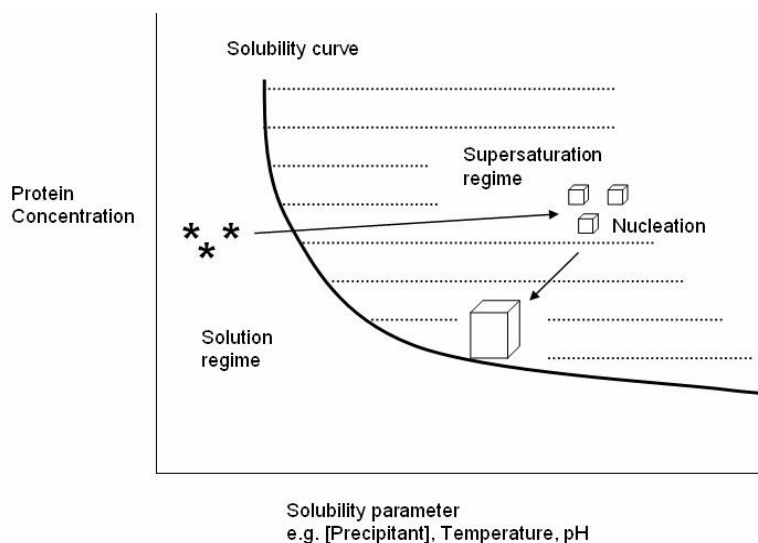


Figure 2.2 Solubility curve and crystallization. The saturation point for a protein is plotted as a function protein concentration and the solubility/precipitation parameter, e.g. precipitant concentration, temperature or pH. For crystallization to occur the protein solution must be shifted from a soluble or equilibrium solution state into a supersaturated non-equilibrium state where crystal nucleation may occur. Nuclei will grow as the system relaxes to an equilibrium state.

Although this basic scenario is widely accepted, the practical implementation of it depends on many parameters. Finding the optimal conditions is actually a search

2. Macromolecular Crystallography

in a multi-dimensional phase space of interdependent parameters. These include protein concentration, the concentration and type of precipitant, pH and buffer type, temperature and any additives, such as detergents in the case of solubilised membrane proteins.

There are a number of methods for crystallizing proteins by variation of the parameters mentioned above. One of the most often applied techniques is vapour diffusion. In this approach a small amount of the protein solution is mixed with small amount of a precipitant solution in which one of the above parameters is varied. The droplet mixture is sealed in a container with a larger reservoir of the precipitant solution and allowed to equilibrate via the vapour phase at constant temperature. Water diffuses from the less concentrated protein:precipitant solution to the higher concentration reservoir solution. In this way the concentrations of protein and precipitant in the droplet will increase and the system will hopefully be brought to a state of supersaturation where nucleation may occur. A typical vapour diffusion setup is shown in Figure 2.3.

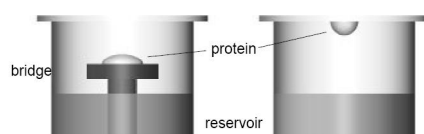


Figure 2.3. Vapor diffusion setup. A typical vapour diffusion setup for screening protein crystallization conditions. Two common formats are sitting drop, left and hanging drop, right.

Initial screens often use an incomplete factorial approach (Jancarik, J. and Kim, S.-H.,1981). These screens vary parameters that have been empirically identified as being most successful in crystallizing other proteins. If initial screens identify conditions where crystalline precipitate or small malformed crystals appear, these may be optimized by systematic fine grid screening of the parameters. In some cases where a suitable metric can be identified, e.g. crystal volume it has been possible to utilize response surface statistical approaches to systematically optimize parameters and crystal growth.

Empirical experience shows that the purity of the protein plays a significant role in the success of crystallization attempts. It is often necessary to purify a protein

to nearly 100% purity and homogeneity. There are however anecdotes of “over purified” proteins not crystallizing presumably due to loss of some important small molecule acting as an additive. The vapor diffusion approach consumes a certain amount of sample. Typically 20ug are required, that is 2ul of a 10mg/ml per condition screened. In general hundreds if not thousands of screens are performed in order to find optimal crystallization conditions. In lucky situations a few milligrams of sample will suffice, but often tens of milligrams if not more will be required. In this respect a protein crystallography project will place considerable demands on the source of purified protein. These demands can be met by suitable overexpression and efficient purification of the target protein

As mentioned above protein crystals, in contrast to small molecule crystals contain a relatively large amount of solvent, ~30%-70% or higher. The protein molecules are therefore more loosely packed and crystal packing forces are weaker than in small molecule cases. This forces the requirement of maintaining crystals in a hydrated state during diffraction data collection, either by mounting in a glass capillary, free mounting with a humid stream of air maintaining the hydrated state or rapid quenching to cryogenic vitreous states with liquid N₂ or He. The cryogenic approach has the distinct advantage of reducing radiation damage during data collection.

Protein crystals are typically viewed not as an ideal, monolithic single crystal but rather as a mosaic of many small crystalline domains. These domains can be slightly misoriented with respect to each other leading to a mosaic spread that can be 0.1° to several degrees. One disadvantage of the cryogenic approach is often an increase in the mosaic spread.

A further property of macromolecular crystals relative to small molecule crystals is the considerably larger unit cells adopted. The larger unit cell implies less unit cells overall for a crystal of the same size as that of a small molecule crystal. This factor along with mosaicity or lower degree of order and the higher solvent content leads overall to weaker diffraction power of protein crystals. This raises the requirement for more powerful X-ray sources.

2.2 X-ray Production

In order to take advantage of the diffraction process a suitable wavelength must be used. The wavelength used must not be larger than the order of the distance

2. *Macromolecular Crystallography*

between the objects to be resolved. In order to resolve atomic distances one needs to use wavelengths of $\sim 1\text{\AA}$ or shorter. The X-ray region of the electromagnetic spectrum is $\sim 1000\text{-}0.1\text{\AA}$. The interaction of light with matter decreases with wavelength and this makes X-ray detection more difficult, thus X-ray crystallography utilizes radiation in the range of $\sim 2.0\text{-}0.5\text{\AA}$

X-rays can be produced in different manners. In the majority of crystallographic laboratories the rotating anode generator is used. In this source, as with the fixed anode X-ray source electrons are emitted from a hot cathode and accelerated through an electric potential field to high kinetic energy and focused on an anode generally made of Cu or Mo although Ag is sometimes used. The deceleration experienced by the electrons converts some of the kinetic energy upon reaching the anode and leads to X-ray emission over a broad range of wavelengths. This braking effect is also called Bremsstrahlung. The remainder of the kinetic energy is lost due to interactions of the electrons with core electrons of the anode material leading to the ejection of core electrons. Anode electrons from higher orbitals then immediately drop into the unoccupied core orbitals and emit X-rays which are of a wavelength characteristic for the anode material. In the case of Cu anodes K-shell electrons are ejected and filled by L-shell electrons. The characteristic emissions from Cu anodes are the $K_{\alpha 1} = 1.54051\text{\AA}$ and $K_{\alpha 2} = 1.54433\text{\AA}$ and $K_{\beta 1} = 1.3922\text{\AA}$. For most work the $K_{\beta 1}$ emission is filtered with Ni foil to leave an averaged $K_{\alpha 1}$, $K_{\alpha 2}$ emission that is used.

For fixed and rotating anode sources the majority of the X-ray emission is from the characteristic emission. Bremsstrahlung of much greater intensity is obtained from synchrotron sources. In these sources the “braking effect” or change in momentum felt by the electrons in the storage ring is due to centripetal acceleration. Even more intense emission above that due to centripetal acceleration in bending magnets can be achieved by the use of insertion devices such as wigglers and undulators which force the electrons to undergo further periodic deviations from the orbital path.

Whereas the wavelength of X-ray emission from fixed and rotating anodes is determined and effectively fixed by the anode material used, the wavelength of emission from synchrotron sources consists of a broad range. For a particular experiment the wavelength used is in principle tunable through the use of suitable monochromators. This gives synchrotron sources the additional advantage, above the

already significant intensity increase of adjustable wavelength for experiments where this of advantage, for instance in experiments utilizing anomalous dispersion.

2.3 X-ray Diffraction by Single Crystals

When X-rays are brought to bear upon an ordered lattice of atoms or molecules the electrons of the atoms or molecules oscillate at a frequency proportional to the wavelength of the X-rays. The electrons in response emit electromagnetic radiation at the same frequency. The observed emitted radiation is the sum of the radiation emitted from all the electrons and in most directions this sum is zero. Only in specific directions is there constructive interference and is an X-ray or reflection observed. As long as the energy of the incoming X-rays is not near the energy of specific electronic transitions of the atoms or molecules, the scattering of X-rays can be treated as scattering from free electrons. For light atoms such as C, N or O this is always the case. For heavier atoms, X-ray energies near transition or absorption edges lead to anomalous effects. These effects however are useful in XANES and EXAFS as a basis of identifying atom type or as an approach to obtaining phase information as described below.

There are many ways to mathematically describe the scattering process. In analogy to the diffraction of visible light the Bragg equation can be formulated:

$$2d_{hkl} \sin \theta_{hkl} = n\lambda \quad (1)$$

The Bragg equation describes the reflection of an X-ray of wavelength λ by lattice planes with a characteristic spacing of d through an incident angle of θ . Every crystal lattice consists of many such lattice planes which can be identified by Miller indices hkl . In this nomenclature the plane hkl intersects the x-axis at a/h , the y-axis at b/k and the z-axis at c/h , where a, b, c are unit cell length along x, y, z respectively. For example the planes defined by $(1, 1, 1)$ intersect the (x, y, z) axes at (a, b, c) . An index of 0 implies that the particular axis is not intersected, $(h, 0, l)$ indicates lattice planes that run parallel to the y-axis. From the Bragg equation it is seen that constructive interference only occurs when the wave path length difference ($2d \sin \theta$) are integral multiples of the wavelength. The lattice planes of Miller index hkl give rise to the X-ray reflection that is then given the same index hkl . For a fixed incident X-ray beam and a rotated crystal, the direction of a diffracted beam can be determined for any particular orientation

2. Macromolecular Crystallography

A very useful method for interpreting the diffraction process is the use of the reciprocal lattice and Ewald sphere concepts. The reciprocal lattice is a mental construct which is related to the real crystal lattice as follows; each reflection hkl is mapped to a point on the reciprocal lattice with index hkl . Each of these reciprocal lattice points is the end point of the vector \mathbf{S} (vectors are differentiated with bold font here from scalars), which is normal to the lattice plane hkl and has length of $1/d_{hkl}$. It is possible to define a reciprocal lattice cell which consists of axes \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* . Each of these reciprocal axes is normal to the plane formed by the other two real axes, e.g. \mathbf{a}^* is normal to the plane formed by \mathbf{b} and \mathbf{c} . The length of the reciprocal axes is the inverse of the real axis length, e.g. $a^* = 1/a$. The points on the reciprocal lattice can be described in vector notation as, $\mathbf{S} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$.

The construction in Figure 3.4 shows that the conditions above for constructive interference are always met when a reciprocal lattice point intersects the Ewald sphere. The Ewald sphere is viewed as a sphere of radius $1/\lambda$ centered on the sample position where diffraction occurs. As the reciprocal lattice rotates in conjunction with the real crystal lattice, it is possible to bring different reciprocal lattice points into reflection condition with the diffracted X-rays going in vector direction \mathbf{MP} .

This construction reflects several important aspects of crystallography. The reciprocal lattice points and thus the diffraction pattern of a crystal depends only on the unit cell and not the cell contents. Information regarding the unit cell contents, e.g. the protein are encoded in the intensities and phases of the reflection data.

The real lattice or protein crystal and the reciprocal lattice are intrinsically linked to each other by the Fourier transform. This is the basis of determining the spatial structure of the sample by Xray crystallography. The basic equations of this relationship are given below.

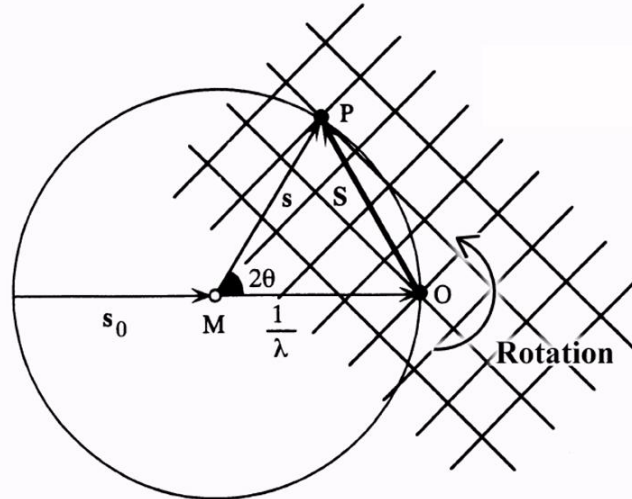


Figure 3.4 The Ewald sphere and reciprocal lattice construction. The Ewald sphere is visualized of radius $1/\lambda$ centered on the sample at M. The reciprocal lattice is visualized as centered on the incident beam axis s_0 and on the Ewald sphere at point O. The reciprocal lattice has unit cell dimensions and angles that are reciprocal to the real lattice at M. The reciprocal lattice rotates in conjunction with the real lattice. The incident beam s_0 is diffracted through and angle 2θ to vector s .

The structure factor F_{hkl} is a combination of N waves in direction hkl diffracted by N atoms in the elementary cell during a diffraction experiment. Each of these waves has an amplitude f_i , the atomic scattering or form factor and a phase α_i which expresses the spatial position of atom i in the unit cell. Each structure factor F_{hkl} in turn also consists of an amplitude and phase and can be described as the following complex number:

$$F_{hkl} = A_{hkl} + iB_{hkl} \quad (2)$$

Where the terms:

$$A_{hkl} = \sum_N f_i \cos 2\pi(hx_i + ky_i + lz_i) \quad (3)$$

$$B_{hkl} = \sum_N f_i \sin 2\pi(hx_i + ky_i + lz_i) \quad (4)$$

The atomic form factor f_i depends on the resolution ($2\sin\theta/\lambda$) because the electron cloud is not point like rather has a spatial extent. With increasing angle the path length difference between electron positions in the cloud increase. Since these distances within the electron cloud are far smaller than the lattice plane spacing d_{hkl} and only a fraction of the wavelength they do not lead to complete destructive

2. Macromolecular Crystallography

interference. The scattering intensity does however decrease with resolution. If one takes into account not only the static spatial extent of the electron cloud but also the dynamic motion of the atom, that is thermal motion about an equilibrium position, then in the simplest case of isotropic motion, the temperature factor B can be introduced.

$$f_{i, dynamic} = f_{i, static} \exp\left(-B_i \frac{\sin^2 \theta}{\lambda^2}\right) \quad (5)$$

It can be shown that B is related to the mean squared deviation \bar{u}^2 of the atom:

$$\bar{u}^2 = \frac{B}{8\pi^2} \quad (6)$$

The structure factor measured in a diffraction experiment can be given as:

$$|F_{hkl}| = \sqrt{A_{hkl}^2 + B_{hkl}^2} \quad (7)$$

The phase α of a particular reflection can be given by:

$$\alpha_{hkl} = \arctan \frac{B_{hkl}}{A_{hkl}} \quad (8)$$

In the reverse direction in order to obtain the electron density distribution within the crystal at position (x, y, z) the following Fourier summation is calculated:

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp[-2\pi i(hx + ky + lz) + i\alpha_{hkl}] \quad (9)$$

The electron density distribution is the primary information about the crystal contents obtained with X-ray crystallography. Within the electron density distribution lies the information about the position of the individual atoms which gave rise to the X-ray reflections.

2.4 Data Collection and Processing

Historically the medium used for capturing the X-ray diffraction data was photographic film. As a result of the relatively low sensitivity, low dynamic range and the intensive development process, film has been replaced by various detection methods. Early electronic detectors were used to measure intensities of reflections one-by-one. These however were only suited to small molecule work. The shear

number of reflections associated with large unit cells and the radiation sensitivity of biological macromolecules obviated the use of such detectors and forced development of other area detectors which collect large numbers of reflections per exposure. To this class of detectors belong the gas discharge, multiwire proportional detectors, fluorescence enhanced video systems, image plate and CCD-detectors. The good performance to price characteristics of the image plate detector led to it being adopted in most home labs. The image plate detector consists of an inorganic phosphor, Ba:F:X:Eu²⁺, where X = Cl, Br which is sensitive to X-ray energy photons. Incident X-rays promote electrons into a metastable excited state which can be read out by use of red laser light even days after exposure. The phosphor can be erased and reused by brief exposure to a bright Xe lamp. In effect the image plate detector functions like re-usable film.

The typical diffraction experiments employ monochromatic radiation to collect reflection data. In the course of the data collection the crystal will be rotated or oscillated during each exposure through a specific angle Φ , usually 0.2°-2.0°. A series of exposures will be collected over a contiguous angular range sufficient to cover a symmetry defined unique sector of reciprocal space.

The next step is to index the reflections of the diffraction data set. That is to assign the correct index hkl to each reflection. Historically the crystal would be mounted and oriented on a screened precession setup allowing for collection of undistorted reciprocal lattice layer images. These could be visually indexed and integrated. The requirement for greater speed led however to the development of the screenless rotation method of data collection. Diffraction patterns collected in this fashion show an arctan dependency and overlap of lattice layer making visual indexing more or less impossible. The theoretical basis of indexing such data collected from randomly oriented crystals is rather complex however the process has been codified in several very powerful computer programs which now allow automatic indexing. Simultaneous to the indexing of reflections is the assignment of crystal lattice to a particular Bravais lattice type. When the lattice type has been correctly assigned and the orientation of the crystal relative to the laboratory frame determined, it is possible to predict all the expected reflections. The actual intensity at these predicted positions can be measured, integrated and corrected for specific factors such as polarization, absorption and Lorentz effects. The data can be reduced after they are scaled for exposure-to-exposure deviation by comparing intensities of

2. Macromolecular Crystallography

symmetry related reflections. Data reduction processes symmetry related reflection intensities to give an average value. The quality of the resulting data set is given by the R-factor, R_{sym} .

$$R_{sym,I} = \frac{\sum_{hkl} \frac{1}{N} \sum_{j=1}^N |I_{hkl} - I_{hkl}(j)|}{\sum_{hkl} I_{hkl}} \quad (10)$$

Here I_{hkl} describes the mean of all observed reflections hkl and $I_{hkl(j)}$ the j -th reflection of N observed reflections.

At this point incorrect indexing and Bravais assignments are identified. Since non-symmetry related or independent reflections will be averaged this leads to higher R_{sym} values than when correctly assigned reflections are merged.

This point also serves to differentiate spacegroup assignments, e.g. P4 versus P4₂2 which belong to the same Bravais lattice and thus have identical reciprocal lattices. The presence of screw axes can also be determined from the systematic absences they cause in the diffraction data. Enantiomorphic spacegroups however can not be discriminated, P4₁22 or P4₃22 will give rise to similar merging statistics and can only be differentiated after phasing and inspection of initial electron density maps. In the case of proteins the presence of left-handed α -helices or D-amino acid conformations would most likely indicate the wrong choice of spacegroup enantiomorph.

The last step in data processing is the calculation of structure factor amplitudes $|F_{hkl}|$ from the measured intensities I_{hkl} and absolute scaling. The structure factor amplitude R-factor R_{sym} or $R_{sym,F}$ is the commonly presented statistic for the quality of the data and is calculated analogous to Eqn. 10

Another variant of data collection is the Laue technique. In this approach the crystal is exposed to a polychromatic, white X-ray beam, i.e. a broad range of wavelengths. This leads to sampling of a large portion of reciprocal space in a single exposure. This is particularly attractive for rapid time-resolved data acquisition for detecting fast dynamic processes such as enzymatic reactions. In practice there are significant problems with reflection overlap leading to incomplete data sets. In the case of enzymatic intermediates it often proves more practical to isolate or trap reaction step intermediates via substrate or transition state analogs, mutations and rapid quench techniques and then applying monochromatic data collection.

2.5 The Phase Problem

As described above reciprocal and real space are directly interrelated by the Fourier transform. The quantities measured in a crystallographic experiment are however only the amplitudes of the structure factors. In order to successfully calculate the electron density distribution by the Fourier summation, it is necessary to recover phase information through other means. A phase sensitive detector would be desirable, but is for various reasons not yet achievable. The inability to measure phase information for the structure factors is known as the crystallographic phase problem.

Along with the methods of multiple isomorphous replacement, multiple anomalous dispersion and molecular replacement, which will be described below, approaches based on direct methods can also be applied in suitable cases. The direct methods can in principle derive phase information from native data alone. These methods are based upon statistical relationships between some certain phases and structure factors along with identification of partial models with the help of Patterson synthesis. In small molecule crystallography the direct methods are used to routinely solve most crystallographic projects, for protein crystallography however the direct method approaches have been less successfully applied due to typically lower resolution and greater number of reflections in these cases. (Hauptmann, 1997; Sheldrick, 1990; Weeks and Miller, 1999).

2.5.1 Isomorphous Replacement

The method of isomorphous replacement (MIR) is based upon local changes or additions to the electron density distribution of the crystal, whereas the remainder of the crystal remains unchanged or isomorphous to the original crystal (Green, *et al.*, 1954). There are a variety of approaches for introducing such changes. Crystals can be placed into a heavy atom containing solution, amino acid residues can be exchanged for heavy atom containing analogs, e.g. S-Met for Se-/Te-Met or metals present in the original crystal can be removed and exchanged for heavier, such as Mg for Mn. Endogenous cofactors may be labelled with heavy atoms. Even rapid exposure to halides such as Cs or I may suffice to derivitize a crystal. The changes must induce large enough intensity changes to be measured above the noise level. At first glance an additional heavy atom in the presence of thousands of light protein atoms might not be expected to contribute to any significant change in intensity, but it

can be shown that a single Hg atom added will significantly change the intensity of reflections due to the coherent scattering of the Hg electrons relative to the light atom electrons. With the substructure coordinates in hand and the structure factors for the native $|F_p|$ and derivative data sets $|F_{ph}|$ it is possible to calculate phases (Figure 5).

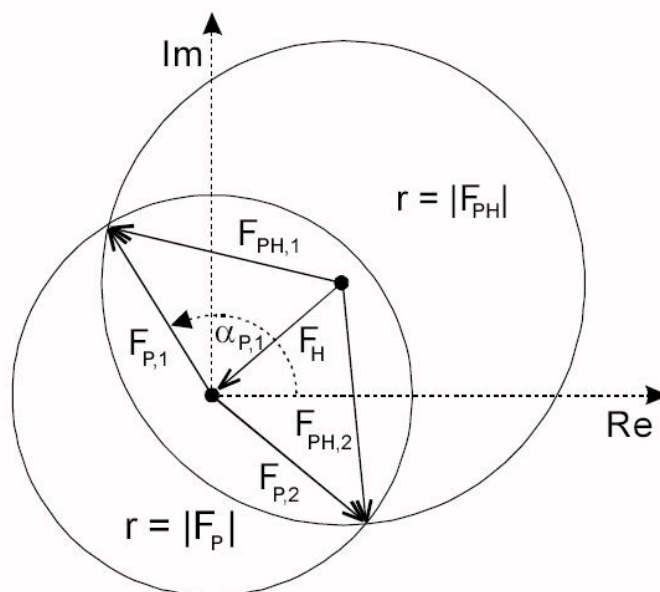


Figure 2.5 Phase determination by Isomorphous Replacement. The heavy atom structure factor is described by vector (F_H). At the beginning and end of F_H a circle is drawn with a radius of the structure factor amplitude for the derivative data set ($|F_{PH}|$) and native data set ($|F_P|$) respectively. The intersections between the two circles yield the possible native protein structure factors ($F_{P,1}$ and $F_{P,2}$) and their phases ($\alpha_{P,1}$ and $\alpha_{P,2}$). There is an ambiguity between possibility 1 and 2 which can be broken with additional derivatives or anomalous dispersion.

The basis of this approach is that the total model structure factor ($F_{tot, hkl}$) is a vector sum of the structure factors of the individual atoms ($F_{i,hkl}$):

$$F_{Total, hkl} = \sum_i F_{i,hkl} = F_{P,hkl} + F_{H,hkl} \quad (11)$$

F_P and F_H are the structure factors of the protein and heavy atom model of the derivatized crystal.

The first step of phase determination is the solution of the heavy atom substructure. The structure factor amplitudes of the heavy atom substructure $|F_H|$ can

be calculated from the structure factor amplitudes of the native $|F_P|$ and the derivative data sets $|F_{PH}|$:

$$|F_H| = |F_{PH}| - |F_P| + \Delta \quad (12)$$

The parameter Δ contains measurement errors, errors that arise due to less than 100% isomorphism between derivative and native datasets. The various errors lead to increase in noise. For most indexes hkl the structure factors $|F_{PH}(hkl)|$ and $|F_P(hkl)|$ are much larger than $|F_H(hkl)|$ and $|F_H(hkl)|$ can be approximated as the difference between $|F_{PH}(hkl)|$ and $|F_P(hkl)|$.

The heavy atom substructure can be calculated after relative scaling of derivative and native datasets by using $|F_{PH}| - |F_P|$ as the structure factor amplitudes of the heavy atom model with either difference Patterson or direct methods. The difference Patterson function is a Fourier summation with coefficients $(|F_{PH}| - |F_P|)^2$ and phase angles equal to zero. The peaks which arise in difference Patterson maps are related to the distances between atoms in heavy atom model. When spacegroup symmetry is considered it is possible to determine x, y, z coordinates for the heavy atom substructure. As a result of centrosymmetry the Patterson function cannot discriminate between coordinates x, y, z and $-x, -y, -z$. Subsequent phase calculations in general use both possible coordinates. The correct coordinates will yield the true electron density map and the inverted coordinates will yield the mirror image with for example left- instead of right-handed α -helices. The correct maps will be selected by visual inspection of the maps.

When initial phases are available, e.g. from initial derivatives, further heavy atom positions can be found via difference Fourier electron density maps. This involves calculation of a Fourier summation with coefficients $|F_{PH}| - |F_P|$ and protein phase angles α_P . Peaks in the difference Fourier density maps correspond to the location of additional heavy atoms. Furthermore the difference Fourier density exhibits a greater contrast than the equivalent Patterson maps. The coordinates of the new atoms are relative to same origin as the derivative from which the phases α_P were used.

The previous discussion assumed free electron behaviour where the phase of the scattered X-ray is exactly 180° to the incoming X-ray phase. This assumption is no longer valid when the energy of the incoming X-ray approaches that of a transition or

2. Macromolecular Crystallography

absorption edge of the atoms in the scattering material. In this case inner core electrons can be excited or even ejected from their orbitals. This leads to phase shifts which are not equal to 180° . Through this anomalous dispersion effect Friedel's law no longer holds, i.e. for these atoms the relationship, $F_{hkl} = F_{-h-k-l}$ no longer applies. The difference between F_{hkl} and F_{-h-k-l} therefore equates to a phase error in the structure factor amplitude of the anomalously scattering atom. Application of this anomalous difference leads to additional vectors in figure 2.5 which allow discrimination of the phase ambiguity. Anomalous differences can also be combined with direct or Patterson methods to obtain the heavy atom substructure. Although the anomalous differences are typically very small relative to the measured intensities and force stringent requirements on data collection and processing, they contain important information since they are collected from a single sample and therefore not subjected to errors and subsequent noise due to intra-sample non-isomorphism.

2.5.2 Multiple Anomalous Dispersion

Multiple anomalous dispersion can be viewed as a special form of isomorphous replacement. A major requirement is the freely adjustable wavelength of the X-ray source, i.e. a synchrotron source and the presence of anomalous scattering atoms in the crystalline sample. These scatters may be endogenous components, such as the Fe of heme cofactors or may be derivatives from heavy atom soaking. A particularly attractive method has been developed over the years and involves the *in vivo* incorporation of the selenium analog of sulfur methionine by auxotrophic expression hosts or by repression of endogenous methionine synthesis. In these cases growth in medium containing selenomethionine often leads to full incorporation of Se at all methionine S sites (Doublí, 1997). The change in the real component of the scattered radiation, or the dispersive difference can be measured by collecting data around the absorption edge of the anomalous scatterer. The absorption edge must be well defined for each sample measured, which is in practice done by fluorescence scans of the sample. Furthermore the anomalous effect at the absorption edge, that is the difference between F_{hkl} and F_{-h-k-l} will be used for phase determination. For samples which contain a naturally occurring anomalous scatterer this approach can spare the difficulties associated with finding suitable heavy atom derivatives.

The form factor for an anomalous scattering atom, f_{anom} can be described through the wavelength dependant $f(\lambda)$ and independent f_0 components. The

wavelength dependant component leads to an intensity decrease due to absorption effects ($f''(\lambda) < 0$) as well as to a phase shift as described above, ($if'''(\lambda)$).

$$f_{anom.} = f_0 + f(\lambda) = f_0 + f'(\lambda) + if'''(\lambda) \quad (13)$$

Both the dispersive difference, which yields different values $f'(\lambda)$ from measurements at different wavelengths and the anomalous difference, which dependant upon the size of $if'''(\lambda)$ leads to the breakdown of Friedel's law, are within phase error proportional to the structure factor amplitude of the heavy atom substructure. Therefore it is possible with these differences to determine the heavy atom substructure and since both terms are orthogonal to each other, uniquely determine the phase. In principle two measurements, at different wavelength would suffice. However to maximize both differences, usually data are collected at maximum f'' and at minimum f'' and at a wavelength far from the absorption edge. The third data set for which f'' is essentially zero serves to maximize the dispersive difference.

As mentioned above the determination of phases from anomalous effects in MAD contain no errors due to non-isomorphism because all data are collected from the same sample. In practice this requires however extremely radiation insensitive samples or cryogenic data collection in order to minimize radiation damage induced non-isomorphism. Even with cryo-techniques phase determination from anomalous dispersion run afoul of radiation damage. This situation has been turned on its head and radiation damage has begun to be used as a source of isomorphous change and phase information. Phases obtained from anomalous methods are usually of excellent quality but success depends on the fact the the measurement error for structure factor amplitudes is small and does not swamp out the small difference signal.

A further advantage of MAD over MIR phasing is in the different resolution dependence of the wavelength dependent and independent components of the atomic structure factor. The independent component, f_0 encompasses all electrons of the anomalous scatterer and exhibits a normal decrease in scattering with resolution, whereas the wavelength dependant terms, $f'(\lambda)$ and $f''(\lambda)$ only involve the inner core electrons of the anomalous scatterer. These terms are nearly constant versus resolution and the relative contribution of these terms and thus the quality of the phases derived, the phasing power, increases with increasing resolution. In the case of MIR the non-isomorphous errors lead to a decrease in phasing power with increasing

2. Macromolecular Crystallography

resolution. The MIR method usually only allows calculation of phases to $\sim 3\text{\AA}$, whereas in MAD the scattering limit of the sample sets the resolution. The only disadvantage of the MAD approach is the need for a tunable X-ray source.

2.5.3 Molecular Replacement

In the molecular replacement method a particular native dataset is phased by determining the correct orientation of a search model with sufficient structural homology to the protein under investigation. The calculated phases from the search model are used with the measured structure factor amplitudes to calculate electron density maps. If the search model possesses sufficient homology the calculated maps should allow the new structure to be modelled and refined. Whether or not the search model is sufficiently similar to allow a solution to be found must be tried anew with each project. Roughly speaking the C α -atom RMSD should not exceed 1 \AA . For protein:substrate complexes or for proteins with high sequence identity, >20%-30% this is often fulfilled. With the enormous increase in the number of known protein structures deposited the chance continues to rise that a suitably similar structure may already be available for molecular replacement.

When a suitable search model is at hand the first step is to find the correct position of the model in the unit cell. There are three rotational and three translational degrees of freedom, i.e. six degrees of freedom in total. It is in principle possible to perform a six-dimensional search, but in practice proves to be too computationally intensive. Application of the Patterson functions (Patterson, 1934) allows for separation of the rotational and translational degrees of freedom which results in splitting the task into two three-dimensional searches, a rotational search followed by a translational search. The computational requirements are reduced in this case (Hoppe, 1957; Rossmann and Blow, 1962; Huber, 1965).

The Patterson function $P(\mathbf{u})$ is the convolution of the electron density distribution $\rho(\mathbf{r})$ in the unit cell with itself. It contains information on all vectors $\mathbf{u}=(u,v,w)$ which describes the interatomic distance between atom at position $\mathbf{r}=(x,y,z)$ and the atom at position $\mathbf{r}+\mathbf{u}$ and can be calculated without knowledge of the phases. The important point is that the intramolecular vectors \mathbf{u} , which describe distances within the molecule only depend upon the rotational orientation of the molecule, whereas the intermolecular vectors depend on both the orientation and the position of

the molecule in the unit cell, i.e. depend on both rotational and translational degrees of freedom.

$$P(u) = \int_r \rho(r) \rho(r+u) dr \quad (14)$$

$$P(u) = P(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}|^2 \cos 2\pi(hu + kv + lw) \quad (15)$$

The Patterson search involves incrementally rotating a search model in a three-dimensional search. For each orientation structure factors, F_{calc} are calculated and a Patterson function calculated. The Patterson density is then compared with the Patterson function calculated from observed structure factors of the target protein data set, F_{obs} and the degree of agreement measured by a correlation coefficient $C(|F_{obs}|^2, |F_{calc}|^2)$.

$$C(P_{obs}, P_{calc}) = \int_u P(|F_{obs}|^2, u) \times P(|F_{calc}|^2, u) du \quad (16)$$

A high degree of correlation between search and target Patterson functions will give rise to a peak in the correlation function. Rotational orientations with high correlation coefficients are then taken to the next step in which the orientated model is moved incrementally through the unit cell in a three-dimensional translational search. Correct positions in the translational search can be determined from the correlation coefficient as in the rotational search as well as from the crystallographic R-factor, given below. In practice the principles of molecular replacement have been codified in various programs.

The separation of the six-dimensional search into two three-dimensional search steps leads to a reduction in contrast of the Patterson rotation search. In some cases it is possible from symmetry considerations to reduce the degrees of freedom of the search such that it becomes computationally possible to search the remaining dimensions directly. The resulting direct search can, through the use of the crystallographic R-factor sometimes find solutions that are not detected by the Patterson method.

2.6 Electron Density Modification

The methods described above do not always deliver an electron density map of sufficient quality for complete model interpretation. This is due in particular to errors

2. Macromolecular Crystallography

associated with the experimentally determined phases, α_{exp} and to errors in the experimentally measured structure factors, $|F_{obs}|$. Electron density modification involves incorporation of additional information into the electron density map. After this the modified electron density is back transformed to give modified phase angles, α_{mod} which are combined with the experimentally measured structure factors, $|F_{obs}|$ to deliver an improved electron density map. Since the experimental structure factors, $|F_{obs}|$ are more accurately measured than the phases, they are always used in the electron density map calculation. The modification cycle is repeated until convergence. The modifications applied are based on various experimental observations.

The first observation used in density modification is the separation of protein and solvent regions. The algorithm derived by Wang (Wang, 1985) determines which regions of the unit cell belong to protein and which to solvent. The electron density associated with solvent regions is then set to a small, constant value or “solvent-flattened”. In order to assign the solvent region it is necessary to know the solvent content of the crystal. This is determined usually from the spacegroup, unit cell dimensions, molecular weight of the protein and the number of molecules per asymmetric unit (Matthews, 1968).

If more than one copy of the molecule are present in the asymmetric unit or the protein can be crystallized in more than one crystal form, then the density can be modified by averaging using non-crystallographic symmetry or inter-crystal averaging (Kleywegt and Read, 1997). It is necessary to know the orientation of the protein in the unit cell and the protein:solvent boundary.

Another mode of electron density modification is based on the fact that the number of points of a particular value in the electron density distribution is dependant only on the resolution of the data set and not the nature of the protein (Zhang and Main, 1990). Adjustment of the experimental electron density histogram, i.e. the number of map points of particular value versus map value can be achieved to fit the optimal histogram. This histogram matching method can lead to improvement in map interpretation.

Additional methods of density modification may utilize, skeletonization (Baker *at al.*, 1993; Swanson, 1994), application of the Sayre equation (Sayre, 1974; Zhang and Main, 1990) and solvent flipping as an extension of solvent flattening (Bricogne, 1974).

The above mentioned density modification approaches are available in a variety of software packages and generally applied in various combinations to obtain density maps of higher interpretability. These approaches are also often applied to phase extension of data from lower to higher resolution.

2.7 Map Interpretation

The map interpretation or model building process is performed primarily on computer graphics systems preferably supporting stereographic display modes. Several model building environments have been developed over the years. These all have in common that they can display electron density maps, calculated from the methods described above at some suitable contour level. Effectively the electron density is displayed on a grid as a contoured tube inside of which the electron density is higher than a particular cutoff and outside the density is below this cutoff. Furthermore the programs all display the standard amino acid residues, nucleic acid residues and various common cofactors. Additional molecular types and covalent modifications can in general be incorporated into the programs library by the user. These objects can be displayed typically as line bond models and moved or rotated at will in order to position the fragment in the electron density. In this way a polypeptide chain can be built up in the electron density map under inspection. The programs provide a variety of additional powerful features for maintaining proper stereochemical restraints and for optimizing the fit of the model to the local electron density region. Stereochemical and rotamer libraries as well as fragment libraries provide rapid means for fitting side chains and even longer segments of the protein model.

When the electron density maps have been derived from the MIR or MAD methods then there is initially no model available and the map is empty. This map is generally skeletonized to yield a trace along the highest density of the map. The map and skeleton is then inspected for typical secondary structure elements such as α -helices and β -sheets. These sections can be initially interpreted as poly-alanine sequences. In cases of poor map quality these sections can be partially refined and used to calculate phases which are combined with the experimental phases to obtain modified phases which may improve the calculated maps and show improved connectivity between secondary structural elements and improve side chain identification. In order to identify side chains and assign a sequence to the model the

2. Macromolecular Crystallography

primary amino acid sequence is consulted and compared to the side chain electron density in map, large side chains in particular aromatic side chains, such as Phe, Tyr and Trp will show characteristic side chain densities that can be used to confidently assign and anchor the sequence of the model. In the case of selenomethionine phased data from anomalous dispersion, methionine sites can be identified. Once anchor points are identified it is possible to use primary sequence knowledge to extend the sequence in the model. With maps of sufficient quality and current model building programs it is possible to directly build the known sequence into the map without the need for poly-Ala fragments.

If the electron density map is derived from molecular replacement methods then the search model is already positioned and can serve as a basis for rebuilding of the new model sequence. Regions of electron density map or difference maps that show significant deviations from the search model can be rebuilt in an effort to obtain map improvement. A problem that arises with molecular replacement solutions is that of model bias. This is a result of the fact that the phases dominate the electron density and the electron density is calculated from search model phases and experimental structure factors. This results in electron density maps that tend to resemble the input search model and not the desired target model. If electron density features can be found that are expected for the target model, i.e. ligands or cofactors which are not present in the search model then one can be confident of the solution.

In order to reduce the effect of model bias on molecular replacement solutions in regions of uncertain electron density omit maps are applied. Omit maps are electron density maps calculated in which a region of the model, e.g. an uninterpretable section of map is omitted. In this way the contribution of the search model is reduced and primarily information from the experimental structure factors is enhanced. In practice composite omit maps are iteratively calculated over the entire molecule in which a small region has been omitted during each step of the calculation. Furthermore simulated annealing methods are applied to shake the remaining coordinates by a small amount thus blurring the contribution of the remaining atoms to the map calculation (Turk, 1992).

In any case cycles of electron density map interpretation are followed by refinement of the current model and then new density maps are calculated with phases calculated from the refined model. Improvements in the model will lead to improved

phases, this in turn leads to improved density maps. The cycle is repeated until convergence.

Recently a variety of programs have been developed which automate various aspects of model building. This includes map skeletonization (Swanson, 1994) which can simplify the detection of secondary structure elements, rotamer libraries for selecting the optimal side chain conformation (Jones, et al., 1991) and procedures for automatically modifying the model along with crystallographic refinement cycles until converged (Lamzin and Wilson, 1997). The last procedure can be used to automatically build waters into a model structure or rebuild models placed by molecular replacement when data of sufficient resolution and phases of high quality are available.

2.8 Model Refinement

After the interpretation of the electron density map and cycles of model rebuilding, in general the model will be refined against the experimental structure factors. During refinement the calculated structure factor amplitudes derived from the model, $|F_{calc}|$ are brought into optimal agreement with the experimentally derived structure factors $|F_{obs}|$. In order to achieve this the positions of model atoms are adjusted, positional refinement along with B-factors, B-factor refinement. In the case of macromolecular crystallography the number of measured parameters, i.e. the number of observed structure factor amplitudes to a particular resolution is not sufficiently greater than the number of total parameters for the model. Therefore the minimization of $|F_{obs}|$ and $|F_{calc}|$ is not enough and additional geometric restraints must be utilized. The geometric restraints are derived from small molecule structure data on individual amino acids and other molecules which have a far higher degree of accuracy (Engh and Huber, 1991). The geometric restraints are introduced as force field restraints or targets in the refinement programs. Furthermore neighboring atoms should exhibit similar dynamics which can be introduced as a thermal B-factor restraint between neighboring atoms.

2. Macromolecular Crystallography

The quality of agreement between $|F_{obs}|$ and $|F_{calc}|$ is expressed through the crystallographic R-factor:

$$R = \frac{\sum_{hkl} \left| |F_{obs,hkl}| - k |F_{calc,hkl}| \right|}{\sum_{hkl} |F_{obs,hkl}|} \quad (17)$$

Here k is a scale factor which brings the measured structure factor amplitudes and the model calculated structure factor amplitudes to the same scale. A model can be viewed as correct if the R-factor is near or below 20% with sensible geometry. An additional free R-factor (R_{free}) is calculated in order to assess model quality (Brünger, 1992a). The R_{free} is calculated for a small set, 5%-10% of experimental structure factor amplitudes which have been set aside and not used in the model refinement process. The agreement between these reflections and the same set calculated from the model is used to determine if model has actually been improved and reflects reality better, or if the improvement in agreement between $|F_{calc}|$ and $|F_{obs}|$ is more likely due to modeling of noise and measurement errors, i.e. over fitting of the model. In general the difference between R and R_{free} should not exceed 5% for a well refined model.

In order to improve the radius of convergence of the refinement process, i.e. reduce the chances of being trapped in a local minimum molecular dynamic simulation has been incorporated into the refinement process. In these simulated annealing approaches the model is heated while maintaining restraints set by crystallographic R-factor and stereochemical targets. Individual atoms are set in motion and their positions calculated from Newtonian dynamics. Typically the model is brought to 3000-4000°K and then slowly cooled. The atoms will then slowly decrease the degree of motion around their equilibrium positions. The increase in convergence radius over least square approaches lies in the ability of the molecular dynamics process to lift the atoms out of local minima.

2.9 Crystallographic Terminology

Completeness - Number of actual measured reflections compared to the number of theoretically possible reflections.

Dataset - Result of processing a series of diffraction images where the measured reflections are given as numerical values, structure factor amplitudes $|F_{hkl}|$

Derivative Dataset - A reflection dataset from a protein crystal modified by heavy atoms which give rise to intensity changes and phase information.

Difference Fourier - A Fourier summation in which differences between measured structure factor amplitudes are used as coefficients. The resulting density maps show molecular differences produced by the structure factor amplitude differences.

Difference Patterson - Patterson function in which differences between measured structure factor amplitudes are used as coefficients.

F_{calc} - Structure factor amplitudes calculated from a molecular model.

F_{obs} - Experimentally observed structure factors amplitudes.

Figure of Merit - The quality of the experimentally determined phase, described as the cosine of the average phase error. For a given structure factor, F_{hkl} the FOM is:

$$m = \int_{\alpha} \{ P_{hkl}(\alpha) \exp[i\alpha] \} d\alpha ; 0 \leq m \leq 1.$$

P_{hkl} is calculated during the phasing process and gives the probability that the structure factor F_{hkl} has phase angle α .

Multiplicity - The average value of the number of times a reflection is measured for a given crystal symmetry.

Native Dataset - A dataset obtained from an unmodified crystal sample.

Number of Reflections - The number of crystallographically independent reflections or structure factor amplitudes.

R-factor - A description of the quality of a refined crystal structure as a function of the agreement between $|F_{calc}|$ and $|F_{obs}|$ as in Eqn. (17)

R_{free} - The free R-factor calculated for a small percentage of reflections withheld from the model refinement process.

R_{sym} - The quality of a measured dataset described as the agreement of symmetry related reflections. This is calculated from the measured structure factor amplitudes $|F_{hkl}|$:

2. Macromolecular Crystallography

$$R_{sym} = \frac{\sum_{hkl} \frac{1}{N} \sum_{j=1}^N ||F_{hkl}| - |F_{hkl}(j)||}{\sum_{hkl} |F_{hkl}|}$$

Where $|F_{hkl}|$ is the average of the j symmetry related reflections, $|F_{hkl}(j)|$ of the structure factor amplitude of reflection hkl .

Reflection - A diffracted X-ray beam and the resulting measured structure factor amplitude.

Resolution - Smallest resolvable distance between objects for the wavelength of light used. The distance between lattice as described in Eqn. (1)

Rigid Body - The refinement of the whole molecule as a rigid, inflexible body.

RMS or r.m.s. - Root mean square. Used to describe deviations or differences.

In contrast to the geometric average the RMS weights larger differences heavier:

$$\text{RMS deviation} = \frac{1}{n} \sqrt{\sum_n (x_n - \bar{x})^2}$$

Where n is the number of measurements, x_n the n -th measurement and \bar{x} the average.

Simulated Annealing - A refinement process involving molecular dynamic simulation and computational heating of the system to increase the radius of convergence of the refinement.

Solvent Region - The volume of a protein crystals not occupied by the protein molecule. The region occupied by solvent molecules.

Solvent Content - The amount of the protein crystal occupied by solvent.

Test Set - The subset of structure factor amplitudes withheld from the refinement process in order to calculate a R_{free} value.

Working Set - The set of reflection used in the refinement process. The reflections which do not belong to the test set.

3. Results and Discussion

3.1. MkFTR Structure from PEG Crystal form P.

M. kandleri FTR was cloned and expressed in *E.coli*. Purified MkFTR yielded three crystal forms from crystallization screens. Table 3.1 shows relevant data for crystal forms, M (MPD), P (PEG) and S, ((NH₄)₂SO₄). (Shima S, *et al.* 1996)

Crystal Form	Conditions	Crystal Parameters
M	20% 2-Methyl-2,4-pentanediol 0.7M Na Cl 0.2M Na.Acetate pH 4.6 4°C	P2 ₁ a=87.0Å b=75.4Å c=204.7Å α=90° β=113.9° γ=90° V _m =2.5Å ³ /Da Solvent content=50% Assuming a tetramer per AU
P	22% PEG 8000 0.3M (NH ₄) ₂ SO ₄ 0.1M MOPS pH 7.0 4°C	I4 ₁ 22 a=157.5Å b=157.5Å c=242.1Å α=90° β=90° γ=90° V _m =3.0Å ³ /Da Solvent content=58% Assuming a tetramer per AU
S	1.2M (NH ₄) ₂ SO ₄ 1% PEG 8000 0.1M Na.Citrate pH 5.6 4°C	I4 ₁ 22 a=151.3Å b=151.3Å c=249.5Å α=90° β=90° γ=90° V _m =2.8Å ³ /Da Solvent content=56% Assuming a tetramer per AU

Table 3.1 Relevant data for MKFTR crystal forms, M (MPD), P (PEG) and S, ((NH₄)₂SO₄).

Crystals of form P appeared within two weeks and grew to typical dimensions of 0.6 x 0.6 x 0.8 mm³ at pH 7.0 at 4°C. Crystals adopted the spacegroup I4₁22 Å with cell parameters a=157.5 Å, b=157.5 Å, c=242.1Å, α=90°, β=90° and γ=90° and diffracted to 1.73 Å resolution. Assuming four monomers per asymmetric unit the V_m value was 3.0 Å³/Da giving a solvent content of 58%.

3.1.1 MIR ,Model Building, Refinement and Model Quality

The structure of MkFTR from crystal form P was solved by MIR phasing from two derivatives. Table 3.2 gives the relevant statistics for the data collection and

3. Results and Discussion

phasing. One site from the lead derivative could be identified in the difference Patterson map with SHELXS (Sheldrick, G.H., 1991). Remaining sites were identified via Fourier difference maps and sites were refined with MLPHARE (CCP4). The initial MIR map was improved by density modification with DM (CCP4) by applying solvent flattening, histogram matching and four-fold NCS averaging. Model building was performed with O (Jones, T.A. *et al*, 1991). The initial model was built lacking a poorly defined loop region from residue 210-230. Preliminary rounds of refinement improved the interpretability of this loop region and allowed proper model building.

Dataset	Native 1	Native 2	Mersalylic Acid	(CH₃)₃Pb
Resolution (Å)	2.7	1.73	3.5	3.2
R _{sym} (%)	7.0	5.4	13.9	8.9
Completeness (%)	87.6	87.8	67.0	90.3
Multiplicity	2.8	3.0	1.8	2.9
R _{merge} (%)	N/A	N/A	31.4	13.5
Sites / monomer	N/A	N/A	2	2
R _{cullis} (%)	N/A	N/A	0.79	0.64
Phasing Power	N/A	N/A	1.1	1.2

Table 3.2 Relevant statistics for the data collection and phasing of MkFTR crystal form P.

The electron density maps revealed a D₂ tetramer in the asymmetric unit of the crystal. According to light scattering analysis MkFTR is present in solution at least partially in a monomeric form (Shima *et al* 1998). Apparently the crystallization conditions employed were sufficient to drive MkFTR to adopt a tetrameric state even in the absence of high salt concentration. Alternatively the crystal nucleation and growth steps may have (1) involved only the tetrameric MkFTR sub-population or (2) fortuitously allowed the monomeric sub-population to interact at the crystal growth surfaces only in orientations analogous to tetrameric interactions or (3) a mixture of these processes. Since the monomers evolved under the selective pressure that they form homo-tetramers then the most favourable, low energy interactions adopted during crystallization should be expected to lead to tetramers in the crystal lattice.

The model of MkFTR from crystal form P was refined with X-PLOR (Brünger, A.T., 1992). Table 3.3 give relevant refinement statistics for the P form MkFTR model. Conventional conjugate-gradient coordinate refinement was applied

with coolstage molecular dynamics. Initially the model was refined under full NCS restraints against native data to 2.0 Å. After completion of poorly defined regions of the model the refinement was carried out with data to 1.73 Å resolution. Full resolution refinement required removing NCS restraints as this led to run-time failures with X-PLOR.

Parameter	
Unit-Cell (Å,°)	157.5, 157.5, 242.1, 90, 90, 90
Spacegroup	I4 ₁ 22
Resolution (Å)	10.0-1.73
Number of reflections	
Total	406062
Unique	135354
R _{sym} (%)	5.4
Completeness (%)	87.8
Multiplicity	3.0
R _{free}	0.248
R _{work}	0.196
RMSD bond length (Å)	0.008
RMSD bond angle (°)	0.973
Mean coordinate error from Luzzati plot (Å)	0.30-0.35 (Free reflection set) 0.25-0.30 (Work reflection set)
Diffraction data precision indicator, $\sigma_w(x)$ (Å)	0.14 (Free reflection set) 0.10 (Work reflection set)
Number of residues	296*4
Number of atoms	9711
Number of waters	827
Real-space fit	0.86
Average B-factor	0.22 Å ²
Ramachandran plot, number of non-Glycine residues in,	
Disallowed regions	0 (0.0 %)
Generous regions	3 (0.3 %)
Allowed regions	74 (7.5 %)
Favoured regions	911 (92.2%)
Alternate conformations	C*58 Y*60
Residues in poor electron density	E*111, E*133, E*136 E*232, D*233, E*238, K*241, E*254

Table 3.3 Relevant refinement statistics for the P form MkFTR model

During interactive inspection of maps and model within O, the individual monomers were superimposed along with maps. In this way regions of similarity could be maintained and allowed manual enforcement of NCS for similar segments. Deviations

3. Results and Discussion

of the superimposed chains indicated that monomers 1 and 3 exhibited similar backbone conformations and formed a group slightly different from monomers 2 and 3 which shared greater backbone conformation similarity to each other. These two groups result from crystal contacts which cause slight deviations of the backbone structures.

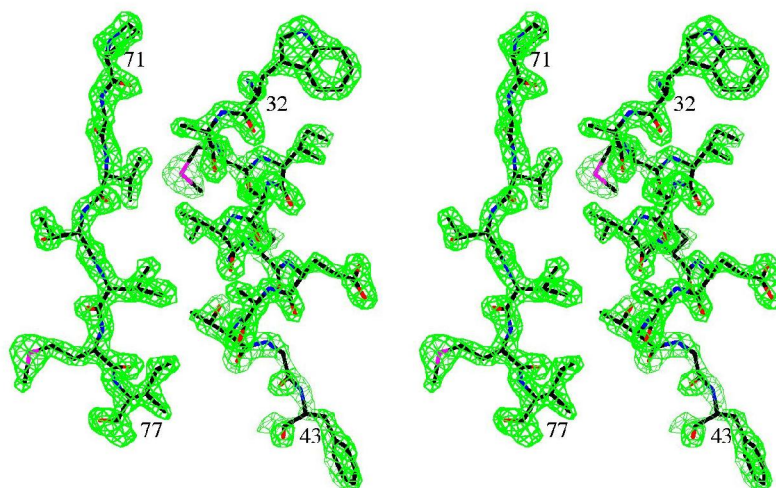


Figure 3.1 Stereoview of the MkFTR model and electron density highlighting the quality of the final electron density map

The MkFTR model was refined to a conventional R_{work} and R_{free} of 19.2% and 24.3% respectively in the resolution range of 10.0 – 1.73 Å. Figure 3.1 shows the quality of the final electron density map. All 296*4 amino acid residues, 8884 atoms could be modeled along with 188, 150, 181 and 163 water molecules for monomer chains A, B, C, and D respectively. Residue sidechains Cys58 and Tyr60 exhibit alternate conformations, Figure 3.2. Poor sidechain electron density was seen for residues, E*111, E*133, E*136, E*232, D*233, E*238, K*241, and E*254. The root-mean-square deviations (RMSD) of the model from ideal stereochemistry was 0.008 Å and 0.973° in bond-length and bond-angle respectively. The average B-factor was 0.22 Å² and mean-coordinate error determined from the Luzzati plot was 0.25 Å and 0.30 Å for R_{work} and R_{free} datasets respectively. Since the Luzzati analysis is based on assumptions which do not exactly apply to real crystallographic models the Cruikshank Diffracton Precision Indicator (Cruikshank, D.W.J., 1996) was calculated and yielded a mean-coordinate error of 0.11 Å. The Ramachandran plot (Ramachandran G.N. 1963). All non-Glycine residues of the P form MkFTR model

adopt favourable dihedral, ϕ - ψ angles. The real-space correlation (Jones, T.A. *et al*, 1991) between model and $2|F_o|-|F_c|$ map was 0.86 and no residues having values below 0.7. Superposition of all four monomer chains using only C_α atom coordinates showed nearly identical conformation exhibiting a low overall RMSD of 0.25 Å.

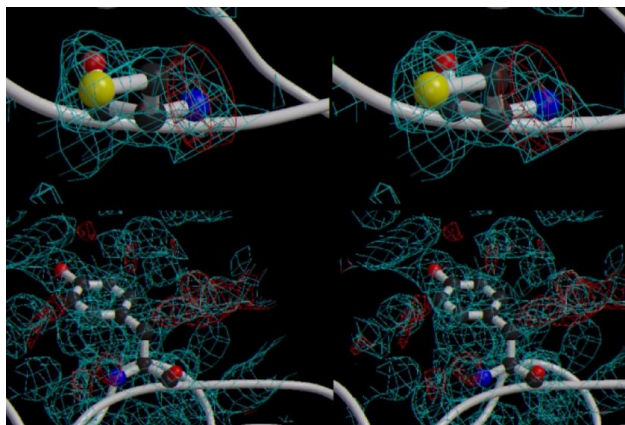


Figure 3.2 Stereoview of alternate conformations observed for Cys58 (top) and Tyr60 (bottom).

3.1.2 Monomer Structure

Overall Fold: The MkFTR monomer has rough dimension of 30 Å x 40 Å x 70 Å. Figure 3.3 (left) shows a C_α backbone trace of the MkFTR monomer color coded blue to red from N- to C-terminus with every tenth residue numbered. Figure 3.3 (right) shows the MkFTR monomer with secondary structure elements labeled and highlighted in green, helices, red, strands and yellow coil regions. Figure 3.4 highlights the topology of the MkFTR secondary structure elements.

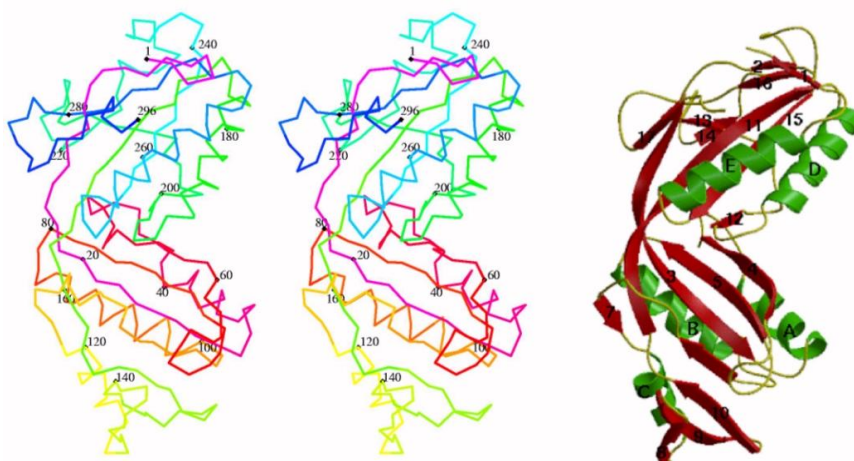


Figure 3.3 Stereoview (left) of the MkFTR C_α trace color coded from blue to red from N- to C-terminus with every tenth residue C_α labeled. Ribbons model of the MkFTR monomer indicating secondary structure elements α -helices in green and β -strands in red

3. Results and Discussion

The structure can be divided into two tightly associated lobes. Residues 1-17 and 163-296 comprise the proximal lobe and residues 18-162 form the distal lobe. The N- and C-termini fall in the proximal lobe

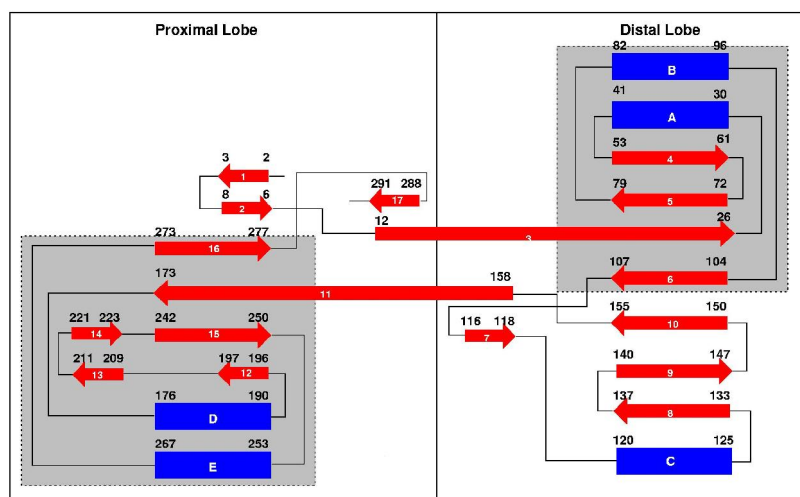


Figure 3.4 Topology diagram of the MkFTR monomer highlighting the secondary structure elements and the pseudo two-fold nature of the N- and C-terminal lobes

Both lobes belong to the $\alpha+\beta$ class of structures. The proximal lobe consists primarily of a four-strand anti-parallel β -sheet extended by two small strands (1 and 2) near the N-terminus. The central β -sheet is surrounded by two helices (D and E) at the front which pack against the distal lobe and the so-called insertion region, a region which is loosely associated with the back of the central β -sheet. The unusual insertion region occurs between residues 198 and 241 and contains only two short anti-parallel β -strands, 13 and 14. The distal lobe consists of a seven-stranded β -sheet composed of a four-stranded anti-parallel β -sheet combined with a β -meander consisting of a three-stranded anti-parallel β -sheet. The whole sheet is flanked by three α -helices, A, B and C on the backside and partly by the proximal lobe on the front side.

The extended and highly twisted β -strands 3 and 11 span both lobes and serve as a covalent linker between them. They form part of the central β -sheet at their C-terminal ends, form a common sheet in the center of the strands and strand 3 joins with strand 7, and strand 11 joins with strand 17 at their N-terminal ends. In addition to this covalent linkage, the two lobes associate strongly through a series of non-covalent interactions either directly or mediated by water molecules.

Both lobes contain hydrophobic core regions, each consisting of about 20 nonpolar side-chains. The hydrophobic cluster of the proximal lobe is located between

the β -strands 11, 12, 15 and 16 and the flanking α -helices, D and E. Residue Phe169 is in the center of this cluster and is in van der Waals contact with eight other aliphatic residues of which six are isoleucines. A preference for isoleucine was described for a thermostable glutamate dehydrogenase. The hydrophobic core in the distal lobe is formed by residues of the β -strands 3, 5, 6, 9 and 10 and by α -helices A, B and C.

Pseudo-Symmetry of the Monomer: The MkFTR monomer can be roughly divided into two similar, tightly associated lobes. Residues 1-17 and 163-296 comprise the proximal lobe and residues 18-162 form the distal lobe. The N- and C-termini occupy the proximal lobe. Figure 3.5 shows the best superposition of the lobes using the program GA_FIT. The RMS deviation for equivalent C_{α} carbons was 1.37Å.

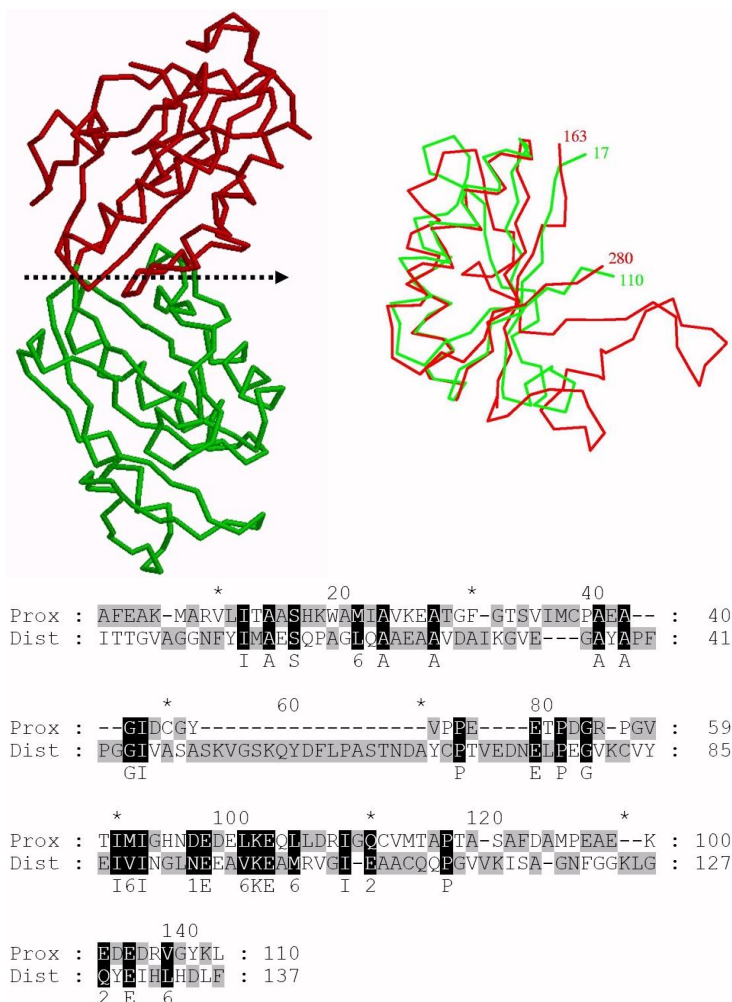


Figure 3.5 Pseudo-symmetry of the MkFTR monomer. View of the best fit superposition (top and left) of the MkFTR monomer lobes and the pseudo two-fold axis (black dotted arrow, top and right). The structural homology of the core regions is evident although the primary sequence indicates no homology (bottom).

3. Results and Discussion

Putative P-loop Walker A-type: The initial analysis of the MkFTR sequence indicated a putative Walker A-type phosphate binding loop, -IVASASKVGSK-. The sequence is most similar to the sequence found in Phosphoglycerate Kinases, (A/V)XXGG(A/S)KVXXXK (Saraste *et al.* 1990). In the crystal structure of yeast Phosphoglycerate Kinase the first lysine of the Phosphoglycerate Kinase phosphate binding loop points towards the α -phosphate of the ATP and the second lysine is near the β -phosphate. The motif is also found in the MkFTR sequences from *M.fervidus* (IVASGSKVGSK) and *M.thermoautotrphicum* (VVASGSKVGSN). In addition the MkFTR from *M.Fervidus* is known to bind ATP-agarose. Since the activity of MkFTR is not affected by the presence of ATP or GTP and the substrates formylmethanofuran (FMF) and tetrahydromethanopterin (H₄MPT) bear no structural similarity to NTPs, the presence of a P-loop consensus sequence is interesting. Perhaps the P-loop sequence is purely coincidental or MkFTR is derived from an ancestral protein that had an ATP or GTP binding function. The crystal structure analysis of MkFTR shows that the P-loop region is in the vicinity of the large cleft feature on the surface of the MkFTR tetramer, see below, however the conformation of the loop differs from that seen in known NTP-binding protein structures and additionally the lysine side chain of the MkFTR P-loop points away from the cleft and is therefore not in a suitable position to interact with ATP or GTP if these substrates were to be bound in the cleft. Figure 3.6 shows the superposed P-loops from Adenylate Kinase, p21-Ras and MkFTR. The MkFTR sequence clearly deviates from the other two structures.



Figure 3.6 Comparison of MkFTR putative P-loop region (red) with other P-loop conformations from Adenylate Kinase (blue), p21-Ras (green)

Although MkFTR does not seem to interact with NTPs the fact that the intracellular osmolyte in *M. kandleri* seems to be cyclic 2,3-diphosphoglycerate the P-loop might play a role in interacting with this compound. In any case the conformation of the MkFTR sequence region is not similar to canonical P-loops.

Fold Similarity: The fold of both lobes of the MkFTR monomer characterize it as a member of the $\alpha+\beta$ family forming two α/β -sandwiches. As seen in other α/β sandwich structures, MkFTR packs the hydrophobic core between the layer of β -strands and α -helices. The core fold of both lobes can be classified as double intertwined split α/β -folds, which is a subgroup of α/β -sandwiches which is found in several functionally distinct proteins. At the time of the structure determination of MkFTR the most closely related structures found with the DALI fold similarity server were; malonyl-CoA:acyl carrier protein transacylase (MLA), D-3-phosphoglycerate dehydrogenase (PSD) and the signal transducing protein P₁I (PIL). Figure 3.7 shows the least square fitted overlay of the MLA structure to MkFTR using the program GA_FIT (May, A., and Johnson, M. 1994).

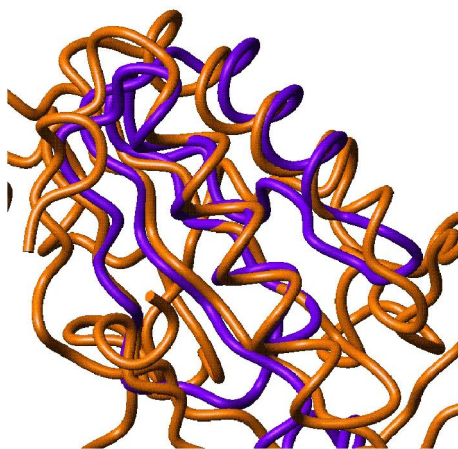


Figure 3.7 Least square fit of MLA (1MUC) structure (blue) to the MkFTR structure (gold).

The RMS deviation between equivalent residues in MkFTR and MLA, PSD and PIL was 1.4Å, 1.6 Å and 1.7 Å respectively in which 53, 55 and 58 of 68, 74 and 87 residues were equivalenced respectively. An additional α/β motif in MkFTR is the $\alpha\beta\beta\beta$ or $\beta\beta\beta\alpha$. meander fold which is connected to the intertwined split $\beta\alpha\beta$ fold forming the distal lobe. Although the overall structure of MkFTR is unique, the

3. Results and Discussion

secondary structure topology and tertiary structure of the individual building blocks are well known.

3.1.3. Quaternary Structure

Under physiological conditions MkFTR adopts a tetrameric form. The configuration seen in crystal supports the structure of MkFTR as a homo-tetramer with dimensions of 60 Å x 75 Å x 80 Å. The tetramer has D_2 , 222 symmetry and each monomer contacts all three symmetry-related partners. The monomers are numbered 1, 2, 3 and 4 and the amino acids of each subunit are indexed with a superscript corresponding to the subunit in order to describe intersubunit interactions.

Parallel to the two-fold axis relating subunits 1 and 4 (2 and 3) is a channel of approximately 27 Å which runs through the entire molecule. In the center of the tetramer the channel widens forming a cavity of 12 Å in diameter. An entrance to the channel of approximately 6 Å is formed by the crevices between monomers 1 and 4 on one side and 2 and 3 on the other side. The channel is nearly completely filled with tightly bound water molecules.

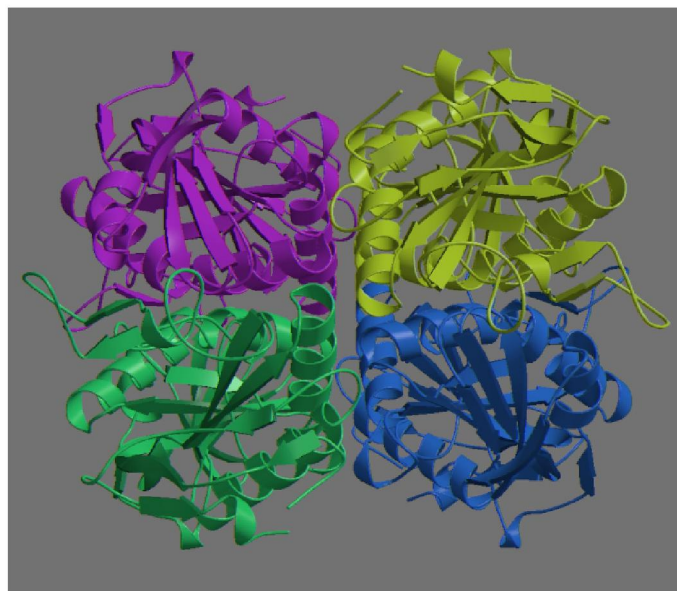


Figure 3.8 View of the MkFTR tetramer structure down the local NCS four-fold tetramer Z-Axis

An additional check of model quality is the correlation of real space fit as determined with the program O (Jones, T.A., *et al.* 1991) to the thermal parameter of the model. Figures 3.9 a-d show the real space correlation with respect to B-factor

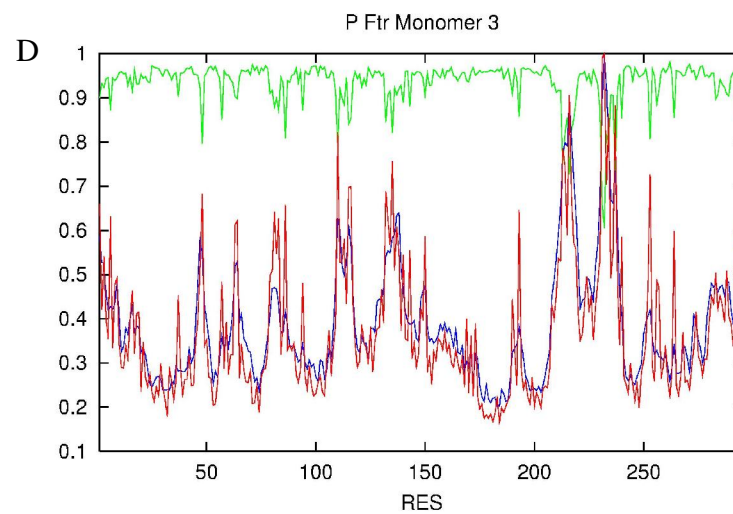
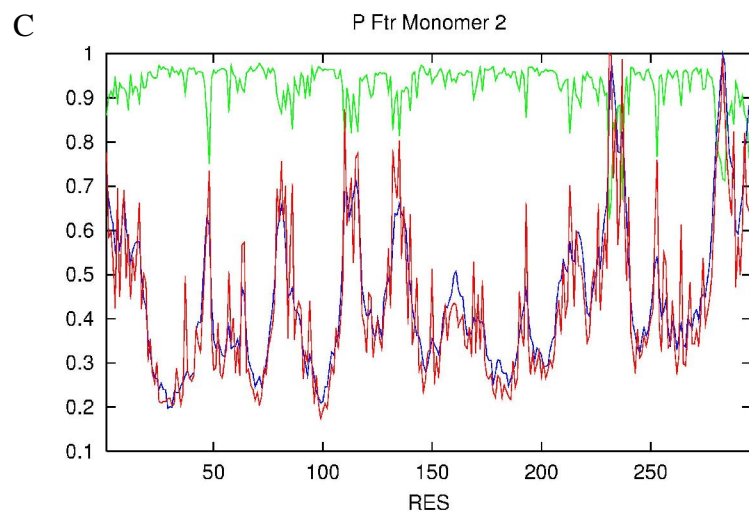
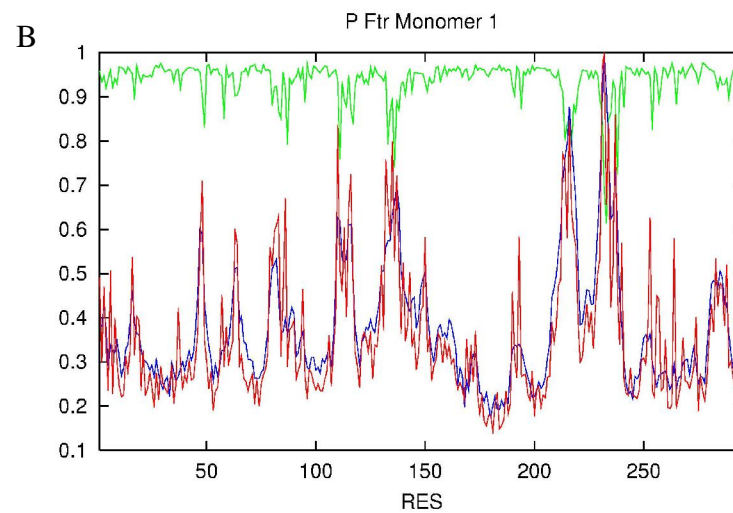
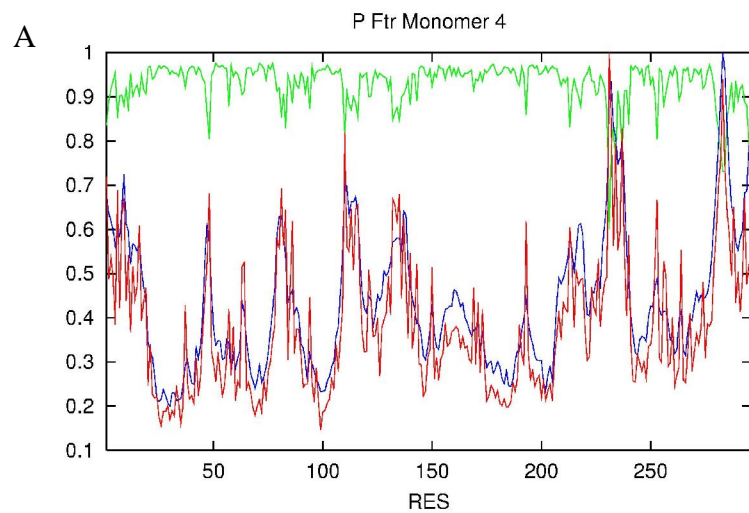


Figure 3.9 Plot of the RS-Fit residuals and mainchain (red) and sidechain (blue) B-factors for the MkFTR monomers. Regions of poorer RS-Fit value correlate with regions of higher thermal motion. Crystal form P MkFTR monomer 1, 2, 3 and 4 are shown in panels A, B, C and D respectively

3. Results and Discussion

along the monomer chains 1, 2, 3, and 4 respectively. In all cases the real space coefficient varies inversely with the thermal parameter as expected. Regions of poor real space fit correlate generally with regions of high B-factor and with regions not associated with secondary structure elements such as helix or sheet.

Comparison of the plots indicates that monomers 1,3 and 2,4 are similar in their real space coefficient/B-factor properties. The differences between the groups 1,3 and 2,4 are due to crystal packing, deviations from NCS, loop coil regions and N-/C- termini. This is seen also in a comparison of the root mean square deviation of the monomers from each other. Figure 3.10 shows the RMS deviation for the overlaid monomers 1:2, 1:3 and 1:4 respectively indicating the dimer of dimer nature of the MkFTR tetramer. The differences between the monomers are due to crystal packing, deviations from NCS, loop coil regions and differences at the N-/C- termini

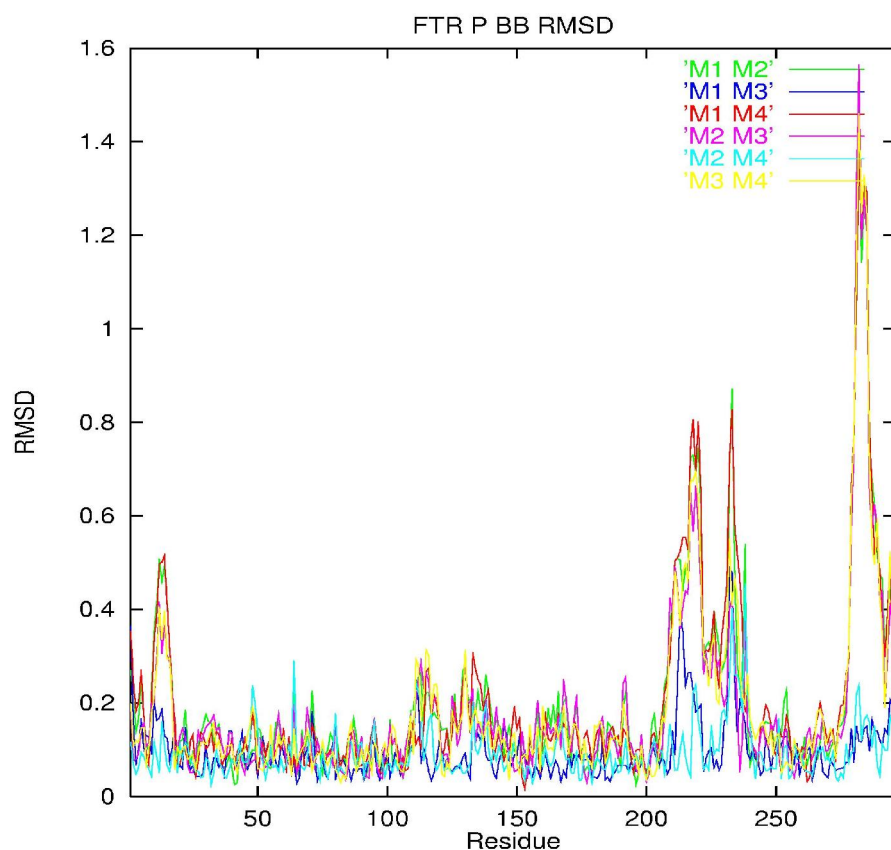


Figure 3.10 Plot of Backbone RMS Deviation between the four MkFTR monomers. The smallest deviation occurs between 1,3 and 2,4. The differences between the monomers are due to crystal packing, deviations from NCS, loop coil regions and N-/C- termini.

Subunit interfaces: The surface area buried by formation of the MkFTR tetramer is $\sim 4000 \text{ \AA}^2$ this implies that about 30% of the solvent exposed area of the monomer is buried after oligomerization. The overall interface is 66% hydrophobic, 18% neutral polar and 16% charged. This compares favorably with values of 65%, 22% and 13% respectively for a set of 23 oligomeric structures. The interfaces of the MkFTR tetramer differ considerably in the amount of area buried, the secondary structure contacts involved and the nature of the interacting forces. Table 3.4 shows statistics for accessible and buried surface areas and composition for the three interfaces in the MkFTR tetramer.

Accessible area (\AA^2)	Overall	Unpolar	Polar	Charged
Monomer	13,622	7,411	2,998	3,213
Dimer M1-M2 M3-M4	10,728	5,452	2,378	2,898
Dimer M1-M3 M2-M4	13,347	7,257	2,717	3,373
Dimer M1-M4 M2-M3	12,870	6,942	2,837	3,092
Tetramer	9,675	4,807	2,174	2,694
Buried area (\AA^2)	Overall	Unpolar	Polar	Charged
Monomer	N/A	N/A	N/A	N/A
Dimer M1-M2 M3-M4	2,894	1,959	620	315
Dimer M1-M3 M2-M4	275	154	75	46
Dimer M1-M4 M2-M3	752	467	127	158
Tetramer	3,947	2,604	824	519

Table 3.4 Statistics for accessible and buried surface areas per monomer and composition for the three interfaces in the MkFTR tetramer

Figure 3.11 a-c shows the relative sizes and compositions of the three interfaces of the tetramer using the program GRASP (Nicholls *et al.* 1993). Subunits 1 and 2 (3 and 4) bury the largest amount of surface area, 2900 \AA^2 which accounts for about 75% of the buried area of the tetramer. As a result one can view the MkFTR

3. Results and Discussion

tetramer structure as a dimer of dimers. The buried area is composed of 68% nonpolar, 21% neutral polar and 11% charged atoms.

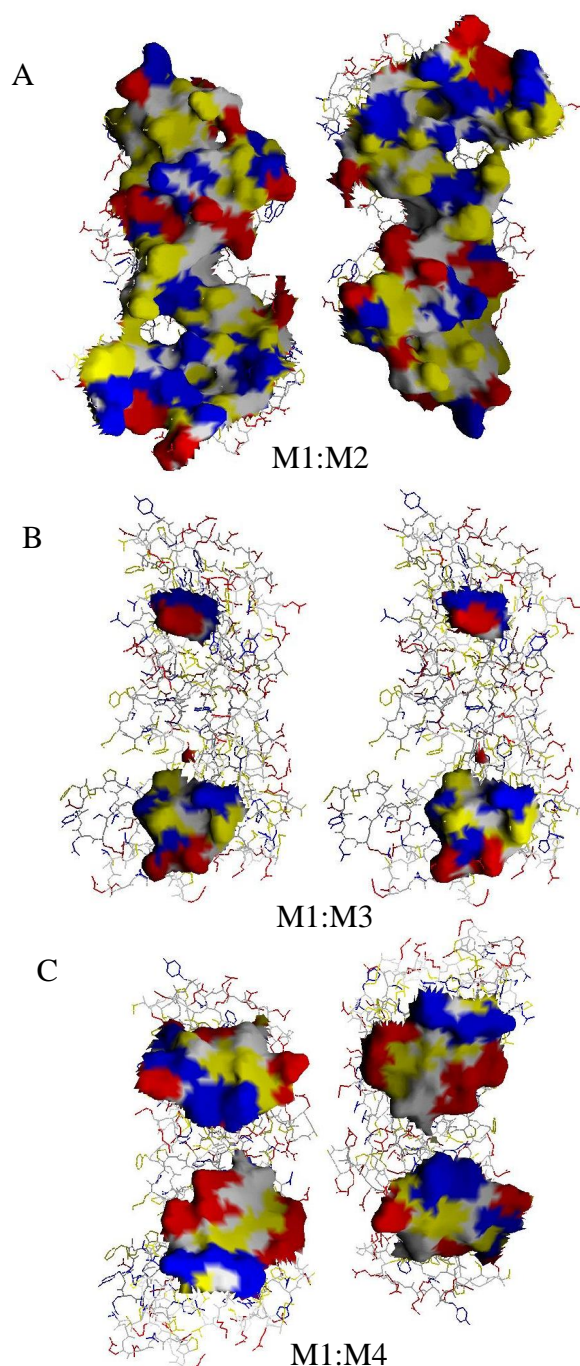


Figure 3.11 Graphical representation of MkFTR subunit interfaces, (a) M1:M2 interface, (b) M1:M3 interface and (c) M1:M4 interface. Surface vertices within 5Å were selected and displayed in GRASP (Nicholls, A., *et al* 1993). Color coding reflects the underlying nearest residue, white mainchain, red non-polar sidechain, blue polar sidechain, yellow hydrophobic sidechain.

Assembly of the monomers generates a contact that spans from the insertion region via the cores of both lobes to the β -meander. The monomers are oriented such that three loop segments of the β -meander and the loop following helix B of subunit 2 pack against the insertion region of subunit 1 and conversely on the other side of the interface. Hydrophobic interactions dominate the the interface contact but specific polar interactions play a role too. At the protein surface an ion-pair cluster is formed between residues Lys210¹, Lys214¹, Asp225¹ and Asp129² extending out from the loop connecting α -helix B and β -strand 6. Asp129² also anchors the insertion region with three hydrogen bonds to the side-chain of Ser213¹ and the peptide amino groups of Lys214¹ and Gln215¹. A hydrophobic patch consisting of Trp32¹, Ile35¹, Pro101¹, Leu180², Ile204² and Tyr244² ties together a distinct regions of the proximal and distal lobes by van der Waals interactions and connects the hydrophobic cores of the two subunits.

At the center of the interface helices A and B pack against the loops following helices A and D, and vice versa. This contact is characterized by a hydrophobic cluster consisting of Phe43¹, Pro51¹, Arg92¹, Phe43², Pro51² and Arg92². The guanidinium group of Arg96¹⁽²⁾ and the phenyl group of Phe43²⁽¹⁾ form a *II-II*-stacking interaction.

The interface between subunits 1 and 3 (2 and 4) is very small (275 Å² per monomer) and represents only 2% of the total surface area of the monomer. There is only one contact point formed between α -helix D and, the following loop and the equivalent segments of the symmetry related partner. At the center of the interface is a small hydrophobic contact between Ala178¹ and Pro177¹ and their symmetry related partners. Polar interactions occur at the edges of the interface.

The area buried between subunits 1 and 4 (2 and 3) is about ~750 Å². This is 6% of the monomer surface area. The interface is relatively polar in comparison to the average chemical composition of interfaces. Helices D and E of the proximal lobe pack against the loop between β -strands 4 and 5 of the distal lobe, and vice versa, forming two contacts. There are eight inter-subunit hydrogen bonds. The interaction between Asp188¹ and the peptide nitrogen of His30⁴ is particularly interesting, because the negatively charged aspartate stabilizes the N-terminal side of α -helix A. The polar nature of the interaction is further reinforced by several firmly bound water molecules.

3.1.4. Surface Topography and Putative Active Site

Often the most prominent surface features of a structure such as clefts, channels or cavities are features associated with the active site of the enzyme. Analysis of the MkFTR tetramer surface with SURFNET (Laskowski, R.A., 1991) indicated a four large tri-furcated clefts symmetrically distributed over the surface (Figure 3.12)

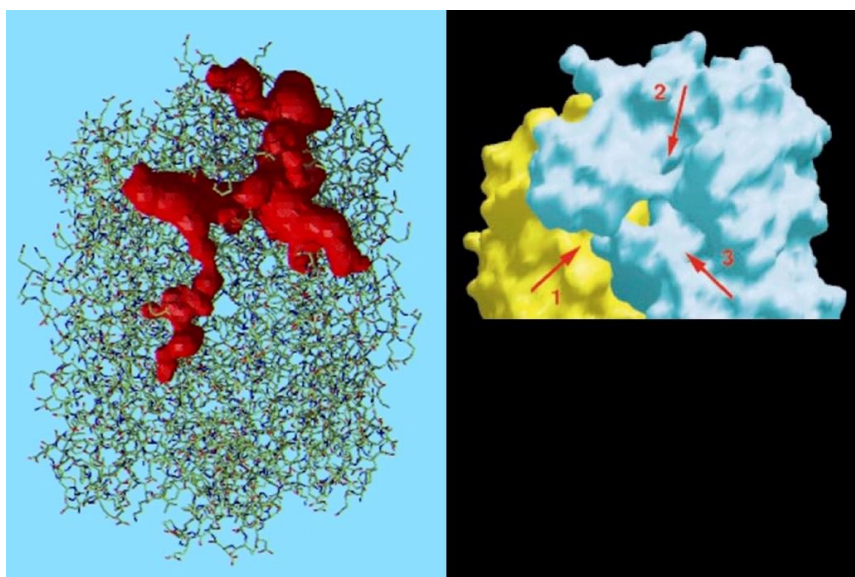


Figure 3.12 One surface cleft of the MkFTR tetramer as determined with SURFNET (left). Only one cleft is shown (red) for clarity. Right GRASP surface rendering with arrows indicating entrances to the cleft

Furthermore this cleft was associated with unidentified electron density. This density is similar to that seen in the deposited structure data for the *M. barkeri* FTR. The *A. fulgidus* FTR data deposition failed to process and give interpretable electron density maps, perhaps due to errors in the mmCIF reflection file. It was originally believed to be the fH_4MPT analog folate which may have been bound and retained after heterologous expression in *E. coli* but analysis of the true ligand positions from deposited MkFTR structure containing bound substrate. (Acharya P *et al.* 2006) (Figure 3.13) shows that the density corresponds very closely with the formylmethanofuran position. Perhaps some furan-like metabolite from *E. coli* can mimic FMF and occupy this position. The nature of the bound small molecule ligand could have been identified by mass spectrometry of crystalline samples.

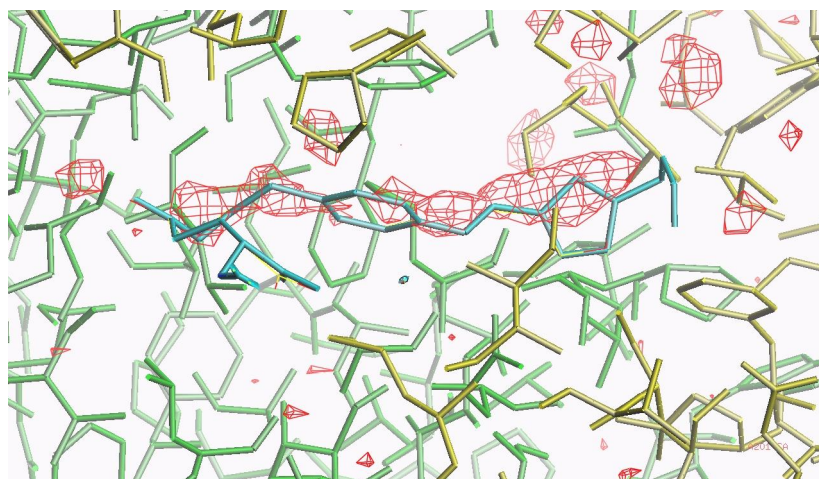


Figure 3.13 View of the MkFTR cleft feature and associated unidentified electron density (red mesh). One MkFTR subunit is in yellow and the other in green. The original assumption that the density might be the H4MPT analog folate are wrong as the density clearly co-locates to the FMF (cyan) site

Attempts to co-crystallize MkFTR with the substrate FMF under conditions of high salt and at temperatures above 4°C yielded crystals which may have been of P222 spacegroup symmetry, however high mosaicity and/or twinning undermined attempts to process the data. The failure was ascribed at the time to this author, however it appears that successful co-crystallization of required both substrates and suitable anaerobic conditions and darkness. (Acharya P, *et al.* 2006). The possibility that the large surface cleft of MkFTR was actually the active site could have been tested by generating site-directed mutants of conserved residues in this region however this was not undertaken.

The association of the cleft feature with the subunit interface offers an explanation for the dependency of enzymatic activity on the oligomeric state of MkFTR. Only in the tetramer is the cleft fully formed and thus activity possible.

3.1.5. Thermal Stability Factors

The crystal structure analysis of MkFTR identifies a number of structural features which may play a role in contributing to the extreme thermostability of the enzyme

N-Terminus / C-Terminus Proximity (N- and C-terminus in the same lobe): In the MkFTR monomer the amino- and carboxy-termini are closely associated in a

3. Results and Discussion

hydrophobic cluster along with Ile3 and Ile 8. This may contribute additionally to the thermostability of the structure.

Helix Dipole Stabilization: The MkFTR tetramer consists of 21 hydrogen bond pairs between monomers 1 and 2. The alignment of the individual peptide unit dipoles leads to a net positive charge on the N-terminus and net negative charge on the C-terminus end of an alpha-helix. The MkFTR monomer displays potential stabilization of helix dipole moments through the incorporation of acidic negatively charged residues near N-termini and basic positively charged residues near C-termini. Glu253 and Glu254 form an acidic group in the vicinity of helix B.

II-II Interactions: Aromatic amino acid side-chains often interact in such a way as to form stacked configurations. The aromatic side-chain of Phe43¹⁽²⁾ and the guanidinium group of Arg92²⁽¹⁾

Hydrogen-bonding: The MkFTR tetramer consists of 21 hydrogen bond pairs between monomers 1 and 2. Between monomer 1 and 3 there are two intersubunit hydrogen bonds and between monomer 1 and 4 there are four hydrogen bonds. Figure 3.14 shows the 2-fold related hydrogen bonds along the z-axis between monomers 1,3 and 2,4 respectively.

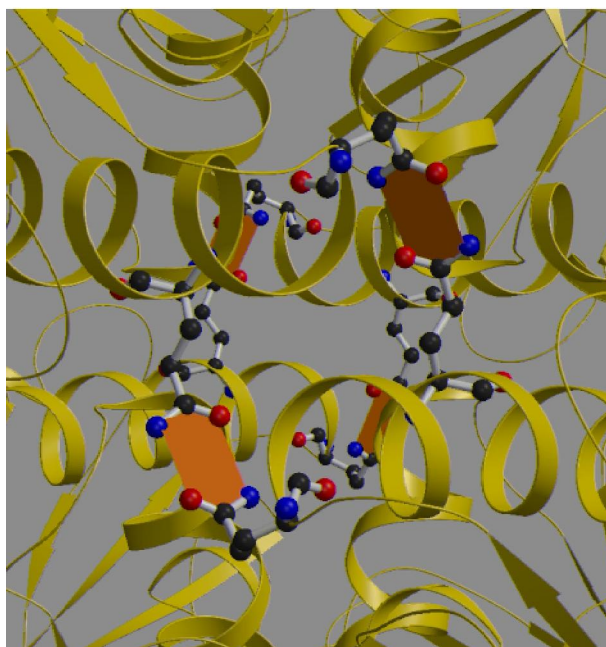


Figure 3.14 Two-fold related hydrogen bonds along the z-axis between MkFTR monomers 1,3 and 2,4 respectively.

3. Results and Discussion

N Mainchain								
Residue	Total	# Buried	# H-Bonds	0	1	2	3	Mean # H-bonds
Ala	35	23	3	20	-	-		
Arg	6	4	-	4	-	-		
Asn	8	3	2	1	-	-		
Asp	18	13	3	10	-	-		
Cys	6	6	-	5	1	-		
Gln	9	6	-	6	-	-		
Glu	34	13	1	12	-	-		
Gly	32	19	2	17	-	-		
His	4	0	-	-	-	-		
Ile	20	19	1	18	-	-		
Leu	13	12	1	11	-	-		
Lys	15	11	-	10	1	-		
Met	9	7	-	7	-	-		
Phe	13	10	4	6	-	-		
Pro	15	10	1-	-	-	-		
Ser	11	8	4	3	1	-		
Thr	12	10	4	6	-	-		
Trp	2	2	-	2	-	-		
Tyr	9	5	-	5	-	-		
Val	25	20	1	19	-	-		
Total	296	201	36 (17.9%)	162	3	-		0.84

O Mainchain								
Residue	Total	# Buried	# H-Bonds	0	1	2	3	Mean # H-bonds
Ala	35	19	2	15	2	-		
Arg	6	5	-	5	-	-		
Asn	8	3	1	2	-	-		
Asp	18	4	-	2	2	-		
Cys	6	2	-	2	-	-		
Gln	9	2	-	2	-	-		
Glu	34	12	-	11	-	1		
Gly	32	14	1	12	1	-		
His	4	2	-	2	-	-		
Ile	20	13	-	10	3	-		
Leu	13	5	-	5	-	-		
Lys	15	4	-	4	-	-		
Met	9	5	-	4	1	-		
Phe	13	7	-	7	-	-		
Pro	15	3	-	2	1	-		
Ser	11	7	-	7	-	-		
Thr	12	4	-	3	1	-		
Trp	2	1	-	1	-	-		
Tyr	9	5	-	4	-	1		
Val	25	10	-	9	1	-		
Total	296	127	4 (3.1%)	109	12	2		1.10

Table 3.5 Mean number of buried intersubunit hydrogen bonds for the MkFTR monomer, NACCESS = Buried and HBPLUS = H-bonds. The cutoff distance for hydrogen bonds was 3.3 Å.

3. Results and Discussion

Table 3.5 gives the mean number of buried H-bonds for the MkFTR monomer. Values calculated for MkFTR are actually lower than those of a representative set of structures (McDonald, I.K. and Thornton, J.M., 1994)

A suggested factor leading to increased thermostability is an increase in the number of buried hydrogen bonds relative to a reference set of structures. The MkFTR monomer in fact exhibits a decreased number of buried hydrogen bonds relative to the reference dataset. In the case of MkFTR an increase in the number of buried hydrogen bonds cannot be invoked as a means of achieving greater thermostability.

Salt bridges: The MkFTR structure contains 35 ion-pairs per monomer for an average of 0.12 ion-pair per residue. This proportion of ion-pairs per residue is higher than the average of 0.04 ion-pair per residue observed in a set of 38 high resolution protein structures. The structural analysis of other hyperthermophilic protein structures indicate that the formation of ion-pair clusters may be an important factor in the thermostability of the structure. In this respect it seems that MkFTR has adopted a similar route to thermal stabilization by increasing the number of salt bridges.

Monomer 1	Monomer 2	Distance (Å)
Lys31 N ζ	Glu184 O γ 2	2.7
Glu39 O ϵ 1	Phe43 N	3.3
Glu39 O ϵ 1	Gly44 N	3.0
Glu39 O	Phe43 N	3.0
Thr45 O γ 1	Gln95 O	2.9
Met49 N	Gln95 N ϵ 2	3.1
Gln95 N ϵ 2	Ser46 O γ	3.2
Gln95 N ϵ 2	Ser46 O	3.1
Thr99 O	Ala206 N	2.9
Phe127 O	Thr223 N	2.9
Asp129 O δ 1	Ser213 O γ	2.5
Asp129 O δ 1	Lys214 N	3.1
Asp129 O δ 1	Gln215 N	3.0
Asp129 O δ 2	Lys210 N ζ	2.8
Asp129 O δ 2	Lys214 N	3.0
Gln132 N ϵ 2	Tyr216 OH	2.9
Pro146 O	Thr223 O γ 1	2.7
Val148 O	Val243 N	2.9
Glu149 N	Ala206 O	3.2
Glu149 O ϵ	Tyr244 OH	2.8
Gly150 O	Thr230 O γ 1	3.2
Monomer 1	Monomer 3	Distance (Å)
Gln176 O ϵ 1	Gln181 N ϵ 2	2.9
Gln176 N ϵ 2	Gln181 O ϵ 1	3.0
Monomer 1	Monomer 4	Distance (Å)
His30 N	Asp188 O δ 2	2.9
Tyr60 OH	Asp188 O	2.5
Glu64 O ϵ 2	Arg261 NH ₁	2.9
Glu64 O ϵ 2	Arg261 NH ₂	3.0

Table 3.6 Intramolecular ionic-bonds within the crystal form P MkFTR tetramer

Surface area minimization and globularity: Richards pioneered the application of the concepts of molecular surface area and volume to the analysis of macromolecular structures (Richards, F.M., 1977). From this work, a general picture of water-soluble globular proteins has emerged emphasizing (i) the relatively smooth, polar surface that minimizes the surface energy in an aqueous environment, and (ii) the efficiently packed, relatively nonpolar interior. In general, minimization of the ratio of surface area to volume increases the stability of an object by simultaneously reducing the unfavorable surface energy and increasing the attractive interior packing interactions.

The possibility that minimization of the surface area to volume ratio might contribute to the stability of MkFR was assessed (Chan, M.K., et al. 1995). The relative surface area was obtained by first calculating the accessible surface area, A_0 with NACCESS, and then dividing by the expected surface area $A_c = 15.0 N^{0.866}$ for a protein of the size of MkFTR. An estimate of the extent of surface area minimization is given by the ratio A_0/A_c . Variations in the ratio from the average value 1 reflect

3. Results and Discussion

changes in the molecular shape or surface roughness that alter the accessible surface area of the protein. The top half of table 3.7 indicates that the MkFTR tetramer seems to minimize the surface area ratio.

A similar analysis based on the globularity index or the ratio of the accessible surface area, A to the expected surface area for a protein of a given molecular weight A/A_G where $A_G = 11.1 Mw^{2/3}$ is given in the lower half of table 3.7. This does not seem to indicate a significant trend of the MkFTR structure to maximize globularity or minimize surface area.

Monomer(s)	A_C	A_O	A_O/A_C
M_1	11863.21	13694.4	1.154
M_2	"	13573.8	1.144
M_3	"	13644.0	1.150
M_4	"	13753.0	1.159
M_1M_2	21617.71	21405.8	0.990
M_1M_3	"	26713.6	1.236
M_1M_4	"	25858.8	1.196
$M_1M_2M_3M_4$	39408.17	38700.6	0.982

Monomer(s)	A_G	A_O	A_O/A_G
M_1	11105.27	13694.4	1.233
M_2	"	13573.8	1.222
M_3	"	13644.0	1.229
M_4	"	13753.0	1.238
M_1M_2	17628.51	21405.8	1.214
M_1M_3	"	26713.6	1.515
M_1M_4	"	25858.8	1.467
$M_1M_2M_3M_4$	27983.52	38700.6	1.383

Table 3.7 Surface area ratios for the MkFTR crystal form P monomer and oligomers.

3.1.6 High Salt Solubility Factors

The surface of the MkFTR tetramer is characterized by extensive patches of acidic residues, Glu and Asp with some basic Lys and Arg residues interspersed for charge compensation. This highly acidic nature leads to a strongly negative

electrostatic character at low ionic strength (Figure 3.15) . Figure 3.16 highlights a particularly acidic cluster in the MkFTR structure.

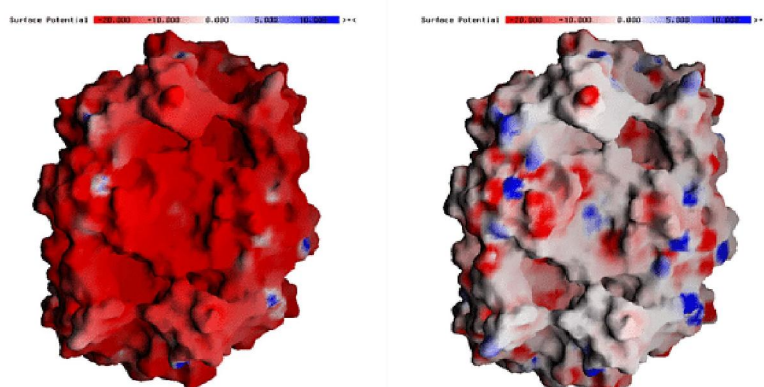


Figure 3.15 GRASP surface representation of MkFTR with electrostatic potential mapped to the surface. At low ionic strength (left) the highly acidic charged nature of the surface leads to strong negative charging. At high ionic strength characteristic of physiological conditions the surface charge takes on a more neutral character (right).

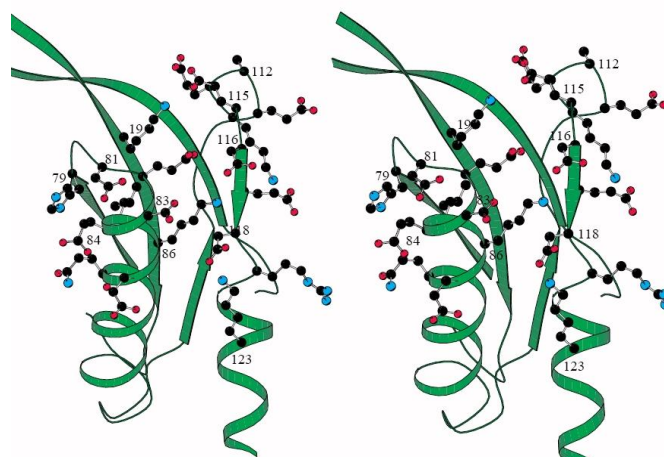


Figure 3.16 Stereoview showing a surface region of MkFTR where basic and, in particular, acidic residues are clustered in the segments 81–88 and 111–119. Lys86 is surrounded by four acidic residues, Glu82, Asp83, Asp116 and Asp118. Atoms are shown in standard colours and labelled.

Dependance of activity on salt concentration and oligomerization: The highly charged nature of the MkFTR structure offers an explanation of the dependency of enzymatic activity on the salt induced oligomerization of MkFTR. Only at high ionic strength is the surface charge compensated to the point where oligomerization can occur leading to formation of the active sites at the subunit interfaces.

3.2. MkFTR Structure from Salt Crystal Form S

3.2.1. Phasing, Refinement and Model Quality

The SALT crystal form S of MkFTR was obtained from 1.2M (NH₄)₂SO₄, 1% PEG 8000 and 0.1M Na.Citrate pH 5.6 at 4°C. (Table 3.1) Phases were determined by molecular replacement, AMORE (Navaza, J. 1994) and the structure refined with X-PLOR (Brünger, A.T., 1992b). Table 3.7 gives the data and refinement statistics for the S form MkFTR model.

Parameter	
Unit-Cell (Å,°)	157.5, 157.5, 242.1, 90, 90, 90
Spacegroup	I4 ₁ 22
Resolution (Å)	10.0-1.9
Number of reflections	
Total	408,443
Unique	137,108
R _{sym} (%)	5.4
Completeness (%)	88.7
R _{free}	0.205
R _{work}	0.180
RMSD bond length (Å)	0.009
RMSD bond angle (°)	0.965
Mean coordinate error from Luzzati plot (Å)	0.30-0.35 (Free reflection set) 0.25-0.30 (Work reflection set)
Diffraction data precision indicator, σ _w (x) (Å)	0.14 (Free reflection set) 0.10 (Work reflection set)
Number of residues	296*2
Number of atoms	4856
Number of waters	1186
Real-space fit	0.89
Average B-factor	0.18 Å ²
Ramachandran plot, number of non-Glycine residues in,	
Disallowed regions	0 (0.0 %)
Generous regions	2 (0.3 %)
Allowed regions	36 (7.7 %)
Favoured regions	554 (92.0%)

Table 3.7 Data and refinement statistics for the S form MkFTR model

Although the Vm analysis for the S crystal form implied a tetramer in the asymmetric unit, only a dimer solution was made available for refinement and analysis. The author did inquire as to why only a dimer was given for the solution and was told that symmetry generates the tetramer. Crystallographic symmetry indeed generates the expected biologically relevant tetramer and due to inexperience the author accepted this at face value. Refinement of the dimer model converged to good residuals of 18.0% and 20.5 %, R_{work} and R_{free} respectively.

The dimer solution unfortunately is missing 50% of the model and indeed it is hard to imagine, although not entirely impossible to obtain diffraction beyond 2 Å from crystals with ~75% solvent content. At a later date the MkFTR crystal form S data were revisited and subjected to molecular replacement with EPMR. EPMR returned a second dimer position that did not overlap with the dimer originally provided to the author and application of crystallographic symmetry yielded a full non-overlapping lattice. Unfortunately it proved impossible to jointly refine both dimers against the reflection dataset given to the author. Each dimer model alone refined to acceptable R values of below 25%, however jointly the refinement does not descend below 45%. Perhaps the data from crystal form S are misindexed and the true space group was orthorhombic masquerading as tetragonal.

Certainly the author will be criticized for this situation, for instance by assuming to have simply lost half the tetramer, the author can only point out that the question regarding only a dimer was raised and was assured that everything was in order. Had the author been informed that there was uncertainty in the solution there would have been a motivation to examine all the data again to resolve this.

The remainder of the analysis is based on the dimer structure. It can be assumed that the actual tetramer structure would not change the results significantly.

3.2.2. Comparison to PEG Structure

Analysis of the MkFTR structure obtained from conditions of high ionic strength might be expected to shed light on the structural features responsible for solubility of MkFTR at high physiological ionic strength.

Monomer comparison: Comparison of RMS deviations between monomers from P and S form MkFTR reveals deviations mainly in coil regions outside of

3. Results and Discussion

secondary structure elements. Additional deviations are due to NCS differences, crystal contacts and at N- and C-termini (Figure 3.17 a-d).

Tetramer comparison: Comparison of the P and S form tetramers shows a slight contraction along the local NCS z-axis of tetramer of less than 5%. The buried interfaces show a small ~5% decrease in the size of the A:C/B:C interface buried decreasing from 792\AA^2 to 761\AA^2 . This would be contrary to the expectation that at higher ionic strength and charge compensation that the interface would increase and the tetramer would become more compact.

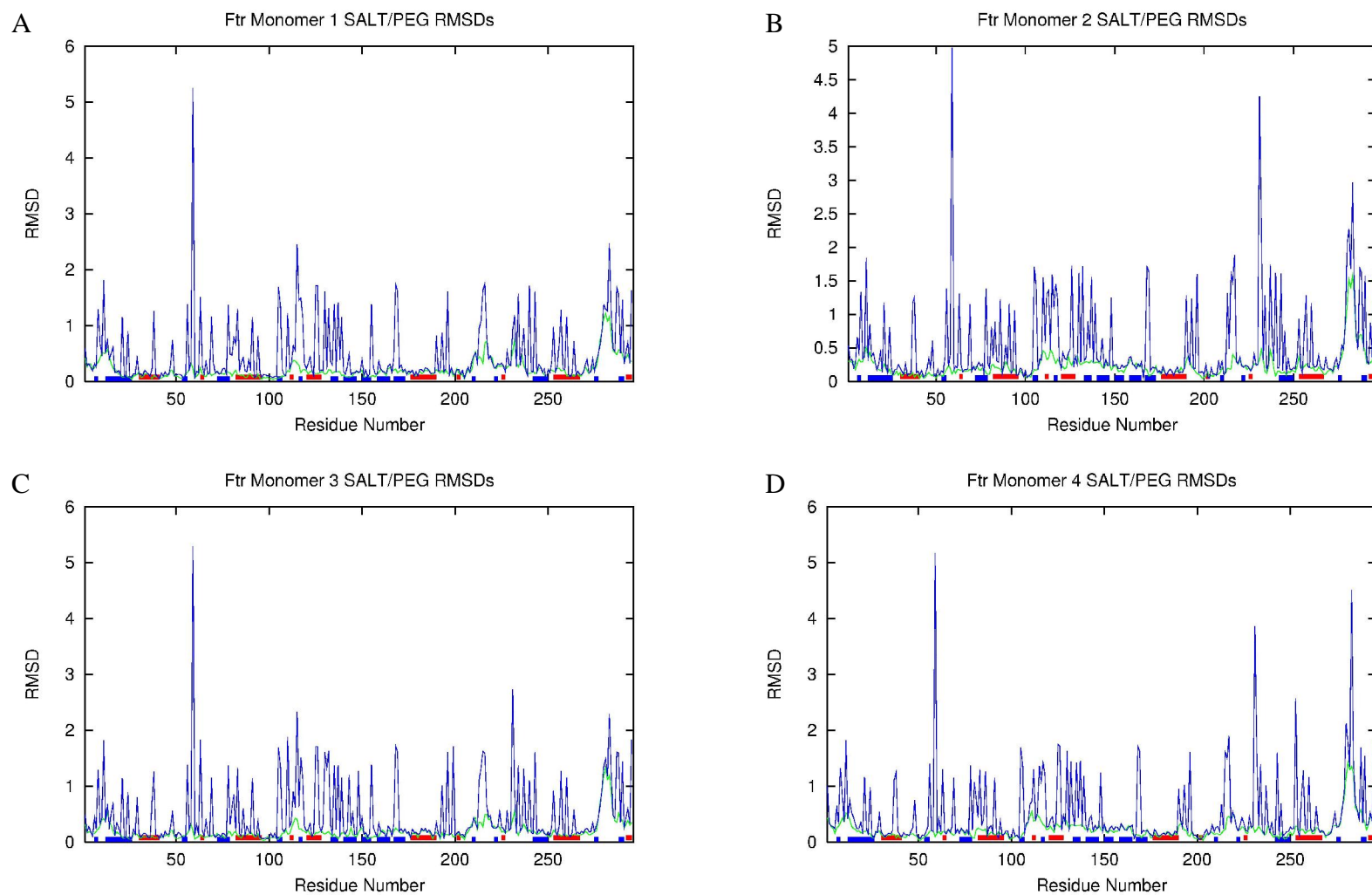


Figure 3.17 RMS deviation between S and P crystal forms of MkFTR monomers, mainchain (green) and sidechain (blue). Regions of secondary structure are shown as bars on abscissa where alpha-helix (blue) and beta-strand (red). Crystal form S to form P MkFTR monomer 1, 2, 3 and 4 comparisons are shown in panels A, B, C and D respectively

3. Results and Discussion

3.2.3 Solubility at high Salt Concentration B

A potential structural feature related to increase solubility of MkFTR at high ionic strength might lie in the fact that Glutamate residues are able to bind more H₂O (Kuntz, I.D. 1971). Binding more solvent would effectively increase the hydrophilic character of the surface relative to the hydrophobic area. More solvent bound would allow the enzyme to remain soluble at high ionic strength.

Analysis of the S form of MkFTR with respect to the P form shows indeed an increase in the amount of water bound by Glu residues. (Table 3.8)

Glu	N	O	OE1	OE2	MC	SC	Total		
PEG		44	72	78	113	116	191	307	
Monomer		11.00	18.00	19.50	28.25	29.00	47.75	76.75	
SALT		31	61	58	81	92	139	231	
Monomer		15.50	30.50	29.00	40.50	46.00	69.50	115.50	
Change		1.41	1.69	1.49	1.43	1.59	1.46	1.50	
%Change		40.91	69.44	48.72	43.36	58.62	45.55	50.49	
Asp	N	O	OD1	OD2	MC	SC	Total		
PEG		14	42	44	42	56	86	142	
Monomer		3.50	10.50	11.00	10.50	14.00	21.50	35.50	
SALT		7	30	36	32	37	68	105	
Monomer		3.50	15.00	18.00	16.00	18.50	34.00	52.50	
Change		1.00	1.43	1.64	1.52	1.32	1.58	1.48	
%Change		0.00	42.86	63.64	52.38	32.14	58.14	47.89	
Lys	N	O	NZ						
PEG		13	42	57		55	57	112	
Monomer		3.25	10.5	14.25		13.75	14.25	28	
SALT		3	29	47		32	47	79	
Monomer		1.5	14.5	23.5		16	23.5	39.5	
Change									
%Change		-53.85	38.10	64.91		16.36	64.91	41.07	
Arg	N	O	NE	NH1	NH2				
PEG		8	4	4	14	15	12	33	45
Monomer		2	1	1	3.5	3.75	3	8.25	11.25
SALT		4	2	1	8	6	6	15	21
Monomer		2	1	0.5	4	3	3	7.5	10.5
Change									
%Change		0.00	0.00	-50.00	14.29	-20.00	0.00	-9.09	-6.67

Table 3.8 Overall changes in amount of H₂O bound by acidic and basic residues between P and S form MkFTR.

Analysis of bound H₂O changes associated with individual residues reveals some Glu residues which bind significantly more H₂O in the S form (Table 3.9). Along with an increase in the amount of bound water there is a decrease in the thermal parameter, B-factor of the residues which might be expected if more water is

bound and the residues become more ordered. This in turn might contribute further to the thermal stability of the enzyme due to increased rigidity.

FTR M1	Glu #	Bf SC			nge			Bf SC			nge			Bf SC			nge			Bf SC			nge		
		PEG	SALT	MC	SC	Tot	PEG	SALT	MC	SC	Tot	PEG	SALT	MC	SC	Tot	PEG	SALT	MC	SC	Tot	PEG	SALT	MC	SC
1	2	31.26	40.22	1	-1	0	48.11	41.32	-1	3	2	42.39	40.22	-1	0	-1	49.22	41.32	1	3	4				
2	7	32.25	38.76	0	0	0	43.41	40.41	1	2	3	41.01	38.76	-1	2	1	48.07	40.41	1	2	3				
3	9	25.47	31.23	1	-2	-1	37.71	31.65	1	-1	0	30.34	31.23	1	-2	-1	43.22	31.65	1	0	1				
4	14	19.02	19.00	0	-3	-3	32.64	18.80	2	2	4	20.75	19.00	1	2	3	34.28	18.80	2	2	4				
5	17	34.37	24.94	-6	1	-5	41.26	29.06	-4	4	0	29.63	24.94	-1	4	3	41.95	29.06	-2	4	2				
6	39	18.23	10.55	0	2	2	16.96	10.50	0	4	4	17.73	10.55	0	3	3	16.47	10.50	0	4	4				
7	53	17.04	13.11	2	-5	-3	23.92	14.41	0	-1	-1	16.16	13.11	2	-5	-3	21.63	14.41	0	-1	-1				
8	64	38.11	36.63	3	-1	2	34.77	35.07	1	2	3	38.57	36.63	3	0	3	35.31	35.07	2	2	4				
9	65	35.70	32.58	-1	-1	-2	35.39	34.23	-1	-1	-2	39.54	32.58	0	0	0	35.99	34.23	-1	-1	-2				
10	82	37.94	24.74	2	1	3	47.35	29.26	1	5	6	41.63	24.74	1	0	1	48.04	29.26	2	5	7				
11	84	40.66	37.60	-3	1	-2	44.57	37.84	0	-1	-1	40.63	37.60	0	2	2	44.84	37.84	1	0	1				
12	87	43.06	32.50	0	1	1	44.78	33.33	1	1	2	42.81	32.50	0	-1	-1	43.12	33.33	1	0	1				
13	111	53.62	46.45	2	-1	1	54.52	51.59	1	2	3	53.38	46.45	2	2	4	56.59	51.59	1	2	3				
14	113	28.56	29.28	0	1	1	40.90	32.47	-1	2	1	34.14	29.28	0	-1	-1	36.93	32.47	-1	1	0				
15	115	23.13	15.43	-1	4	3	29.82	28.37	0	-1	-1	25.45	15.43	1	2	3	32.98	28.37	0	-1	-1				
16	117	46.22	38.81	0	1	1	48.17	41.47	1	1	2	44.81	38.81	0	0	0	44.59	41.47	0	1	1				
17	133	48.51	44.32	0	0	0	49.05	40.28	1	3	4	44.73	44.32	0	-1	-1	46.39	40.28	1	2	3				
18	134	41.87	43.87	1	1	2	44.30	39.92	2	3	5	39.75	43.87	0	0	0	45.20	39.92	1	2	3				
19	136	50.91	46.35	1	1	2	50.17	44.75	1	2	3	48.56	46.35	1	3	4	47.10	44.75	1	3	4				
20	149	25.36	15.86	3	-1	2	19.37	14.55	3	-1	2	25.83	15.86	3	0	3	20.50	14.55	2	0	2				
21	151	36.79	32.64	-3	1	-2	31.88	28.83	-2	2	0	37.26	32.64	-3	3	0	35.15	28.83	-3	1	-2				
22	155	17.96	14.99	2	3	5	20.75	11.21	2	0	2	20.17	14.99	2	0	2	20.51	11.21	2	0	2				
23	174	23.01	20.65	-2	-1	-3	30.27	24.84	-2	4	2	24.57	20.65	-1	-2	-3	28.39	24.84	-1	2	1				
24	184	14.33	9.78	2	-1	1	15.31	9.66	2	-5	-3	14.04	9.78	2	-1	1	14.56	9.66	2	-4	-2				
25	194	37.31	35.56	-2	0	-2	41.56	38.89	1	-1	0	42.00	35.56	0	0	0	43.04	38.89	0	0	0				
26	232	61.23	59.37	-1	1	0	62.12	62.04	3	1	4	63.23	59.37	1	1	2	68.88	62.04	3	1	4				
27	235	53.80	49.14	2	0	2	53.26	51.91	2	-1	1	55.39	49.14	2	0	2	54.34	51.91	1	0	1				
28	238	55.03	51.60	0	2	2	62.10	58.64	-1	1	0	57.48	51.60	1	2	3	57.30	58.64	0	1	1				
29	245	18.15	11.86	0	1	1	22.92	16.06	0	2	2	18.66	11.86	0	3	3	21.36	16.06	0	0	0				
30	253	24.92	26.90	4	3	7	39.45	30.44	1	2	3	31.00	26.90	3	1	4	36.76	30.44	2	1	3				
31	254	40.33	37.37	1	-1	0	47.77	38.70	-1	-1	-2	47.33	37.37	-1	2	1	46.70	38.70	1	0	1				
32	258	27.91	24.02	0	1	1	32.77	24.63	0	1	1	30.65	24.02	0	1	1	33.95	24.63	0	0	0				
33	265	37.22	29.43	4	-3	1	38.52	29.06	3	-2	1	38.64	29.43	4	-3	1	38.48	29.06	3	-1	2				
34	289	32.93	38.28	1	1	2	51.35	41.55	-1	0	-1	31.76	38.28	-3	0	-3	44.67	41.55	-3	0	-3				

Table 3.9 Changes in B-factor and bound H2O for Glu residues in MkFTR S versus P

Figure 3.18 shows some of the more significantly changing Glu residues with the associated change in bound water.

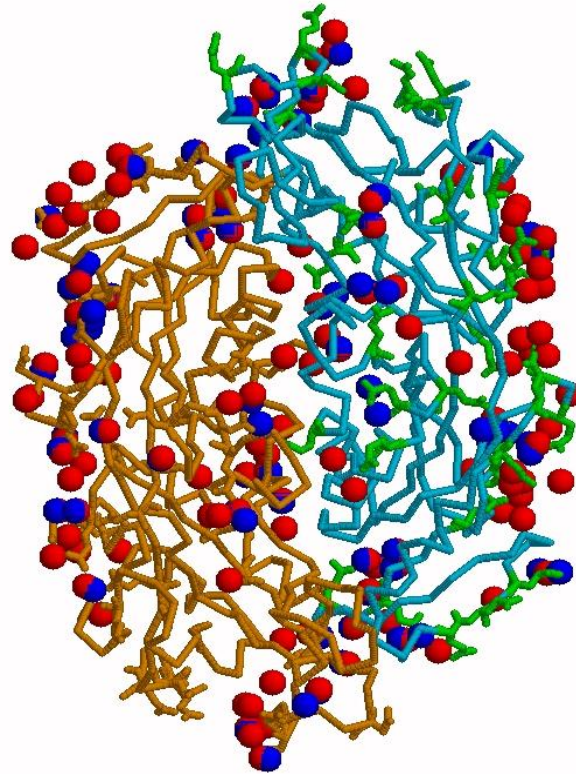


Figure 3.18 Glutamate residues showing the largest change in bound water with the associated change in bound waters. Waters from P form MkFTR are blue and from S form are red.

Further analysis of the B-factor trends of S form MkFTR indicate however that these results are inconclusive at best. In fact all residues in S form MkFTR exhibit nearly equal decreases in B-factor relative to the P form. (Figure 3.19)

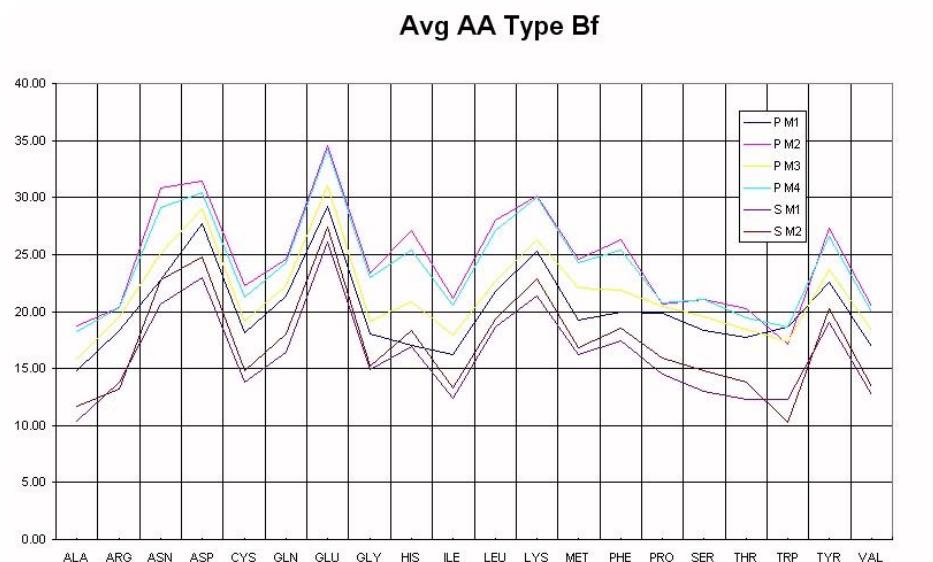


Figure 3.19 Residue average B-factors for all amino acid types in S and P form MkFTR. The lower two curves from S form residues clearly show that all residues exhibit a similar decrease thermal parameter.

In order to generalize these observations somewhat thaumatin was chosen for comparison. Thaumatin structures are available at roughly the same resolution as MkFTR and at different ionic strengths. The thaumatin results are furthermore inconclusive. Going from PEG (1THV) to 0.5M NaCl (1PP3) shows no clear B-factor decrease. However going to 0.5M Na.Tartrate (1LY0) does lead to systematic B-factor reduction. Perhaps the ionic strength regime is too low or the extent of acidic surface residues is too low to show similar trends

3. Results and Discussion

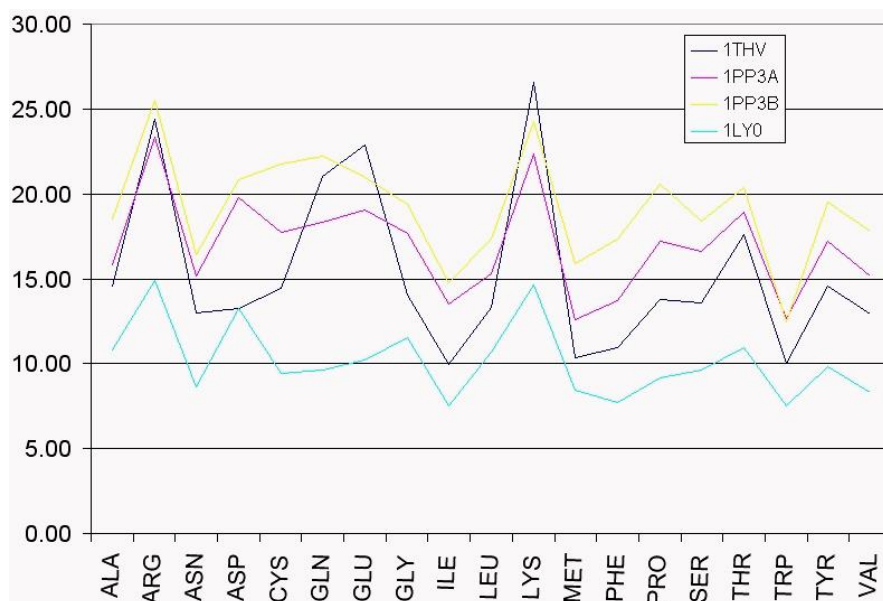


Figure 3.20 Residue average B-factors for thaumatin structures at various ionic strengths.

Other comparisons are probably available using lysozymes or myoglobins structures at varying ionic strength, but an extensive survey of PDB entries for amino acid:H₂O binding and B-factor versus ionic strength trends is beyond the scope of this work. How much the decrease in B-factor between low and high ionic strength depends specifically on increases in H₂O bound by Glu is difficult to determine. The general increase in the ordering of water at high ionic strength may lead in general to an increased ordering of solutes, i.e. proteins will be forced to lower B-factors. The additional problem of whether H₂O or counter ions such as Na⁺ is a further complication that is difficult to determine. Perhaps a detailed calorimetric analysis with series of counter ions, e.g. Cs⁺, Rb⁺ and co-crystallization with such salts might indicate if specific ions are bound rather than H₂O.

4. Materials and Methods

4.1. Cloning, expression and purification

The *fir* gene for *M. kandleri* FTR was cloned and expressed in *Escherichia coli* BL21 (DE3) under control of the T7 promoter. (Shima *et al.*, 1995) Purification was performed by column chromatography and concentration of the enzyme was performed at 4°C under aerobic conditions. Protein concentration was determined by the method of Bradford using Bio-Rad reagents and ovalbumin as a standard. Enzyme after purification was diluted with 50 mM MOPS pH 7 containing 2 mM dithiothreitol and concentrated to 12-16 mg/ml by ultrafiltration.

4.2 Crystallization and data collection

Initial crystallization experiments were performed in hanging drop vapor diffusion format using sparse matrix crystallization screens (Jancarik and Kim, 1991) Screening kits were from Hampton Research. Three sets of crystallization conditions were found and improved by systematic search yielding three crystal forms of MkFTR (M, P and S).

The best crystals of form M were obtained from 2 μ l of enzyme solution and 2 μ l of reservoir solution containing 20% 2-methyl-2,4-pentanediol, 0.7M NaCl, and 0.1M sodium acetate, pH 4.6. Drop equilibration was against 1 ml of reservoir solution at 4°C. Plate-shaped crystals appeared within 2 weeks and reached a maximum size of 0.1 x 0.3 x 0.6 mm³ and diffracted beyond a resolution of 2Å. The crystals displayed space group P₂ with unit cell dimensions a = 87.0Å, b = 75.4Å, c = 104.7Å, and β = 113.9° as determined by autoindexing oscillation photographs with the program REFIX (Kabsch, 1988). The presence of four molecules per asymmetric unit gave a crystal volume per protein mass V_m of 2.5 Å³/Da (Matthews, 1968) implying a solvent content of 50%. Crystals of form M displayed relatively high mosaicity and rapid radiation decay.

Crystals of form P were obtained at 4°C, using 4 μ l of enzyme solution and 4 μ l of reservoir solution containing 22% polyethylene glycol (PEG) 8000, 0.3M (NH₄)₂SO₄ and 0.1 M MOPS pH 7.0. Bipyramidal crystals appeared within 2 weeks and grew to a size of 0.6 x 0.6 x 0.8 mm³. These diffracted to 1.7 Å resolution and displayed the space group I4₁22 with unit cell parameters of a = b = 157.5Å and c =

242. Å. Assuming four molecules per asymmetric unit the V_m value was $3.0 \text{ \AA}^3/\text{Da}$, and the solvent content was 58%.

The best crystals of form S grew to a size of $0.4 \times 0.5 \times 0.5 \text{ mm}^3$ when the reservoir solution consisted of 1.2M $(\text{NH}_4)_2\text{SO}_4$, 1% PEG 8000, 0.1M sodium citrate, pH 5.6. The obtained rhombohedral crystals diffracted at least to a resolution of 2.2 \AA . The space group was determined to be $I4_122$ and the unit cell parameters to be $a = b = 151.3 \text{ \AA}$ and $c = 249.5 \text{ \AA}$ consistent with four, five, or six molecules in the asymmetric unit. The identical space group and the similar unit cell dimensions of crystal forms P and S suggested a related crystal packing.

Crystals of form P were used for structure determination. Native data set 1 and all derivative data sets were collected in house with $\text{CuK}\alpha$ radiation from an R-AXIS II image-plate detector equipped Rigaku rotating anode generator. Data processing and scaling were carried out with the program MOSFLM (Leslie, A.G.W., 1992) and the CCP4 package (Collabrative Computational Project, Number 4. (1994). Native data set 2 was collected with synchrotron radiation ($\lambda=1 \text{ \AA}$) from three crystals of crystal at the Max-Planck-beamline, DESY (Deutsches Elektronensynchrotron, Hamburg) using a MAR-Research image-plate detector. Intensities were evaluated with the program MOSFLM and internally scaled with programs of the CCP4 package. 408,443 reflections were measured and reduced to a final data set of 137,108 unique reflections. The resulting R_{sym} was 5.4% and the completeness 87.7% in the resolution range $1.73\text{--}10.0 \text{ \AA}$.

4.3. Phase determination

4.3.1. MkFTR crystal form P phasing

The structure of MkFTR in crystal form P was solved by the MIR method. Two heavy atom derivatives provided suitable phasing parameters. One site of the Pb derivative was localized in difference Patterson maps using SHELXS (Sheldrick, 1991). The resulting heavy-atom positions were refined and SIR phases calculated with the program MLPHARE (Otwinowsky, Z., 1991). Further heavy-atom sites were found with difference Fourier maps. Final phasing was performed with one Pb and one Hg derivative giving a figure of merit of 0.38, between $10\text{--}3.2 \text{ \AA}$ resolution. The resulting MIR electron-density map was considerably improved by the solvent flattening (solvent content 60%) and the histogram matching with DM (Cowtan, K.D.,

1994) resulting in an R_{free} of 32.3% from 10–3.2 Å resolution. Fourfold molecular averaging with phase extension to 2.8 Å was carried out with RAVE (Kleywegt, G.A. & Jones, T.A., 1994). After ‘skeletonization’ of the electron density and separation of a single subunit with the programs MAPMAN (Kleywegt, G.A. & Jones, T.A., 1996) and O (Jones, T.A. *et al*, 1991) a molecular mask of this subunit was generated with MAMA (Kleywegt, G.A. & Jones, T.A., 1994). The noncrystallographic symmetry (NCS) operators were calculated from four heavy-atom sites per subunit within O and improved with the program IMP (Kleywegt, G.A. & Jones, T.A., 1994).

4.3.2. MkFTR crystal form S phasing

The S crystal form of MkFTR was phased by molecular replacement, AMORE (Navaza, J. 1994) using the crystal form P MkFTR structure as search model. The crystal form S model was refined with X-PLOR (Brünger, A.T., 1992b). Although the V_m analysis for the S crystal form implied a tetramer in the asymmetric unit, only a dimer solution was provided to the author for refinement and analysis. Crystallographic symmetry generates the expected biologically relevant tetramer. Refinement of the dimer model converged to residuals of 18.0% and 20.5 %, R_{work} and R_{free} respectively.

4.4 Substrate co-crystallization

Extensive screening trials were performed with MkFTR enzyme and the substrate FMF under aerobic incubation at 37°C. Crystals however were of poorly formed habit and diffraction when obtained was disordered, highly mosaic and of low resolution.

4.5. Model building and refinement

Solvent flattened electron-density maps at 3.2Å resolution and fourfold averaged electron-density maps at 2.8Å resolution were suitable for building in a Ca trace with the option ‘Baton’ in O (Jones, T.A. *et al*, 1991). Ultimately 282 of 296 amino acids could be incorporated into the electron density. Major problems during model building were separating closely attached parts of different subunits and fitting the sequence to the electron density, as the N- and C-terminal ends are in van der Waals contact to each other and not easily detected. Refinement was performed

within X-PLOR (Brünger, A.T., 1992b) using the force field parameters developed by Engh and Huber (Engh, R. A. and Huber, R., 1991). During initial molecular dynamic minimization using the slow-cooling protocol (3000–300°K) and a subsequent positional and temperature factor refinement the R_{work} and the R_{free} factors dropped to 24.3% and 30.0%, respectively, in the resolution range 10–2.5 Å. Following this missing residues between 214 and 227 were visible in the 2Fo–Fc map. The second and following rounds of refinement utilized all data from 10.0–1.73Å resolution. In the first rounds, equivalent atoms of the four subunits related by fourfold NCS symmetry were treated as restrained, in later stages the NCS restraints were omitted. Water molecules were built in when the 2Fo–Fc and the Fo–Fc electron density map were higher than 1σ and 3σ , respectively. Moreover, the map peak had to be roughly spherical and lie within 3.3Å from a hydrogen-bond donor/acceptor. Water molecules with temperature factors greater than 50Å^2 were removed from the model. After several rounds of alternate automatic refinement and manual inspection of the electron density the R factor converged.

4.6. Analysis of the structure

Model errors were assessed using PROCHECK (Laskowski et al. 1993), which verifies deviations of geometric parameters (from ideal values) and dihedral angles (from allowed regions) for the mainchain and sidechain atoms. The fit of individual coordinates to the electron density was checked by calculating the real-space correlation (Jones, T.A. et al, 1991) between them. Structural superposition was either performed with GA_FIT (May and Johnson 1994) or within O (Jones, T.A. et al, 1991), where the least-squares fit algorithm of Kabsch (Kabsch, W., 1978) is implemented. Similarities between the fold of Ftr and that of other structurally characterized proteins were identified with the DALI WWW server (Holm,L. & Sander,C., 1993). Ionic interactions were designated as salt bridges or ion pairs when their distances were below 4.0 Å (Barlow,D.J. & Thornton,J.M., 1983). Accessible surface areas were calculated with NACCESS (Hubbard,S.J. & Thornton, J.M.,1993) and the classification of the area into hydrophobic, uncharged polar and charged was performed according to Miller et al. (Miller S., et al, 1987).

5. References

- Aguilar, C.F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M. and Pearl, L.H., (1997), Crystal structure of the β -glycosidase from the hyperthermophilic Archaeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. *J Mol Biol* **271**:789-802.
- Acharya, P., Warkentin, E., Ermler, U., Thauer, R.K. and Shima, S. (2006) The Structure of Formylmethanofuran: Tetrahydromethanopterin Formyltransferase in Complex with its Coenzymes *J. Mol. Biol.* **357**, 870–879
- Auerbach, G., Huber, R., Grättinger, M., and Zaiss, K. (1997). Closed structure of phosphoglycerate kinase from *thermotoga maritima* reveals the catalytic mechanism and determinants of thermal stability. *Structure* **5**, 1475-1483.
- Baker, D., Bystroff, C., Fletterick, R.J., and Agard, D.A. (1993) PRISM: Topologically constrained phase refinement for macromolecular crystallography. *Acta Cryst.* **D49**, 429-439.
- Barlow, D.J. and Thornton, J.M. (1983) Ion-pairs in proteins *J. Mol. Biol.*, **168**, 857-885
- Bernal, J.D. and Crawford, D. (1934) X-ray Photographs of Crystalline Pepsin. *Nature (London)*, **133**, 794-795.
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, **72**, 248-254
- Bragg, W.L., (1913) The structure of some crystals as indicated by their diffraction of X-rays. *Proc. Roy. Soc.* **A89**, 248-277
- Breitung, J., Boerner, G., Scholz, S., Linder, D., Stetter, K.O., and Thauer, R.K. (1992) Salt dependence, kinetic properties and catalytic mechanism of N-formylmethanofuran:tetrahydromethanopterin formyltransferase. *Eur. J. Biochem.*, **210**, 971-981
- Bricogne, G. (1974) Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Cryst.* **A30**, 395-405.
- Brünger, A.T. (1992) The Free R Value: a Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures. *Nature*, **355**, 472-474.
- Brünger, A.T. (1992b) X-PLOR Version 3.1: A system for X-ray Crystallography and NMR Yale University Press, New Haven, CT, USA.
- Bult, C.J., and Venter, J.C. (1996) Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii*. *Science*, **273**, 1058-1073.
- Burggaf, S., Stetter, K.O., Rouviere, P. and Woese, C.R. (1991) *Methanopyrus kandleri*: an archeal methanogen unrelated to all other known methanogens. *Systematic appl. Microbiol.*, **14**, 346-351.

5. References

- Carter, Jr. C.W. and Yin, Y. (1994) Sparse matrix sampling: a screening method for crystallization of proteins. *Acta. Cryst.*, **D50**, 572
- Chan, M.K., Mukland, S., Kletzin, A., Adams, M.W.W., and Rees, D.C. (1995) Structure of a Hyperthermophilic Tungstopterin Enzyme, Aldehyde Ferredoxin Oxidoreductase. *Science*, **267**, 1463-1469.
- Cheah, E., Carr, P.D., Suffolk, P.M., Vasudevan, S.G., Dixon, N.E. & Ollis, D.L. (1994) Structure of Escherichia coli signal transducing protein PII *Structure*, **2**, 981-990
- Collaborative Computational Project, Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763
- Cowtan, K.D. (1994) 'DM': an automated procedure for phase improvement by density modification. In Joint CCP4 and ESR-EACBM Newsletter on Protein Crystallography, 31, 83-91
- Creighton, T. (1993). Proteins: Structure and molecular properties. New York: W. H. Freeman and Co. S. 9-10.
- Cruikshank, D.W.J., (1996) Protein precision re-examined: Luzzati plots do not estimate final errors. In Proceedings of the CCP4 Study Weekend. Dodson, E., Moore, M., Ralph, A. & Bailey, S., eds. pp.11-22. Daresbury Laboratory, Warrington, UK.
- Danson MJ, Hough DW: (1997) The structural basis of protein halophilicity. *Comp Biochem Physiol A* **117A**:307-312.
- DeDecker, B.S., O'Brien, R., Fleming, P.J., Geiger, J.H., Jackson, S.P. & Sigler, P.B. (1996) The Crystal Structure of a Hyperthermophilic Archaeal TATA-Box Binding Protein *J. Mol. Biol.*, **264**, 1072-1084
- Delboni, L.F., Mande, S., Rentier-Delrue, F., and Mainfroid, V. (1995). Crystal structure of recombinant triosephosphate isomerase from *Bacillus stearothermophilus*. An analysis of potential thermostability factors in six isomerases with known three-dimensional structures points to the importance of hydrophobic interactions. *Protein Science* **4**, 2594-2604.
- Delboni, L.F., and Hol, W.G.H. (1995) Crystal structure of recombinant triosephosphate isomerase from *Bacillus stearothermophilus*. An analysis of potential thermostability factors in six isomerases with known three-dimensional structures points to the importance of hydrophobic interactions *Protein Sci.*, **4**, 2594-2604
- Delboni, L.F., Mande, S., Rentier-Delrue, and F., Mainfroid, V. (1995). Crystal structure of recombinant triosephosphate isomerase from *Bacillus stearothermophilus*. An analysis of potential thermostability factors in six isomerases with known three-dimensional structures points to the importance of hydrophobic interactions. *Protein Science* **4**, 2594-2604.
- DiMarco, A.A., Bobik, T.A., and Wolfe, R.S. (1990) Unusual coenzymes of methanogenesis. *Annu. Rev. Biochem.*, **59**, 355-394

5. References

- DiMarco, A.A., Sment, K.A., Konisky, J. and Wolfe, R.S. (1990) The formylmethanofuran:tetrahydromethanopterin formyltransferase from *Methanobacterium autotrophicum* DeltaH, Nucleotide sequence and functional expression of the cloned gene. *J. Biol. Chem.* **265**, 472-476.
- Doublié, S. (1997) Preparation of Selenomethionyl proteins for phase determination. *Meth. Enzymol.* **276**, 523-530
- Drenth, J. (1994). Principles of Protein X-ray Crystallography. Springer-Verlag, New York - Berlin.
- Dym, O., Mevarech, M. and Sussman, J.L. (1995). Structural Features That Stabilize Halophilic Malate Dehydrogenase from an Archaeobacterium. *Science* **267**, 1344-1346.
- Engh, R. A. and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.* **A47**, 392-400.
- Ermler, U., Merckel, M., Thauer, R.K., and Shima, S. (1997). Formylmethanofuran:tetrahydromethanopterin formyltransferase from *Methanopyrus kandleri* - new insights into salt-dependence and thermostability. *Structure* **5**, 635-646.
- Ewald, P. (1923). Kristalle und Röntgenstrahlen. Springer-Verlag, Berlin.
- Ewald, P.P., (1921) Das 'reziproke Gitter' in der Strukturtheorie. *Zeitschrift für Kristallographie.* **56**, 129-156.
- Flocco, M.M. and Mombray, S.L. (1994) Planar stacking interactions of arginine and aromatic side chains in proteins *J. Mol. Biol.*, **235**, 709-717
- Friedrich, W., Knipping, P. and Laue, M., (1912) Interferenz-Erscheinungen bei Röntgenstrahlen. *Sitzungsbericht der Mathematisch-Physikalischen Klasse der königlichen Bayerischen Akademie der Wissenschaften zu München*, 303-322.
- Frolow F., Harel M., Sussman J.L., Mevarech M. and Shoham M. (1996): Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. *Nat Struct Biol*, **3**:452-457.
- Gewirth, D. (1995). The HKL Manual. A Description of the programs DENZO, XDISPLAYF, and SCALEPACK: An Oscillation DATA Processing SUITE for Macromolecular Crystallography. The Scalepack Manual. Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, USA.
- Gewirth, D. (1995). The HKL Manual. A Description of the programs DENZO, XDISPLAYF, and SCALEPACK: An Oscillation DATA Processing SUITE for Macromolecular Crystallography. The Denzo and XDisplayF Manual. Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, USA.

5. References

- Green D.W., Ingram, V.M., and Perutz, M.F. (1954) The structure of haemoglobin. IV. Sign detection by the isomorphous replacement method. *Proc. Roy. Soc. London*, **A225**, 287-307.
- Hauptmann, H. (1997) Phasing methods for protein crystallography. *Curr. Opin. Struct. Biol.* **7**, 672-680.
- Hennig, M., Darimont, B., Sterner, R., and Kirschner, K. (1995). 2.0 Å structure of indole-3-glycerol phosphate synthase from the hyperthermophile *Sulpholobus solfataricus*: possible determinants of protein stability. *Structure* **3**, 1295-1306.
- Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices *J. Mol. Biol.* **233**, 123-138
- Hoppe, W. (1957) Die Faltmolekülmethode: Eine neue Methode zur Bestimmung der Kristallstruktur bei ganz oder teilweise bekannten Molekülstrukturen. *Acta Cryst.* **10**, 750-751.
- Hough, D.W. and Danson, M.J. Extremozymes. (1999) *Curr Opin Chem Biol.* **3**:39-46.
- Hubbard, S.J. and Thornton, J.M. (1993), 'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London."
- Huber, R. (1965) Die automatisierte Faltmolekülmethode. *Acta Cryst.* **19**, 353-356
- Huber, R., Kurr, M., Jannasch, H.W. and Stetter, K.O. (1989) A novel group of abyssal methanogenic archaeobacteria (*Methanopyrus*) growing at 100°C *Nature*, **342**, 833-834.
- Hubbard, S.J., Campbell, S.F., and Thornton, J.M. (1991) Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase inhibitors. *J. Mol. Biol.* **220**, 507-530
- Ignatova, Z. and Gierasch, L.M. (2006) Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. *Proc. Natl. Acad. Sci. USA* **103**, 13357-61
- Jaenicke, R. (1996) Stability and folding of ultrastable proteins: eye lens crystallins and enzymes from thermophiles. *FASEB J.*, **10**, 84-92
- Jancarik, J. and Kim, S.-H. (1981) Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.*, **24**, 409-411
- Janin, J., Miller, S. and Chothia, C.H. (1988) Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.*, **204**, 155-164
- Jolley K.A., Russell, R.J.M., Hough, D.W. and Danson, M.J.: (1997) Site-directed mutagenesis and halophilicity of dihydrolipoamide dehydrogenase from the halophilic Archaeon *Haloferox volcanii*. *Eur J Biochem*, **248**:362-368.

5. References

- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119
- Kabsch, W. (1978) A discussion of the solution for the best rotation to relate two sets of vectors *Acta Cryst.* **A34**, 827-828
- Kabsch, W. (1988) Automatic indexing of rotation diffraction patterns. *J. Appl. Crystallogr.* 2157-71
- Kabsch, W. & Sander, C. (1983) Dictionary of protein secondary structures: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637
- Kelly, C.A., Nishiyama, M., Ohnishi, Y., Beppu, T. & Birktoft, J.J. (1993) Determinants of protein thermostability observed in the 1.9Å crystal structure of malate dehydrogenase from the thermophilic bacterium *Thermus favus*. *Biochemistry*, **32**, 3913-3922
- Kendrew J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.C. and Davies, D.R. (1960) Structure of myoglobin; a three-dimensional fourier synthesis at 2 Å resolution. *Nature*, **185**, 422-427.
- Kleywegt, G.J. and Read, R.J. (1997) Not your average density. *Structure*, **5**, 1557-1569.
- Kleywegt, G.A. and Jones, T.A. (1994). 'Halloween — masks and bones.' In *Proceedings of the CCP4 Study weekend*. (Bailey, S., Hubbard, R. & Waller, D., eds), pp. 59–66, Daresbury Laboratory, Warrington, UK.
- Kleywegt, G.J., and Jones, T.A. (1996) xdlMAPMAN and xdlDATAMAN--programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets *Acta Cryst.* **D52**, 826-828
- Korndorfer, I., Steipe, B., Huber, R., Tomschy, A. and Jaenicke, R. (1995) The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5Å resolution. *Structure*, **3**, 1295-1306.
- Korndörfer, I., Steipe, B. Huber, R. and Tomschy, A. (1995). The Crystal Structure of Holo-glyceraldehyde-3-phosphate Dehydrogenase from the Hyperthermophilic Bacterium *thermotoga maritima* at 2.5 Å Resolution. *J. Mol. Biol.* **246**, 511-521.
- Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures *J. Appl. Cryst.*, **24**, 946-950
- Kunow, J., Shima, S., Vorholt, J.A. and Thauer, R.K. (1996) Primary structure and properties of the formyltransferase from the mesophilic *Methanosarcina barkeri*: comparison with the enzymes from thermophilic and hyperthermophilic methanogens. *Arch. Microbiol.* **165**, 97-105
- Kuntz, I.D. (1971) Hydration of macromolecules IV. Polypeptide conformation in frozen solutions *J. Am. Chem. Soc.*, **93**, 516-518

5. References

- Kurr. M., Huber, R., König. H., Jannasch, H. and Stetter, K.O. (1991) *Methanopyrus kandleri* and sp. nov. represents a novel group of hyperthermophilic methanogens growing at 110°C. *Arch. Microbiol.*, **156**, 239-247.
- Lamzin, V.S. and Wilson, K.S. (1997) Automated Refinement for Protein Crystallography. *Meth. Enzymol.* **277**, 269-305.
- Laskowski, R.A. MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures *J. Appl. Cryst.*, **26**, 283-291
- Laskowski, R.A. (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph.* **13**, 323-30
- Lehmacher, A. (1994) Cloning, sequencing and transcript analysis of the gene encoding the formylmethanofuran:tetrahydromethanopterin formyltransferase from the hyperthermophilic *Methanothermus fervidus*. *Mol. Gen. Genet.*, **242**,73-80.
- Lentzen G. And Schwarz T, (2006) Extremolytes: Natural compounds from extremophiles for versatile applications *Appl Microbiol Biotechnol.* **72**, 62
- Leslie, A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 and ESR-EACBM Newsletter on Protein Crystallography, 26, 83-91
- Lim, J.-H., Yu Y.G., Han, Y.S., Cho, S.-j., Ahn, B.-Y., Kim, S.-H. & Cho, Y. (1997) The Crystal Structure of an Fe-Superoxide Dismutase from the Hyperthermophilic *Aquifex pyrophilus* at 1.9Å Resolution: Structural Basis for Thermostability. *J. Mol. Biol.*, **270**, 259-274
- McDonald, I.K. and Thornton, J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J Mol Biol.* **238**, 777-93.
- Mamat B., Roth A., Grimm C., Ermler U., Tziatzios C., Schubert D., Thauer R.K. and Shima S., (2002), Crystal structures and enzymatic properties of three formyltransferases from archaea: Environmental adaptation and evolutionary relationship *Protein Science* **11**, 2168–2178
- Maestrojuán, G., Boone, J., Mah, R. and Menaia, J. (1992). Taxonomy and Halotolerance of Mesophilic Methanosarcina Strains, Assignment of Strains to species, and Synonymy of Methanosarcina mazei and Methanosarcina frisia. *Int. J. Syst. Bacteriol.* **42**, 561-567.
- Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
- May, A., and Johnson, M. (1994). Protein structure comparison using a combination of a genetic algorithm, dynamic programming and least-squares minimization. *Protein engineering* **7**, 475-485
- Merritt, E.A. and Murphy, M.E.P. (1994) Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Cryst.* **D50**, 869-873

- Miller, S., Lesk, A.M., Janin, J., and Chothia, C. (1987) The accessible surface area and stability of oligomeric proteins. *Nature*, **328**, 834-836
- Navaza, J. 1994. AMoRe: An automated package for molecular replacement. *Acta Cryst. A* **50**: 157-163.
- Nicholls, A., Bharadwaj, R. and Honig, B. (1993). GRASP: graphical representation and analysis of surface properties. *Biophys J.* **64**, 166-170.
- Orengo, C.A., and Thornton, J.M. (1993) Alpha plus beta folds revisited: some favoured motifs. *Structure*, **1**, 105-120
- Otwinowsky, Z. (1991) Maximum likelihood refinement of heavy atom parameters. In Proceedings of the CCP4 Study Weekend, Wolf, W., Evans, P.R., & Leslie, A.G.W., eds. pp.80-86, Daresbury Laboratory, Warrington, UK.
- Patterson, A.L. (1934) A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals. *Phys. Rev.* **46**, 372-376.
- Perutz, M. and Raidt, H. (1975). Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature* **255**, 256-259.
- Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G. and North, A. (1960) Structure of haemoglobin; a three-dimensional fourier synthesis at 5.5 Å resolution, obtained by x-ray analysis. *Nature*, **185**, 416-422.
- Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configuration. *J. Mol. Biol.* **7**, 95-99
- Richards, F.M., (1977) *Annu. Rev. Biophys. Bioeng.* **6**, 151
- Röntgen, W.C., (1895) Ueber eine neue Art von Strahlen. *Sitzungsberichte der Physikalisch-Medizinischen Gesellschaft zu Würzburg*, 132-141.
- Rossmann, M.G. and Blow, D.M. (1962) The detection of subunits within the crystallographic asymmetric unit. *Acta Cryst.* **15**, 24-31.
- Rosper, S. and Stetter K.O., (1991) Methyl-coenzyme M reductase and other enzymes involved in methanogenesis from CO₂ and H₂ in the extreme halophile *Methanopyrus kandleri*. *Arch. Microbiol.*, **156**, 49-55.
- Russell, R.J.M., Ferguson, J.M.C., Hough, D.W., Danson, M.J. and Taylor, G.L. (1997) The crystal structure of citrate synthase from the hyperthermophilic Archaeon *Pyrococcus furiosus* at 1.9 Å resolution. *Biochemistry*, **36**:9983-9994.
- Salminen, T., Teplyakov, A., Kankare, J., Cooperman, B.S, Lahit, R. and Goldman, A. (1996) An unusual route to thermostability disclosed by the comparison of *Thermus thermophilus* and *Escherichia coli* inorganic pyrophosphatases *Protein Sci.*, **5**, 1014-1025

5. References

- Saraste, M., Sibbald, P.R. and Alfred Wittinghofer (1990), The P-loop; a common motif in ATP- and GTP-binding proteins, *Trends Bioch. Sci.*, **15**, 430-434
- Sayre, D. (1974) Least-Squares refinement. II: High resolution phasing of a small protein. *Acta Cryst.* **A30**, 180-184
- Schuller, D.J., Grant, G.A., and Banaszak, L.J. (1995) The allosteric ligand site in the cooperative enzyme phosphoglycerate dehydrogenase *Nat. Struct. Biol.*, **2**, 69-76
- Serre, L., Verbree, E.C., Dauter, Z., Stuitje, A.R. & Derewenda, Z.S. (1995) The *Escherichia coli* malonyl-CoA: acyl carrier protein transacylase at 1.5 Å resolution. Crystal structure of a fatty acyl synthase. *J. Biol. Chem.*, **270**, 12961-12964
- Sheldrick, G.M. (1990) Phase Annealing in *SHELX-90*: Direct Methods for Larger Structures, *Acta Cryst.* **A46**, 467-473.
- Sheldrick, G.H. (1991). Heavy atom location using *SHELXS-90*. In *Proceedings of the CCP4 Study Weekend*. (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 23-38, Daresbury Laboratory, Warrington, UK.
- Shima, S., Weiss, D.S., and Thauer, R.K., (1995) Formylmethanofuran: tetrahydromethanopterin formyltransferase (Ftr) from the hyperthermophilic *Methanopyrus kandleri*: Cloning, sequencing and functional expression of the ftr gene and one-step purification of the enzyme overproduced in *E.coli*. *Eur. J. Biochem.*, **230**, 906-913.
- Shima, S., Thauer, R.K., Michel, H. and Ermler, U. (1996) Crystallization and preliminary X-ray diffraction studies of formylmethanofuran:tetrahydromethanopterin formyltransferase from *Methanopyrus kandleri* *Proteins*, **26**, 118-120.
- Shima, S., Tziatzios, C., Schubert, D., Fukada, H., Takahashi, K., Ermler, U. and Thauer, R.K. (1998) Lyotropic-salt-induced changes in monomer/dimer/tetramer association equilibrium of formyltransferase from the hyperthermophilic *Methanopyrus kandleri* in relation to the activity and thermostability of the enzyme. *Eur. J. Biochem.* **258**, 85-92.
- Shima, S., Héroult, D.A., Berkessel, A., and Thauer, R.K. (1998). Activation and thermostabilization effects of cyclic 2,3 diphosphoglycerate *Arch. Microbiol.* **170**: 469-472
- Stetter, K.O. in: J. and K. Trần Thanh Vân, J.C. Mounolou, J. Schneider, C. McKay (Eds.) *Frontiers of Life - Colloque Interdisciplinaire du Comité National de la Recherche Scientifique*, Proc. 3rd "Recontres de Blois", Chateau de Blois, France 1991, Editions Frontières, Gif-sur-Yvette, France, 1992, pp. 195.
- Swanson, S. M. (1994) Core tracing: Depicting connections between features in electron density. *Acta Cryst.* **D50**, 695-708.
- Thauer, R.K. (1998) Biochemistry of methanogenesis: a tribute to Marjory Stephenson. *Microbiology* **144**: 2377-2406

- Thauer, R.K. (1996) Biodiversity and unity in biochemistry *Anatomie van Leewenhoek*, **71**, 21-32
- Turk, D. (1992) Weiterentwicklung eines Programms für Molekülgraphik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklärungen. Dissertation TU München, Deutschland.
- Vetriani, C, Maeder, D.L., Tolliday, N., Yip, K.S.P., Stillman, T.J., Britton, K.L., Rice, D.W., Klump, H.H. and Robb, F.T., (1998) Protein thermostability above 100°C: A key role for ionic interactions *Proc. Natl. Acad. Sci. USA* **95**, 12300-12305
- Wallace, A.C., Laskowski, R.A. and Thornton, J.M. 1995 LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Prot. Eng.*, **8**, 127-134.
- Wallon, G., Kryger, G. Lovett, S. and Oshima, T. (1997). Crystal Structures of *Escherichia coli* and *Salmonella typhimurium* 3-Isopropylmalate Dehydrogenase and Comparison with their Thermophilic Counterpart from *Thermus thermophilus*. *J. Mol. Biol* **266**, 1016-1031.
- Walker, J. Wonacott, A. and Harris, J. (1980). Heat Stability of a Tetrameric Enzyme, d-Glyceraldehyde-3-Phosphate Dehydrogenase. *Eur. J. Biochem.* **108**, 581-586.
- Wang, B.-C. (1985) Resolution of phase ambiguity in macromolecular crystallography. *Meth. Enzymol.* **115**, 90-112.
- Weeks, C. M. and Miller, R. (1999) Optimizing Shake-and-Bake for proteins. *Acta Cryst.* **D55**, 492-500.
- Woese, C.R., Kandler, O. and Wheelis, M.L., (1990). *Proc. National Academy of Sciences, USA* **87**, 4576
- Wolfe, R.S. (1991) My kind of biology. *Annu. Rev. Microbiol.*, **45**, 1-35.
- Yip, K., Stillman, T., Britton, K., Artymiuk, P. and Rice, D.W. (1995) The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure*, **3**, 1147-1158.
- Yip, K.S.P., Britton, K.L., Stillman, T.J., Lebbink, J., de Vos, W.M., Robb, F.T., Vetriani, C, Maeder, D. and Rice, D.W.: (1998) Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family. *Eur J Biochem*, **255**:336-346
- Zaccai, G. and Eisenberg, H. (1990) Halophilic proteins and the influence of solvent on protein stabilization *Trends Biochem. Sci.*, **15**, 333-337
- Zhang, K.Y.J. and Main, P. (1990) The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. *Acta Cryst.* **A46**, 377-381.

Structure Analysis of *Rps. viridis* Terbutryn Resistant Mutant T3 Reaction Centre.

Several terbutryn herbicide resistant *Rps. viridis* mutants were available. One double mutant, T3 was of interest because one mutation, V-M263-F was associated with the menaquinone, Q_A binding site. Since Q_A was not expected to exchange appreciably, it was unclear why resistance to herbicide should develop there.

Crystals of T3 reaction centres were available from Dr. Günther Fritsch. Diffraction data were collected at Hasylab/DESY and in house. Processing of the data showed no indication of positive density for a Phe residue at the position M263. Rebuilding the model with a Phe at the site, refinement of the model and map calculation in X-PLOR revealed a strong negative density at the Phe site, clearly showing that the Phe was not present in the sample measured. Since the reversion rate was expected to be low, the bacteria for the sample were not grown in the presence of herbicide, Dr. Günther Fritsch personal communication. Perhaps the lack of evidence for Phe at this site can be attributed to the sample being native.

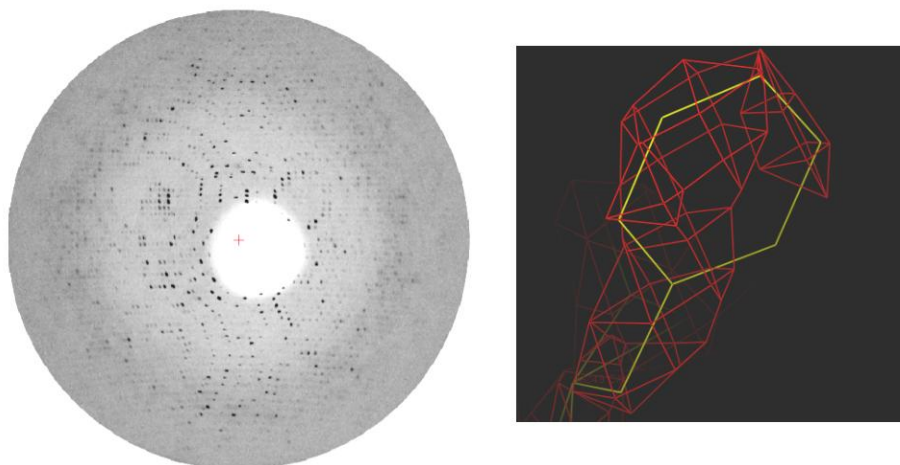


Figure 1 Sample diffraction pattern (left), negative difference electron density map (right) of the V263F model indicating strong negative density due to the Phe model.

Crystallization and structure of the photosynthetic reaction centres from *Rhodobacter sphaeroides* - wild type and mutants.

Fritzsche G, Ermler U, Merckel M and Michel H.

In: The Reaction Center of the Photosynthetic Bacteria, pp 3-13,
Michel-Beyerle ME (ed)
Springer-Verlag, Berlin, Heidelberg, New York. (1996)

**Crystallization and Structure of the Photosynthetic Reaction Centres
from *Rhodobacter sphaeroides* - wild type and mutants**

Günter Fritzsche, Ulrich Ermler, Michael Merckel and Hartmut Michel

Max-Planck-Institut für Biophysik, Abteilung Molekulare Membranbiophysik,
D-60528 Frankfurt a.M., Germany

Abstract. Trigonal crystals of the photosynthetic reaction centres from the purple bacterium *Rhodobacter sphaeroides* have been grown with potassium phosphate as precipitant. These crystals diffract to 2.6 Å and are suitable for a detailed structure determination. The cofactor binding is similar to that observed in the reaction centres of *Rhodospseudomonas viridis*. A 23 Å long water chain connects the secondary quinone Q_B with the cytoplasm. The structures of the mutants Thr M222 → Val, Trp M252 → Phe, and Trp M252 → Tyr show minor differences compared with the wild type. In the carotenoidless strain R26 the carotenoid-binding site is occupied by a detergent molecule.

Keywords. Photosynthesis, reaction centre, crystallization, structure

1. The orthorhombic crystal form

The photosynthetic reaction centre (RC) from the purple bacterium *Rhodobacter (Rb.) sphaeroides* is an integral membrane protein with an $M_r = 102,000$. In the past, its molecular structure has been studied by analysis of orthorhombic crystals [1,2,3,4,5,6]. When solubilized with N,N-dimethyldodecylamine N-oxide (LDAO) or octyl β-glucoside orthorhombic crystals of this RC can be grown with poly(ethyleneglycol) (PEG⁴⁰⁰⁰) as the precipitant in the presence of sodium chloride (300 to 500 mM). Indispensable for the formation of these crystals is the addition of ≥ 3% (w/v) heptanetriol as was found for the crystallization of the RC's from *Rhodospseudomonas (Rps.) viridis* [7].

Applying the vapour diffusion technique orthorhombic crystals appear after a few days or weeks. They are needle-like and may grow to 4 mm long, the thickness is usually a few tenths of a mm, sometimes almost 1 mm (Fig. 1). In spite of these huge dimensions the crystals lack compactness. Generally, considerable parts inside the needles are hollow. When exposed to X-rays the orthorhombic crystals are of low quality. The resolution is about 3 Å in the horizontal direction of the X-ray film. In the vertical direction, however, it is only 4 Å or even worse. This anisotropic diffraction of the reflections is due to fact

that the contacts in the crystal lattice are relatively strong in direction to the long crystal axis (z-axis) but rather weak in the directions perpendicular to the z-axis. The observed number of reflections (ca. 22,000 in [4]) is much less than required in order to refine the three spatial coordinates and one isotropic atomic B-factor for each of about 7,000 atoms. The completeness of the data sets (independent reflections) is less than 70% in all cases. For these reasons and for the lack of heavy metal derivatives the structure of the RC from *Rb. sphaeroides* could only be solved with the information from the known RC structure of *Rps. viridis* [8]. Due to the limited resolution the minor differences between both RC's could not be detected by analysis of the orthorhombic crystals.

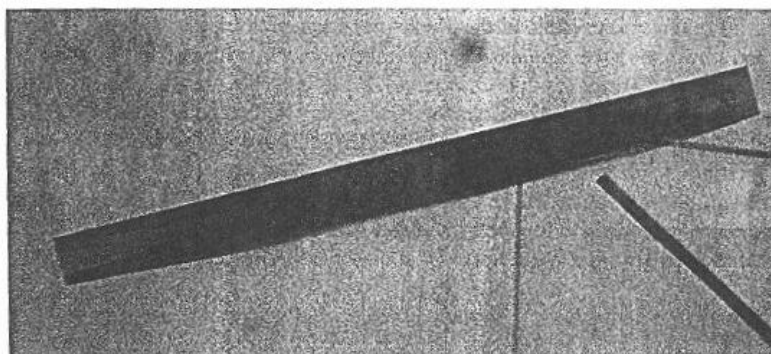


Fig. 1. Orthorhombic crystal grown from PEG⁴⁰⁰⁰. The length of the crystal is 3.5 mm, the thickness is 0.3 mm. The light, wedge-shaped areas at the right and left part of the crystal indicate low compactness.

2. The trigonal crystal form

With the introduction of potassium phosphate as precipitant another crystal form of this protein could be grown [9]. The favorable concentration of RC is 100 μ M (10.6 mg/ml). Again LDAO is a proper detergent with an optimal concentration of $0.1 \pm 0.05\%$ (v/v). The suitable concentration of potassium phosphate in the mother solution is 0.5 to 1.0 M whereas 1.4 to 1.8 M are optimal for the bottom solution. Heptanetriol is absolutely necessary for this crystallization. Its favourable concentration is 1 to 2% (w/v) with an optimum at about 1.8% (w/v). The addition of 1% to 10% dioxane supports the growth of single trigonal crystals, probably due to a decrease of the dielectric constant. Crystals appear after one or several weeks. They are compact, up to 3 mm long and sometimes more than 1 mm thick (Fig. 2). They belong to the trigonal space group $P3_121$ with unit cell dimensions of 141.4 \AA , 141.4 \AA , 187.2 \AA . The solvent content is 77%.

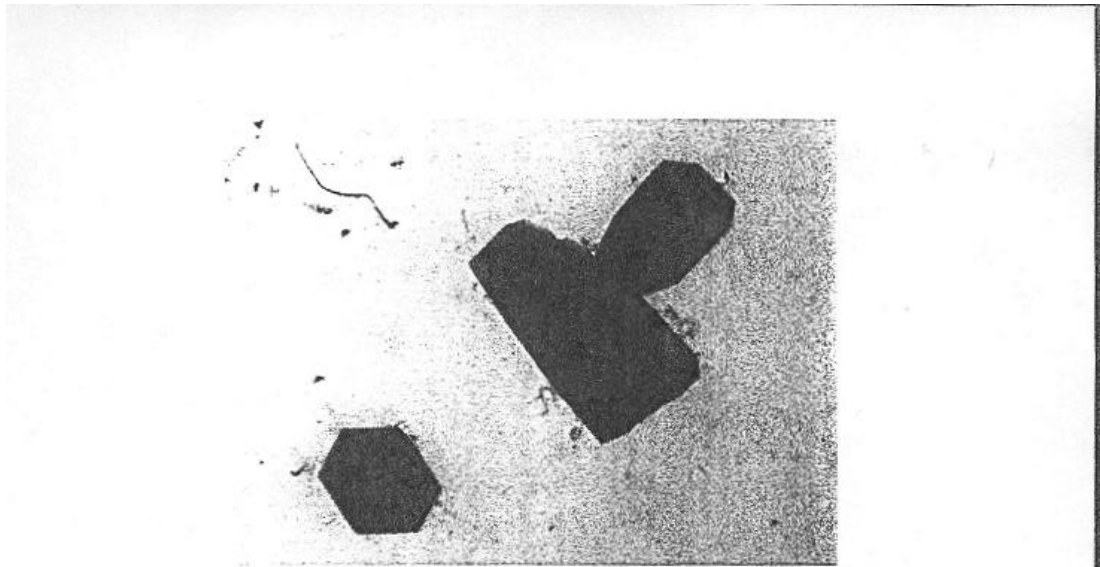


Fig. 2. Trigonal crystals from RC of *Rb. sphaeroides* grown from potassium phosphate. The length of the biggest crystal is 0.8 mm, its thickness 0.5 mm. The crystal with the hexagonal shape is seen along its z-axis.

With potassium phosphate as precipitating agent the trigonal crystals can coexist with the orthorhombic ones in the same crystallization sample (Fig. 3).

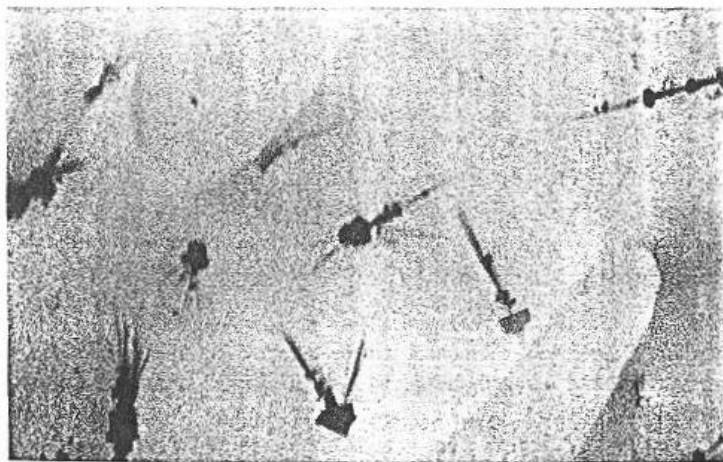


Fig. 3. Coexisting trigonal and orthorhombic crystals. Here, 5 μ l of mother solution have been put on a glass slide. The thin, needle-like structures are orthorhombic - the flat structures trigonal crystals.

The needle-like orthorhombic crystals resemble in their low diffraction power the crystals obtained with PEG⁴⁰⁰⁰. For this reason the conditions in the mother solution must be chosen in such a way that preferentially trigonal crystals grow. This can be achieved 1) by applying concentrations of heptanetriol less than 2.0% and 2) by choosing pH < 6.7. The quality of the trigonal crystals in the X-ray beam, however, is less good at pH < 7.0, but it is much better at pH values around 7.5. So, a pH ≥ 7 should be chosen for the crystallization, even if orthorhombic needles appear and compete with the trigonal crystals. Up to now, the results of the crystallization attempts are not yet reproducible and the physical-chemical background of crystal formation is not understood very well.

The crystal packing density is rather low. In Fig. 4 a view along the crystallographic z-axis is shown where the protein density is represented in black and the solvent space (buffer plus detergent) in white. Fig. 5 shows the crystal packing along the x-axis.

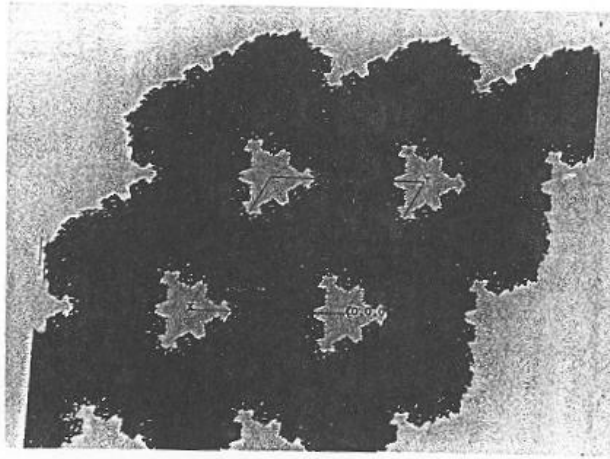


Fig. 4. A view into the crystal along the z-axis. The protein is represented in black, the solvent-containing "holes" in white. The unit cell is marked with thin lines.

More than 10 detergent molecules can be detected crystallographically. They are located around the hydrophobic parts of the L- and M-subunit and replace the lipid molecules (Fig. 6).

Recently, a third crystal form of the RC of *Rb. sphaeroides* has been grown [10]. These crystals have a tetragonal space group and diffract to 2.8 Å resolution. Thus, three different crystal forms of the *Rb. sphaeroides* RC are now available [Table 1].

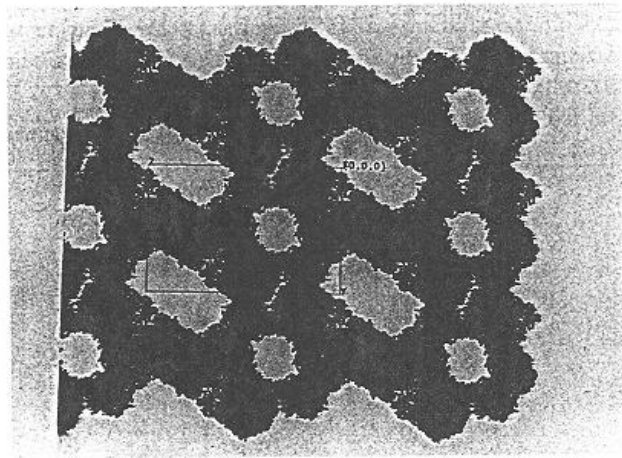


Fig. 5. A view into the crystal along the x-axis. Further explanations see Fig. 4.

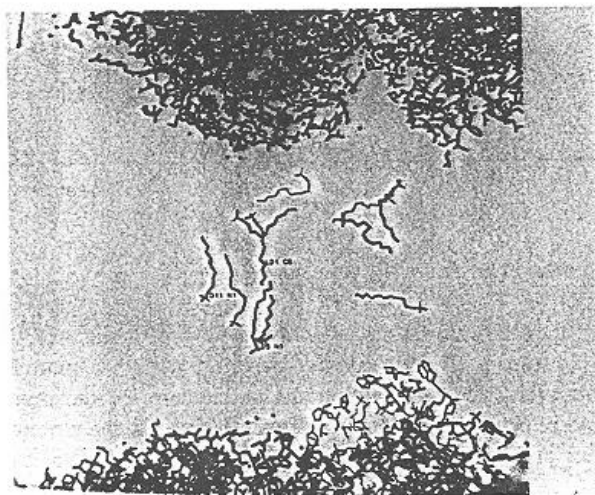


Fig. 6. The crystallographically detectable LDAO molecules. In the upper and lower part of the figure the protein density in a crystal is shown. In the middle part, one RC molecule is omitted (ghost molecule), merely its LDAO molecules are presented. The ghost molecule is enveloped by the LDAO's, its H-subunit touches the residues of the RC molecule shown below. The white regions left and right from the LDAO's are holes filled with buffer (see Fig. 4 and 5).

Table 1. Crystal types of RC from *Rb. sphaeroides*

	Orthorhombic	Trigonal	Tetragonal
Precipitant	PEG	Potassium phosphate	PEG
Detergent	LDAO	LDAO	β -OG
Space group	P2 ₁ 2 ₁ 2 ₁	P3 ₁ 21	P4 ₁ 32 ₁ 2 ₁
Number of unique reflections	23,000	63,000	51,000
Number of molecules in the asymmetric unit	1	1	2
Resolution	3 - 4 Å	2.6 Å	2.8 Å
References	[1,2,3]	[11, unpubl.]	[10]

3. Structural features of the wild type RC

The structure of the RC from *Rb. sphaeroides*, wild-type strain ATCC 17023, has been analyzed by means of the trigonal crystals [11]. With more than 56,000 reflection points and 2.65 Å resolution the architecture of the protein-cofactor interaction could be studied in more detail than before. In the structure based on the orthorhombic crystal form, some differences in cofactor binding between the RC's from *Rps. viridis* and *Rb. sphaeroides* were found [1,2,3,4,5,6]. Most of these differences could not be confirmed by the analysis of the structure from the trigonal crystals. On the contrary, the protein-cofactor interactions along the electron transfer pathway is rather similar to the *Rps. viridis* RC.

Generally, the distances between the centres of mass of the cofactors along the active branch (A branch) are slightly larger (ca. 0.2 Å) in the *Rb. sphaeroides* than in the *Rps. viridis* RC's. It is unclear to what extent these small deviations are a consequence of the different crystallization conditions like ionic strength or

dielectric constant. The major difference in the structures of both RC's is observed at the Q_B -binding site. Whereas in the *Rps. viridis* RC the secondary quinone Q_B is located 8 Å away from the non-heme iron it is 12.3 Å in the *Rb. sphaeroides* RC. Further refinement of mutated resp. carotenoidless RC from *Rb. sphaeroides* shows an ambiguous pattern of the Q_B position (Fig. 7). The nearest Q_B is about 10 Å away from the non-heme iron. Detailed studies of these discrepancies are under way.

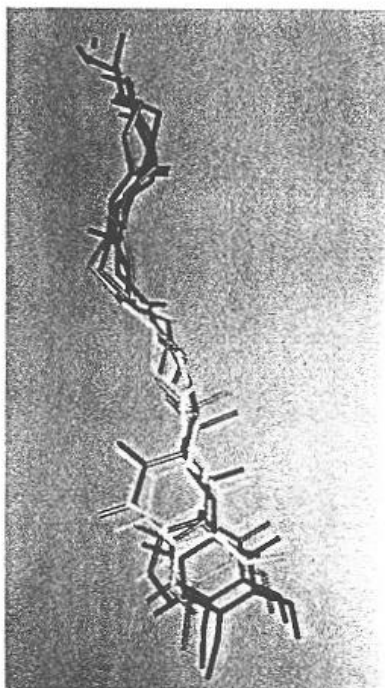


Fig. 7. Superposition of Q_B molecules from five different RC structures (wild type, R26, mutants Trp M252 → Tyr, Ser M222 → Val and Trp M252 → Phe). There is no common site for the five head groups. They are distributed over a distance of about 3 Å. The Q_B head group in the wild type is 12.3 Å away from the non-heme iron (black) whereas the one from Trp M252 → Phe is about 10 Å away.

Between the Q_B -binding site and the cytoplasm a water chain of 23 Å length has been found. It consists of 14 fixed water molecules which are close enough to each other to form a linear structure of water molecules connected by hydrogen bonds (Fig. 8). Along this chain mostly acidic and basic amino acid residues are

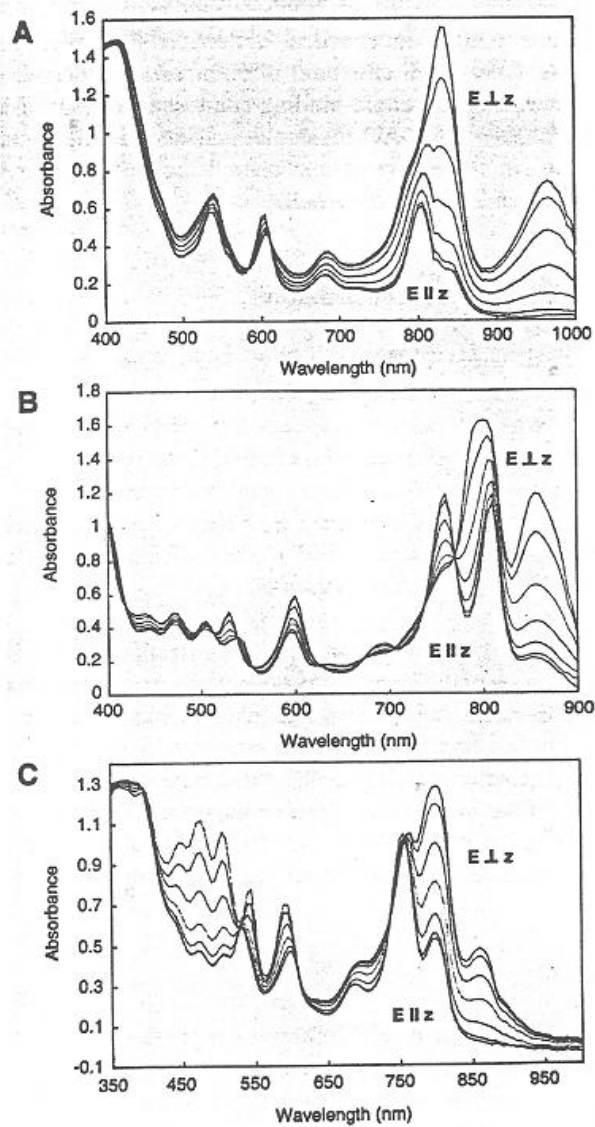


Fig. 9. The polarized absorption spectra of tetragonal crystals from *Rps. viridis* RC (A), of the orthogonal (B) and trigonal (C) crystals from *Rb. sphaeroides* RC. E || z (E ⊥ z): The direction of the polarized light is oriented parallel (perpendicular) to the long axis (z-axis) of the crystal. In each case, a strong dichroism is observed in the region of the special pair absorption (960 nm for the RC from *Rps. viridis* and 860 nm for the RC from *Rb. sphaeroides*). If the light is polarized perpendicular to the z-axis the special pair absorption for each of the three crystal forms is much stronger than for light polarized parallel to the z-axis.

Good crystals could also be grown from the carotenoidless strain R26. A data set with more than 63,000 independent reflections has been collected. The resolution is 2.6 Å. Only small structural differences are observed for the major part of the RC. Even the carotenoid-binding site resembles that of the wild type. A ligand, most probably an LDAO molecule, is located in the central part of the carotenoid binding site. The ring of one phenylalanine residue is tilted by 90° filling a gap due to the missing carotenoid.

5. Crystal growth and cofactor orientation

Up to now, four different crystal forms of RC's from *Rps. viridis* and *Rb. sphaeroides* are available: A tetragonal form from the RC from *Rps. viridis*; an orthorhombic, a trigonal and a tetragonal form, respectively, from *Rb. sphaeroides*. The polarized absorption spectra of three types are known (Fig. 9). They share a striking feature: The special pair absorption (960 nm with the *Rps. viridis* RC, 860 nm with the *Rb. sphaeroides* RC) is maximal when the actinic light is polarized perpendicular to the long axis (z-axis) of the crystals. This dichroism is clearly pronounced in each case.

This result leads to the conclusion that the direction of the fastest crystal growth is oriented perpendicular to the planes of the bacteriochlorophyll dimers. Thus, a relationship exists between special pair orientation and crystal growth. This is due to the fact that the transition moment of the major special pair absorption band is oriented parallel to the membrane between the polar surface parts at both sides of the membrane. These polar surface domains are responsible for the formation of the crystal lattice so that a natural tendency of preferred crystal growth perpendicular to the special pair transition moment exists.

Acknowledgements

We thank Dr. D. Oesterhelt and Dr. U. Stolz for supplying the mutants as well as Dr. H. Scheer and M. Meyer for cooperation with the crystallization of the RC from R26. We acknowledge help in use of the facilities at DESY, Hamburg, by the staff of the EMBL outstation. This work was supported by the Max-Planck-Gesellschaft and the Fonds der Chemischen Industrie.

References

1. Allen, J.P., Feher, G., Yeates, T.O., Rees, D.C., Deisenhofer, J., Michel, H. and Huber, R. (1986) Proc. Natl. Acad. Sci. USA 83, 8589-8593.
2. Chang, C.-H., Tiede, D., Tang, J., Smith, U., Norris, J. and Schiffer, M. (1986) FEBS Lett. 205, 82-86.

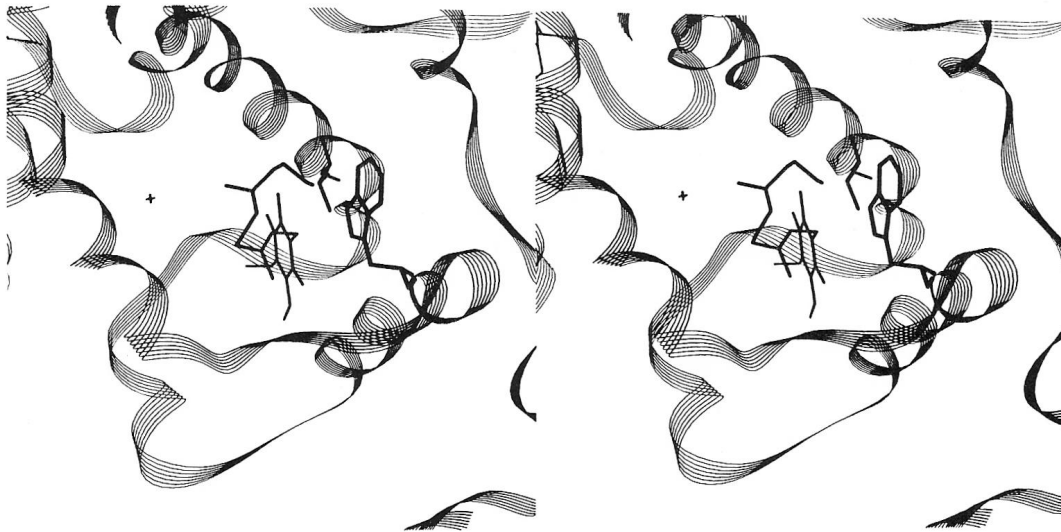
3. Arnoux, B., Ducruix, F., Reiss-Husson, F., Lutz, M., Norris, J., Schiffer, M. and Chang, C.-H. (1989) *FEBS Lett.* 258, 47-50.
4. Yeates, T.O., Komiya, H., Chirino, A., Rees, D.C., Allen, J.P. and Feher, G. (1988) *Proc. Natl. Acad. Sci. USA* 85, 7993-7997.
5. Allen, J.P., Feher, G., Yeates, T.O., Komiya, H. and Rees, C.D. (1988) *Proc. Natl. Acad. Sci. USA* 85, 8487-8491.
6. El-Kabbani, O., Chang, C.-H., Tiede, D., Norris, J. and Schiffer, M. (1991) *Biochemistry* 30, 5361-5369.
7. Michel, H. (1982) *J. Mol. Biol.* 158, 657-572.
8. Deisenhofer, J., Epp, O., Miki, R., Huber, R. and Michel, H. (1985) *Nature* 318, 618-624.
9. Buchanan, S.K., Fritsch, G., Ermler, U. and Michel, H. (1993) *J. Mol. Biol.* 230, 1311-1314.
10. Allen, J.P. (1994) *Proteins Struct. Funct. Genet.* 20, 283-286.
11. Ermler, U., Fritsch, G., Buchanan, S.K. and Michel, H. (1994) *Structure* 2, 925-936.
12. Stolz, U., Finkle, U., Holzappel, W., Lauterwasser, C., Zinth, W. and Oesterhelt, D. (1994) *Europ. J. Biochem.* 223, 233-242.

Appendix A2: Photosynthetic reaction centre mutants from *Rb. sphaeroides*.

Kristallographische Verfeinerung von Mutanten des Reaktionszentrums von *Rb. sphaeroides*.

In dem Reaktionszentrum (RZ) aus Purpurbakterien werden Elektronen von dem "special pair" Bakteriochlorophyll-a-Dimer über das "bridging" Bakteriochlorophyll-a und Bakteriopheophytin-a zum Quinone-a übertragen. Die M-Untereinheit des RZ bindet Quinone-a. In allen bekannten Aminosäuresequenz der RZ aus Purpurbakterien sind die beiden Reste M252 tryptophan (TRP) und M222 threonine (THR) erhalten. Es wird vermutet, daß die Reste eine wichtige Rolle bei der Bindung des Quinone spielen und zu den electron transfer-Eigenschaften des Quinones beitragen.

Drei Mutanten der Quinone-a Stelle des RZ von *Rb. sphaeroides* wurden mittels gerichtete Mutagenese produziert, M252 TRP->TYR, M252 TRP->PHE und M222 THR->VAL. Die Kristallstrukturen wurden durch Differenz-Fourier Methoden aufgeklärt. Die Ergebnisse der Verfeinerung den Strukturen werden vorgestellt.



Stereo Abbildung der Qa stelle von *Rb. sphaeroides* wild-typ.

Max-Planck-Institut für Biophysik

Günter Fritsch

10. Juli 1997

Lieber Michael,

ich darf hoffentlich beim Du bleiben.

Das Paper ist noch in Arbeit. Es hatte trotz häufiger Nachfrage ungefähr ein Jahr bei Hartmut geruht, was ich Dir, glaube ich, auch öfters gesagt hatte. Hartmut war mit unserem Text nicht sehr einverstanden, ich mußte und muß es noch umschreiben. Ich weiß, daß ich damit hinterherhänge. Ich werde Dir die neue Fassung vor dem Einreichen bei einer Zeitschrift noch einmal vorlegen.

Es tut mir leid, daß diese Sache so lang dauert. Entschuldige bitte.

Welche Vorgänge die Verzögerung vor allem verursacht haben, werde ich Dir gern mündlich mitteilen.

Mit besten Grüßen



Structure Analysis of Glucose:Fructose Oxidoreductase

The anaerobic Gram-negative bacterium *Zymomonas mobilis* occurs naturally in sugar-rich growth media. The bacterium ferments glucose, fructose and sucrose, utilizing the Entner–Doudoroff pathway, with ethanol and carbon dioxide as the principal products. Glucose-fructose oxidoreductase (GFOR) from *Z. mobilis* is a periplasmic, soluble enzyme that forms a homotetramer of 160 kDa with one NADP(H) cofactor per subunit that is tightly, but noncovalently, bound.

Well diffracting crystals of GFOR in spacegroup $P2_12_12$ were available. Extensive heavy atom screening failed to yield useful derivatives for phase determination. Native GFOR was very sensitive to mercurial compounds. An active site cysteine mutant, C194S was very robust toward mercurial compounds. Extensive screening and co-crystallization attempts failed to yield derivatives with this mutant.

A group of six cysteine mutants in the C194S background were generated with the hope of obtaining suitable heavy atom label sites. Mutagenesis of the *gfo* gene was carried out in M13 and sequence confirmation of the products was obtained. Following the preparation of the mutants the GFOR structure was solved by a competing group.

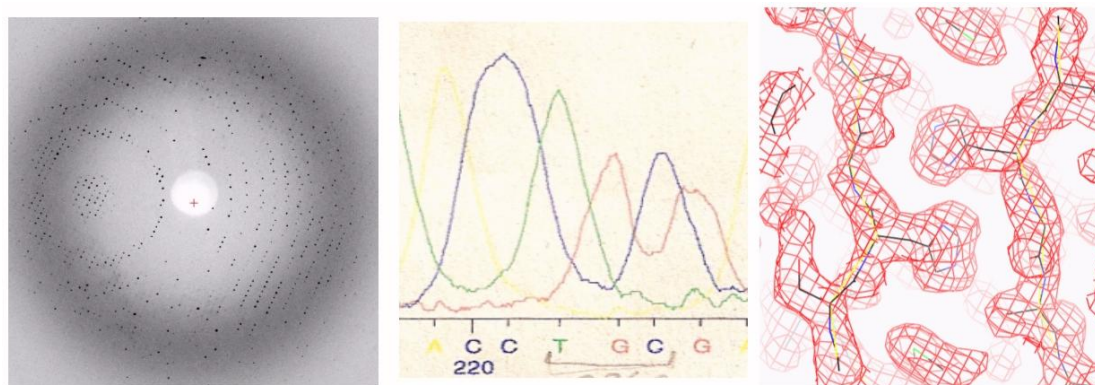


Figure 1 Structure analysis of GFOR. Typical diffraction pattern for GFOR crystal, left. Example ALF sequencer output for one mutant, T236C in the C194S background (T(U)GC=Cys)

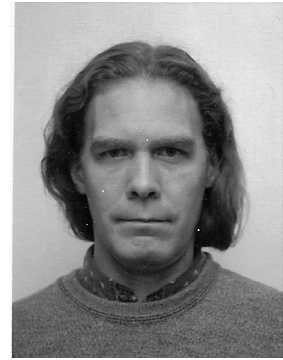
A consistent problem with the analysis was a large peak in native Patterson maps. This may have been due to misindexing two crystal forms differing in the *b* axis length, ~ 94 Å versus ~ 284 Å giving rise to a pseudo-translation peak in the native Patterson. These two forms were properly identified by the competing group. The coordinates from this group were used as a molecular replacement search model

against the data available to obtain a solution. Although there was discussion to proceed with analysis of certain mutant forms of GFOR the project was ended

Curriculum Vitae

Personal

Michael Christopher Merckel
Cornwall, New York, USA
3 June 1961



Education:

1966-1973 Groton Heights Elementary School, Groton, Connecticut, USA.

1973-1976 Robert E. Fitch Junior High School, Groton, Connecticut, USA.

1976-1979 Robert E. Fitch Senior High School Groton, Connecticut, USA.

Studies:

1979-1983 University of Connecticut, Storrs, Connecticut, USA.
B.S. Biophysics

1983-1984 University of Connecticut, Storrs, Connecticut, USA.
B.S. Chemistry

1985-1991 University of Connecticut, Storrs, Connecticut, USA.
M.Sc. Biophysical Chemistry

Thesis:

1991-1997 Johann Wolfgang Goethe-Universität
Max-Planck-Institut für Biophysik, Frankfurt am Main, Germany.
Ph.D. thesis under the supervision of Dr. Ulrich Ermler.

Research:

1998-1999 Turku University, Turku, Finland
Biotechnology Institute
Researcher

1999-2008 Helsinki University, Helsinki, Finland
Structural Biology & Biophysics/Biomedicum
Researcher

Publications

Polianskyte Z, Peitsaro N, Dapkunas A, Liobikas J, Soliymani R, Speer O, Seitsonen J, Butcher S, Cereghetti GM, Linder M, Merckel M, Thompson J, & Eriksson O. A module in the submitochondrial architecture manifested by the obesity-linked protein LACTB. *submitted*

Merckel M.C., Huiskonen J.T., Bamford D.H., Goldman A. and Tuma R. (2005). The Structure of the Bacteriophage PRD1 Spike Sheds Light on the Evolution of Viral Capsid Architecture. *Molecular Cell* **18**(2), 161-170.

Kainov D.E., Simonov V., Bamford D.H., Tuma R., Gottlieb P., Wei H., Walsh M.A., Belrhali H. and M.C. Merckel (2004). Crystallization and preliminary X-ray diffraction analysis of bacteriophage phi12 packaging factor P7. *Acta Cryst.* **D60**(12), 2368-2370.

Psencik J., Ikonen T.P., Laurinmaki P., Merckel M.C., Butcher S.J., Serimaa R.E. and Tuma R. (2004). Supramolecular Organization of Chlorosomes, the Light Harvesting System of Green Bacteria. *Biophys. J.*, **87**, 1165-1172.

Nummelin H., Merckel M.C., Leo J.C., Lankinen H., Skurnik M. and Goldman A. (2004). The *Yersinia* adhesin YadA collagen-binding domain structure is a novel left-handed parallel β -roll. *The EMBO Journal*, **23**(4), 701–711.

Nummelin H, Merckel MC, el Tahir Y, Ollikka P, Skurnik M, Goldman A. (2003). Structural studies of *Yersinia* adhesin YadA. *Adv Exp Med Biol* **529**:85-88.

Merckel M.C., Tanskanen J., Edelman S., Westerlund-Wikström B., Korhonen T.K. and Goldman A. (2003). The Structural Basis of Receptor-binding by *Escherichia coli* Associated with Diarrhea and Septicemia. *J. Mol. Biol.* **331**(4), 897-905.

Sen S., Jaakola V.-P., Heimo H., Merckel M., Kivelä P., Lundstrom K., Scheinin M. and Goldman A. (2003). ScintiplatTM assay for a GPCR: a novel method for solid phase binding studies. *Perkin Elmer Technical Note*.

Nummelin H., Merckel M., el Tahir Y., Ollikka P., Skurnik M. and Goldman A. (2002). Structural Studies of a Collagen Binding Fragment of *Yersenia* Adhesin YadA. *Acta Cryst* **A58** (Supplement), C297.

Paavilainen V.O., Merckel M.C., Falck S., Ojala P.J., Pohl E., Wilmanns M. and Lappalainen P. (2002). Structural conservation between the actin monomer-binding sites of twinfilin and ADF/cofilin. *J. Biol. Chem.* **277**(45), 43089–43095.

Kajander T., Merckel M.C. , Thompson A., Deacon A.M., Mazur P., Kozarich J.W., and Goldman A. (2002). The Structure of *Neurospora crassa* 3-Carboxy-cis,cis-Muconate Lactonising Enzyme, a β -propellor Cycloisomerase. *Structure*, **10**, 483-492.

Merckel M.C., Kajander T., Deacon A.M., Thompson A., Grossmann J.G., Kalkkinen N. and Goldman A. (2002). 3-Carboxy-cis,cis-muconate lactonizing enzyme from *Neurospora crassa*: MAD phasing with 80 Selenomethionines. *Acta Cryst.* **D58(5)**, 727-734.

Merckel M.C., Fabrichniy I.P., Salminen A., Kalkkinen N., Baykov A.A., Lahti R. and Goldman A. (2001). Crystal Structure of *Streptococcus mutans* Pyrophosphatase: A New Fold for an Old Mechanism. *Structure*, **9**, 289-297.

Leppänen V.-M., Merckel M., Ollis D.L., Wong K.K, Kozarich J.W. and Goldman A. (1999). Pyruvate Formate-Lyase is Structurally Homologous to Type I Ribonucleotide Reductase. *Structure*, **7**, 733-744.

Ermler U., Merckel M.C., Thauer R.K. and Shima S. (1997). Formylmethanofuran: tetrahydromethanopterin Formyltransferase from *Methanopyrus kandleri*-new Insights into Salt-dependence and Thermostability. *Structure*, **5**, 635-646.

Fritsch G, Ermler U, Merckel M and Michel H (1996), Crystallization and structure of the photosynthetic reaction centres from *Rhodobacter sphaeroides* - wild type and mutants. In: Michel-Beyerle ME (ed) The Reaction Center of the Photosynthetic Bacteria, pp 3-13, Springer-Verlag, Berlin, Heidelberg, New York.

Zaki L., Böhm R. and Merckel M. (1996). Chemical labelling of Arginyl-Residues involved in anion transport mediated by human Band 3 protein and some aspects of their location in the peptide chain. *Cellular and Molecular Biology*, **42(7)**, 1053-1063

Acknowledgements

I'm very grateful to Professor Dr. Hartmut Michel and Dr. Ulrich Ermler for the opportunity and support to pursue this work at the Max-Planck-Institut für Biophysik (Frankfurt am Main). Furthermore I thank Professor Dr. Hugo Fasold for accepting the role as official advisor with the Johann Wolfgang Goethe-Universität (Frankfurt am Main) and for his guidance.

I thank Dr. Ulrich Ermler for the introduction to some of the practical aspects of protein crystallography. Frau Barbara Schiller and Dr. Lutz Kampmann I owe thanks for excellent VAX and UNIX computer support during this work. In addition I must give Dr. Lutz Kampmann my gratitude for allowing a degree of personal friendship to develop during my stay.