CRYSTALLOGRAPHIC STUDIES OF FOLYLPOLYGLUTAMATE SYNTHETASE AND RECOMBINANT HUMAN LACTOFERRIN

by

Xiaolin Sun

A dissertation submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in the

Department of Biochemistry

Institute of Molecular BioSciences

at

MASSEY UNIVERSITY, NEW ZEALAND

December, 1998

ABSTRACT

This thesis is written in two parts. In the first (chapters 1-4), crystallographic studies on the enzyme folylpolyglutamate synthetase from *Lactobacillus casei*, both in complex with $MgATP^{2-}$ and in its apo form, are presented. In the second part (chapters 5-8), a structural analysis of recombinant diferric human lactoferrin is reported.

Folylpolyglutamate synthetase (FPGS) is an ATP-dependent enzyme from eukaryotic and bacterial sources. It catalyzes the addition of glutamate residues to folate to produce folylpolyglutamates which are required for effective intracellular retention of folate and are the preferred substrates for the enzymes of one-carbon metabolism. The crystal structures of *L. casei* FPGS in both the $MgATP^{2-}$-bound and apo forms have been determined by the methods of multiple isomorphous replacement and molecular replacement, and refined by restrained least squares method using data to 2.4 Å resolution. The structural analysis of MgATP-FPGS reveals that folylpolyglutamate synthetase is a modular protein consisting of two domains, one with a typical mononucleotide-binding fold and the other strikingly similar to the folate-binding enzyme, dihydrofolate reductase (DHFR). The ATP-binding site is located in a interdomain cleft and a presumed mode of folate-binding has also been suggested for FPGS by analogy to the structure of DHFR. An unexpected structural similarity has been discovered between FPGS and the UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase (MurD). It is proposed that FPGS and MurD might carry out their biological functions in a very similar way and the structural comparison suggests that a possible domain movement could be involved in the catalytic reaction of FPGS. Two disordered loop regions in the MgATP-FPGS structure are well defined in apo-FPGS, allowing analysis of the interactions between these loops and surrounding structures of the protein.

Human lactoferrin (hLf) has considerable potential as a therapeutic agent. Over-expression of hLf in the fungus *Aspergillus awamori* has resulted in the availability of large quantities of this protein. Here the crystal structure of the recombinant human

lactoferrin (rhLf) has been determined by X-ray crystallography at 2.2 Å resolution. Superposition of the rhLf structure on to the native milk hLf shows a very high level of correspondence, and their dynamic properties, as indicated by the B factor distribution, also agree closely. This demonstrates that the structure of the protein is not affected by the mode of expression or the use of strain improvement procedures.

ACKNOWLEDGEMENTS

This thesis is not only the result of three and a half years work, but also a reflection of a great amount of help and support shown to me over these years by a large number of people.

In particular I thank my principal supervisor, Professor Ted Baker for giving me the chance to be his PhD student, for his guidance and kind advice throughout the years, and for his constructive criticism during the preparation of this thesis. His support is valued and much appreciated. I also thank Mrs. Heather Baker for her understanding and constant encouragement.

I am especially grateful to Dr. Clyde Smith, my associate supervisor, for his daily assistance throughout the course of my research, from crystallographic computing to making nice figures for the manuscript. He is always there wherever I need help. Thank you very much, Clyde.

My thanks also go to my other associate supervisor, Dr. Geoffrey Jameson, for his excellent advice and helpful discussions on structure refinement.

I would like to thank Mr. Steven Shewry for his help in my protein purification experiments, for his crystallization of recombinant human lactoferrin and allowing me to use his purification protocol.

I also would like to thank Dr. Bryan Anderson for his excellent advice on using computing programmes of crystallography. And I also thank Dr. Peter Metcalf for his assistance with computer and typing part of this thesis.

Thanks to the members of the laboratory of structural biology at both Massey and Auckland University for their help in many aspects. Stanley, Gillian, Jakki, Maria, Todd, Richard, Andrew McCarthy, Andrew Suther-land Smith, Allan, Ross, Neil, Bin and all others.

The work on the folylpolyglutamate synthetase was carried out in collaboration with Dr. Andrew Bognar (University of Toronto, Canada), I thank him for kindly providing the protein samples and useful result of biochemical experiments.

I thank Massey University for the award of a doctoral scholarship which enabled me to carry out this work.

Finally, I am indebted to my family for their love, support and encouragement all the time.

TABLE OF CONTENTS

Chapter 2

MATERIALS AND METHODS

Chapter 3

STRUCTURE AND FUNCTION OF FPGS

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

ABBREVIATIONS

ADK adenylate kinase

ADP adenosine-5'-diphosphate

ALL acute lymphoblastic leukaemia

AMPCPP α,β-methylene-adenosine-5'-triphosphate

AMPPCP β,γ-methylene-adenosine-5'-triphosphate

AMPPNP β,γ-imido-adenosine-5'-triphosphate

ATP adenosine-5'-triphosphate

CHO Chinese hamster ovary

CTP cytidine-5'-triphosphate

DDATHF 5,10-dideaza-5,6,7,8-tetrahydrofolate

DHFS dihydrofolate synthetase

DHFR dihydrofolate reductase

DMSO dimethyl sulfoxide

DNA Deoxyribonucleic acid

cDNA complementary DNA

FeNTA ferric nitrilotriacetate

EF-Tu elongation factor Tu

EXAFS Extended X-ray absorption fine structure

FOM figure of merit

FPGS folylpoly-γ-glutamate synthetase

FPLC fast performance liquid chromatography

GARFT glycinamide ribonucleotide formyltransferase

GTP guanosine-5'-triphosphate

HEPES N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]

HPLC high pressure liquid chromatography

H_2PteGlu 7,8-dihydrofolate

H_4PteGlu 5,6,7,8-tetrahydrofolate

hLf diferric human lactoferrin

rhLf recombinant diferric human lactoferrin

IgG Immunoglobulin G

K_m Michaelis constant

MIR multiple isomorphous replacement

MurD UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase

PEG4000 polyethylene glycol 4000

PMN polymorphonuclear leucocytes

PteGlu pteroylmonoglutamic acid (folic acid)

$PteGlu_n$ folylplolyglutamate

rms root mean square

SDS PAGE sodium dodecyl sulphate-polyacrylamide gel electrophoresis

TEMED NNN'N'-Tetramethylethylenediamine

TMV tobacco mosaic virus

Tris Tris(hydroxymethyl)aminomethane

TS thymidylate synthase

UDP uridine-5'-diphosphate

UK uridylate kinase

UMA UDP-N-acetylmuramoyl-L-alanine

UTP uridine-5'-triphosphate

V_{max} maximal velocity

RELATED PUBLICATIONS

Some of the material presented in this thesis has already been published, or has been accepted for publication.

Sun, X., Bognar, A. L., Baker, E. N., and Smith, C. A. (1998). Structural homologies with ATP- and folate-binding enzymes in the crystal structure of folylpolyglutamate synthetase. ***Proc. Natl. Acad. Sci. USA*** **95**, 6647-6652.

Sun, X., Baker, H. M., Shewry, S. C., Jameson, G. B., and Baker, E. N. Crystal structure of recombinant human lactoferrin, expressed in *Aspergillus awamori*. ***Acta Cryst.*** in press.

PART A

CRYSTALLOGRAPHIC STUDIES OF FOLYLPOLYGLUTAMATE SYNTHETASE FROM *LACTOBACILLUS CASEI*

INTRODUCTION

1.1 Structure and natural occurrence of folates

The folates are a family of vitamins that serve as active coenzymes in a variety of reactions in both bacterial and mammalian tissues. In these reactions, known as one-carbon metabolism, the folate coenzymes act as acceptors or donors of one-carbon units which are attached to coenzymatically active tetrahydrofolate and are utilized in the biosynthesis of various cellular components, including purine, thymidine, and amino acids. It has been known for many years that folates exist naturally primarily as polyglutamate derivatives with the glutamate moieties linked via γ-carboxyl peptide bonds. The naturally occurring pteroylpolyglutamates constitute a large family of closely related compounds arising from modifications of three parts of the parent compound, pteroylglutamatic acid, which is depicted in Fig. 1.1-I.

The chemical modifications include three states of reduction of the pteridine moiety (R), various kinds of one-carbon substituents (R1), and the elongation of the glutamate chain (R2). For the purpose of simplicity, the name "polyglutamates" will be used in this thesis to refer to folyl-(γ-glutamyl)$_n$ -glutamate where the chain lengths, the one-carbon substituent, and the state of reduction of the pyrazine ring are not specified; the abbreviation PteGlu refers to pteroylmonoglutamic acid or folic acid, H_2PteGlu and H_4PteGlu to 7,8-dihydro- and 5,6,7,8-tetrahydro-PteGlu, respectively; and 5-methyl, 10-formyl, 5,10-methylene etc., preceding the abbreviation, refer to the nature and position of the one-carbon substituent. A subscript, as in PteGlu$_n$, denotes the number of glutamic acid residues in γ-peptide linkage.

As predominant intracellular forms of folate coenzymes in all cells, polyglutamate derivatives play an important role in the regulation of folate homeostasis and are absolutely required for the normal functioning of the metabolic cycles of one-carbon metabolism. Folylpolyglutamates are the physiological substrates for the enzymes of one-carbon metabolism and are potential regulators of different cycles of one-carbon metabolism. The polyglutamate chain is required for normal cellular retention of folates, partially because the long anionic chain of folylpolyglutamates limits their ability to diffuse through membranes (Shane and Stokstad, 1976). In contrast, the

folylmonoglutamates are the transport form of folates, since they are effectively transported across the cell membrane (Goldman, 1971).

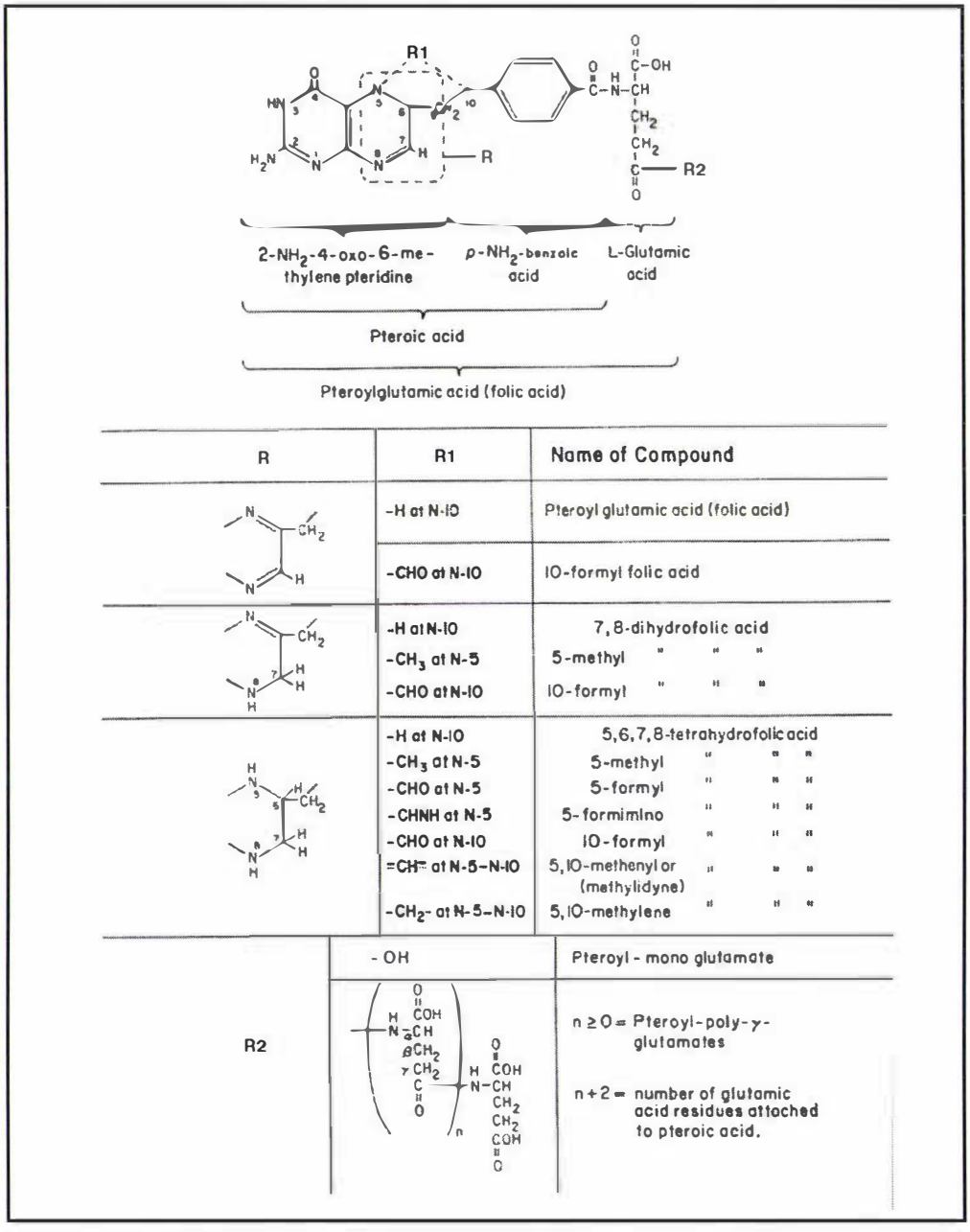


Fig. 1.1-I Structure and nomenclature of the naturally occurring folates. Taken from Krumdieck et al. (1983).

In mammalian organisms, the major pathway for utilization of folate as a nutritional requirement involves absorption of the monoglutamate form across the intestinal mucosa into the blood stream. Nutritional sources of folylpolyglutamates need to be

converted first to the monoglutamate form by γ-glutamyl hydrolase before they are transported across the membrane (McGuire and Coward, 1984). Once inside the cell, the glutamate chain length is extended through an ATP-dependent reaction. In both bacterial and mammalian cells, this polyglutamylation reaction is catalyzed by a single enzyme, folylpolyglutamate synthetase which is the subject of this thesis (Cichowicz et al., 1981). In general, the length of chain varies from 2 to 8 glutamates in total under normal conditions, although some derivatives contain up to 12 glutamates (Nakamura and Kozloff, 1978). There is generally a distribution of chain lengths, with a predominant form which varies between tissues and between species. *E. coli* contains predominantly pentaglutamates with significant amounts of tetra- and hexaglutamate. Yeast contains predominantly heptaglutamates with lesser amounts of hexa- and octa-glutamates. Mammalian cells contain pentaglutamates as the predominant intracellular folate.

The importance of intracellular folylpolyglutamate formation was first clearly demonstrated in cultured Chinese hamster ovary (CHO) cell mutants which lack folylpolyglutamate synthetase activity (Taylor and Hanna, 1977). Although folate transport by these cells is unimpaired, the mutant cells contained low intracellular folate levels and they were folylmonoglutamates due to an inability to synthesize folylpolyglutamates, The mutant cells required exogenous methionine, glycine, purine, and thymidine for growth, whereas the wild type grew in the absence of these compounds provided sufficient folate, vitamin B_{12}, and homocysteine were supplied in the medium (McBurney and Whitmore, 1974; Taylor and Hanna, 1975). Normalizing intracellular folate levels in the mutant cells, by increasing the medium folate content, did not eliminate the mutant phenotype, further indicating that the phenotype was due to a lack of folylpolyglutamates rather than low intracellular folate levels.

1.2 Physiological functions of folylpolyglutamates

The observation that all organisms have the capacity to synthesize folylpolyglutamates, including organisms which are unable to biosynthesize folic acid, suggests that folylpolyglutamates are vital to cellular metabolism (McGuire and Bertino, 1981). Since the discovery that cells contain only the polyglutamate forms of the coenzyme, intensive studies have been carried out to determine the physiological roles

of folylpolyglutamate, and hence why the cell expends energy to add the additional glutamate residues. So far, several proposed physiological functions for the folylpolyglutamates have been generally accepted. Firstly, folylpolyglutamates are often more effective substrates or inhibitors for the enzymes in one-carbon metabolism, and in many cases are the true substrates or inhibitors. Moreover, the observation that a polyglutamate substrate for one metabolic cycle can be an effective inhibitor for another metabolic cycle means that the relative levels of polyglutamate substrates and inhibitors could regulate folate enzymes, and the flux of one-carbon units, through different metabolic pathways (Shane, 1989). Secondly, folylpolyglutamate substrates can be channelled from one active site to another in a multifunctional enzyme without release of the intermediate products, a phenomenon that is not observed with monoglutamate substrates (Schirch and Strong, 1989). Thirdly, polyglutamylation of folates results in forms which are not transported across the cell membrane, thus helping in the retention of the folates inside cells (McGuire and Bertino, 1981).

1.2.1 Folylpolyglutamates as substrates and inhibitors for folate-dependent enzymes in one-carbon metabolism

Many studies both *in vitro* and *in vivo* demonstrate that folylpolyglutamates are the preferred substrates or inhibitors for folate dependent enzymes. As noted above, polyglutamates are in most cases much better substrates the monoglutamates. The effectiveness of polyglutamate substrates varies with the chain length, the highest activities generally being achieved with the length of polyglutamate which predominates in the organism used as the enzyme source (McGuire and Bertino, 1981). The various reactions of one-carbon metabolism in the cytoplasm are shown in Fig. 1.2.1-I. These reactions form key steps in various metabolic cycles including the cycles of methionine, thymidylate and purine syntheses. It is obvious from these that the physiologically active forms of the folate coenzyme are the tetrahydrofolylpolyglutamates ($H_4PteGlu_n$). The species $H_4PteGlu_n$ serves to harbour various one-carbon units which are bound to it at either the N-5 or N-10 positions of the pteridine and the p-aminobenzoate groups, or bridging these nitrogen atoms (cf. Fig. 1.1-I). The coenzymes also act as cosubstrates in these reactions. Consequently, folate metabolism and its regulation are interwoven with the regulation of the synthesis of products of one-carbon metabolism, and factors that regulate any one cycle of one-carbon metabolism

would be expected to influence folate availability for the other cycles of one-carbon metabolism. Various $H_4PteGlu_n$ derivatives serve as a one-carbon donor for the biosynthesis of amino acids, thymidylate and purine (Blakley, 1969). The key intermediate in these reactions is 5,10-methylene-$H_4PteGlu_n$. This important one-carbon derivative is metabolized in three different metabolic pathways; reduction to 5-methyl-$H_4PteGlu_n$ for methionine biosynthesis, reduction and transfer to deoxyuridylate to form thymidylate, and oxidation to 10-formyl-$H_4PteGlu_n$ for purine ring biosynthesis. Each of these pathways regenerates $H_4PteGlu_n$, and it is converted back to 5,10-methylene-$H_4PteGlu_n$ by the conversion of serine to glycine catalyzed by serine hydroxymethyltransferase, which is the major route for one-carbon entry into the folate pool in cells. Two other minor sources of one-carbon groups are formate and the catabolism of histidine. These enter the one-carbon pool in the cycle utilizing formyl-$H_4PteGlu_n$.

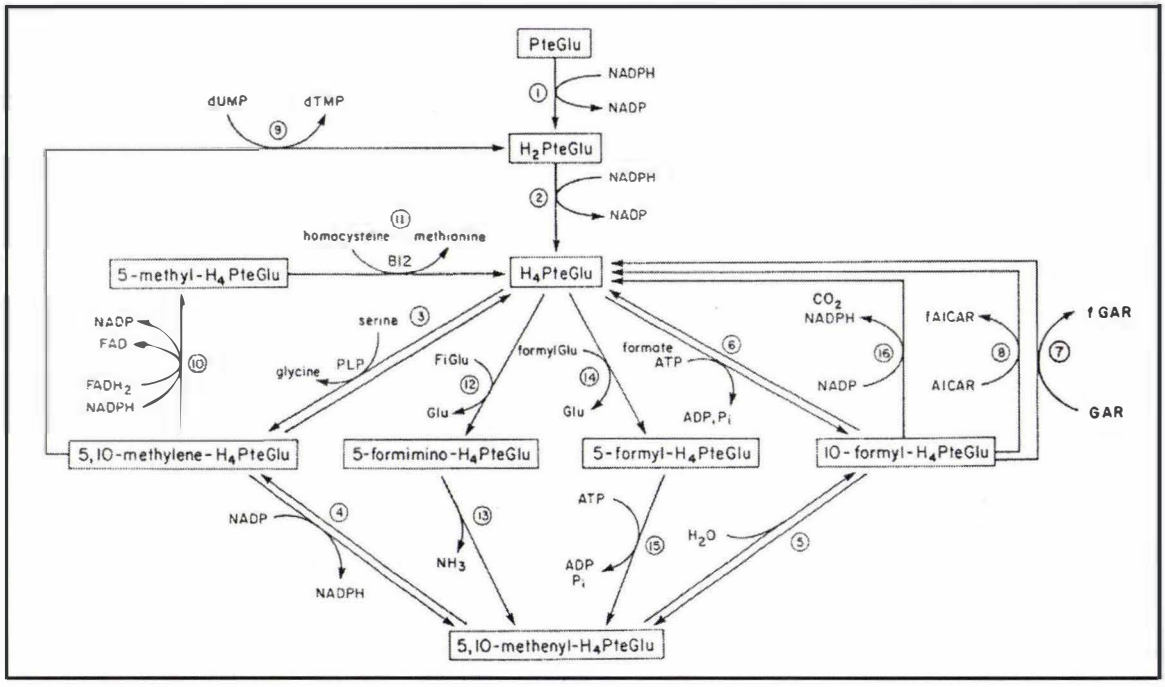


Fig. 1.2.1-I Cytoplasmic pathways of one-carbon metabolism. Numbers refer to the following enzymes: 1 and 2, dihydrofolate reductase; 3, serine hydroxymethyltransferase; 4, 5,10-methylenetetrahydrofolate dehydrogenase; 5, 5,10-methenyltetrahydrofolate cyclohydrolase; 6, formyltetrahydrofolate synthetase; 7, glycinamide ribonucleotide transformylase; 8, 5-amino-4-imidazolecarboxamide ribonucleotide transformylase; 9, thymidylate synthase; 10, flavoprotein methylenetetrahydrofolate reductase; 11, methionine synthase; 12, formiminoglutamate formiminotransferase; 13, formiminotetrahydrofolate cyclodeaminase; 14, N-formylglutamate transformylase; 15, methenyl-tetrahydrofolate synthetase; 16, 10-formyl-tetrahydrofolate dehydrogenase.

Serine-glycine interconversion is one of the essential amino acid metabolic cycles, by which the β-carbon from serine is transferred to $H_4PteGlu_n$ to generate 5,10-methylene-$H_4PteGlu_n$ and glycine (Fig. 1.2.1-I, reaction 3). This reaction is catalyzed by serine hydroxymethyltransferase, a key enzyme in the generation of one-carbon units. The enzyme shows a strong preference for utilization of the polyglutamate substrate, $H_4PteGlu_3$, which is a much more effective substrate than $H_4PteGlu_1$ and $H_4PteGlu_2$ (Matthews et al., 1982). In addition, the potent inhibition of this enzyme by 5-methyl-$H_4PteGlu_n$, the folate substrate for methionine synthesis (see below), suggests a mechanism for regulating the flow of one-carbon units into the folate pool. 5-Methyl-$H_4PteGlu_n$ accumulates in tissues when the vitamin B_{12}-dependent methionine synthesis is inhibited (Matthews, 1984), which should cause increased inhibition of serine hydroxymethyltransferase. Under these conditions, all other metabolic cycles of one-carbon metabolism, including thymidylate synthesis, are depressed. A major contributing factor to this derangement of one-carbon metabolism may be the decreased one-carbon flux into the folate pool caused by high levels of 5-methyl-$H_4PteGlu_n$ (Matthews et al., 1982).

The methionine cycle, a major cycle of one-carbon utilization, involves the reduction of 5,10-methylene-$H_4PteGlu_n$ to 5-methyl-$H_4PteGlu_n$ (Fig. 1.2.1-I, reaction 10) followed by the transfer of the methyl group to homocysteine to form methionine and to regenerate $H_4PteGlu_n$ (Fig. 1.2.1-I, reaction 11). 5,10-Methylene-$H_4PteGlu_n$ reduction is catalyzed by methylenetetrahydrofolate reductase, and methionine synthase catalyzes the following methionine synthesis. The methionine synthase reaction is the sole reaction whereby the methyl group of 5-methyl-$H_4PteGlu_n$ can be metabolized in mammalian cells. Under conditions of vitamin B_{12} deprivation, loss of the B_{12}-dependent enzyme results in the trapping of folate in the 5-methyl-$H_4PteGlu_n$ form (Shane and Stokstad, 1985) and increased levels of homocysteine, which has been implicated in heart disease and some forms of neural disorders (Mills et al., 1996; Verhoef et al., 1998). Both methylentetrahydrofolate reductase and methionine synthase from bacterial and mammalian cells show a very marked preference for folylpolyglutamate substrates and inhibitors.

Substantial decreases in K_m values (Michaelis constants) are observed with increasing polyglutamate chain length for almost all enzymes, indicating they are more efficient substrates. Pig liver methylenetetrahydrofolate reductase, for instance, displays

a seventy-fold decrease in K_m for the 5,10-methylene-H_4PteGlu$_6$ substrate compared to the monoglutamate, although V_{max} values remained relatively constant (Matthews and Baugh, 1980). With pig liver methionine synthase, there is a high degree of specificity for the 5-methyl-H_4PteGlu$_6$ substrate. This high degree of selectivity for the hexaglutamate substrate has important physiological implications.

5-Methyl-H_4PteGlu$_1$ is the major circulating form of folate and is the major form of folate taken up by tissues. Removal of the methyl group, via the methionine synthase reactions, is required before the folate can be utilized in other reactions of one-carbon metabolism. Methyl folates normally account for a significant proportion of the folate one-carbon species in the cell. These data suggest that incorporation of exogenous folate by tissues is repressed by physiological concentrations of 5-methyl-H_4PteGlu$_n$ polyglutamates. Conditions that result in reduction of the 5-methyl-H_4PteGlu$_n$ pool, such as inhibition of methylenetetrahydrofolate reductase or folate depletion, should stimulate the accumulation of exogenous folate by tissues. Expansion of the 5-methyl-H_4PteGlu$_n$ pool, caused for example by vitamin B_{12} deficiency, would be expected to resist the accumulation of folate by tissues (Matthews et al., 1987). It has been reported that the vitamin B_{12} independent methionine synthases from various sources, such as *Bacillus subtilis* (Salem et al., 1972) and *Candida utilis* (Salem and Foster, 1972), all require a triglutamate for activity as well. It has been shown that H_2PteGlu is an inhibitor of the pig liver methylenetetrahydrofolate reductase (Matthews and Haywood, 1979). The inhibition increases with polyglutamate length up to hexaglutamate, which is 500-fold more inhibitory than the monoglutamate. Since the Ki of the hexaglutamate (13 nM) is well within the physiological range of H_2PteGlu$_n$ concentration, the one-carbon flux through the reductase reaction should be regulated by the relative amounts of the substrate 5,10-methylene-H_4PteGlu$_n$ and the inhibitor H_2PteGlu$_n$ (Matthews and Baugh, 1980). In mammalian cells, 5,10-methylene-H_4PteGlu$_n$ can be used for either methionine or thymidylate biosynthesis. When DNA synthesis occurs and the demand for thymidylate is high in growing cells, the rapid flux through thymidylate synthase could cause a rise in the levels of H_2PteGlu$_n$ which would be expected to decrease the one-carbon flux into methionine synthesis. This could be a mechanism to conserve one-carbon moieties for thymidylate and purine biosynthesis.

Thymidylate synthase catalyzes the transfer of formaldehyde from 5,10-methyl-H_4PteGlu$_n$ to the 5-position of deoxyuridylate to form thymidylate and reduced

$H_2PteGlu_n$ (Fig. 1.2.1-I reaction 9). The increased affinity of polyglutamates for thymidylate synthases from different biological sources suggests that a polyglutamate substrate would be preferentially utilized under conditions in which both a mono- and a polyglutamate substrate were available for the synthetase. It has been shown that the *Lactobacillus casei* thymidylate synthase will preferentially use the hexaglutamate substrate over the monoglutamate (Kisliuk et al., 1981). Studies with human thymidylate synthases also reveal that the K_m for 5,10-methylene-$H_4PteGlu_5$ substrate is considerably lower than with the monoglutamate substrate, while V_{max} is increased slightly (Dolnick and Cheng, 1978). Many folylpolyglutamates have been tested as thymidylate synthase inhibitors. Inhibition by 5-formyl-$H_4PteGlu_n$ was shown to increase 200-fold as the chain length increased up to five (Friedkin et al., 1975). Methotrexate polyglutamates are also potent inhibitors of human thymidylate synthase, marked increases in affinity being observed for pentaglutamate compared to monoglutamate substrate (Allegra et al., 1985).

It is clear from several major one-carbon metabolic cycles mentioned above that in every case, a polyglutamate is a better substrate than monoglutamate, having either a lower K_m, a higher V_{max}, or both. Folylpolyglutamates are quantitatively the most important substrates and are, in fact, essential for cellular one-carbon metabolism. In addition, polyglutamates as a product or substrate of a one-carbon metabolic enzyme can be inhibitors for other one-carbon metabolic enzymes. The concentration at which significant inhibition occurs is often within the physiological range of the polyglutamates. This indicates that one-carbon metabolism could be regulated, at least in part, by the relative amounts of the substrates and the inhibitors. The mechanisms by which polyglutamates are better utilized as substrates are varied. The polyglutamate chain is often used only for increased binding as evidenced by a decreased K_m with no difference in V_{max} (Matthews and Baugh, 1980). In those reactions where either both the K_m and the V_{max} are changed or just the V_{max} is changed (Matthews et al., 1987), the polyglutamate chain could also influence the active site and kinetic mechanism.

1.2.2 Multifunctional complexes and channelling of folylpolyglutamates

In addition to increasing the affinity of folate substrates or inhibitors for the one-carbon metabolic enzymes, folylpolyglutamates can also influence the activity of

enzymes by channelling preferentially through multifunctional proteins. Multifunctional proteins generally consist of a single polypeptide chain, which carries multiple catalytic or binding activities, each activity in the protein being co-ordinately regulated by a single substrate. If two sequential metabolic reactions are involved in this protein, an intermediate could be transferred directly between active sites without release from the protein. This "channelling" may allow rapid establishment of a steady state and increase the effective concentration of the substrate for the enzyme (Gaertner, 1978).

Two multifunctional proteins have been identified in mammalian one-carbon metabolism. The sequential activities of formiminoglutamate formiminotransferase (Fig. 1.2.1-I reaction 12) and 5-formiminotetrahydrofolate cyclodeaminase (Fig. 1.2.1-I, reaction 13) are associated with a single bifunctional protein (Drury et al., 1975). Experimental data suggest that both enzyme activities show a strong preference for folylpolyglutamate substrates and the bifunctional protein may contain one polyglutamate chain-binding site per pair of formiminotransferase-cyclodeaminase sites (MacKenzie et al., 1980). With the monoglutamate substrate, $H_4PteGlu_1$, the protein does not retain the intermediate (5-formimino-H_4PteGlu), and it is detectable in the medium. With $H_4PteGlu_3$ as the substrate, some intermediate appears in the medium, demonstrating only partial channelling. With $H_4PteGlu_5$ as substrate, essentially no intermediate is released, demonstrating almost perfect channelling with this chain length (Mackenzie and Baugh, 1980; Paquin et al., 1985). The experiments suggest that a single-site anchor for the polyglutamate tail on the protein contributes to the transfer of the product of the formiminotransferase reaction to the active site of cyclodeaminase. The pentaglutamate represents the optimal chain length, while slightly shorter or longer chain lengths allow channelling but also allowing some dissociation of the intermediate.

Channeling of folates has also been demonstrated to occur between two of the three enzyme activities of C_1 synthase, a single chain trifunctional protein on which the dehydrogenase, cyclohydrolase, and synthase are associated. 5,10-Methylene-$H_4PteGlu_n$ can be catalytically converted to 5,10-methenyl-$H_4PteGlu_n$ by methylenetetrahydrofolate dehydrogenase (Fig. 1.2.1-I, reaction 4). The latter product is then transformed by methenyltetrahydrofolate cyclohydrolase (Fig. 1.2.1-I, reaction 5) to 10-formyl-$H_4PteGlu_n$, a one-carbon derivative utilized in *de novo* purine biosynthesis. Alternatively, 10-formyl-$H_4PteGlu_n$ can be obtained by the direct formylation of $H_4PteGlu_n$, catalyzed by formyltetrahydrofolate synthetase (Fig. 1.2.1-I, reaction 6). The

C_1 synthase consists of two domains, the smaller one containing the dehydrogenase and cyclohydrolase activities and the larger one containing the synthase activity (Villar et al., 1985). Channelling occurs between the dehydrogenase and cyclohydrolase activities on the small domain. Both monoglutamate and polyglutamate substrates show partial channelling through the dehydrogenase-cyclohydrolase active sites, although channelling of polyglutamates appears to be slightly more efficient than for the monoglutamate substrate (Mackenzie and Baugh, 1980). However, this incomplete channelling might be vital because 5,10-methenyl-H_4PteGlu$_n$ is required for *de novo* purine biosynthesis (McGuire and Bertino, 1981).

Channelling could be an important kinetic property in one-carbon metabolism. It may reflect the specificity and efficiency of a multi-functional protein, and may maintain low concentrations of the labile free coenzyme (Schirch and Strong, 1989). Folylpolyglutamates, especially in the case of formiminotransferase-cyclodeaminase, appear to be essential for channelling of a metabolite from one active site to another in a multifunctional enzyme. The polyglutamate chain plays an important role during the channelling process.

1.2.3 Folylpolyglutamates and intracellular retention of folate

Many experiments, with various organisms, have shown that the folylmonoglutamates are most effectively transported across the cell membrane, but that polyglutamate forms of folate have greatly reduced permeability through the cell membrane and hence polyglutamylation of folates serves to trap or retain folates intracellularly (Goldman, 1971; Hoffbrand et al., 1973; Nahas et al., 1972). Further experiments demonstrate that the cells only transport folates which are smaller than those that predominate within that cell type, and there might be a relationship between the retaining ability and polyglutamate chain length. For example, intracellular folates of *L. casei* have longer chains than folates in the cell growth medium, although both are polyglutamates. This suggests that shorter polyglutamates are preferentially released and longer polyglutamates are preferentially retained, with the degree of retention being directly related to chain length (Buehring et al., 1974; Shane and Stokstad, 1975). *Corynebacterium sp.* growth medium contains a high proportion of triglutamates, while tetraglutamates predominate intracellularly, again suggesting preferential retention of

longer derivatives (Shane, 1980a). Monoglutamates are readily transported into and out of mammalian cells (Nahas et al., 1972). However the permeability of hepatoma cells to polyglutamates of methotrexate, a folate analogue, is extremely low even when the total intracellular methotrexate concentration is a hundred-fold greater than the protein-bound fraction (Galivan, 1979). The long chain of the folylpolyglutamates is obviously essential for the retention of folates and their one-carbon derivatives inside the cells, although the exact effect of the long chain in transport remains to be investigated. The highly anionic nature of the long chain of folylpolyglutamates is believed to limit their ability to diffuse through membranes. Moreover, folylpolyglutamates may also be involved in binding with membrane proteins, which could also contribute to their retention in cells (Shane, 1989).

Although the evidence clearly supports intracellular retention of polyglutamates, different folylpolyglutamtates predominate in different tissues and cells. As noted earlier, *E. coli* contains predominantly pentaglutamates, yeast contains predominantly heptaglutamates and mammalian cells contain pentaglutamates as the predominant intracellular folate. Considerable research has been focused on the distribution of folylpolyglutamates in a variety of bacterial and mammalian cells, as well as possible reasons for the varied distributions of chain length of the polyglutamates in different organisms (Cichowicz et al., 1981). Firstly, some studies suggested that the distribution of the glutamate chain length can be affected by the nutritional state of the cells (Foo and Shane, 1982; McGuire and Coward, 1984). For example, the proportion of total folates in rat liver occurring as long chain polyglutamates (measured as hexa- to octaglutamates) is greatest in folate-deficient rats (Cassady et al., 1980). The length of the chains is inversely related to the total concentration of folates in the liver, suggesting that these concentrations regulate the lengths of the polyglutamate chains.

On the other hand, data on the predominant lengths that occur in a variety of biological sources and on influx and efflux of polyglutamates, suggest that cells themselves have evolved polyglutamylation mechanisms that will synthesize lengths just longer than that which will pass the cellular membrane by either active transport or diffusion (McGuire and Bertino, 1981). For instance, mammalian cells have pentaglutamates as the predominant lengths because triglutamates can pass the membrane, albeit in very low amount (Hoffbrand et al., 1973). In *L. casei*, derivatives up to heptaglutamate still weakly cross the cell membrane, and *L. casei* cells

correspondingly contain predominantly octaglutamates. Only diglutamates or shorter derivatives can cross the *S. faecalis* cell wall (Baugh et al., 1970) and its intracellular folates are mainly tetraglutamates. In this way, cells can retain the necessary levels of folates for cellular growth while avoiding synthesis of excessively long polyglutamates.

This natural selection of polyglutamate length of cells is also reflected in different substrate specificities for the folylpolyglutamate synthetases from different biological sources. In other words, the ability of different chain length folylpolyglutamates to act as substrates for the synthetases from different sources could be related to the distribution of folylpolyglutamates in different biological samples (Cichowicz et al., 1981). Comparison of the distribution of folylpolyglutamates with the polyglutamate substrate specificity of enzymes in different cells reveals that the predominant folylpolyglutamates accumulated *in vivo* are all poor substrates for the organism's folylpolyglutamate synthetase, while shorter glutamate chain length folates which can not be retained effectively *in vivo* are effective substrates for the enzyme (Cook et al., 1987).

It is clear from the above observations that the final distribution of folylpolyglutamate chain length in cells appears to be a function of the properties of the cell, including membrane permeability to polyglutamates and substrate specificity of the folylpolyglutamate synthetase. For a particular cell, the substrate specificity is highly dependent on its membrane permeability to polyglutamates of various chain lengths, such that shorter and more readily transported polyglutamates are all effective substrates and their chain lengths become extended to derivatives that can be retained by the cell. As a special case, folylmonoglutamates are most effectively transported across various cell membranes, and they are generally the best substrates for the folylpolyglutamate synthetases and are readily converted into polyglutamates.

1.3 Biosynthesis of folylpolyglutamates

1.3.1. Folylpolyglutamate synthetase and its occurrence

As discussed above, folylpolyglutamates are the major natural forms of folate found intracellularly. In both bacterial and mammalian cells, the sequential γ-polyglutamylation of folate or its C_1 derivatives is carried out by the following reaction:

$$H_4PteGlu_n + \text{L-glutamate} + MgATP \Rightarrow H_4PteGlu_{n+1} + MgADP + Pi$$

This reaction is catalyzed by the enzyme folylpolyglutamate synthetase (FPGS). Since the chain lengths of the intracellular polyglutamates always exceed the longest form that can be transported efficiently, it follows that all cells must have their own folylpolyglutamate synthetases for synthesis of polyglutamates with a particular chain length distribution. It has been demonstrated by various investigations (Cichowicz et al., 1981; Shane, 1980b; Taylor and Hanna, 1979) that a single enzyme is responsible for synthesis of all polyglutamate chain lengths in both bacterial and mammalian cells, although a few experimental results appear to be in contradiction. More than one enzyme was once thought to be present in *E. coli* because the isolated FPGS only synthesized diglutamates from its preferred monoglutamate substrate, 10-formal-H_4PteGlu, *in vitro*, whereas *E. coli* cells contain predominantly pentaglutamate derivatives (Masurekar and Brown, 1975). Likewise, purified *L. casei* FPGS metabolized its preferred substrate only to a tetraglutamate derivative, compared with the octa- and nonaglutamates normally found *in vivo* (Bognar and Shane, 1983). However, it is likely that the synthesis of longer polyglutamates by isolated enzymes could not be detected given that the reaction conditions, such as pH and type or concentration of folate, were not optimal (McGuire and Bertino, 1981). In addition, the enzyme from *E. coli* may utilize 10-formyl-$H_4PteGlu_1$ to form diglutamates, but utilize $H_4PteGlu_2$ (or other diglutamate folates) produced in the course of metabolism to form longer polyglutamtes. Thus, the failure to detect folates with chain lengths greater than two for the isolated *E. coli* FPGS was due to the change in one-carbon form of the folate substrates required for further polyglutamylation of the glutamate chain (Cichowicz et al., 1981).

FPGS is generally present in relatively low amounts in all biological sources. The tissue distribution of mammalian FPGS has been determined using assay conditions optimal for the rat liver FPGS (McGuire et al., 1979). The highest specific activity was found in the liver, a tissue containing predominantly folylpentaglutamates and the highest amount of total folate. FPGS activity was also detectable in most mammalian tissues such as spleen, stomach etc., but heart and muscle tissue, which contain very low levels of folates, have only negligible activity (Moran and Colman, 1984b). In regard to

the subcellular localization within mammalian liver, most synthetase activity is present in the cytoplasm of cells, although some activity is found in the mitochondrial fraction.

As shown in Table 1.3.1-I, a number of bacterial and mammalian FPGS have so far been purified and characterized, although the marked lability of the protein made these studies difficult. A cDNA encoding human FPGS has been cloned and sequenced and the protein has been overexpressed in *E. coli* (Atkinson et al., 1997; Chen et al., 1996; Garrow et al., 1992); the *L. casei* FPGS gene has also been cloned and overexpressed in *E. coli* (Toy and Bognar, 1990). The gene encoding the *E. coli* FPGS has been cloned and the protein expressed at high levels (Bognar et al., 1987; Bognar et al., 1985). In addition, a cDNA encoding murine FPGS has been cloned recently (Spinella et al., 1996).

Table 1.3.1-I FPGS fully or partially purified and characterized to date

Proteins	References	Proteins	References
Corynebacterium sp	(a)	Rat liver	(i)
L.casei	(b)	Mouse liver	(j)
L. casei recombinant	(c)	Beef liver	(k)
E. coli	(d)	Pig liver	(l)
S. faecalis	(e)	Sheep liver	(m)
N. crassa	(f)	Human liver	(n)
N. gonorrhoeae	(g)	Human recombinant	(o)
Chinese hamster ovary cell	(h)	Pea seeds	(p)

References cited: (a) Shane (1980b); (b) Bognar and Shane (1983); (c) Toy and Bognar (1990); (d)Bognar et al. (1985); (e) Cichowicz et al. (1981); (f) Sakami et al. (1973); (g) Fussenegger and Meyer (1996); (h) Taylor and Hanna (1977); (i) McGuire et al. (1980); (j) Moran and Colman (1984a); (k) Pristupa et al. (1984); (l) Cichowicz and Shane (1987a); (m) Gawthorne and Smith (1973); (n) Clarke and Waxman (1987); (o) Garrow et al. (1992); (p) Imeson and Cossins (1991).

1.3.2 General molecular properties

General characteristics of folylpolyglutamate synthetases from various biological sources have been reported. All human and bacterial enzymes are found as monomeric

Table 1.3.2-I Properties of some selected folylpolyglutamate synthetases[a]

Source	*Corynebacterium*	*L. casei*	*E. coli*	Pig liver	Rat liver	Human
Purification (fold)	7,000	40,000-200,000	93	180,000	–	610
Mr	53,000	43,000	47,000	59,000	69,000	60,000
pH optimum	9.5	9.5	9.4	9.5	8.4	9.6
Monovalent cation	200 mM K^+	200 mM K^+	200 mM K^+	20 mM K^+	20 mM K^+	20 mM K^+
Divalent cation	Mg^{2+}	Mg^{2+}	Mg^{2+}	Mg^{2+}	Mg^{2+}	Mg^{2+}
Reducing agent	None	None	None	Yes	Yes	Yes
DHFS activity[b]	Yes	No	Yes	No	No	No
Monoglutamate substrate	H_4PteGlu	5, 10-methylene-H_4PteGlu	10-formyl-H_4PteGlu	H_4PteGlu, H_2PteGlu	H_4PteGlu, H_2PteGlu	H_4PteGlu, H_2PteGlu
Polyglutamate substrate	5, 10-methylene-H_4PteGlu$_n$	5,10-methylene-H_4PteGlu$_n$	–	H_4PteGlu$_n$	H_4PteGlu$_n$	H_4PteGlu$_n$
Stabilizers	K^+, Pi, Me_2SO	K^+, Pi, Me_2SO	K^+, Pi	Protease inhibitors, Pi, ethylene glycol, mercaptoethanol	–	ethylene glycol, mercaptoethanol

a. Properties listed here are adapted from references: Bognar and Shane (1986); Shane et al. (1993); Cichowicz et al. (1981); McGuire and Bertino (1981); Cichowicz and Shane (1987); some properties not available are marked as "-".
b. Dihydrofolate synthetase activity.

proteins. Enzyme activity is in each case dependent upon the presence of a suitable folate substrate, MgATP, L-glutamate, K^+ and a reducing agent in the case of the mammalian enzyme. Some properties of selected enzymes are listed in Table 1.3.2-I.

a. Instability of FPGS

A feature of the bacterial and mammalian FPGS family is the low abundance in cells, and the instability of the purified enzymes. This has complicated the purification of sufficient quantities of the enzyme for detailed physical analysis, including mechanistic studies. In the past few years, cloning of the FPGS cDNA and subsequent overexpression has allowed the purification of the human (Chen et al., 1996), *L. casei* (Toy and Bognar, 1990) and *E. coli* (Bognar et al., 1985) enzymes to homogeneity, characterization of their properties, and stabilization during purification and storage. The lability of FPGS varies among enzymes depending on the source, concentration, and conditions of storage such as pH and temperature. For instance, recombinant *L. casei* FPGS, which is the subject of this thesis, is very unstable in the stock buffer solution (30 mM Hepes/KOH, 0.1 M $(NH_4)_2SO_4$, pH 7.2 at 4°C) used for crystallization trials. The protein degrades quickly, with several discrete bands at lower molecular weight relative to the 43 kDa band of intact enzyme appearing on sodium dodecyl sulphate gel electrophoresis (SDS PAGE) after about a week. These bands become stronger over time, indicating continued degradation. As an extreme case, the mouse synthetase appears to be the most unstable of all the FPGS studied, as crude extracts lost all activity at 4°C overnight and more highly purified preparations lost activity rapidly unless a stabilizer was added (Moran and Colman, 1984a).

A variety of stabilizers have been tested for FPGS from different sources (Atkinson et al., 1997; Bognar and Shane, 1983; Cichowicz and Shane, 1987a; Shane, 1980b). Some stabilizers can retard the loss of activity of some FPGS to a degree, but no stabilizer has yet been found to completely stop degradation. Each kind of FPGS has particular stabilisers. The bacterial FPGS enzymes are stabilized by high levels of KCl and phosphate. The latter is a product of the enzymatic reaction and consequently is an inhibitor of the reaction. Another more effective stabilizer for bacterial FPGS is dimethyl sulfoxide (DMSO); Around 20-30% (v/v) DMSO is generally used for the purification procedures and the long term storage of this enzyme at 4°C (Shane, 1980b).

Different stabilizers and conditions have been used for the stabilization of purified mammalian enzymes. As shown in Table 1.3.2-I, ethylene glycol (~10% v/v) and mercaptoethanol (50 mM), which increased the lability of the preparation of the enzyme from *Corynebacterium,* can stabilize enzyme activities of mammalian FPGS for several months. Proteinase inhibitors such as benzamidine, which was not useful for the *Corynebacterium* FPGS, are effective stabilizers for the pig liver enzyme (Cichowicz and Shane, 1987a), and some other mammalian FPGS.

Although the exact mechanism of degradation of FPGS is not clear, two putative causes for the lability of FPGS have been suggested. It was found during the purification of pig liver FPGS (Cichowicz and Shane, 1987a) that the enzyme stability was increased when the protease inhibitor benzamidine was added to the homogenization buffer. This suggested that during purification steps, significant proteolysis was occurring. The protease inhibitor treatment greatly stabilized the enzyme in comparison with a variety of initial chromatographic procedures, without added protease inhibitor, which resulted in total loss of enzyme activity. Therefore, proteolysis is probably one of the causes of lability of FPGS, although this has not been tested for FPGS from other sources. In addition, it appears likely that the low protein concentration was another major contributing factor to the instability observed, because highly purified preparations which contained low protein concentrations, were generally more unstable than the crude extract. This is true for both bacterial and mammalian FPGS. The dilute purified protein preparation was believed to bind to glass or plastic containers, resulting in instability (Bognar and Shane, 1986; Cichowicz and Shane, 1987a).

b. Monovalent cation requirement

Folylpolyglutamate synthetases all have an absolute requirement for a monovalent cation for their activity. The greatest activation is achieved with K^+, but Rb^+ and NH_4^+ are also activators, albeit less effective, while Li^+, Na^+ and Cs^+ are ineffective at any concentration. Monovalent cation activation is clearly restricted to ions of ionic radii 1.33-1.47 Å. Moreover, excessively high levels of all monovalent ions inhibit enzyme activity and the extent of inhibition is proportional to ionic radius, suggesting that the inhibition is due to an interaction at a separate site from the activation site (Cichowicz

and Shane, 1987a). Previous studies have suggested that the cation may be required at the catalytic site (Suelter, 1970), or to promote a specific conformational change required for catalysis (Evans and Sorger, 1966), or just for the structural integrity of the protein (Himes and Harmony, 1973). The mechanism of FPGS enzyme activation by monovalent cations is unknown at this stage. It has been reported (Shane, 1983) that K^+ protects the *Corynebacterium* enzyme from inactivation by a variety of proteases. This protection does not require the presence of substrates, suggesting that the monovalent cation may induce a conformational change in the protein, or may stabilize a particular conformation. However, further kinetic experiments and direct structural evidence, such as from the crystal structure of the protein, are necessary to further define the effect of the monovalent cation.

c. Divalent cation requirement

FPGS from all sources requires Mg^{2+} because the $MgATP^{2-}$ chelate is the true substrate for the reaction and free ATP^{4-} is a potent inhibitor. Thus, the ratio of Mg^{2+} to ATP^{4-} in the reaction mixture has a pronounced effect on enzyme activity. Activity increases with the concentration of $MgATP^{2-}$ chelate, but excess Mg^{2+} is inhibitory (Cichowicz and Shane, 1987a; Shane, 1980b). There is no evidence for a requirement for Mg^{2+} apart from that needed for the formation of the substrate $MgATP^{2-}$. Mn^{2+} is the only other divalent cation that can substitute for Mg^{2+} in *E. coli*, *L. casei* and pig liver FPGS, although it is less effective.

d. Effect of pH

As shown in Table 1.3.2-I, all FPGS display high pH optima, ranging from 8.4 to 9.6, and have low or non-existent activity near or below pH 7. Many different buffer solutions give full activity at the optimal pH (McGuire et al., 1980; Shane, 1980b; Taylor and Hanna, 1977), except that phosphate and borate buffers are inhibitory. Phosphate presumably is a product inhibitor, and borate probably complexes with the vicinal diol of ATP, lowering the effective ATP concentration.

The effect of pH on FPGS enzyme activity is directly related to the concentration of glutamate (Cichowicz and Shane, 1987a). At low glutamate concentrations, the K_m for

L-glutamate decreases quickly with increasing pH, and a very sharp alkaline pH optimum is observed. There is little activity at physiological pH (7.4). At high glutamate concentrations the same pH optimum is observed, but the variation with pH is much less, so that enzyme activity at physiological pH is only slightly decreased from that observed at optimal pH. The optimal pH values for all enzymes are similar to the pKa of the amino group of glutamate. This pH effect, and the ability of high glutamate concentrations to overcome it at physiological pH, suggest that the free amine of glutamate is the form of the substrate that binds to the enzyme. At physiological pH (7.4), activity is only slightly reduced provided sufficient unprotonated glutamate is available. It was reported for the pig liver FPGS (Cichowicz and Shane, 1987a) that the K_m for glutamate at pH7.5 (2.3 mM) is similar to the glutamate level in mammalian liver and is higher than the glutamate concentration in peripheral tissues (Krebs et al., 1976).

e. Substrate specificity

The FPGS from all sources requires a nucleotide triphosphate for activity, and MgATP or MgdATP generally give the highest activity, for both bacterial and mammalian enzymes. The substrate MgATP is converted to MgADP and Pi (phosphate), which are both product inhibitors. Other nucleotide triophosphates, such as UTP, GTP, CTP etc., can be utilized as substrates but exhibit lower affinity for the particular enzymes. It appears that the effectiveness of these nucleotides is defined primarily by their ability to bind to the nucleotide binding site in the enzymes. β,γ-Methylene-ATP (AMPPCP) is a very effective inhibitor of the reaction; it exhibits high affinity for the enzymes but lacks catalytic activity. In contrast, α,β-methylene-ATP (AMPCPP) is not inhibitory. These data suggest that the γ-phosphate is directly involved in intermediate formation during catalysis (Cichowicz et al., 1981; McGuire and Coward, 1984).

The specificity of the glutamate binding site of bacterial and mammalian FPGS has been assessed by studies of a wide variety of glutamate analogues, homologues and isomers (Cichowicz and Shane, 1987b; Masurekar and Brown, 1975; McGuire et al., 1980; Shane, 1980b). All FPGS enzymes display nearly absolute specificity for L-glutamate. These analogue studies demonstrate that binding to the glutamate site requires an L-α-amino acid, a carbon skeleton similar in length to glutamate,

unsubstituted α-amino and α-carboxyl groups, and a negative charge in the gamma position (McGuire and Coward, 1984). Almost all modifications of L-glutamate molecule lead to loss of affinity for the enzymes. The only exceptions are L-homocysteate and 4-fluoroglutamate (Bognar and Shane, 1983; Cichowicz and Shane, 1987b; McGuire and Coward, 1985; Shane and Cichowicz, 1983). The former is an alternate substrate for the bacterial enzymes and the latter for both bacterial and mammalian enzymes. Both of these glutamate analogues are less effective and appear to act as chain terminators in that their incorporation into the folate molecule prevents or greatly retards the further addition of glutamate moieties. None of the enzymes can use pre-formed γ-glutamylglutamate as substrates for folylpolyglutamate synthesis, suggesting that the polyglutamate chain is extended by addition of glutamate moieties one at a time (McGuire and Coward, 1984).

The folate substrate specificity of FPGS depends on the source of the enzyme. Among bacterial enzymes, the *E. coli* and *S. faecalis* FPGS preferentially utilize 10-formyl-H_4PteGlu (Masurekar and Brown, 1975), the *Corynebacterium* enzyme utilizes H_4PteGlu (Shane, 1980b) and the *L. casei* enzyme preferentially utilizes 5,10-methylene-H_4PteGlu (Bognar and Shane, 1983) as the monoglutamate substrates. Other monoglutamate substrates are utilized less well, primarily due to weaker binding. However,the bacterial enzymes all use 5,10-methylene-H_4PteGlu$_n$ as their preferred, or only, polyglutamate substrate. The mammalian FPGS have much broader specificity, being able to utilize a variety of folate monoglutamates. The most effective monoglutamate substrates for the pig liver enzyme are H_4PteGlu, H_2PteGlu and 10-formyl-H_4PteGlu (Cichowicz and Shane, 1987b) which are similar to the substrates reported for the human (Chen et al., 1996), rat and mouse liver enzymes (McGuire et al., 1980; Moran et al., 1985). Although mammalian FPGS can use a variety of polyglutamate derivatives as substrates, they all utilize H_4PteGlu$_n$ as their most effective polyglutamate substrate.

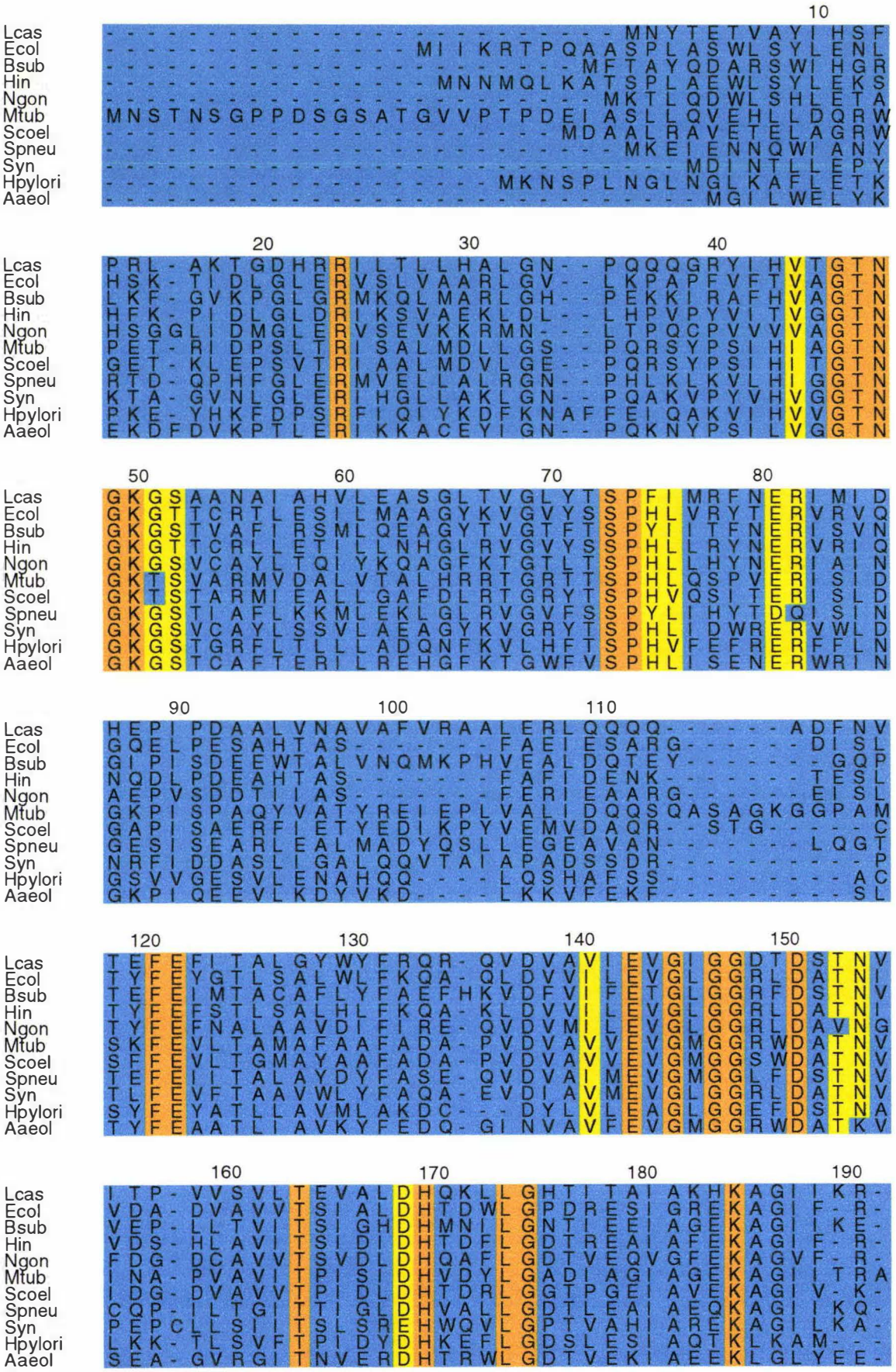
Product analysis for the bacterial enzymes reveals that only a single glutamate is readily added to their monoglutamate substrates. Synthesis of the longer derivatives which predominate in these species is extremely slow compared to the production of diglutamate. However, most bacterial enzymes readily metabolize the polyglutamate 5,10-methylene-H_4PteGlu$_n$ substrates to longer chain length derivatives. In comparison, the mammalian enzymes are able to synthesize long polyglutamates quite readily with

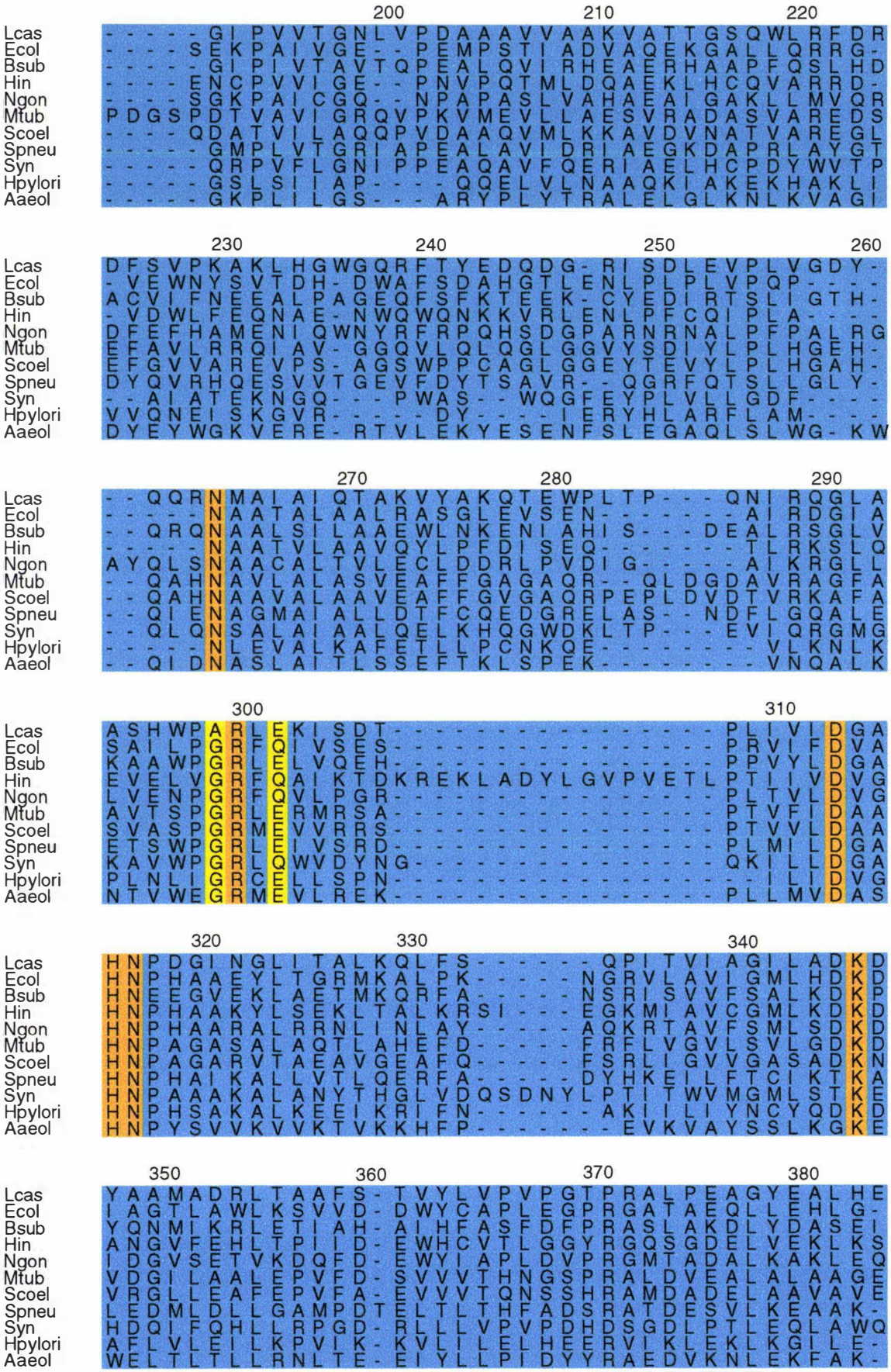
H_4PteGlu as an initial substrate. The polyglutamates can be up to the predominant length of folylpolyglutamates in most mammalian cells, although the substrate activity decreases with increasing chain length, which is probably caused by the difficulty in positioning the γ-carboxyl of the terminal glutamate residue at the active site (Cichowicz and Shane, 1987b). As described in Section 1.2.3, this ability of different chain length folylpolyglutamates to act as substrates could contribute to the *in vivo* distribution of folylpolyglutamates in a particular organism. The studies of folate analogs as substrates for mammalian FPGS indicated that a pteridine-like moiety is necessary, and any change in the terminal glutamate present on the folate leads to loss of activity (Bognar et al., 1983; George et al., 1987; McGuire and Coward, 1984; Moran et al., 1984). Thus, both a heterocyclic portion and a terminal glutamate are essential for binding, and it was also suggested that folate substrate-enzyme interaction may involve two co-operative domains (McGuire and Coward, 1984).

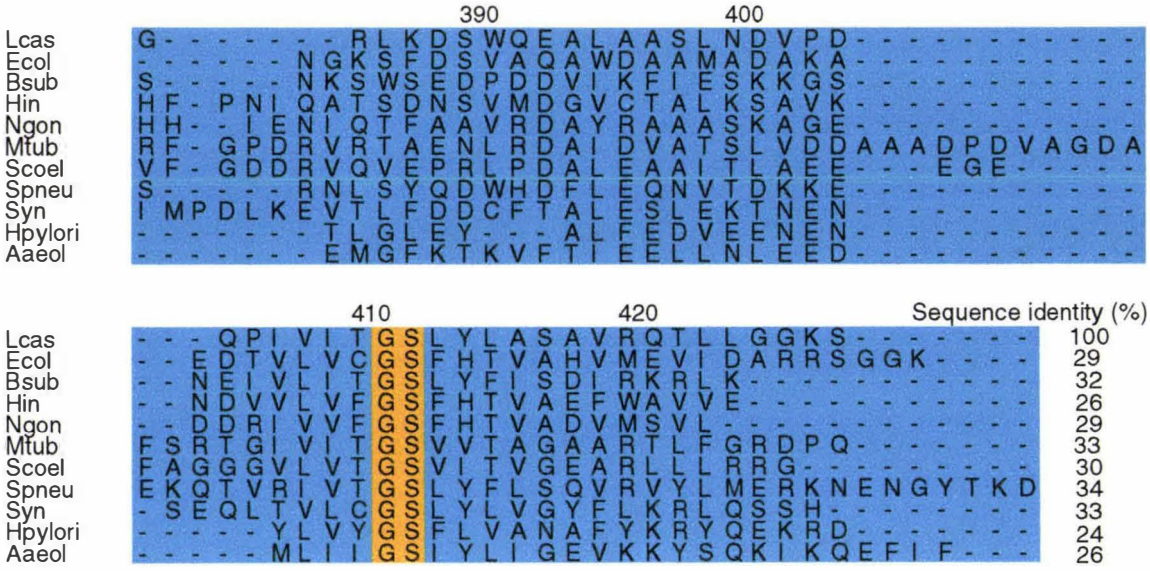
1.3.3 Functional differences and sequence alignment of FPGS family

In addition to the striking difference in folate substrate specificity between bacterial and mammalian enzymes, as described above, some other functional differences between FPGS from different species are clear from Table 1.3.2-I. The properties of all mammalian enzymes are similar and differ from the bacterial proteins in their larger molecular size, their lower requirement for K^+ and their additional requirement for a reducing agent. The eukaryotic proteins generally metabolize folate substrates to longer chain polyglutamates while the bacterial FPGS extend the polyglutamate chain primarily to the tri- and tetraglutamate derivatives (Shane et al., 1993). Some bacterial FPGS, such as from *E. coli, Corynebacterium* and *S. Pneumoniae* etc., also possess dihydrofolate synthetase (DHFS) activity, and are thereby bifunctional enzymes. On the other hand, the *L. casei* and all mammalian FPGS lack this DHFS activity.

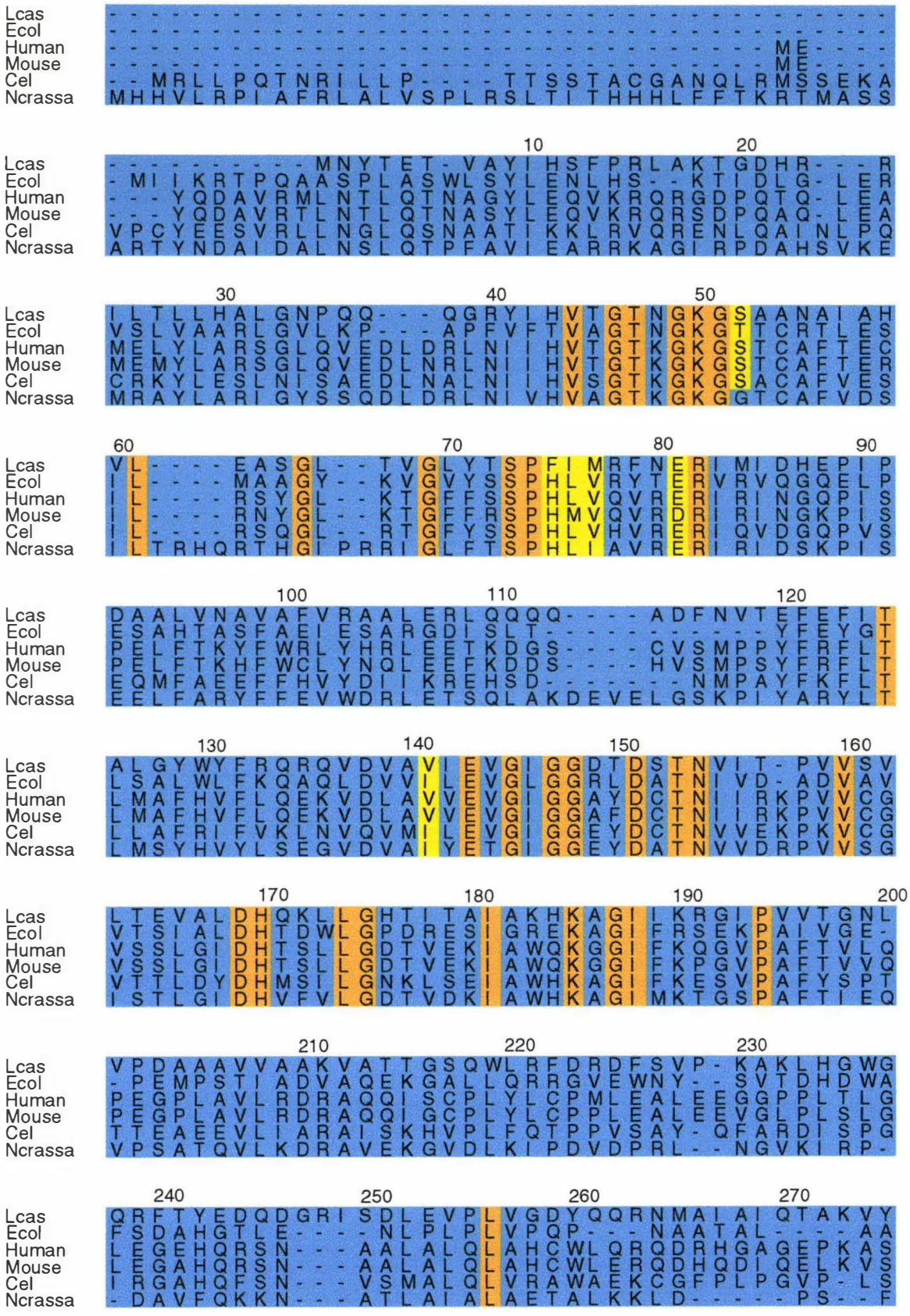
Amino acid sequence comparisons between the FPGS of *L. casei* and other bacterial enzymes, and between *L. casei*, mammalian, worm and fungal proteins are given in Fig. 1.3.3-I, respectively. The sequence alignment using Pairwise Sequence Alignment (Thompson et al., 1994) reveals that the *L. casei* protein shows generally limited sequence identity with other bacterial and mammalian proteins (approximately 30% and 24%, respectively; each pairwise sequence identity is listed at the end of Fig. 1.3.3-I).


Lcas
Ecol
Bsub
Hin
Ngon
Mtub
Scoel
Spneu
Syn
Hpylori
Aaeol
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190


200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
Lcas
Ecol
Bsub
Hin
Ngon
Mtub
Scoel
Spneu
Syn
Hpylori
Aaeol


390
400
Lcas G - - - - - - - R L K D S W Q E A L A A S L N D V P D - - - - - - - - - - - -
Ecol - - - - - - N G K S F D S V A Q A W D A A M A D A K A - - - - - - - - - - - -
Bsub S - - - - - N K S W S E D P D D V I K F I E S K K G S - - - - - - - - - - - -
Hin H F - P N I Q A T S D N S V M D G V C T A L K S A V K - - - - - - - - - - - -
Ngon H H - - I E N I Q T F A A V R D A Y R A A A S K A G E - - - - - - - - - - - -
Mtub R F - G P D R V R T A E N L R D A I D V A T S L V D D A A A D P D V A G D A
Scoel V F - G D D R V Q V E P R L P D A L E A A I T L A E E - - - E G E - - - - -
Spneu S - - - - - R N L S Y Q D W H D F L E Q N V T D K K E - - - - - - - - - - -
Syn I M P D L K E V T L F D D C F T A L E S L E K T N E N - - - - - - - - - - -
Hpylori - - - - - - - T L G L E Y - - - A L F E D V E E N E N - - - - - - - - - - -
Aaeol - - - - - - - E M G F K T K V F T I E E L L N L E E D - - - - - - - - - - -
410
420
Sequence identity (%)
Lcas - - - Q P I V I T G S L Y L A S A V R Q T L L G G K S - - - - - - - 100
Ecol - - E D T V L V C G S F H T V A H V M E V I D A R R S G G K - - - - 29
Bsub - - N E I V L I T G S L Y F I S D I R K R L K - - - - - - - - - - 32
Hin - - N D V V L V F G S F H T V A E F W A V V E - - - - - - - - - - 26
Ngon - - D D R I V V F G S F H T V A D V M S V L - - - - - - - - - - 29
Mtub F S R T G I V I T G S V V T A G A A R T L F G R D P Q - - - - - - - 33
Scoel F A G G G V L V T G S V I T V G E A R L L L R R G - - - - - - - - - 30
Spneu E K Q T V R I V T G S L Y F L S Q V R V Y L M E R K N E N G Y T K D 34
Syn - S E Q L T V L C G S L Y L V G Y F L K R L Q S S H - - - - - - - 33
Hpylori - - - - - Y L V Y G S F L V A N A F Y K R Y Q E K R D - - - - - - - 24
Aaeol - - - - - M L I I G S I Y L I G E V K K Y S Q K I K Q E F I F - - - 26

(a)


Lcas
Ecol
Human
Mouse
Cel
Ncrassa
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270

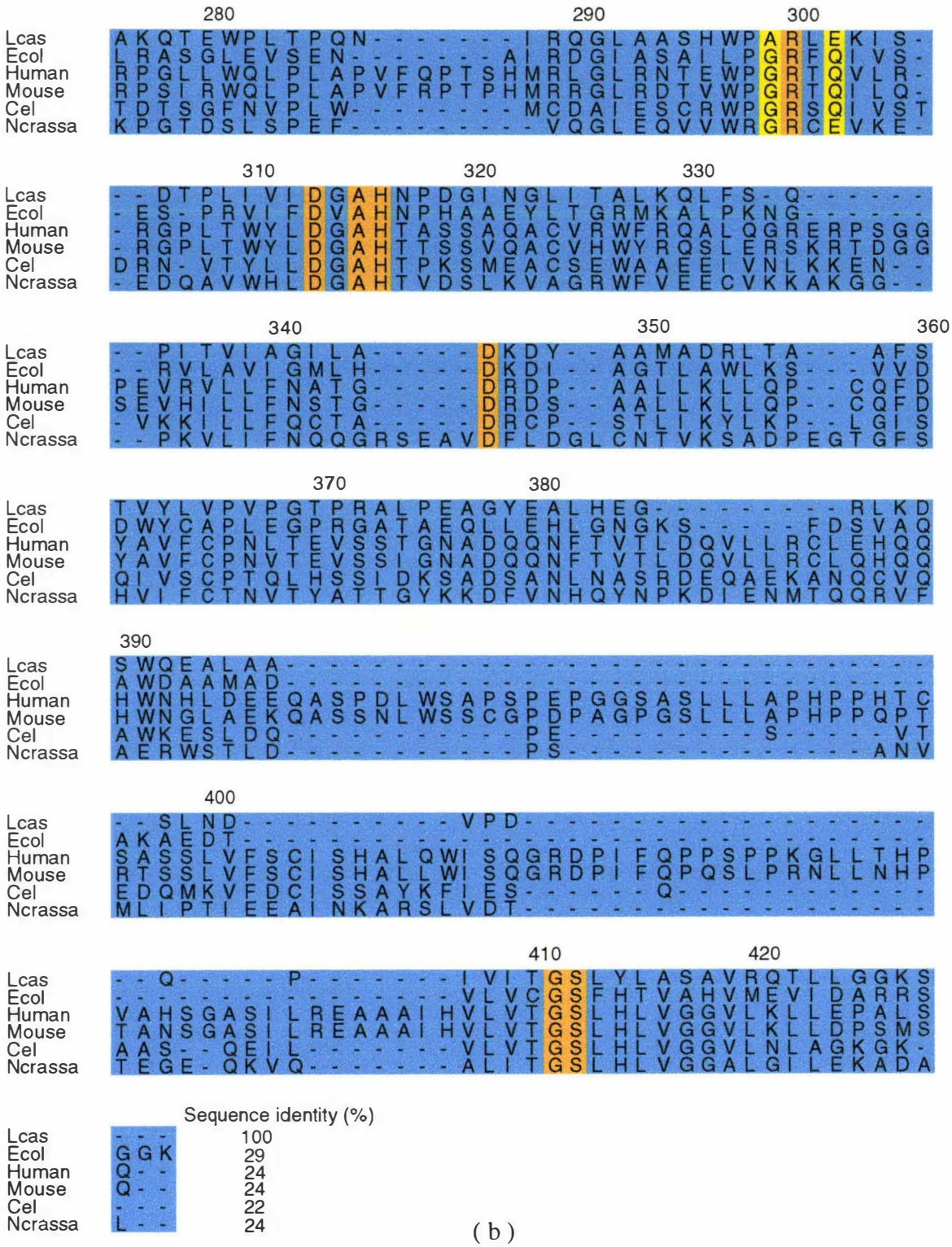


Fig. 1.3.3-I Multiple sequence alignment of FPGS family (a) between *L. casei* and other bacterial enzymes; (b) between *L. casei*, mammalian and fungal enzymes.

The multiple sequence alignment was made by using CLUSTAL W1.72 (Smith et al., 1996). The sequence sources used in alignment are Lcas (*Lactobacillus casei*), Ecol (*Escherichia coli*), Bsub (*Bacillus subtilis*), Hin (*Haemophilus influenzae*), Ngon (*Neisseria gonorrhoeae*), Mtub (*Mycobacterium tuberculosis*), Scoel (*Streptomyces coelicolor*), Spneu (*Streptococcus pneumoniae*), Syn (*Synechocystis sp.*), Hpylori (*Helicobacter pylori*), Aaeol (*Aquifex aeolicus*), Cel (*Caenorhabditis elegans*), Ncrassa (*Neurospora crassa*). The conserved residues are indicated in orange, some conservative substitutions are indicated in yellow. The numbering correspond to sequence of *L. casei* and the sequence identity between *L. casei* and each of other sources are listed at the end.

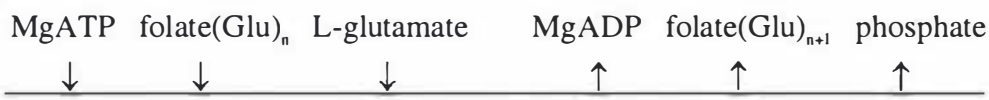
The lower sequence identity between *L. casei* and mammalian proteins may also be related to greater differences in their properties.

Two major conserved regions are obvious from Fig. 1.3.3-I. One of them is at a sequence resembling a P-loop motif for binding the phosphate group of nucleotide substrates, which is located between Gly46 and Ser52 in *L. casei* FPGS. Our structural analysis has confirmed that this region is the P-loop motif of FPGS. Mutagenesis of *L. casei* FPGS revealed that mutations in this sequence region, such as G49A, K50R and G51S, resulted in either inactive or less effective enzymes (Toy and Bognar, 1994), supporting this as a putative nucleotide binding site. The sequence comparisons also suggest that all members of the FPGS family have basically the same binding pocket for the nucleotide substrate. Another major conserved region is between Glu143 and Asn154 (*L. casei* numbering). This sequence is believed to correspond to the nucleotide binding B motif (Garrow et al., 1992) which exhibits conformational changes during nucleotide hydrolysis in other ATP and GTPases (see discussion in Section 3.6).

Despite the limited sequence identity, a striking feature of the multiple sequence alignment is that most of the highly conserved residues, in addition to the two major conserved regions, interact with the putative substrates and/or occupy positions which are presumably important for activity and proper folding of the enzyme. This will be discussed later in detail (Section 3.6), based on the structural model obtained for MgATP-FPGS.

1.3.4 Proposed substrate binding and kinetic mechanism

Kinetic analyses of the folylpolyglutamate synthetase catalyzed reaction have been carried out for both bacterial and mammalian enzymes (Bognar and Shane, 1983; Cichowicz and Shane, 1987b; Shane, 1980c) and the same mechanism was observed for all enzymes. These studies indicated that the mechanism is sequential and of the ordered Ter-Ter type shown as follows:



The order of substrate addition is MgATP, folate$(Glu)_n$ and glutamate, and the order of product release is MgADP, folate$(Glu)_{n+1}$ and phosphate. MgADP and phosphate are both inhibitors. An important implication of this mechanism is that the enzyme cannot catalyze the sequential addition of glutamate to enzyme-bound folate. The folate product has to be released and then rebind to the enzyme before an additional glutamate moiety can be added. A number of investigators (Shane, 1980c; Tang and Coward, 1983) have suggested that this reaction proceeds via phosphorylation of the γ-carboxyl of an enzyme-bound folate, followed by nucleophilic attack by the free amine of glutamate on the acyl-phosphate intermediate which is shown in Fig. 1.3.4-I.

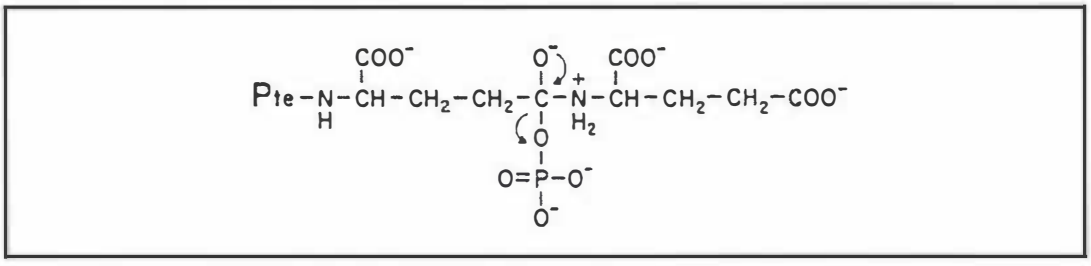


Fig. 1.3.4-I Proposed reaction intermediate for the folylpolyglutamate synthetase reaction.

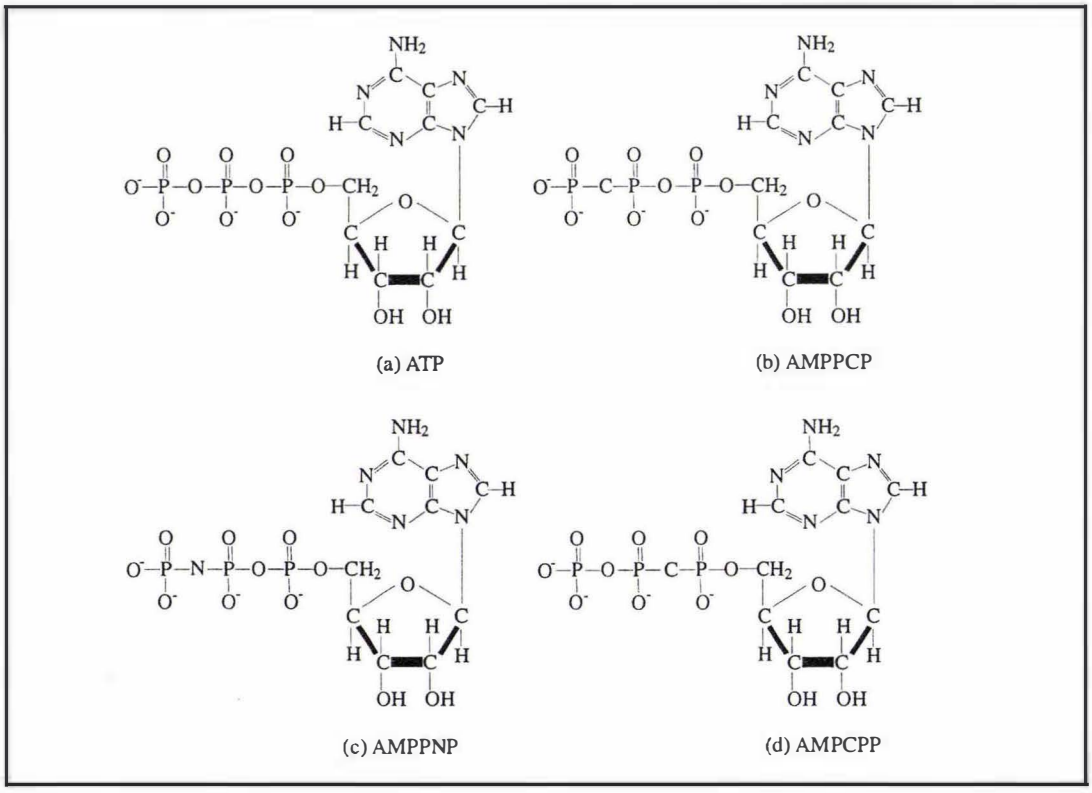


Fig. 1.3.4-II Structures of some ATP analogues: (a) ATP; (b) β,γ-methylene-ATP (AMPPCP); (c) β,γ-imido-ATP (AMPPNP); (d) α,β-methylene-ATP (AMPCPP).

This has been suggested principally by analogy with the mechanism of glutamine synthetase (Meister, 1978), because there is so far no direct structural evidence from any of the folylpolyglutamate synthetases. However, an experiment demonstrating transfer of ^{18}O from [α, γ-^{18}O] folate to phosphate supports the existence of an acylphosphate intermediate (Banerjee et al., 1988), and this acylphosphate intermediate has been isolated by HPLC using an anionic exchanger (Shane et al., 1993). Moreover, the findings that β,γ-methylene-ATP (AMPPCP) and β,γ-imido-ATP (AMPPNP), but not α,β-methylene-ATP (AMPCPP), inhibit the synthetase are all consistent with the phosphoryl transfer. The structures of these ATP analogues are shown in Fig. 1.3.4-II.

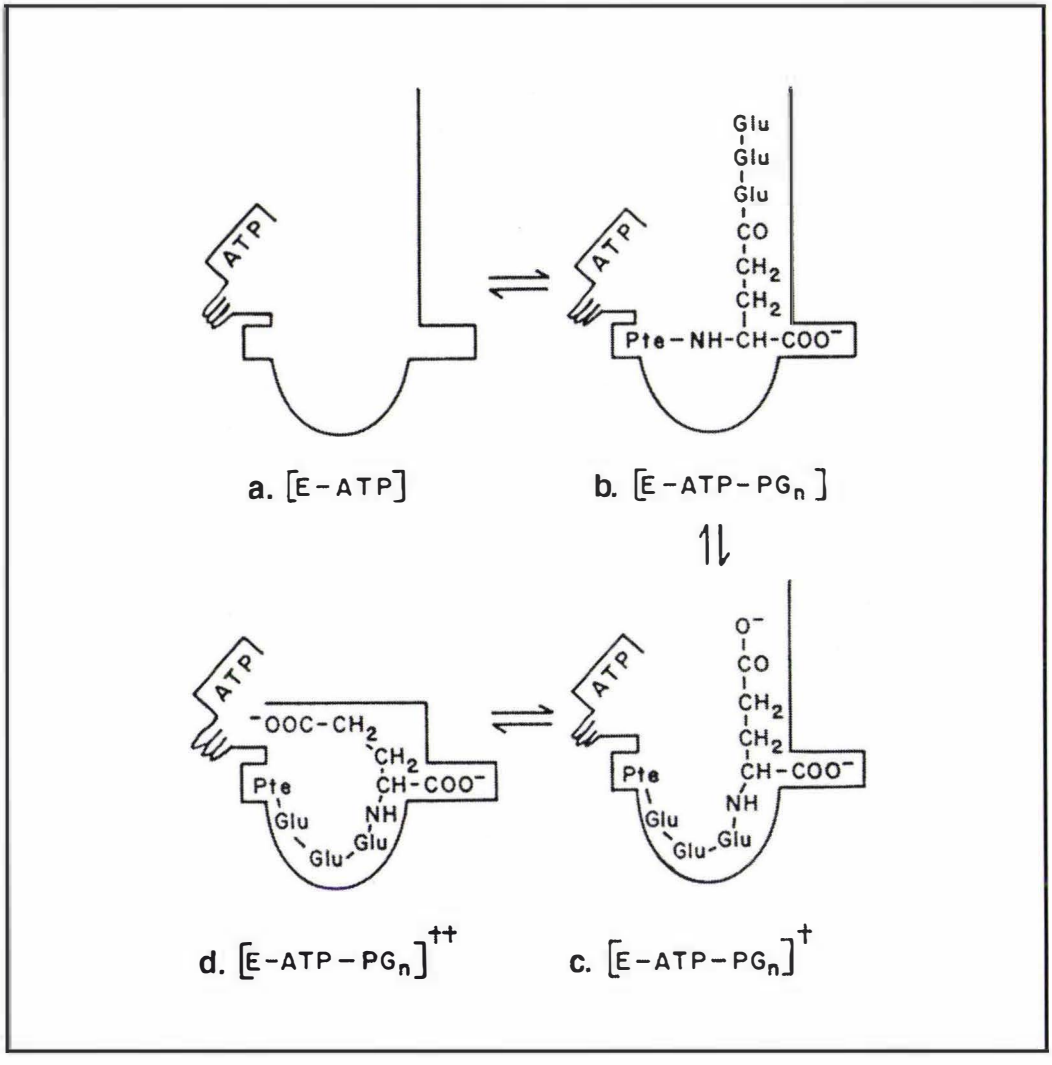


Fig. 1.3.4-III Proposed mechanism for binding of folylpolyglutamate to folylpolyglutamate synthetase. Taken from Shane (1989).

As a result of studies of folate specificity (Cichowicz and Shane, 1987b; George et al., 1987), a mechanism for binding of folylpolyglutamates to FPGS has also been proposed (Shane, 1989). It is suggested that binding of folates to the enzyme reflects a cooperative effect resulting from two low-affinity events, the initial binding of the pterin moiety followed by the binding of the internal glutamate residue of polyglutamate (cf. Fig. 1.3.4 IIIb). The initial glutamate residues move through a temporary glutamate-binding site until the terminal glutamate residue is bound (cf. Fig. 1.3.4-IIIc), after which a conformational change of the enzyme occurs to bring the γ-carboxyl of the terminal glutamate into the active site (cf. Fig. 1.3.4-IIId). This conformational change occurs only when the α-carboxyl of the terminal glutamate is bound, which is consistent with the observation that the terminal α-carboxyl is essential for positioning the terminal γ-carboxyl for reaction (Moran et al., 1984). From this mechanism, the lower effectiveness of longer polyglutamate substrates could be explained by random movement of the polyglutamate chain, which is expected to decrease the likelihood of the terminal residue being positioned correctly in the α-carboxyl binding site (cf. Fig. 1.3.4-IIIc).

1.4 Role of FPGS in antifolate chemotherapy

1.4.1 Folate analogues and cancer chemotherapy

A wide variety of antifolates such as aminopterin, methotrexate and related compounds have been tested for their effect on many tumours during the last four decades. The first report of folate analogues being used to advantage against neoplastic disease was the treatment of leukaemia with aminopterin (Farber et al., 1948). L1210, an acute lymphoid leukaemia, has been extensively used in studies of the anti-tumour action of these antifolates. Methotrexate and its halogen derivatives are the most active analogues, and methotrexate, an inhibitor of dihydrofolate reductase (DHFR), has therefore become representative of antifolates in the chemotherapy of neoplastic disease. It has played a significant role in the overall survival and improved central nervous system chemoprophylaxis of children with acute lymphoblastic leukaemia (Niemeyer et al., 1985). Methotrexate therapy has also been reported to be beneficial to patients with squamous cell carcinoma of the head, neck and lung, and cancer of breast (Hellman et

al., 1964; Ross and Selawry, 1966; Vogler et al., 1965). Chemotherapy with antifolates has taken on increased importance with recent reports of emergence of secondary neoplasms in patients treated with alkylating agents and topoisomerase inhibitors (Boffetta and Kaldor, 1994). In contrast, methotrexate has a particularly good safety record. Lifetime carcinogenicity studies and *in vitro* oncogenic transformation tests have indicated virtually zero threat of carcinogenicity at substantial doses of the drug.

In addition to DHFR-based therapeutic agents, the discovery of the therapeutic activity of folate-based inhibitors of thymidylate synthase (TS) (Jones et al., 1981) established that folate metabolism overall is a potentially rich source of anti-tumour drugs and has focused efforts on several other enzymes within folate pathways, in addition to DHFR. As a result of these efforts, a new generation of potent inhibitors of TS (e.g. CB3717, 10-propargyl-5,8-dideaza-folic acid, and its derivative ZD1694) and of *de novo* purine synthesis via inhibition of glycinamide ribonucleotide formyl-transferase (GARFT) (e.g. DDATHF, 5,10-dideazaterahydrofolate) have been developed (Beardsley et al., 1989; Clarke et al., 1993). Both DDATHF and ZD1694 have been found to be active against human cancers and these two compounds represent the two most potent anti-folates ever to be administered to humans (Clarke et al., 1992; Ray et al., 1993). The chemical structures of some antifolates are shown in Fig. 1.4.1-I.

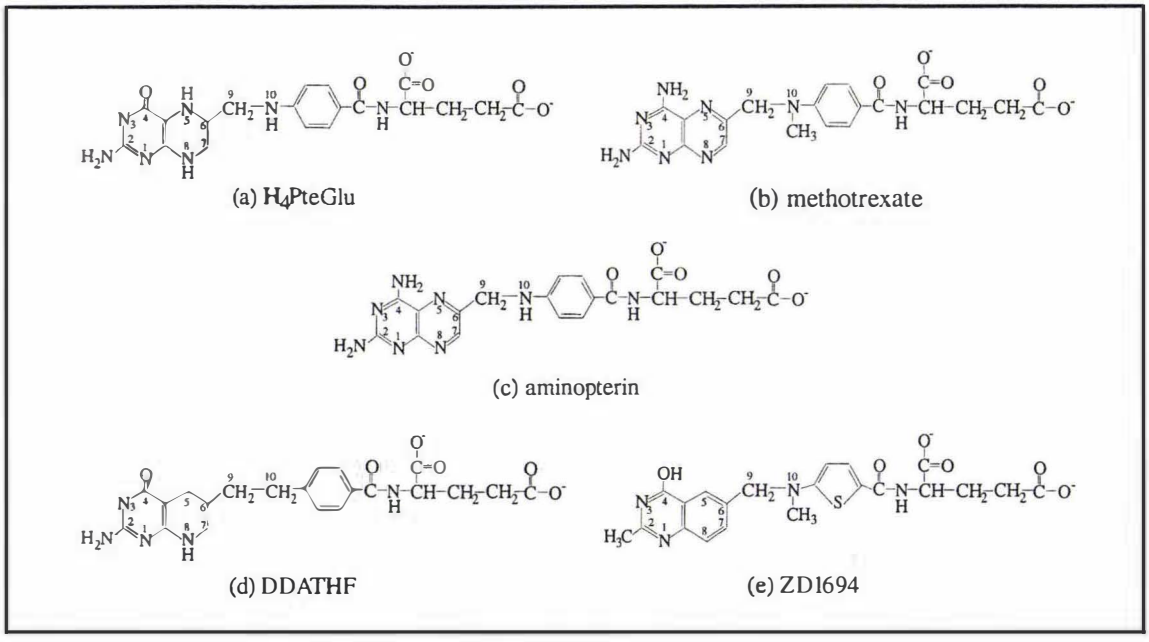


Fig. 1.4.1-I The chemical structures of some antifolates: (a) tetrahydrofolate (H_4PteGlu); (b) methotrexate; (c) 4-amino-4-deoxy folic acid (aminopterin); (d) 5,10-dideaza-5,6,7,8-tetrahydrofolate (DDATHF); (e) ZD1649.

These new classes of cytotoxic antifolates, and many "classical" antifolates such as methotrexate, are substrates for the enzyme FPGS. Increasing *in vitro* and clinical evidence indicates that polyglutamylation by FPGS is central to the therapeutic activity of cytotoxic antifolates. Their metabolism to long chain polyglutamylated species is considered important for maximum pharmacologic effects. In this way, FPGS has come under intensive investigation in the past few years because of its role in the activation of antifolate drugs, and it has become an important target for the development of novel drugs.

1.4.2 Cellular transport and polyglutamylation of antifolates

As described before, the physiologically active coenzymatic forms of reduced folates and pharmacologically active antifolates exist primarily as polyglutamate derivatives in cells. However, only folylmonoglutamates can be efficiently transported into and out of cells by a carrier-mediated system (Goldman, 1971; Nahas et al., 1972). Folylpolyglutamates of chain length equal to or greater than 3 seem to be poorly accepted by this transport carrier in mammalian cells (Baugh et al., 1971; Godwin and Rosenberg, 1975). Therefore, dietary folates in the intestinal lumen are initially hydrolyzed to the monoglutamate form by the enzyme γ-glutamyl hydrolase, and as such are absorbed across the intestinal mucosa. There are two major folate transport systems, i.e., the folate membrane carriers and the folate binding protein (Freisheim et al., 1989; Kamen et al., 1988). Natural folates and antifolates utilize both systems for entry into mammalian cells.

Once natural folates and antifolates enter the cytoplasm, glutamylation occurs through catalysis by the enzyme FPGS. For the antifolate drugs, penta- and hexaglutamates are the predominant species in most mammalian tissues. Recognition of the significance of polyglutamation for the cytotoxicity and therapeutic efficacy of antifolate drugs leads to a new drug development strategy for agents that can be used efficiently as substrates for mammalian FPGS. The substrate specificity of mammalian FPGS for antifolate analogues is expected to play a role in defining the cytotoxic action of these compounds. Many studies have addressed the structural factors that determine the activity of antifolates as substrates for mammalian FPGS (Habeck et al., 1995; McGuire et al., 1995; Moran et al., 1987). These investigations have resulted in some

new antifolates with enhanced activity as substrates for polyglutamation. For example, development of the 10-deazaaminopterin series created analogues with both increased transport capabilities and increased ability to undergo polyglutamylation, both properties being selective for tumour cells over normal tissues (DeGraw et al., 1982; Sirotnak et al., 1984). Synthesis of the analogue ZD1694, which contains a 2-deaza-2-methylquinazoline group and substitution of a thiazole for the benzene ring in the p-aminobenzoylglutamate moiety, also led to dramatic increases in the ability to be polyglutamylated (Jackman et al., 1991a; Jackman et al., 1991b).

The metabolism of methotrexate to potentially retainable polyglutamate derivatives is a slow process in mammalian cells, compared with reduced folates, because di- and/or triglutamate derivatives of methotrexate are very poor substrates for the FPGS enzyme (Cook et al., 1987). This again indicates that the antifolate substrate specificity of FPGS could have a large effect on its polyglutamylation and therefore its retention. Obviously, modifications of antifolates that increase their substrate effectiveness for FPGS would result in more effective cytotoxic agents.

Polyglutamylation and retention of antifolates are also very sensitive to differences in FPGS levels in various tissues. Some studies have documented different levels of FPGS activity in various normal and neoplastic tissues (Barredo and Moran, 1992; Li et al., 1992). These studies suggest that expression of FPGS is regulated by two mechanisms, one that is linked to cellular proliferation and another that controls enzyme activity after differentiation and is tissue-specific. The studies performed in tumour tissues also show that expression of FPGS is heterogeneous among various human neoplasms, and is proliferation-dependent, and under tissue-specific control (Goker et al., 1993; Lenz et al., 1994). Furthermore, accumulated evidence indicates that the control of expression of FPGS in normal and various malignant tissues plays a crucial role with respect to the cytotoxicity and therapeutic selectivity of antifolates (Fry et al., 1983; Pizzorno et al., 1991; Rumberger et al., 1990; Synold et al., 1996). For instance, Whitehead et al. (1990) have reported a significant survival advantage for children with acute lymphoblastic leukaemia (ALL), treated with methotrexate on the basis of initial ability of their lymphoblasts to accumulate methotrexate and its polyglutamates. Very high levels of FPGS activity reported for lymphoblasts from patients with ALL (Barredo and Moran, 1992) would help to explain the sensitivity of this disease, as well as other tumours, to therapy with antifolates. Thus it is possible that the level of

expression of FPGS in different tumour types or the properties of the enzymes being expressed could have significant implications in the clinical use of antifolates and the development of novel, more selective, therapeutic applications of antifolates.

1.4.3 Pharmacologic role of antifolate polyglutamylation

The importance of polyglutamylation in the pharmacology of antifolates is reflected in two aspects. Firstly, the polyglutamylated forms of various antifolates are retained in cells longer, increasing their cytotoxity by extending the length of exposure, i.e, increasing retention of drug within the tumour cells. Secondly, polyglutamylated antifolates often have a higher affinity for their target enzymes than the parent drugs.

As mentioned before, there are three classes of clinically relevant antifolates available now, i.e, inhibitors of dihydrofolate reductase DHFR (methotrexate, aminopterin), inhibitors of thymidylate synthase TS (CB3717 and ZD1694), and inhibitors of *de novo* purine biosynthesis through glycinamide ribonucleotide formyltransferase (GARFT) (DDATHF). After exposure of proliferating mammalian tumour cells to these drugs, essentially all intracellular drugs are converted to polyglutamate forms (Fry et al., 1983; Jackman et al., 1991b; Pizzorno et al., 1991). For all three classes of drugs, longer chain polyglutamates are responsible for intracellular retention in tissues that express sufficient levels of FPGS. Many studies have addressed the formation and retention of methotrexate polyglutamates and their relationship to drug effect. The cytotoxicity of methotrexate does not require its conversion to polyglutamates but cellular accumulation of the drug is highly dependent on FPGS activity (Kim et al., 1993; Lin et al., 1996). These studies clearly show that the level of FPGS can be a major determinant of methotrexate retention, and consequently cytotoxicity, even though its target in the cell is DHFR and effective inhibition of DHFR does not require polyglutamylation of methotrexate.

Like methotrexate, the new antifolates aimed at TS and GARFT, such as DDATHF and ZD1694, depend on polyglutamylation for their intracellular retention which in turn contributes to their cytotoxic activity. DDATHF (Dideazatetrahydrofolate), a potent inhibitor of the first folate dependent enzyme in the *de novo* purine synthetic pathway, GARFT, is rapidly converted to its polyglutamate forms *in vitro*, particularly DDATHF-Glu_5 and DDATHF-Glu_6, which are retained intracellularly even after prolonged

incubation with drug-free media. ZD1694, a very potent TS inhibitor and excellent substrate for FPGS, is also readily polyglutamylated (Jackman et al., 1991b). More importantly, both new drugs also depend on polyglutamylation for their highly increased inhibition activity against the target enzymes. The polyglutamates of DDATHF bind to GARFT up to 100 times tighter than DDATHF itself (Baldwin et al., 1991). The tetra- and pentaglutamates of ZD1694 are approximately 100 times more potent inhibitors of their target enzyme than the monoglutamate form (Jackman et al., 1991a). Further evidence of the importance of FPGS has also been reflected in recent reports indicating that low or altered FPGS activity can be a mechanism of resistance to cytotoxic antifolates. For example, a methotrexate resistant mutant cell line derived from CCRF-CEM cells has been obtained, which is deficient in FPGS (McCloskey et al., 1991). Another mutant cell line deficient in FPGS has been derived from the human ileocecal adenocarcinoma cell line HCT-8 which is resistant to ZD1694 (Lu et al., 1995). In contrast, Chinese hamster ovary (CHO) cells that express higher FPGS activity are more sensitive to methotrexate (Kim et al., 1993).

FPGS has thus been demonstrated to be important for the cytotoxic efficacy of antifolates. Various studies suggest that the potency of antitumour activity of antifolate drugs could be correlated with the efficiency of activation by FPGS enzyme. The highly increased cytotoxicity of antifolate drugs in their polyglutamylated forms could be due to either enhanced intracellular retention or enhanced binding of polyglutamated inhibitor to target enzymes, or both.

1.5 Aims of this project

As described in the previous sections, a large number of investigations have focused on FPGS and its role in the synthesis of folylpolyglutamates. Recombinant DNA approaches have resulted in the expression and purification of various FPGS, in sufficient quantities for detailed characterization. The biochemical and kinetic data accumulated for FPGS have given information on the substrate specificity and on the metabolism of folates and antifolates. Although a putative catalytic mechanism involving an acylphosphate intermediate has been proposed, little is known about the molecular structure of FPGS as there is no crystal structure yet available. A number of

important questions remain to be answered for a full understanding of the structure and function of FPGS. For example,

(i) what is the role of the monovalent cation K^+;
(ii) how does this enzyme bind its three substrates;
(iii) how is the growing polyglutamate tail accommodated;
(iv) what triggers the conformational change during catalysis;
(v)what is the nature of the conformational change.

As the first step toward an elucidation of the molecular mechanism of polyglutamylation of folates and antifolates, this project is aimed at the full crystallographic analysis of FPGS from *L. casei*, including the MgATP-FPGS complex and apo-FPGS, so as to provide a structural model for the FPGS family. Although *L. casei* FPGS is different from mammalian FPGS in its folate substrate specificity, identical kinetic behaviour of bacterial and mammalian FPGS indicates that they have the same reaction mechanism. The structural characteristics of *L. casei* FPGS should then represent, at least, the general features of the FPGS family.

MATERIALS AND METHODS

2.1 Protein instability and purification

2.1.1 Protein Instability

Lactobacillus casei folylpolyglutamate synthetase used in this crystallographic study was expressed in *E. coli* and purified by Dr A. L. Bognar as described previously (Bognar and Shane, 1983; Toy and Bognar, 1990). The enzyme was provided as an ammonium sulphate precipitate. As described in Section 1.3.2a, this enzyme proved to be unstable in the stock solution used for crystallization, comprising 10 mg/ml FPGS in 30 mM HEPES/KOH (pH 7.2), 0.1 M ammonium sulphate.

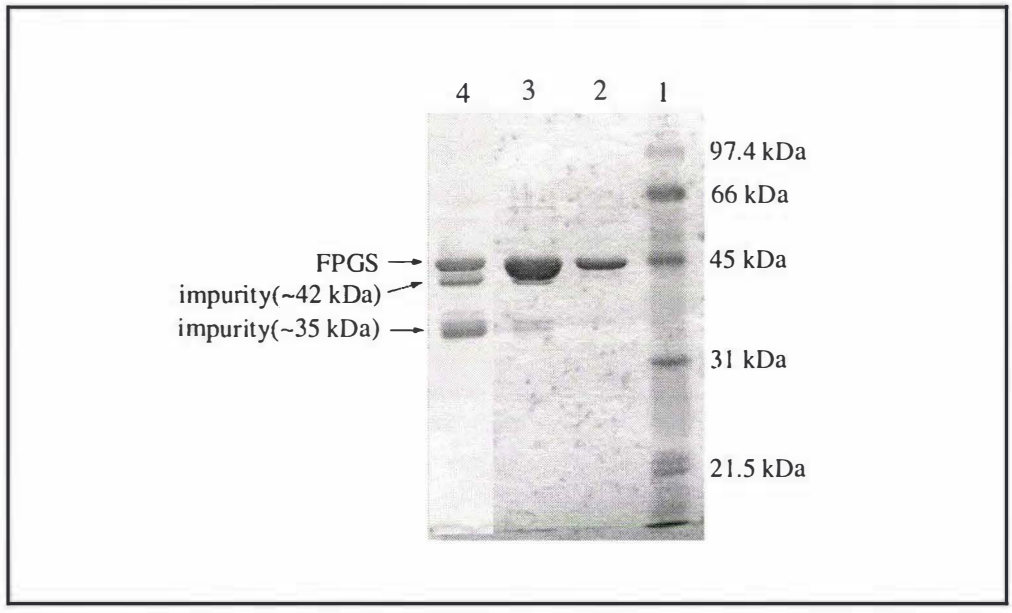


Fig. 2.1.1-I SDS PAGE of *L.casei* FPGS protein over time. Experimental details are described in Section 2.1.2. Protein standards, rabbit muscle phosphorylase B (Mr= 97,400), bovine serum albumin (Mr= 66,000), hen egg white ovalbumin (Mr= 45,000), bovine carbonic anhydrase (Mr= 31,000), soybean trypsin inhibitor (Mr= 21,500), are indicated in *lane 1*. 1 μl protein stock solution was diluted with the same buffer solution (4 μl) and denaturing buffer (5 μl) to give a protein sample (10 μl, ~1 mg/ml FPGS) for the gel electrophoresis. The samples from freshly prepared, 1 week old and 3 week old stock solution are shown in *lane 2, lane 3,* and *lane 4,* respectively. The arrows on left indicate the positions of FPGS and the impurities at both 42 kDa and 35 kDa.

Protein degradation was apparent within a week after each stock solution was prepared. There are two predominant low molecular weight impurities (~42 kDa and ~35 kDa) resulting from this degradation. They are shown in Fig 2.1.1-I. This degradation can ultimately result in the complete breakdown of the FPGS.

The exact cause of the degradation of FPGS is not clear at this stage. A variety of different stabilizing agents have been tested, with 20% DMSO (v/v) being the most effective stabilizer for *L. casei* FPGS (Bognar and Shane, 1983). DMSO is added to the protein stock solution by dialysis against 30 mM HEPES/KOH (pH 7.2), 0.1 M ammonium sulphate, 20% DMSO. However, gel electrophoresis indicates that 20% DMSO only retards the degradation to some extent, with the degradation still continuing but at a lesser rate.

It was found in our experiments that the enzyme sample is more stable if it is kept as an ammonium sulphate precipitate. Therefore, only small amounts of protein stock solution (~60μl) were made for each batch of crystallization trials.

2.1.2 SDS-polyacrylamide gel electrophoresis

0.75 mm slab gels were cast, with a resolving gel consisting of 12% acrylamide, 0.9% bisacrylamide, 0.375 M Tris/HCl pH 8.8, 0.1% SDS, 0.1% (w/v) ammonium persulphate, and 0.07% (v/v)TEMED; and a stacking gel consisting of 4% acrylamide, 0.36% bisacrylamide, 0.2 M Tris/HCl pH 6.8, 0.1% SDS, 0.4% (w/v) ammonium persulphate, and 0.05% (v/v)TEMED. The reservoir buffer was 0.02 M Tris/HCl pH 8.3, 0.192 M glycine, 0.1% SDS. The denaturing buffer contains 0.125 M Tris/HCl pH 6.8 4% SDS, 20% glycerol, 10% 2-mercaptoethanol and a drop of Bromophenol Blue.

Prior to loading, the protein samples were diluted 1:1 with denaturing buffer and heated at ~95°C for 2 minutes. The electrophoresis was carried out under constant current, using 8mA until the dye front was through the stacking gel, and 15mA through the resolving gel.

Gels were stained using Coomassie brilliant blue R-250 (0.125% Coomassie blue R-250, 50% methanol, 10% acetic acid), and destained using destaining I solution (50% methanol, 7% acetic acid) followed by destaining II solution (5% methanol, 7% acetic acid).

2.1.3 Protein purification

The degradation of the enzyme during prolonged storage caused a serious problem as the crystallographic studies proceeded. The crystallization conditions used previously to obtain good crystals could not be repeated or gave poor quality crystals (see Section 2.2.2), and attempts were therefore made to purify the degraded enzyme.

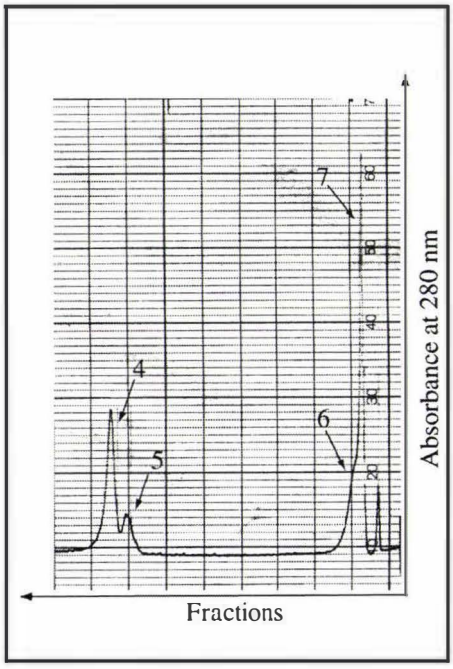


Fig. 2.1.3-I The elution profile of degraded FPGS sample from the Mono-S column. Experimental detail is described in the text. The fractions of elution peaks marked as number 4-7 were checked by SDS PAGE, which are shown in lanes marked with the same number (4-7) in the Fig. 2.1.3-II.

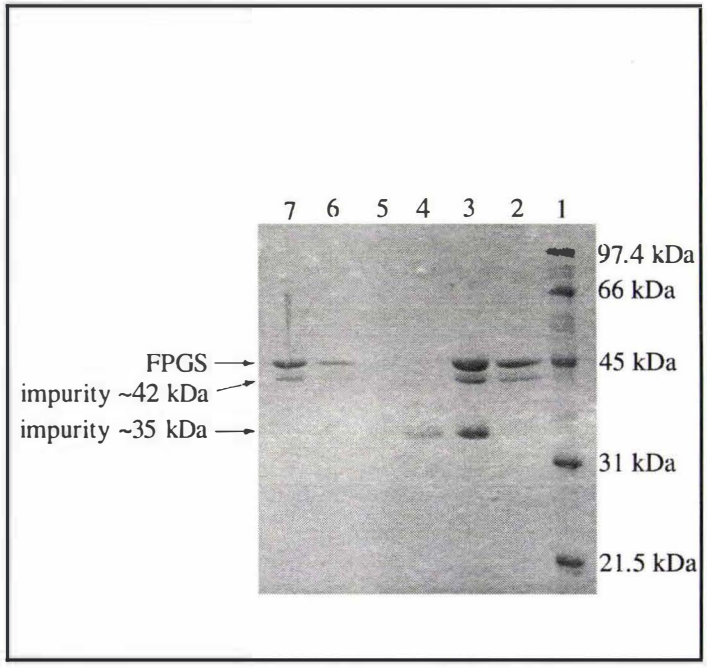


Fig. 2.1.3-II SDS PAGE of the partially purified FPGS. The experimental details are described in Section 2.1.2. Protein standards, as in Fig. 2.1.1-I, are indicated in *lane 1*. The degraded FPGS sample before the Mono-S column is shown in *lane 3*. The fractions 4-7 from the Mono-S column are shown in *lanes 4-7*. The pooled protein (fraction 6 and 7) is shown in *lane 2*. The arrows on left indicate the positions of FPGS and impurities at both 42 kDa and 35 kDa.

Gel filtration was tried initially using a Superdex-75 column equilibrated with 50 mM HEPES/KOH (pH 7.0), 0.1 M ammonium sulphate. Gel electrophoresis results for each fraction collected showed that the degraded impurities eluted together with FPGS under three different flow rates (1 ml, 0.5 ml, 0.1 ml/min). Subsequent experiments

focused on various ion exchange columns with different buffer solutions and pH. An optimum purification scheme was obtained as follows: The enzyme samples were loaded on to a Mono S cation exchange column (Pharmacia Biotech HR5/5, 5 mm x 50 mm) equilibrated with 50 mM Succinic/KOH, pH 4.5, 0.1 M NaCl and the column was washed using the buffer solution until the eluent showed no absorbance at 280 nm. The FPGS was eluted using a 0.1-1.0 M NaCl gradient (250 ml: 250 ml). Fractions of each peak were collected and checked using SDS-polyacrylamide gel electrophoresis.

The elution profile from the Mono-S column is shown in Fig. 2.1.3-I. The corresponding SDS PAGE results are shown in Fig. 2.1.3-II. It is clear from these two figures that the degradation product at 35 kDa (fraction 4 and 5) has been removed, while the impurity running at 42 kDa elutes with the full-length FPGS. However, subsequent crystallization trials using pooled protein (*lane 2* in Fig. 2.1.3-II) containing both FPGS and degraded impurity at 42 kDa, resulted in good quality crystals similar to those grown using fresh FPGS (see Section 2.2.2). This indicates that this partial purification is adequate for the purpose of crystallization.

2.1.4 Determination of protein concentration

The concentrations of FPGS solutions used in crystallization were estimated from the A_{280} absorbance using the formula:

$$C = (A_{280}/\varepsilon_{280})\bullet[(800\mu l+X\mu l)/X\mu l]$$

where A_{280} is absorbance at 280 nm; 800(μl), buffer volume; X(μl), protein sample solution volume; ε_{280}, extinction coefficient, 0.71 ($ml \cdot mg^{-1} \cdot cm^{-1}$) determined for FPGS; and C is protein concentration in mg/ml.

The extinction coefficient ε_{280} for FPGS was derived from a Bio-Rad Protein Assay (Bradford, 1976). A standard curve for IgG was provided by Dr C. A. Smith and is shown in Fig. 2.1.4-I. The concentration of an FPGS sample solution with known absorbance at 280 nm was determined using this standard curve and the value of ε_{280} was derived. A typical absorption spectra for an FPGS sample solution is shown in Fig. 2.1.4-II. The concentration of FPGS in the solution is calculated as 0.056 mg/ml.

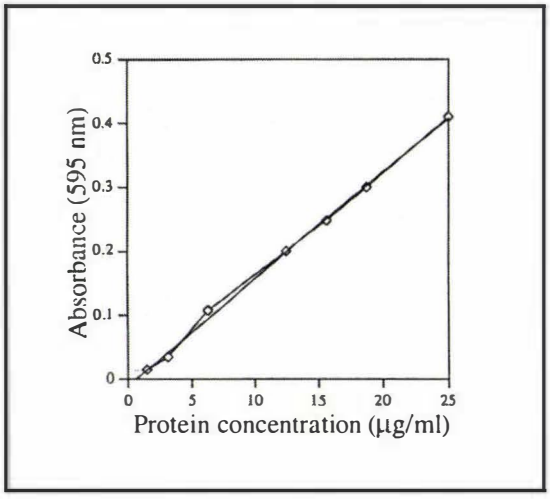


Fig. 2.1.4-I The standard curve of the concentration for IgG.

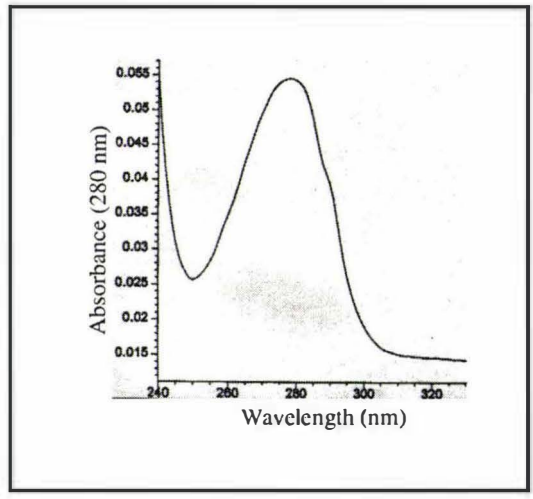


Fig. 2.1.4-II The absorption spectra for an FPGS sample.

2.2 Protein Crystallization

2.2.1 Hanging drop trials

Crystallization of *L. casei* FPGS was first carried out using the hanging-drop vapour diffusion method (McPherson, 1982). Initial crystal screening through Crystal Screen I (Jancarik and Kim, 1991) indicated promising crystallization conditions from PEG4000 solution with acetate buffer. Further variations on these conditions resulted in growth of small block-like crystals from 15% PEG4000(w/v), 0.1 M Acetate/KOH (pH 5.3), 0.2 M ammonium sulphate and 20 mM KCl at 4°C. The protein concentration was ~10 mg/ml. Similar crystallization conditions have also been reported by others (Cody et al., 1992). Crystals obtained under this condition always sit in precipitate and often stack together. They were not good enough for data collection but were used as crystal seeds to grow large single crystals by the micro-seeding method.

2.2.2 Micro-seeding procedure

All diffraction quality crystals used in native and heavy-atom derivative data collections were grown by the micro-seeding technique. In general, there are two

methods for transferring micro-crystal seeds into supersaturated protein/precipitant solution (Ducruix and Giege, 1992): (i) dilution of a stock microcrystal solution (usually a 10^3- or 10^4-fold dilution), and addition of 1-2μl of this diluted microcrystal solution into protein/precipitant drops; (ii) directly transferring microcrystal seeds from microcrystal stock solution to protein/precipitant drops by a whisker. We chose the latter method because the supersaturation of the protein/precipitant drops is unchanged and it was shown to be successful in our initial experimental trials. The micro-seeding procedure is depicted in the following Fig. 2.2.2-I.

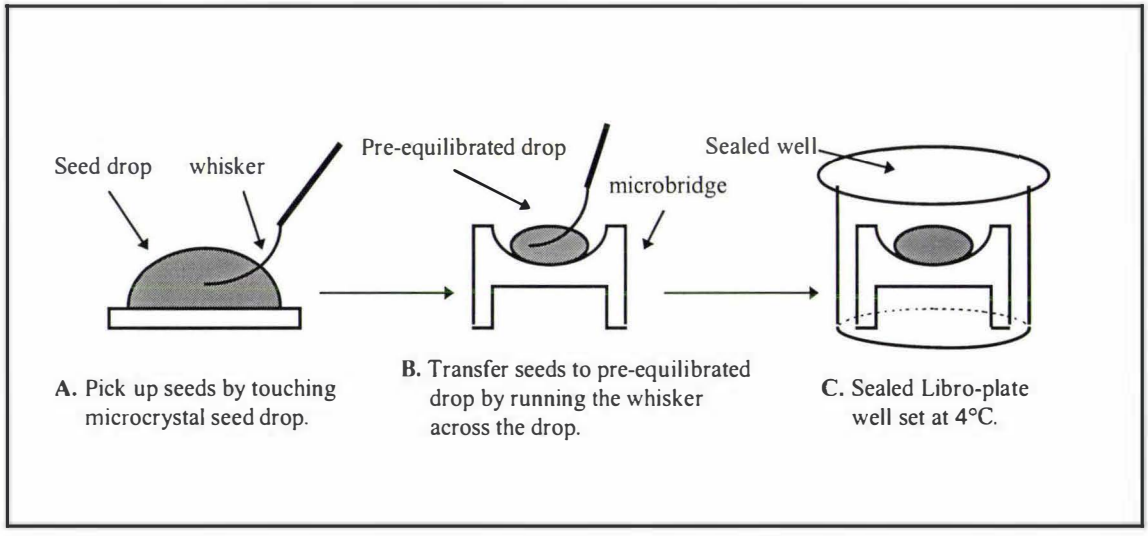
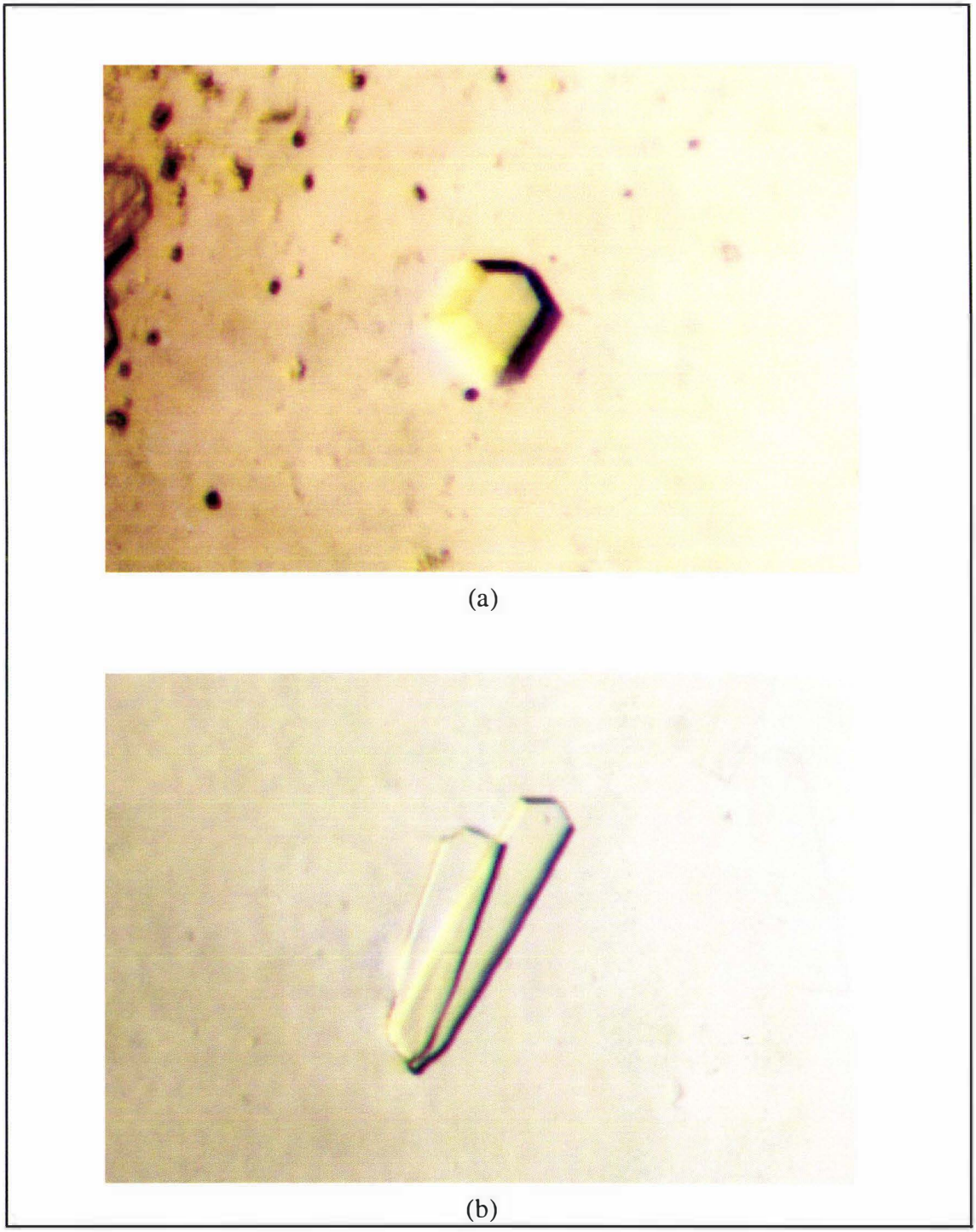


Fig. 2.2.2-I The micro-seeding procedures used for crystallization of *L. casei* FPGS.

In **A**, the microcrystal seeds solution drop was prepared by crushing block crystals as small as possible into microcrystals in the precipitant solution drop which contains the same components as in the solution drop in **B,** but without protein. In **B**, ~15μl of supersaturated protein/precipitant solution was set in a microbridge, which contains 4%-6% (w/v) PEG4000, 40 mM acetate/KOH(pH 5.3), 2~3 mg/ml protein for growing apo-enzyme crystals; 4%~6% (w/v) PEG4000, 40 mM acetate/KOH(pH 5.3), 5 mM ATP, 5 mM $MgCl_2$ and 2~3 mg/ml protein for growing MgATP-FPGS complex crystals. After seeding, the microbridge was sealed in one well of a Linbro plate in **C**, and set in the cold room at 4°C. Both apo-enzyme and FPGS-MgATP complex single crystals appear

in 2 or 3 days, and the crystals reached their optimum size within 2 weeks (see Fig. 2.2.2-II a,b).



(a)

(b)

Fig. 2.2.2-II Crystals of FPGS for (a) MgATP-FPGS, crystal shown in the photo is ~0.5x0.5x0.4 mm; (b) apo-FPGS, crystal shown in the photo is ~ 0.8x0.2x0.1 mm.

The number of single crystals in the microbridge can be generally controlled by the amount of contact between the whisker and the drop. One quick touch of the whisker usually resulted in the growth of 12-15 single crystals, while a prolonged streaking of the whisker typically gave a shower of smaller crystals. Under this micro-seeding method, crystallization was highly sensitive to the concentration of PEG4000, pH, concentration of acetate buffer and, more importantly, purity of the enzyme FPGS.

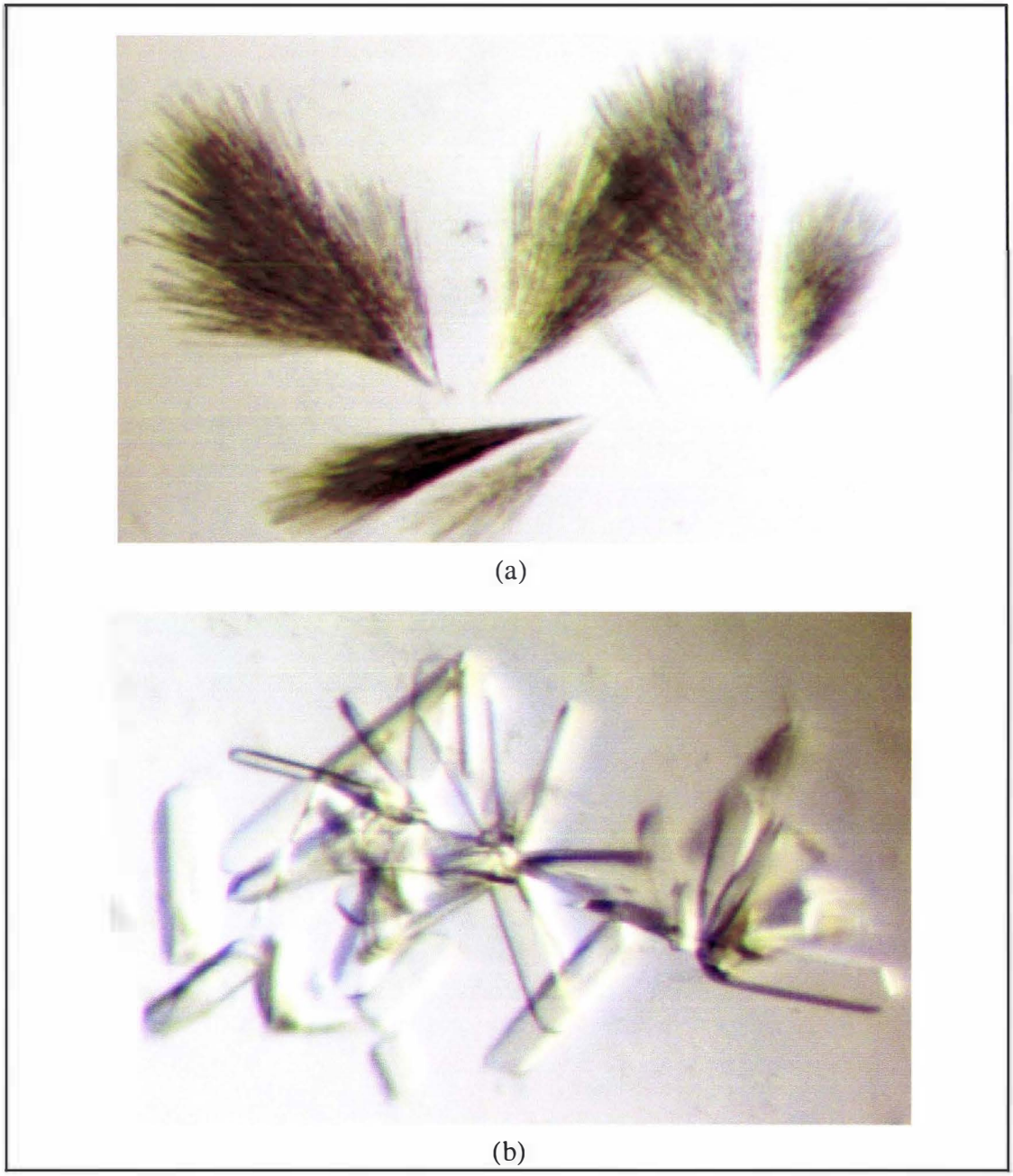


Fig. 2.2.2-III Crystals of apo-FPGS, (a) with degraded enzyme sample and; (b) with repurified protein sample.

When all other conditions were kept the same as described in the micro-seeding procedure, increasing the PEG4000 concentration (>8%(w/v)) resulted in a shower of small crystals; increasing pH (≥ 6.5) resulted in a sharp increase in the solubility of the protein and clear protein/precipitate drop; increasing acetate buffer concentration (≥ 60 mM) resulted in a cluster of long, thin needles (only tested for apo-enzyme).

The crystallization trials showed that the purity of the enzyme is a key factor for growing good quality crystals. None of the optimized crystallization conditions for apo-enzyme or MgATP-FPGS complex were effective when the enzyme sample was degraded to a great extent. With degraded enzyme sample, several clusters of downy crystals (Fig. 2.2.2-IIIa) were grown, which strongly contrasts to the nice single crystals (Fig. 2.2.2-IIIb) grown with repurified protein samples using the method described in Section 2.1.3.

2.3 X-ray data collection and processing

2.3.1 Data collection

Diffraction data from both MgATP-FPGS and apo-FPGS crystals were collected using the oscillation method on an R-axis IIC system (Sato et al., 1992) utilizing the Fuji imaging plates as the X-ray detector and Cu-Kα radiation from a Rigaku RU-200 rotating anode generator operated at 50kV and 100mA. All data were collected at room temperature. Crystals were sealed in a 0.5 mm diameter thin-walled glass capillary with mother liquor at both ends.

Both apo-FPGS and MgATP-FPGS crystals suffered radiation damage, with an accompanying decrease in diffraction intensity. At the time, no cryo-equipment was available to freeze crystals during data collection, a procedure which often greatly retards radiation damage (Nave, 1995). A complete data set for the MgATP-FPGS complex was collected from two single crystals. The data set for apo-FPGS was collected from one single crystal and was 82% complete to 2.5 Å resolution.

The MgATP-FPGS crystals diffracted to a maximum resolution of 2.2 Å in initial still diffraction images, and to 2.4 Å for final data set. The apo-FPGS crystal diffracted

to a resolution of 2.4 Å in the initial still diffraction images, with the final data set extending to a resolution of 2.5 Å. Both MgATP-FPGS and apo-FPGS crystals belong to space group $P2_1$ with cell dimensions a=54.02 Å(54.18), b=46.05 Å(45.89), c=84.85 Å(85.54) β=107.29°(107.82°) for MgATP-FPGS and apo-FPGS crystals (numbers in brackets refer to apo-FPGS). It is clear that the MgATP-FPGS crystals and the apo-FPGS crystals are essentially isomorphous, despite their different morphologies (Fig. 2.2.2-II). Based on the molecular weight of 43 kDa for FPGS, the Matthews coefficient (V_m) was determined to be 2.62 Da/Å^3 for both crystals, assuming one FPGS molecule per asymmetric unit, indicating a solvent content of 53% (Matthews, 1968).

2.3.2 Data Processing

All diffraction images were processed using the program DENZO (Otwinowski, 1993) to generate profile-fitted relative intensities. The scaling and merging of data was carried out using the programs ROTAVATA and AGROVATA from the CCP4 program suite (Collaborative Computational Project No.4, 1994). Structure factor amplitudes were obtained from the intensity measurements employing the Bayesian treatment of French and Wilson (French and Wilson, 1978). This procedure also serves to put the data on an approximately absolute scale via a conventional Wilson plot (Wilson, 1949). The program SCALEPACK (Otwinowski, 1993) was also employed for data sets of apo-enzyme to check the estimates of the mosaicity of the crystal and to obtain reliable cell dimensions.

The data processing statistics are listed in Table 2.3.2-I and Table 2.3.2-II. The data set for MgATP-FPGS is 100% complete with a redundancy of 2.5. The overall $\langle I/\sigma \rangle$ for the final data set is 9.2, and the overall R_{merge} is 7.3%. In the outer resolution shell (2.48-2.40 Å), $\langle I/\sigma \rangle$ is 2.8 with 100% completeness, and R_{merge} is 27.2%. As shown in Fig. 2.3.2-I and Fig. 2.3.2-II, both data sets have a normal intensity distribution over resolution, in which $\langle I/\sigma \rangle$ decreases while R_{merge} increases along with increasing resolution.

Table 2.3.2-I Statistics of data collected for MgATP-FPGS complex

Dmin(Å)	Nmeas	Nref	Comp(%)	Multipl.	I/σ(I)	R_{merge}
9.15	655	259	98.4	2.2	21.6	0.026
6.52	1202	460	98.6	2.4	17.5	0.031
5.34	1580	588	100.0	2.4	16.9	0.039
4.63	1829	687	99.0	2.4	16.1	0.040
3.79	2309	867	100.0	2.4	12.9	0.050
3.28	2750	1023	100.0	2.5	10.0	0.072
2.94	3114	1146	99.9	2.5	6.5	0.115
2.68	3328	1230	99.7	2.5	4.2	0.188
2.48	3602	1323	99.6	2.5	3.1	0.247
2.40	3729	1384	100.0	2.5	2.8	0.272
Total	39117	15866	100.0	2.5	9.2	0.073

Nmeas: number of data measured; Nref: number of unique reflections;
Comp(%): completeness; Multipl.: multiplicity;
$R_{merge}=\Sigma|I_j(hkl)-\langle I(hkl)\rangle|/\Sigma|I_j(hkl)|$, where $I_j(hkl)$ are intensity measurements for a reflection and $\langle I(hkl)\rangle$ is their weighted mean value.

Table 2.3.2-II Statistics of data collected for apo-FPGS

Dmin(Å)	Nmeas	Nref	Comp(%)	Multipl.	I/σ(I)	R_{merge}
9.52	554	207	73.8	2.7	13.6	0.033
6.79	949	339	72.7	2.8	14.5	0.036
5.56	1230	461	78.0	2.7	13.6	0.048
4.82	1445	533	77.2	2.7	14.0	0.047
3.95	1779	684	79.7	2.6	11.7	0.056
3.42	2126	828	83.6	2.6	9.1	0.082
3.06	2402	943	84.6	2.5	5.7	0.136
2.79	2609	1045	85.5	2.5	3.8	0.203
2.59	2800	1150	87.4	2.4	2.7	0.289
2.50	2913	1197	87.3	2.4	2.3	0.331
Total	30029	11786	83.6	2.5	7.8	0.090

Nmeas: number of data measured; Nref: number of unique reflections;
Comp(%): completeness; Multipl.: multiplicity;
$R_{merge}=\Sigma|I_j(hkl)-\langle I(hkl)\rangle|/\Sigma I_j(hkl)$, where $I_j(hkl)$ are the intensity measurements for a reflection and $\langle I(hkl)\rangle$ is their weighted mean value.

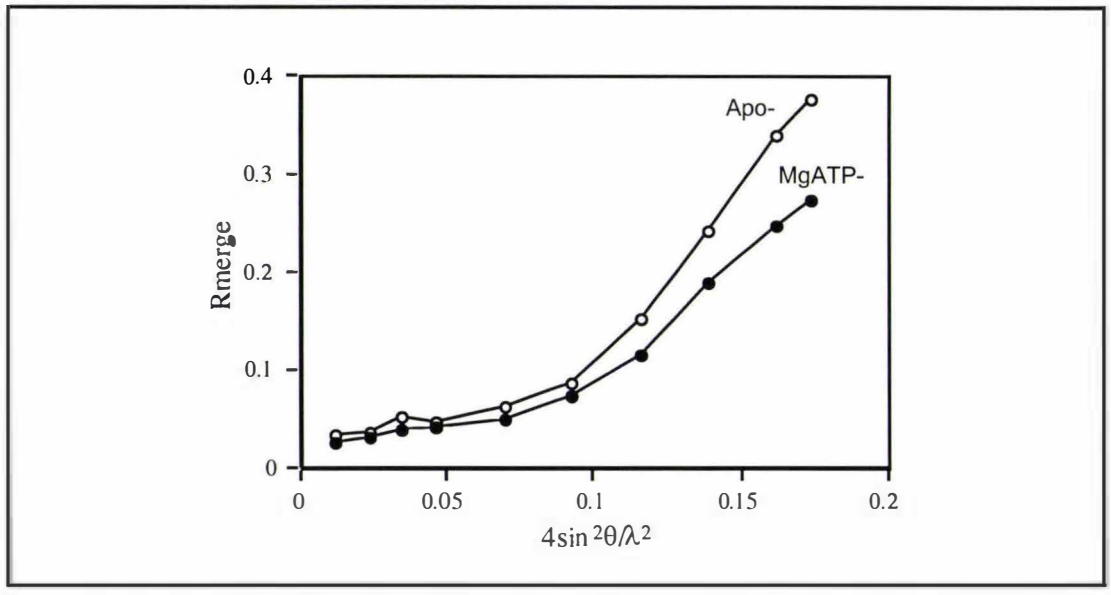


Fig. 2.3.2-I R_{merge} distribution as a function of resolution for both apo- and MgATP-FPGS.

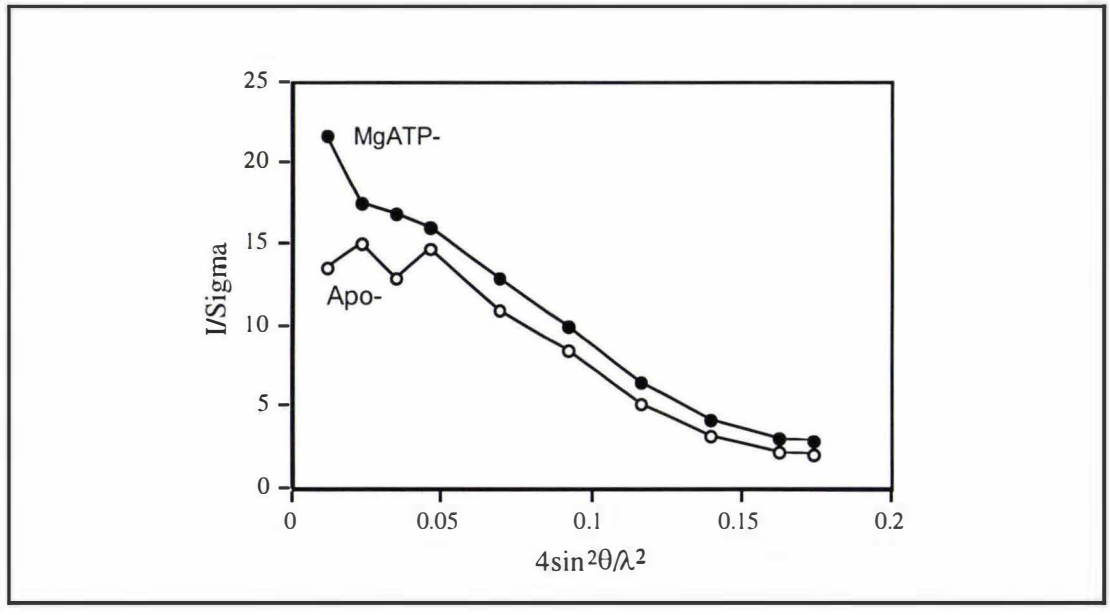


Fig. 2.3.2-II $I/\sigma(I)$ distribution over resolution for both apo- and MgATP-FPGS.

The scaling of MgATP-FPGS to apo-FPGS shows a significant difference in intensity between the two data sets, with R_{iso} of 20.7% ($R_{iso}= \Sigma|F_{ATP}-F_{apo}|/\Sigma F_{apo}$, where F_{ATP} and F_{apo} are the MgATP-FPGS and apo-FPGS structure factors, respectively). The scaling was carried out using the program SCALEIT from the CCP4 program suite.

This difference may result from the absence of Mg^{2+} and the ATP from the apo-FPGS structure and from thermal motion of the loop regions; three disordered loops in the N-domain of the MgATP-FPGS structure were found to be rigid in the apo-FPGS structure (see discussion in Chapter 4). Furthermore, the unit cell volume of the apo-FPGS crystal is approximately 0.6% larger than that of the MgATP-FPGS, indicating that the crystal packing may not be exactly the same between the two crystals. This could also contribute to the R_{iso} value.

Compared to the data for MgATP-FPGS, the data for apo-FPGS, which was processed to 2.5 Å, is significantly weaker. The overall $\langle I/\sigma \rangle$ for final data set is 7.8, and overall R_{merge} is 9.0%. In the outer resolution shell (2.59-2.50 Å), the $\langle I/\sigma \rangle$ is 2.3 with 87% completeness, and R_{merge} is 33.1%. The quality comparison between MgATP-FPGS and apo-FPGS crystals is also shown in Fig. 2.3.2-I and Fig. 2.3.2-II. The data set of MgATP-FPGS is better in both R_{merge} and $\langle I/\sigma \rangle$ than that of the apo-FGPS crystal. Therefore, the MgATP-FPTS complex crystals were used as the native crystals in subsequent MIR (Multiple Isomorphous Replacement) experiments.

2.4 Multiple Isomorphous Replacement

2.4.1 Heavy atom soaking experiments

Although the preparation of heavy atom derivatives remains essentially a "trial and error" process, some general approaches towards heavy-atom attachment to proteins have been discussed (Blundell and Johnson, 1976; Drenth, 1991; Petsko, 1985), based on previous structure analyses. Some biochemical information on the protein usually gives a few clues with regard to suitable heavy atom reagents. The amino acid sequence shows that *L. casei* FPGS contains no cysteine residues, which means that the normal mercuration of cysteine residues is not possible with this enzyme. Therefore, efforts to obtain derivatives were first focused on other commonly used heavy-atom reagents such as platinum, lead and rare earth elements. Under the soaking conditions required by the mother liquor (pH 5.3), the nucleophilicity of histidine would be decreased to some extent, but a large number of carboxylate groups of glutamate or aspartate in the enzyme are still expected to strongly bind most cations. Alternatively, the MgATP binding site

of the FPGS was expected to provide an alternative site for the binding of heavy atom compounds such as 2'-I-ATP, K_2OsO_4 and Na_2WO_4.

The heavy-atom compounds tried in this structural analysis are listed in Table 2.4.1-I. In all of the experiments, native MgATP-FPGS crystals grown in acetate buffer (pH 5.3) were soaked in mother liquor solutions containing the heavy-atom compound at 4°C. All derivative data were collected on the same equipment and under the same conditions used for data collection of native crystals.

Table 2.4.1-I Conditions tested for heavy-atom soaking experiment

Compounds tested	Soaking concentration range (mM)	Soaking time range	No. of data sets collected[a]
K_2PtCl_4	10 - 0.2	1 week - 2 hours	8
$Pt(terpy)Cl_2$[b]	5 - 1	2 weeks - 2.5 hours	4
$Me_3Pb(acetate)_2$	50 -5	14 - 2.5 hours	4
2'-I-ATP	10 - 5	3 weeks - 1 week	2
K_2OsO_4	5 - 2.5	20 - 3.5 hours	2
Na_2WO_4	2	6 hours	1
$NaAuCl_4$	1	10 - 2 hours	2
$Tm(NO_3)_3$	40 - 0.5	20 - 2 hours	4
$K_3UO_2F_5$	5 - 2.5	24 - 11 hours	2
EtHgCl	2.5 - 1	11- -2 hours	3

a. The data were only collected for those crystals diffracting to at least 4 Å.
b. Chloroterpyridylplatinum chloride.

All the platinum soaks resulted in non-isomorphous derivatives with $R_{iso} > 50\%$ ($R_{iso} = \Sigma|F_{PH}-F_P|/\Sigma F_P$), even though the concentration of the heavy-atom compound was decreased to 1 mM at only 2 hours soaking time. Soaking in 2'-I-ATP, an ATP analogue, did not cause any significant intensity differences, which might reflect a steric effect of the large iodine atom. In addition, attempts to co-crystallize FPGS together with the 2'-I-ATP produced only clusters of poor quality crystals. Although derivatives soaked with the tetrahedral anion, OsO_4^{2-} and WO_4^{2-}, have reasonable diffraction intensity changes (R_{iso} ~20%), their difference Patterson maps were very noisy and gave

no indication of heavy atom binding. Four useful derivatives were finally obtained and used in the calculation of experimental phases. Details of these derivatives are given in Table 2.4.1-II. It was noted that the last heavy metal compound tried, EtHgCl, bound to histidine residues of FPGS under the relatively low pH (5.3) and occupied 3 totally different crystallographic sites from the other three derivatives, thus making an important contribution to the initial electron density map.

Table 2.4.1-II Derivative data collection and phasing statistics

	Trimethyl-lead acetate(1)	Trimethyl-lead acetate(2)	Potassium uranyl fluoride	Ethylmercury chloride
Formula	$Me_3Pb(Ac)_2$	$Me_3Pb(Ac)_2$	$K_3UO_2F_5$	EtHgCl
No. of crystals	1	1	1	1
Cell dimensions	a=54.04Å(54.02)*	a=54.24Å	a=53.52Å	a=54.09Å
	b=45.84Å(46.05)	b=45.97Å	b=46.15Å	b=46.10Å
	c=84.84Å(84.85)	c=85.02Å	c=84.69Å	c=85.51Å
	β=107.95°(107.29)	β=107.94°	β=106.8°	β=107.71°
Temperature	ambient	ambient	ambient	ambient
Resolution (Å)	2.5	3.0	3.0	3.0
No. of measured ref.	17080	9264	11019	12415
No. of unique ref.	9145	6290	7456	6156
Completeness (%)	64.5	78.6	93.1	75.1
Multiplicity	1.9	1.5	1.5	2.0
R_{merge} (%)	9.1	9.9	8.3	7.5
R_{ano} (%)	9.9	11.6	13.2	6.7
R_{iso} (%)	23.3	21.1	30.6	15.0
Soaking conc. (mM)	20.0	5.0	5.0	1.0
Soaking time (hours)	2.5	4.0	11.0	2.5
No. of binding sites	3	3	6	3
Phasing Power	1.49 (1.25)	1.17 (0.93)	1.70 (1.25)	1.20 (0.93)
R_{cullis}	0.73 (0.71)	0.81 (0.80)	0.69 (0.73)	0.81 (0.74)

$R_{merge}=\Sigma|I_j(hkl)-\langle I(hkl)\rangle|/\Sigma I_j(hkl)$, where $I_j(hkl)$ is the symmetry equivalent intensity measurement for a reflection and $\langle I(hkl)\rangle$ is their weighted mean value.

$R_{ano}=\Sigma|F_{(+)}^2-F_{(-)}^2|/\Sigma(F_{(+)}^2+F_{(-)}^2)$, where $F_{(+)}^2$ and $F_{(-)}^2$ are Bijvoet pairs.

$R_{iso}=\Sigma|F_{PH}-F_P|/\Sigma F_P$, where F_{PH} and F_P are the derivative and native structure factors respectively.

$R_{cullis}=\Sigma|F_{PH}-|F_P+F_H||/\Sigma|F_{PH}-F_P|$, where F_H is the calculated heavy atom structure factor. The value in brackets is for centric reflections.

Phasing Power= $\Sigma\langle F_H\rangle/E$, where $\langle F_H\rangle$ is the root-mean-square heavy-atom structure factor and E is the residual lack of closure error. The value in brackets is for centric reflections.

* Unit cell parameters for the native crystal are shown in parentheses.

Both R_{iso} and R_{ano} values listed in Table 2.4.1-II show that there were significant intensity changes for all four derivatives, which is a positive indication of heavy-atom binding to the enzyme. At the same time, the unit cell dimensions of the derivatives were all similar to those of the native protein, with a maximum difference of 0.9% in unit cell volume in the case of EtHgCl derivative. This indicates that the derivatives were still isomorphous with the native protein.

2.4.2 Search for heavy atom positions and experimental phases

a. Crystallographic methods

All scaling of the derivative data to the native was performed using the program SCALEIT in the CCP4 suite, according to the method of Kraut et al. (Kraut et al., 1962). A difference Patterson function was calculated with the coefficients $(|F_{PH}|-|F_P|)^2$, where $|F_{PH}|$ is the derivative structure factor amplitude, and $|F_P|$ the native structure factor amplitude. This function is sensitive to lack of isomorphism and shows heavy atom positions on Harker sections (Phillips, 1966). An anomalous difference Patterson function was calculated with the coefficients $(\Delta|F|ano)^2$, which is much more noisy than the isomorphous difference Patterson even though it contains the expected heavy atom vector peaks (Rossmann, 1961). This function provides an useful way of confirming the heavy atom positions deduced from difference Patterson synthesis. The most powerful way to search for minor binding sites of all four derivatives is difference Fourier synthesis. Cross-phase difference Fourier syntheses have coefficients $m(|F_{PH}|-|F_P|)\exp(i\alpha_p)$, where $|F_{PH}|$ and $|F_P|$ have the same meaning as for difference Pattersons, and the figure of merit m and the phases α_p are calculated from one or more other derivatives (Stryer et al., 1964). The cross-phase difference Fourier syntheses are particularly useful if there are many sites of heavy atom substitution, where the difference Patterson syntheses become difficult to interpret. It is important that the phases, α_p, must not be calculated from any model for the heavy atom derivative being tested, because Fourier synthesis tends to be determined by the phases rather than the magnitudes (Dickerson et al., 1967). This was demonstrated by the very noisy difference Fourier peaks after a Fourier synthesis, $m(F_{PH}-(F_P+F_H))$, was calculated for the

$Me_3Pb(Ac)_2$ derivative to try to find the second site, with phases from its first Pb binding site.

All difference Pattersons and difference Fourier syntheses were calculated with the program FFT in the CCP4 suite (Collaborative Computational Project No. 4, 1994). Refinement of the heavy atom positions, occupancies and B factors, and computation of phase estimates was done with the program MLPHARE (Otwinowski, 1991).

b. Heavy atom positions searching

A schematic procedure used for searching for heavy atom positions is shown in Fig. 2.4.2-I. A difference Patterson function was first calculated for each of four derivatives. A major heavy atom position was readily identified from the strongest peak for all four derivatives (the second strongest peak for the EtHgCl derivative was also a Harker peak and identified as another mercury position). Testing of some additional potential Harker peaks as possible minor positions for each derivative indicated that these additional peaks on the Harker section were not heavy atom binding sites as they led to a decrease in the figure of merit (FOM) and a negative occupancy in MLPHARE phase refinement.

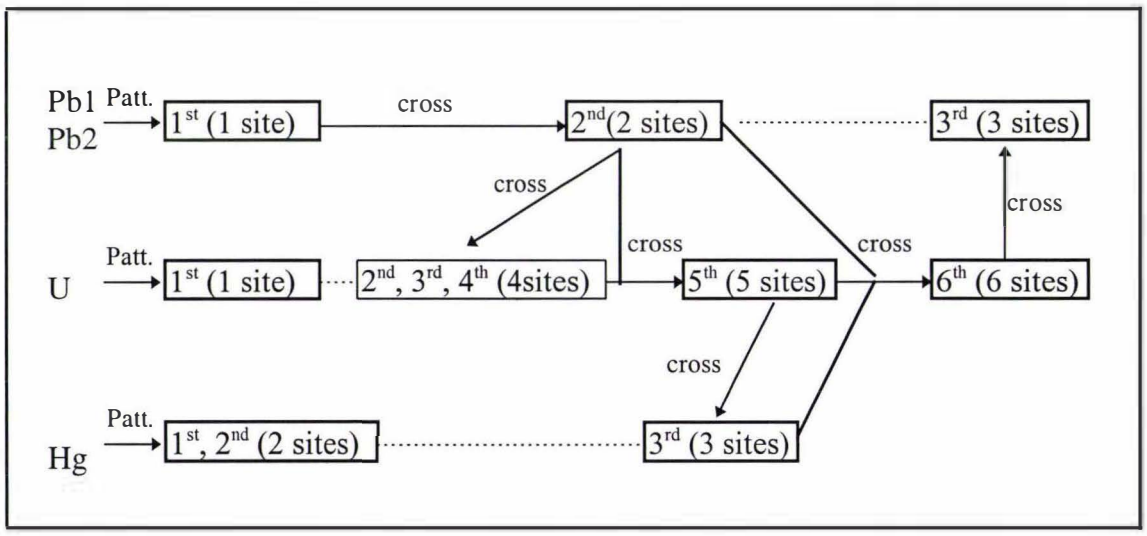


Fig. 2.4.2-I Schematic procedures for searching for heavy-atom positions.
Cross difference Fouriers deduce: 2nd site of lead by 1st site (Pb1,Pb2); 2nd,3rd and 4th sites of uranyl by 2 sites (Pb1,Pb2); 5th site of uranyl by 4 sites (Uranyl) and 2 sites (Pb1); 3rd site of Hg by 5 sites (Uranyl); 6th site of uranyl by 2 sites (Pb1),3 sites (Hg) and 5 sites (Uranyl); 3rd sites of lead by 6 sites (Uranyl).

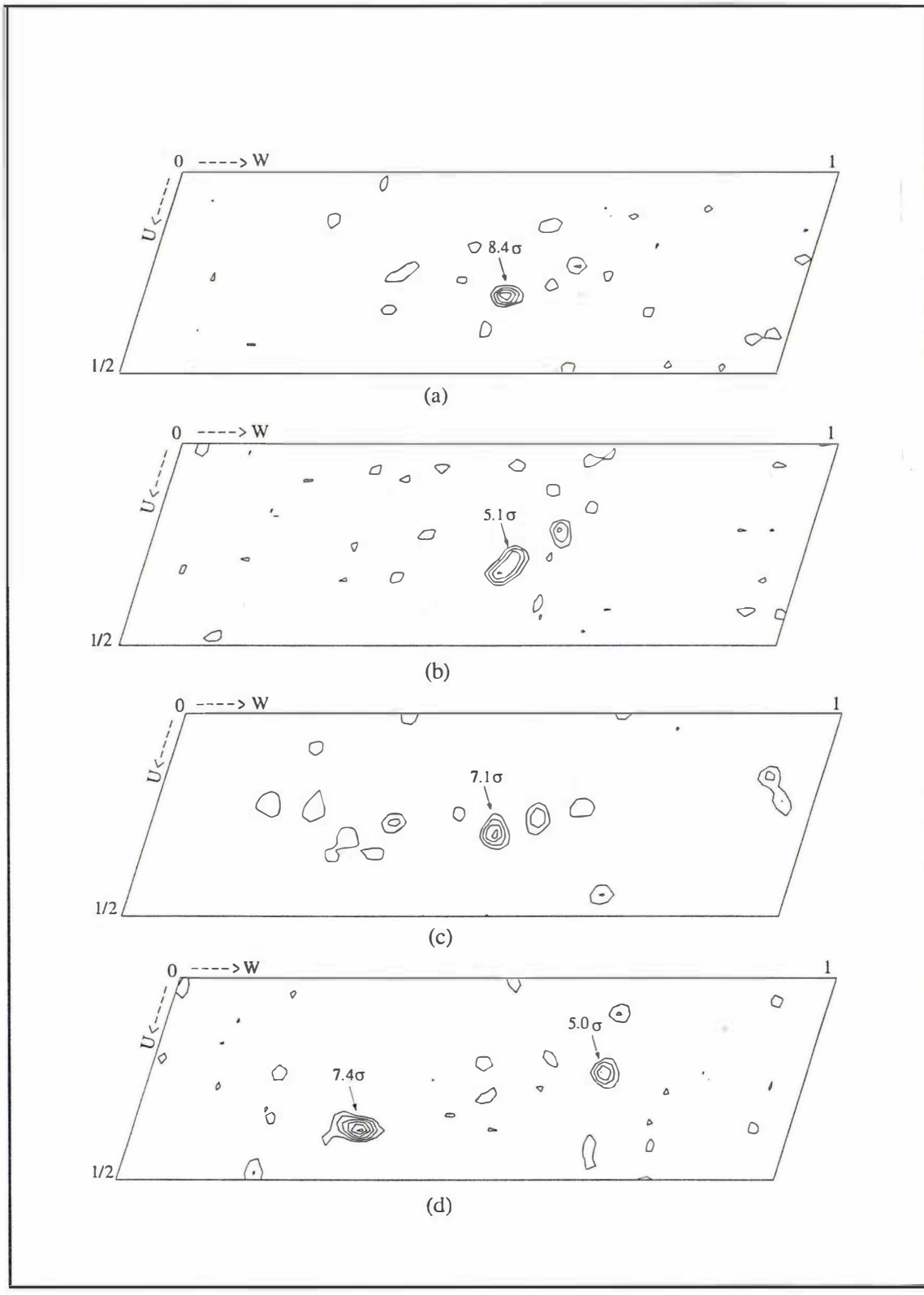


Fig. 2.4.2-II Harker section of the isomorphous difference Patterson for (a) $Me_3Pb(Ac)$-1; (b) $Me_3Pb(Ac)$-2; (c) $K_3UO_2F_5$; (d) EtHgCl. The major positions for each of the derivatives are indicated by arrows. In the case of EtHgCl, the first two strongest peaks indicated correspond to the two major binding positions.

Difference Pattersons both with and without the low resolution terms (d>20 Å) were computed for each derivative to make sure that the strongest peaks assumed to be heavy atom vectors returned in the Patterson map, since heavy atom vector maps can be dominated by low angle terms, and movement of solvent or disordered heavy atoms often results in large differences between F_{PH} and F_P (Blundell and Johnson, 1976).

Harker sections of the isomorphous difference Patterson for all four derivatives are depicted in Fig. 2.4.2-II; the strongest peak indicated by an arrow in each figure corresponds to the major position. Anomalous difference Pattersons were calculated and were unsurprisingly noisy. A peak corresponding to the major site could eventually be found for each derivative, well down in the peak list, which is, however, supportive of the assignment of the major position obtained.

All other heavy atom positions in each derivative were deduced by a series of cross phase difference Fouriers as shown in Fig. 2.4.2-I. In most cases, the difference Fourier syntheses produced both known heavy atom position peaks and some new minor peaks. Most of these minor peaks were cross-checked against the difference Patterson syntheses and were incorporated into phase refinement. The correct minor positions improved the FOM, phasing power and Rcullis.

In the latter stages, cross-phase difference Fouriers were also used to check suspicious minor positions. For example, the third minor positions for both $Me_3Pb(Ac)_2$-1 and $Me_3Pb(Ac)_2$-2 derivatives were originally deduced by using two known sites for each of the two derivatives in cross phase difference Fourier syntheses. However, once the 3 sites for Pb-1 and Pb-2, the 3 sites for Hg and 6 sites for U were included in the refinement, the 3rd sites for both lead derivatives refined to negative occupancies and high B values. They did not appear in the difference Fourier synthesis with cross phases from 6 sites of $K_3UO_2F_5$ derivative. Instead new peaks assumed to be the real third site in each of the two lead derivatives were identified. The new peaks were incorporated into the phase refinement and were confirmed by an improvement in the FOM, Phasing Power and by their positive occupancies and low B factors. It is evident that the first two heavy atoms in both lead derivatives occupy the same positions (see Table 2.4.2-I), indicating that little new phase information has been included in the previous cross phase difference Fourier for lead derivatives.

c. Experimental phase calculation

The final heavy atom parameters and locations are listed in Table 2.4.2-I. The first (U1) site of the $K_3UO_2F_5$ derivative was used to define the origin and all other heavy atom coordinates were related to it through cross difference Fouriers. The refinement of the anomalous occupancies for all heavy atoms was carried out in the later stages of the phase refinement. The resultant anomalous occupancies all turned out to be negative, indicating the incorrect hand of the heavy atom arrangement in the derivatives (Otwinowski, 1991). Therefore, the y coordinates of all heavy atoms have been set negative to choose the correct enantiomorph. This resulted in positive anomalous occupancies for all heavy atoms, and the correct hand of the solution as confirmed in following structural model building.

Table 2.4.2-I Heavy-atom parameters and locations

Heavy-atom	X	Y	Z	Occ[a]	Aocc[b]	B (Å^2)	Binding sites
U1	0.147	-0.249	0.259	0.723	0.633	14.2	E143
U2	0.278	-0.011	0.147	0.618	0.506	16.7	E122
U3	0.112	-0.262	0.317	0.399	0.328	17.6	ADP or ATP
U4	0.590	-0.230	0.374	0.361	0.223	42.2	D401
U5	0.878	-0.031	0.768	0.431	0.341	25.4	E81
U6	0.451	-0.108	0.423	0.437	0.484	45.4	E393, D389
Pb(1)1	0.154	-0.242	0.274	0.618	0.370	4.1	E143
Pb(1)2	0.189	-0.196	0.314	0.275	0.232	11.4	ADP or ATP
Pb(1)3	0.590	-0.220	0.365	0.131	0.085	19.1	D401
Pb(2)1	0.153	-0.240	0.274	0.562	0.474	7.7	E143
Pb(2)2	0.187	-0.166	0.300	0.306	0.279	57.0	ADP or ATP
Pb(2)3	0.237	-0.138	0.326	0.152	0.043	8.8	ADP or ATP
Hg1	0.190	-0.038	0.670	0.501	0.419	40.2	H234
Hg2	0.885	-0.487	0.162	0.292	0.327	13.9	H59
Hg3	0.103	-0.090	0.747	0.130	0.121	14.4	H296

a. Occupancy. b. Anomalous occupancy.

Several heavy atom binding sites of the enzyme are shared by different heavy atoms. U1, Pb(1)-1 and Pb(2)-1 bind to the carboxylate group of E143; Pb(1)-2 and Pb(2)-2 bind to ADP or ATP; U4 and Pb(1)-3 bind to the carboxylate group of D401. All three mercury atoms bind to histidine residues at different sites. All heavy atom positions are confirmed by: (i) most positions having corresponding Harker peaks in difference Patterson maps, although some Harker peaks of the minor positions are very weak; (ii) sites in each derivative are consistent with others through cross difference Fouriers; (iii) every site has reasonable occupancies and B values, and contributes positively to the Phasing Power, Rcullis and figure of merit; and (iv) the sites later turned out to be in structurally reasonable positions.

Final experimental phases were calculated and refined in the resolution range 10-3 Å, giving an overall figure of merit (FOM) for the resultant MIR phases of 0.594.

2.5 Density modification methods

In the past decade, some new methods have been developed to improve phases. They all apply real space constraints based on known features of a protein electron density map in order to improve the approximate phase information obtained from experimental sources. These new methods have also been incorporated into many density-modification techniques. A recent review on density modification has been given (Kleywegt and Read, 1997).

The initial MIR phases (to 3.0 Å resolution) for MgATP-FPGS were improved and extended (to 2.4 Å resolution) using real space density modification procedures (solvent flattening, histogram matching, phase extension). In the early stages of model building, phases from the partial model and the MIR phases were combined to 3.0 Å resolution, and the same density modification procedures then were applied to improve and extend the phases. Some of the density modification procedures used for the MgATP-FPGS complex are described below.

2.5.1 Solvent flattening (Wang, 1985)

Protein crystals contain between about 20% and 80% solvent with a typical solvent content of 50%. Most solvent is disordered in time and space and therefore does not contribute to the measured diffraction intensities. This means that the solvent regions should appear relatively flat and empty. Also the protein regions should have no negative density. Given that, it can be assumed that significant deviations of a real electron density map from these criteria are due to errors. Thus, some of the errors can be eliminated by levelling the density in the solvent regions and truncating negative intensities in the protein regions. The modified map is then back transformed and the resulting phases, appropriately weighted, are combined with the original phases. In solvent flattening, the density in the solvent regions is typically replaced by its average value at each cycle of density modification.

The solvent flattening procedure for MgATP-FPGS was implemented in the DM program suite (Cowtan, 1994). The program used the automatic Wang molecular envelope mask calculation based on input solvent content value.

2.5.2 Histogram matching

In the histogram matching technique (Lunin, 1993; Zhang and Main, 1990), it is assumed that the distribution of density values for a protein is known and depends only on resolution and a temperature factor (B). In this way, a match can be found which transforms the histogram of density values from the current electron density into a known one. Prior information concerning protein density distribution is therefore incorporated during phase refinement. Since the histogram matching procedure is applied to points in the protein region, solvent flattening together with the histogram matching provide real space constraints on the electron density over the entire volume of the protein crystal.

Histogram matching for MgATP-FPGS was implemented together with solvent flattening using DM (Cowtan, 1994). The calculation of scale factors and B-factors for the data are automatic. This is performed by comparison with an empirically derived

database of map variance at different resolutions, and is more reliable than the conventional Wilson plot.

2.5.3 Phase extension

Phase extension was first used in the structure determination of tobacco mosaic virus (TMV) (Bloomer et al., 1978). This technique has also been incorporated into various density modification programs to propagate phase information from low resolution to a higher resolution that is typically near the diffraction limit. A concise formulation of the theory of phase extension has been given (Lawrence, 1991).

The phase extension technique in the program DM (Cowtan, 1994) was used for phase extension of the data. A specific "COMBINE OMIT" mode, in which a reflection omit calculation is used to reduce dependency between initial and modified structure factors, was chosen for the phase extension and combination of modified structure factors. All unphased data were introduced during phase extension by taking thin isotropic resolution shells of data and adding them to the current set of phased data. The shell width at different resolutions was adjusted so that approximately equal numbers of reflections were added at each step. For the MgATP-FPGS data, there were 8451 of 15725 reflections initially phased by MIR. Phase extension was from 2.99 to 2.40 Å, and the shells were of width 0.04-0.02 Å over the extension range. This resulted in the introduction of between 300 and 400 reflections at each extension step.

2.5.4 Phase combination with MIR phases

Another frequently used phase improvement procedure is phase combination. The phases calculated from the partial structure in early modelling stages were combined with the MIR phases at 3.0 Å resolution, using the program SIGMAA (Read, 1986).

It was proposed (Rossmann and Blow, 1961) that the combination of different sources of phase information could be realized by simple multiplication of the corresponding phase probability distributions using the following formula:

$$P(\alpha) = P_{iso}(\alpha)P_c(\alpha) \qquad \text{Eq. 2.1}$$

where $P(\alpha)$ is the combined phase probability distribution, and $P_{iso}(\alpha)$ and $P_c(\alpha)$ are the isomorphous and partial structure phase probability distributions, respectively. In SIGMAA, a computationally convenient phase combination method is used, depending on the Hendrickson and Lattman (Hendrickson and Lattman, 1970) formulation of the phase probability profile for a phase α:

$$P(\alpha) = \exp(A\cos\alpha+B\sin\alpha+C\cos2\alpha+D\sin2\alpha) \qquad \text{Eq. 2.2}$$

where A, B, C, D are the phase coefficients. In this way, the phase information from different sources can be combined by a simple addition of the phase coefficients from each determination.

The phase probability distributions from MIR were calculated by the method of Blow and Crick (Blow and Crick, 1959), and the phase probability distributions for the partial model structure were based on a weighting scheme proposed by Sim (Sim, 1959; Sim, 1960). As reported in other structure determinations (Rice, 1981; Rice et al., 1988), this phase combination technique proved very useful in the density modification of the MgATP-FPGS structure.

2.5.5 Overview of approach taken

The density modification techniques described above were used in a cyclic manner. The initial map with MIR experimental phases was first modified by solvent flattening, histogram matching and phase extension in DM. After the first partial structural model was obtained, phase combination of the MIR phases with the phases calculated from the partial model was carried out to produce much better maps upon which the DM density modifications were again applied. The same procedures were iterated until a virtually complete structural model had been obtained. The striking improvements of density maps by the DM and phase combination techniques for the MgATP-FPGS structure can be clearly seen in Fig. 2.5-I.

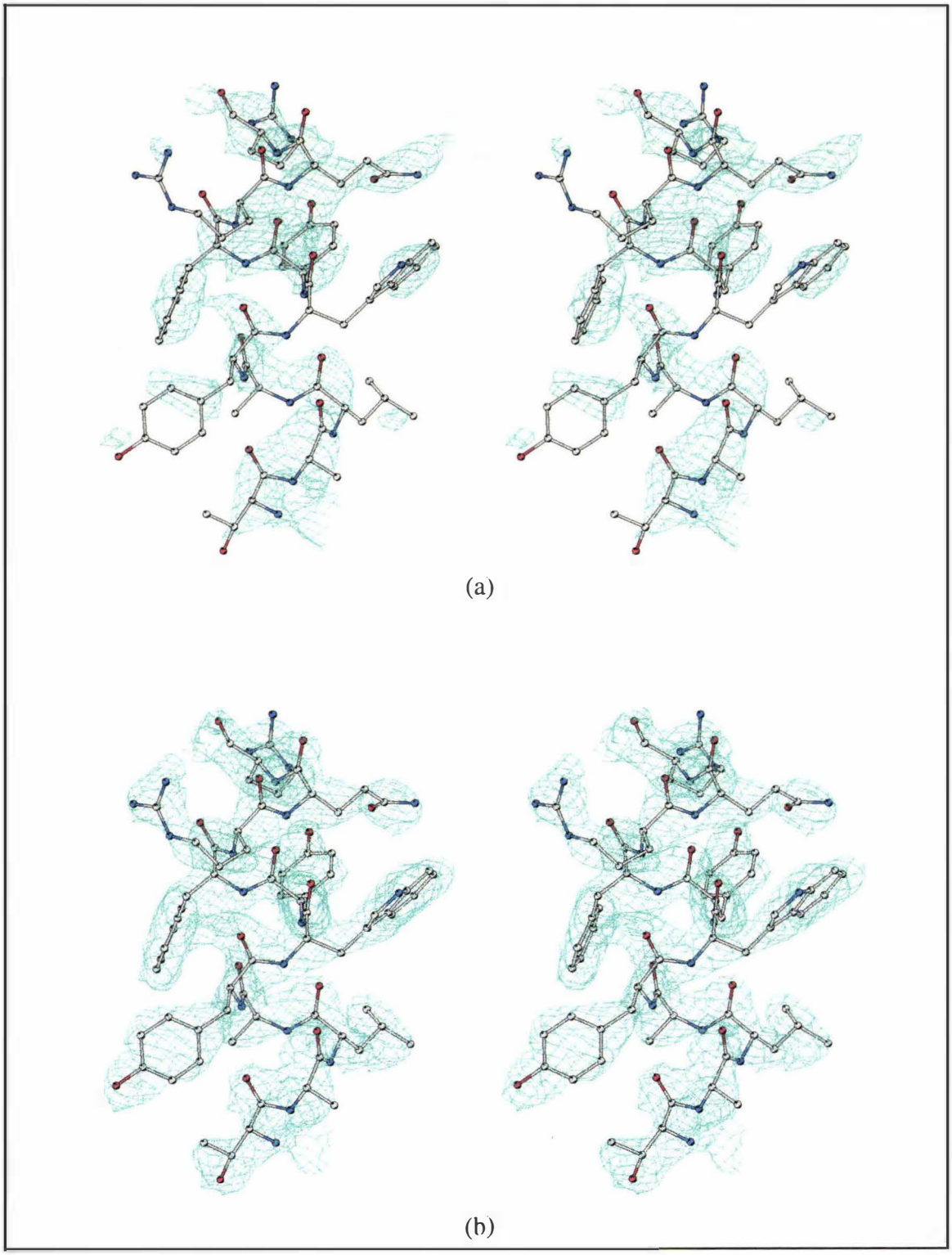


Fig. 2.5-I The improvement of electron density map by the DM and phase combination techniques for the MgATP-FPGS structure. The density maps for part of helix A5 are calculated from, (a) the initial MIR phases at 3.0 Å resolution; (b) the combined phases after density modification and extension to 2.4 Å. This region contains the four consecutive aromatic residues YWYF (129-132) identified firstly in the sequence assignment (see Section 2.6.4). The density maps are contoured at 1.5 σ.

2.6 Structural model building and refinement

The treatment of the initial MIR map with density modification techniques resulted in a partially interpretable map from which the first partial model was obtained. Successive phase combination and model refinement allowed tracing of the polypeptide chain of the protein and building of side chains. Sequence assignment was then established unambiguously. A flow diagram illustrating the processes of structural model building is shown in Fig. 2.6-I.

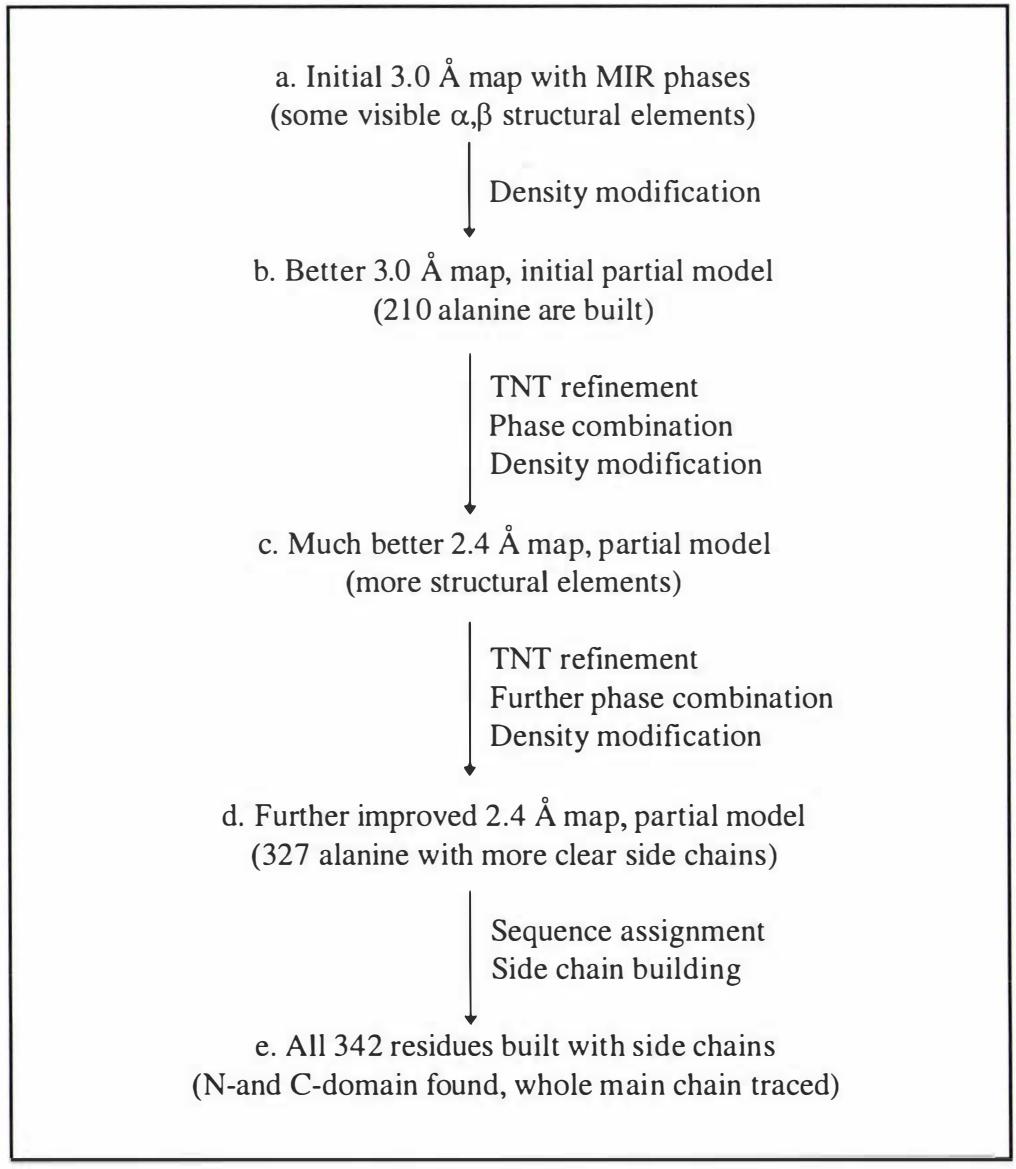


Fig. 2.6-I Flow diagram illustrating the processes of structural model building.

2.6.1 Initial model building (a and b in Fig. 2.6-I)

Once the heavy-atom positions had been obtained for both uranyl and lead derivatives, an electron density map with MIR phases was calculated (FOM 0.53 in resolution range of 10-3.0 Å) and density modification applied. The resultant map showed five short β-strands and an α-helix, but density was often broken. In general, it was a poor quality map, and no attempt was made to build a model into it. The inclusion of phase information from the mercury derivative into the MIR phases resulted in a significant improvement of the map quality, even though the FOM was not greatly increased. As discussed earlier, the three different mercury positions compared with uranyl and lead provided entirely new phase information, which was responsible for the map improvement.

From the inspection of the electron density map obtained after density modification (FOM 0.594 in the resolution range 10-3.0 Å, followed by phase extension to 40-2.4 Å), twelve β-strands and seven α-helices were readily identified. All helices were right-handed, confirming the correct hand of the MIR solution (DeGraw et al., 1982). Although the complete tracing of the polypeptide chain of the molecule was still not possible at this stage, the first initial structural model, consisting of polypeptide backbone in the regions of secondary structural elements, was built. The model building was performed on an SGI workstation using the molecular graphics program TURBO-FRODO (Roussel, A., Inisan, A.-G. and Cambillau, C., AFMB; Biographics, Marseilles, France).

Polyalanine fragments were placed in the electron density map, corresponding to regular α-helices and β-strands. There were 210 alanine residues built in this initial model. During this process, it became apparent from the outline of the molecule that the protein consists of two domains, one larger and another smaller.

Not surprisingly, broken density was seen in the initial electron density map, even in the regions of β-strands and α-helices. Some strong density was evident between some of the β-strands, resulting from hydrogen bonding contacts, yet the direction of the strands was often unknown. Special caution was taken to cope with the poorly-defined density, and in all cases prior knowledge of regular secondary structure elements was used to ensure that the model was built with a reasonable conformation.

2.6.2 Initial model refinement

A restrained least squares refinement method was used to refine the initial partial model. This refinement was carried out using the TNT program suite (Tronrud, 1992; Tronrud et al., 1987). The RFACTOR module was used during the refinement, in which the function minimized is:

$$f(x) = \sum_{hkl} 1/\sigma_o^2(hkl) \bullet (|F_o(hkl)| - |F_c(hkl, \chi)|)^2 \qquad \text{Eq. 2.6-1}$$

where $|F_o(hkl)|$ and $|F_c(hkl, \chi)|$ are the observed and calculated structure factor amplitude, respectively. χ is shift vector and $\sigma_o^2(hkl)$ is standard deviation of the measurement.

In the early stages of initial model refinement, the resolution range was chosen to exclude both the low resolution reflections and weak high resolution reflections, e.g., 10-2.6 Å. The overall temperature factor (B) was kept constant and set close to that calculated from the Wilson plot. Ksol and Bsol (scaling factors for bulk solvent correction, see Section 2.7.4) were set at 0.8 and 150 Å^2, respectively, which are normal values for most protein structures. This was because the program could not calculate these parameters accurately in the initial stages because of the incompleteness of the model and exclusion of low resolution data.

Suitable weights for bond length and angle restraints were chosen by running some cycles of refinement and adjusting the weights to ensure that the rms deviations for both bond lengths and angles were reasonable. The RFACTOR weight was varied during the process of the refinement. Since the initial model was constructed by building alanine residues into the density map, such models generally have many misplaced atoms and many places where there is poor agreement with the geometric restraints. Thus, large shifts will be required to correct the model. It has been suggested (Tronrud, 1994) that in such a situation the best procedure is to adapt the model to the diffraction data first, and then bring the model into agreement with ideal geometry. This is done by running a series of cycles with a high RFACTOR weight and another series with a lower weight. In the case of MgATP-FPGS, an RFACTOR weight of 0.01 was set for the first series of cycles, then values of 0.05 and 0.001 were chosen in later cycles, such that the rms

deviations for bond lengths and angles at the end were generally near the target values (0.02 Å for bond lengths, 3.0° for bond angles).

As an alternative, a real space refinement module is also included in the TNT program (Tronrud, 1992). Real space refinement was mainly used for the initial model refinement, in which the function minimized is

$$f(x) = \Sigma_{space}(\rho_o(r) - \rho_c(r,\chi))^2 \qquad \text{Eq. 2.6-2}$$

where ρ_o is the observed electron density, ρ_c the calculated electron density from partial model, r a model vector, and χ an shift vector. The phases were fixed during the refinement process.

2.6.3 Phases and initial model improvement (c and d in Fig. 2.6-I)

After the least squares refinement was carried out, the initial model was then used to derive partial model phase information. This initial model contained 210 alanine residues (out of a total of 428), constituting about half of the number of residues included in the final model. A significant proportion of the initial electron density map needed to be improved and interpreted. This was done by combination (see Section 2.5.4) of the original MIR phase information with phases obtained from the partial structural model (MIR phases of data to 3.0 Å were combined in each phase combination procedure). The map calculated with the combined phases was then further improved using density modification and phase extension (to 2.4 Å resolution) procedures described previously (see Section 2.5). The map calculated with these phases showed new electron density corresponding to α-helices, β-strands, and some connecting loops. After more residues were built in, the new partial model was refined by the restrained least squares using TNT (Tronrud, 1992; Tronrud et al., 1987), in order to improve the agreement with the experimental observations. This procedure was repeated for two more cycles, and most of the missing residues were built in.

A total of 327 alanine residues in 18 fragments were built into the partial model during the cycles of rebuilding and phase combination. The quality of the map with combined phases showed significant improvement, which was reflected in the increased

mean FOM (0.594 for initial model in (b), 0.679 for partial model in (c) and 0.759 for partial model in (d) of Fig. 2.6-I). Some wrongly built residues were identified and corrected. In the last cycle, the density for some side chains became clearer, and many residues such as tryptophan and phenylalanine were readily visible in the map.

2.6.4 Sequence assignment (e in Fig. 2.6-I)

Although the partial model consisted of many fragments, the electron density map was good enough to allow the alignment of the sequence with the partial structure. The most obvious side chain was tryptophan with its large flat electron density. There are a total of 6 tryptophan residues in the FPGS sequence: Trp130, Trp219, Trp236, Trp281, Trp297, and Trp391. Most of these side chains were well defined and identified by their characteristic density. In addition, other aromatic side chains such as tyrosine, phenylalanine and histidine were quite often used to confirm the correct sequence. Proline residues, which are often on turning points of the main chain and have side chain density that bulges from the main chain, also contributed to the sequence assignment.

The well-defined Trp130 was the first residue to be assigned to the model. This residue appears in the sequence (···GYWYFR···), residues 128-133. This section of the sequence was readily identified through the typical densities of both the tryptophan and its neighbouring aromatic residues. From this starting point, the sequence was then traced in both directions, with some connecting residues between fragments being built into the model. Both aromatic side chains and proline residues were frequently used to check the sequence assignment. The second well-defined Trp219 was identified by its characteristic density and confirmed by several nearby aromatic side chains. This resulted in the complete sequence assignment of the longest fragment (residues 177-314) despite the density for several connecting residues being quite weak. At this stage, the sequence for most of the residues in the N-terminal domain had been assigned except for a short helical fragment which was separated from the rest of the structure. However, it was able to be assigned as residues 1-14 based upon its helical structure (see discussion below) and a particular sequence (···YIHSFPR···).

Trp391 was recognized in the small C-terminal domain and the residues 390-425 were subsequently assigned. Two short β-strands in the C-domain were separated from any other nearby residues. They were identified as residues 332-340 and residues 361-368, based on their β-strand structure (see discussion below) and density particular to proline residues which exists in both strands.

```
MNYTETVAYIHSFPRLAKTGDHRRILTLLHALGNPQQQGRYIHVTGTNGKGSAANAIAHV  60
-----HHHHHH----H------HHHHHHHHH----------EE---------HHHHHHHH

LEASGLTVGLYTSPFIMRFNERIMIDHEPIPDAALVNAVAFVRAALERLQQQQADFNVTE 120
HH----EE-------EHHH--HEEE----------HHHHHHHHHHHHHHHHH---------

FEFITALGYWYFRQRQVDVAVIEVGIGGDTDSTNVITPVVSVLTEVALDHQKLLGHTITA 180
-EHHHHHHHHHH------EEEEEE------------EEEEEEEHHHHHHHHH--HHHHH

IAKHKAGIIKRGIPVVTGNLVPDAAAVVAAKVATTGSQWLRFDRDFSVPKAKLHGWGQRF 240
HHHH----EE---EE---------HHHHHH-HHHHH-HHHHH--------------HHH

TYEDQDGRISDLEVPLVGDYQQRNMAIAIQTAKVYAKQTEWPLTPQNIRQGLAASHWPAR 300
-H-------------------HHHHHHHHHHHHHHH----------H-HHHHHHH---HH

LEKISDTPLIVIDGAHNPDGINGLITALKQLFSQPITVIAGILADKDYAAMADRLTAAFS 360
HH------EEEE----------HHHHHHH-----EEEEE-----HHHHHHHHHHHHHHH-

TVYLVPVPGTPRALPEAGYEALHEGRLKDSWQEALAASLNDVPDQPIVITGSLYLASAVR 420
--EE----------HHHHHHHHH----HHHHHHHHHH---------EEEE--HHHHHHHH

QTLLGGKS 428
HHH-----
```

Fig. 2.6.4-I The secondary structures predicted for *L. casei* FPGS. H = helix, E = strand, - = no prediction.

A useful reference used during this sequence assignment was theoretical secondary structure prediction. Several different prediction programs were run and the program NNPREDICT, which is based on a neural network (Kneller et al., 1990), was finally chosen because most of its secondary structure predictions were demonstrated to be in agreement with our sequence assignment. The secondary structures predicted for *L. casei* FPGS are shown in Fig. 2.6.4-I. The structure prediction proved to be quite useful, particularly for the short α-helix or β-strand fragments, as some were isolated from the rest of the structure and contained no clearly recognisable side chains. However, their α-helical or β-strand structures provided a clue to their sequence

assignment when the secondary structure prediction was taken into account. As mentioned above, after most of the sequence for the N-terminal residues (23-314) was assigned, a short isolated α-helix fragment remained. It was obvious by referring to the secondary structure prediction (Fig. 2.6.4-I) that this α-helical fragment could be assigned as residues 1-14. Similarly, there were two isolated short β-strand fragments in the C-terminal domain, which came between residues 315-390 (all other residues having been assigned). The secondary structure prediction (Fig. 2.6.4-I) indicated that there were three α-helices and two short β-strands predicted in this region. Attention was then focused on the two predicted β-strand regions, and the two isolated, short β-strand fragments were readily assigned.

Clearly, accuracy of the prediction is essential for the successful application of this method. It was noted from the results of the predictions for FPGS using different programs that the accuracy varies with the program used. Fortunately, NNPREDICT worked very well in this case.

Following sequence assignment, there were a total of 342 residues in the partial model for which the side chains were built unambiguously. The direction of several short fragments of β-strand had to be reversed so that they were in agreement with the sequence. The sequence assignment for the linker region between the N-terminal and the C-terminal domain, was checked carefully and confirmed the correct location of both domains. The N-terminal domain of the molecule comprises residues Met1-His296; with the C-terminal domain from residues Trp297-Ser428.

2.7 Model rebuilding and final refinement

The structural model that resulted after sequence assignment was largely complete, although there were three short breaks totalling 20 residues in the N-terminal domain and two long breaks containing 38 missing residues in the C-terminal. The crystallographic R-factor for this partial model was 30% for data in the range of 30-2.4 Å. However, further model refinement and rebuilding were required to correct many misplaced residues and locate the missing residues. All refinement was carried out using the restrained least squares program TNT (Tronrud et al., 1987), with geometry

restraints (Engh and Huber, 1991). The general refinement strategy, as described above (Section 2.6.2), was to initially relax restraints on the geometry, and then tighten them to bring the structure back to an ideal geometry. However, the degree of both relaxing and tightening was varied with progress of the refinement so that the Gradient Length (Tronrud, 1994) for both the RFACTOR and GEOMETRY modules was kept at the same level. The resolution range of data for refinement was gradually extended to the high resolution limit and then to low resolution in the latter stages. The free R-factor (Kleywegt and Brunger, 1996) using a randomly selected 5% of the data, was monitored during the course of refinement. Individual temperature factors were kept constant in the early stages and were refined when the crystallographic R-factor dropped below 25%. At first two B-factors per residue were refined (for the main chain and side chain atoms respectively), then individual isotropic B-factors were employed in the final stages of refinement, using B factor restraints (Tronrud, 1996a).

At the end of each round of refinement, difference Fourier electron density maps, both $2F_o$-F_c and F_o-F_c, were calculated for model rebuilding. Initially, two sets of density maps were calculated using (i) phases from the partial model only and (ii) combined phases (partial model + MIR). The maps were used at the same time in the early stages of model building because the map calculated with combined phases showed more new density. At later stages, phase combination was not carried out, and the calculated phases were used directly. Model rebuilding was carried out using the TURBO-FRODO.

In all iterative cycles of rebuilding and refinement, difference Fourier syntheses were calculated to produce new electron density maps for subsequent model rebuilding. Maps were computed with Fourier coefficients of the form $(2m|F_o|-D|F_c|)$ and $(m|F_o|-D|F_c|)$ using SIGMAA weighting (Read, 1986), where $|F_o|$ is the native structure factor amplitude, $|F_c|$ is the calculated structure factor amplitude, m and D are the figure of merit and the coordinate errors estimated for the partial structure, respectively; they are defined by Read (Read, 1994).

2.7.1 Omit map and model errors correction

At this stage, it was noted that there was significant model bias in the maps, with electron density for poorly defined side chains disappearing after the side chain was

removed from the partial model. Hence, a series of omit maps, in which regions with poor density and high B values, or any other regions in doubt were omitted from the model, were calculated. Difference Fourier maps were used to check model errors and establish the correct conformation or position of poorly defined side chains and loops without interference from model bias (Bhat, 1988). Major rebuilding of the structure was carried out by utilizing such omit maps during the earlier stages of model refinement and rebuilding.

A number of model errors were then readily identified and corrected by inspection of the omit maps. They included some wrongly built configurations of peptides, misplaced side chains and wrong directions of two short β-strand fragments. For example, the short β-strand fragment in the N-terminal domain (residues 157-165) has the sequence TPVVSVLTE. Residue Ser161 was built into the density correctly, but the other residues had previously been built in the wrong direction, which presumably resulted from the similarity in side chains (Val160 to Val162, Val159 to Leu163, Pro158 to Thr164), relatively weak density and the disordered sidechain of Glu165. Also the model bias contributed to conceal the poor fit of the model to the density except for the residues at both ends of the fragment. An omit map (all residues of the fragment were omitted) showed clear electron density for this region and the residues were then unambiguously fitted. As another example, residues Ser73 and Pro74, which are both close to the ATP binding site and located in a well-defined region, did not fit particularly well in the density. An F_o-F_c density map did not give a clear indication of the correct conformation. This problem was finally solved through the use of an omit map in which residue Pro74 was able to be fitted into well-defined density in a cis-peptide configuration.

In later stages of model rebuilding, a model with less errors resulted in much clearer difference Fourier maps. This made it possible to identify residues which might have been sequenced incorrectly or mutated during expression in *E. coli*. Residue Gly128 always showed extra positive density in F_o-F_c maps near the α-carbon atom. The substitution Gly128→Ala128 was made, followed by cycles of refinement, in which the side chain of Ala128 was fitted into strong density (at 2σ level in $2F_o$-F_c map). This assignment was also confirmed in an omit map with whole residue omitted. Therefore,

Ala128, instead of Gly128, is identified as the correct sequence, from the crystal structure.

2.7.2 Location of the phosphate binding site

The location of the phosphate binding site was not established until the later stages of the refinement, although a difference map calculated between the MgATP-FPGS and apo-FPGS data sets showed some very strong density (>10σ) in the cleft between the N- and C-terminal domains. The two largest peaks in the F_o-F_c maps were consistently in same place and about 3.3 Å apart. The shape of the electron density in both $2F_o$-F_c and F_o-F_c maps resembled a pyrophosphate (PPi), with the remainder of the nucleotide not immediately visible. Some weak, discontinuous density extended from the PPi and projected into the solvent region of the cleft. This density could account for the ribose and adenine groups, however these have not been modelled at this stage. A third relatively weak peak near the two phosphate peaks was putatively assigned as the third phosphate, but it has not been modelled either. High B values for the PPi suggested that the nucleotide probably had only partial occupancy, and this was confirmed by further refinement using an occupancy of 0.5 for the PPi atoms. This resulted in B values similar to the surrounding protein atoms and well-defined density for the PPi. Two possible reasons can be given for the low occupancy of the nucleotide. Firstly, the MgATP-FPGS crystals were grown at pH 5.3. It is likely that the ATP could be hydrolyzed significantly at low pH (Miller and Westheimer, 1966), leading to a low concentration of ATP in solution. Secondly, at this pH (5.3) the enzyme is essentially inactive. The affinity of FPGS for MgATP is relatively low (K_m=5.6 mM), even at the optimum pH (9.7) (Bognar and Shane, 1983), and may be lower at pH 5.3. These factors would contribute to poor nucleotide binding to FPGS in acidic conditions.

On the opposite side of the PPi, density for a potential metal ion was identified. This has been putatively modelled as Mg^{2+} with an occupancy of 50%.

2.7.3 Modelling ordered water structure

In protein crystals, solvent can occupy between ~20% and ~80% of the crystal (Matthews, 1968). Some solvent molecules are well ordered such that their electron

density can be readily resolved by both X-ray and neutron diffraction methods (Karplus and Faerman, 1994; Levitt and Park, 1993). These ordered solvent molecules make significant contributions to the structure factors, and are generally included in a structure determination.

The ordered solvent molecules were not included in the structural model until late in refinement, by which stage a lot of potential solvent sites, particularly inside the protein molecule, were clearly seen. The solvent molecules, all treated as water, were fitted into peaks in difference Fourier maps (typically contoured at 2σ in F_o-F_c maps; 1σ in $2F_o$-F_c maps). Each potential water molecule was inspected on the graphics terminal and included into the model only if it made hydrogen bonds to protein atoms or other water molecules. Those water molecules in regions where the protein structure was still ambiguous, i.e., poorly defined or not yet built, were not included.

It is believed (Otting et al., 1991) that the ordered water molecules in proteins have short occupancy times at a specific site, and are constantly in rapid exchange with the bulk solvent. During refinement, the occupancy of the water molecules were set at 1.0, and their isotropic B factors were refined. There were no restraints placed on the positional or B factors of the water molecules. Those water molecules for which both the B factor refined to high values (>70 $Å^2$) and the density was poor, were removed, and the water model was re-evaluated after another series of refinement cycles. In the final round of refinement, several water molecules with high B factors (>70 $Å^2$) but reasonable electron density were set to an occupancy of 0.5.

2.7.4 Bulk solvent correction

In contrast to the ordered water structure, most solvent is disordered in the solvent regions between the protein molecules in the crystal lattice. The effective scattering from the disordered solvent regions (known as bulk solvent), particularly at low resolution, causes large discrepancies between the observed and calculated structure factors, leading to calculated structure factor amplitudes which are systematically much larger than the observed structure factor amplitudes at a resolution below ~5 Å (Vijayan, 1980). This systematic deviation leads to severe problems in scaling, in least-squares refinement, and in difference Fourier map calculations. It was common practice in the

past to avoid these problems by omitting the low resolution data during refinement. However, doing so causes series termination errors in Fourier syntheses (Tronrud, 1996b).

Two bulk solvent models are currently available to solve this problem. One of them is the exponential scaling model based on the application of Babinet's principle (Moews and Kretsinger, 1975). This simple model can be expressed in a computationally convenient fashion, and thus it is implemented in most of the crystallographic refinement programs, including TNT (Tronrud, 1992; Tronrud et al., 1987). Therefore this approach was first adopted in the refinement of the FPGS structure. By application of Babinet's principle, it is assumed that the structure factors of the bulk solvent electron density are directly proportional to the structure factors of the protein electron density with strictly opposite phases. From this approximation, the total calculated structure factor amplitude can be derived as follows (Moews and Kretsinger, 1975):

$$F_c = F_m\,[1-K_{sol}\exp(-B_{sol}\sin^2\theta/\lambda^2)] \qquad \text{Eq. 2.7-1}$$

where F_c is the calculated structure factor amplitude of the protein model including bulk solvent model, and F_m is the calculated structure factor amplitude of the protein model. The scale factor K_{sol} reflects the ratio of the solvent electron density to the protein electron density with typical values 0.75-0.95 and B_{sol} restricts the down-scaling of F_m to resolutions below ~5 Å with typical values 150~250 Å^2. Thus the bulk solvent modelling has been reduced to a scaling problem. With an absolute scale factor K, and an overall B factor, F_o can be scaled to F_m by the formula:

$$F_o = 1/K\bullet\exp[-B\bullet\sin^2\theta/\lambda^2)F_m(1-K_{sol}\exp(-B_{sol}\bullet\sin^2\theta/\lambda^2)] \qquad \text{Eq. 2.7-2}$$

All four scaling parameters (K, B, K_{sol}, B_{sol}) were used in program TNT for the refinement of the FPGS model. The estimates of the scale factor K and overall B factor are based on Wilson statistics (Wilson, 1942). K_{sol} was determined by the program and B_{sol} was set as 150 Å^2 in earlier stages of refinement without including low resolution data. Both scale factors were calculated by the program in the later stages of refinement, with all data (40.0-2.4 Å).

Due to the assumption of strictly opposite phases, this exponential scaling model underestimates the contrast between protein and bulk solvent electron density. Another more realistic mask model has also been developed and is implemented in X-Plor version 3.851 (Jiang and Brunger, 1994). In this mask bulk solvent model, the protein molecules are placed on a grid in the unit cell, and all grid points outside the protein region are filled with bulk solvent. The calculated structure factors of the bulk solvent electron density, scaled with a factor K_{sol} and smoothed with an artificial temperature factor B_{sol}, are then vectorially added to the calculated structure factors of the protein model to give the total calculated structure factors as follows (Kostrewa, 1997):

$$\overrightarrow{Fc} = \overrightarrow{Fm} + K_{sol}\,\overrightarrow{F_{sol}}\exp(-B_{sol}\,\mathrm{Sin}^2\theta/\lambda^2) \qquad \text{Eq. 2.7-3}$$

The two scaling parameters K_{sol} and B_{sol} are determined by a least squares refinement. Typical refined values for these two parameters are K_{sol} =0.3-0.4, reflecting the electron density of the crystallization buffer, and B_{sol} =15-40 Å^2, which produces a rather steep fall-off of the bulk solvent electron density with minimum overlap with the protein electron density. No assumption is made about the phase relationship between the protein structure factors and the bulk solvent structure factors.

As an alternative approach to correct the effects of bulk solvent, this mask model was also adopted late in the refinement of MgATP-FPGS structure by using X-Plor 3.851 (Jiang and Brunger, 1994). The mask bulk solvent correction gave a better approximation to the observed structure factor amplitudes than the exponential scaling bulk solvent correction.

The crystallographic R-factor and free R-factor for the final model of MgATP-FPGS are 18.6% and 25.7% with the mask bulk solvent model, compared to 19.4% and 27.7% with the exponential scaling bulk solvent model, differing largely in the low resolution range. The difference Fourier maps calculated with the mask bulk solvent correction were also employed in late stages of model rebuilding. Some new density in the regions of missing structure were clearly seen in difference maps with the mask model correction in contrast to that with the exponential scaling model correction (see Fig. 2.7.4-I), although it was still difficult to locate all of missing residues, probably because of inherent disorder.

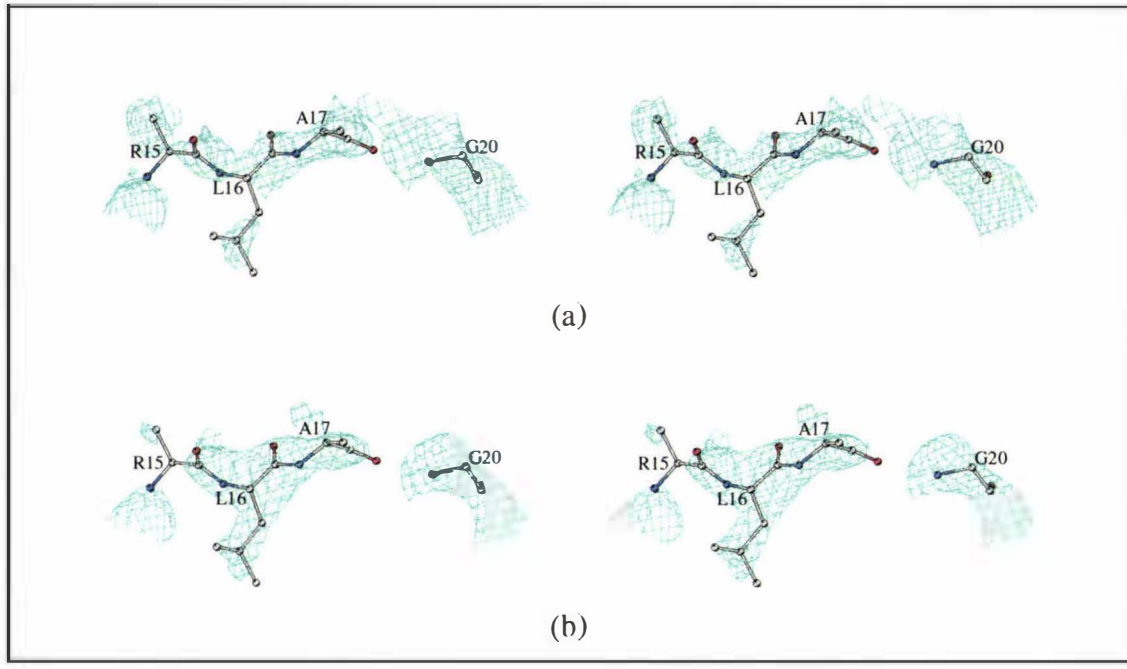


Fig. 2.7.4-1 The $2F_o$-F_c maps (at 1.0 σ) calculated with different bulk solvent correction models: (a) the exponential scaling model and (b) the mask model correction. A disordered loop region, residues (15-20) which were omitted in the maps, is shown.

2.7.5 Dummy-atom procedure for building missing structure

In the final round of refinement, five missing parts of the structural model had still not been located, corresponding to five loop regions located on the surface of the molecule. In some of these regions, some weak (usually 0.5σ level in $2F_o$-F_c, 1σ level in F_o-F_c maps), discrete electron density could be observed, by which the polypeptide chain could be roughly traced, but the building of residues was impossible. Thus a dummy-atom method (Urzhumtsev, 1996) was tried to improve the local electron density of difference Fourier maps.

This method is based on the assumption that an electron density map of at least medium resolution already has been interpreted everywhere except in a small region with poor density. A fine grid covers this region, and every node is occupied by a dummy atom. These dummy atoms are usually hydrogen atoms with initial parameters corresponding to weak density (B factors are 10-40 $Å^2$, occupancies are 0.1-0.5).

The occupancy and position of the dummy atoms are then refined without any restraints, using structure factors which are calculated from the known protein partial

structure and the dummy atom model. Dummy atoms which refine to lower occupancies are removed from the model. Structure factor amplitudes are then calculated from the refined model (protein atoms and dummy atoms) and are used to estimate phase errors in the corresponding calculated phases (Lunin and Skovoroda, 1995). These values are used to calculate weighted electron density maps. This procedure is similar to the omit maps as described in Section 2.7.1, except that in the given region a dummy-atom model is constructed, and new structure factors are calculated from a mixed model. The resultant maps with phases calculated from both the partial structure and the refined dummy-atom model hopefully are more readily interpretable in the missing part compared with the usual omit maps.

The dummy-atom procedure was carried out for MgATP-FPGS using X-Plor (Brunger, 1992) and the programs SPHERXPL and RFLEXPL (Urzhumtsev et al., 1996) for generation of a dummy-atom model and phase error estimations. The new maps from this procedure showed improved density in some regions. Unfortunately, this procedure did not significantly improve the density for the missing parts of the MgATP-FPGS model. Although this method, coupled with inspection of conventional difference maps, provided a little more new information for several missing parts, the situation remained ambiguous, and the missing structure could not be determined with any confidence. As mentioned earlier, this difficulty of recovering missing structure most probably reflects the intrinsic disorder of the surface loops.

2.7.6 The quality of final model

The final MgATP-FPGS model contains 393 out of 428 residues, 2988 non-hydrogen protein atoms, 170 water molecules, one Mg^{2+}, and a pyrophosphate. Five loops are missing (residues 18-19, 146-150, 170-176, 342-347, and 373-384). By using the exponential bulk solvent correction, the final crystallographic R-factor for the final model is 19.4% for all data (15356 reflections) between 20 and 2.4 Å, with a free R-factor of 27.7% (810 reflection). As a comparison, the crystallographic R-factor and free R-factor for the same final model are 18.6% and 25.7% respectively, with the mask bulk solvent correction. The model has good geometry with rms deviations from bond lengths and bond angles of 0.008 Å and 1.43°, respectively (see Table 2.7.6-I).

Table 2.7.6-I Refinement statistics for MgATP-FPGS complex

Refinement statistics		TNT stereochemistry, rms deviations	
Resolution range:	20.0-2.40 Å	Bond length :	0.008 Å
Number of reflections:	15356	Bond angle :	1.4°
Number of protein atoms:	2988	Torsion angle :	20.5°
Number of solvent (water):	170	Trigonal planes :	0.010 Å
Other groups modelled:	1 Mg^{2+}, 1 PPi	General planes :	0.013 Å
Crystallographic R-factor:	0.194 (0.260, 1479 ref.)[a]	Bad contacts :	0.052 Å
free R-factor (810 ref.):	0.277 (0.370, 87 ref.)[a]	B-value correlations :	2.0 $Å^2$

a. For outermost shell (2.53-2.44 Å).

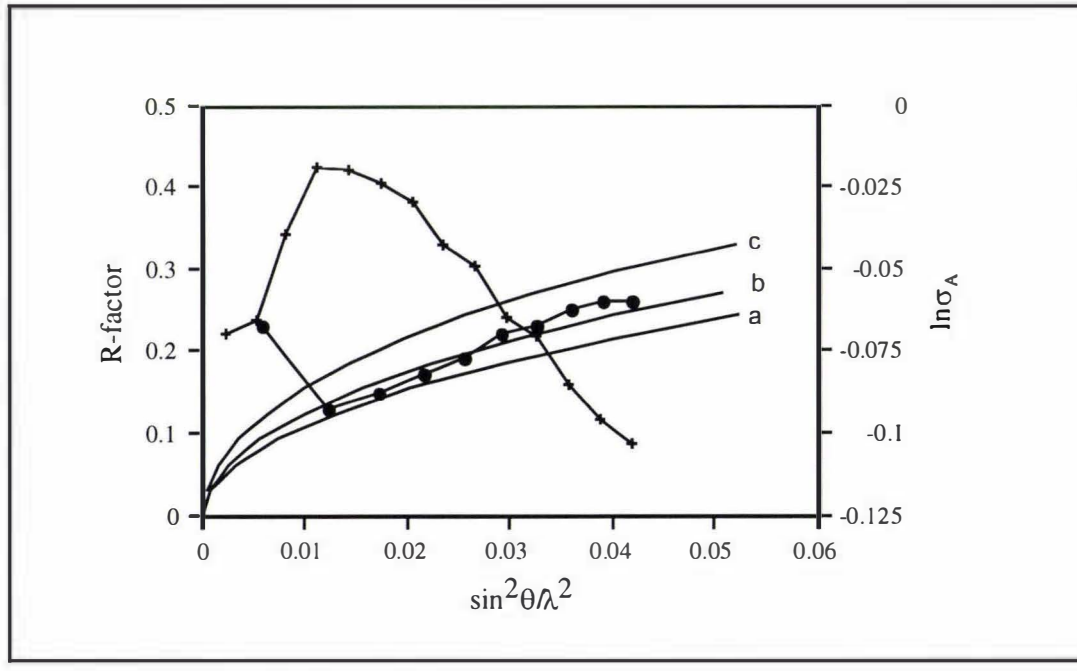


Fig. 2.7.6-I Plots of the variation of the crystallographic R factor (•) and $\ln\sigma_A$ (+) with resolution. The theoretical variation of the R factor for coordinate errors of 0.175, 0.20 and 0.25 Å (curves a, b, and c, respectively) is shown.

The average maximum coordinate error in the MgATP-FPGS model, estimated from a Luzzati plot (Luzzati, 1952), is between 0.20 and 0.25 Å, while a σ_A plot (Read, 1986) gives a similar but more accurate value of 0.23 Å. Both plots are shown in Fig. 2.7.6-I. The values obtained from the plots represent the maximum average error, and the error in the well-defined internal regions such as secondary structure elements would be less than this.

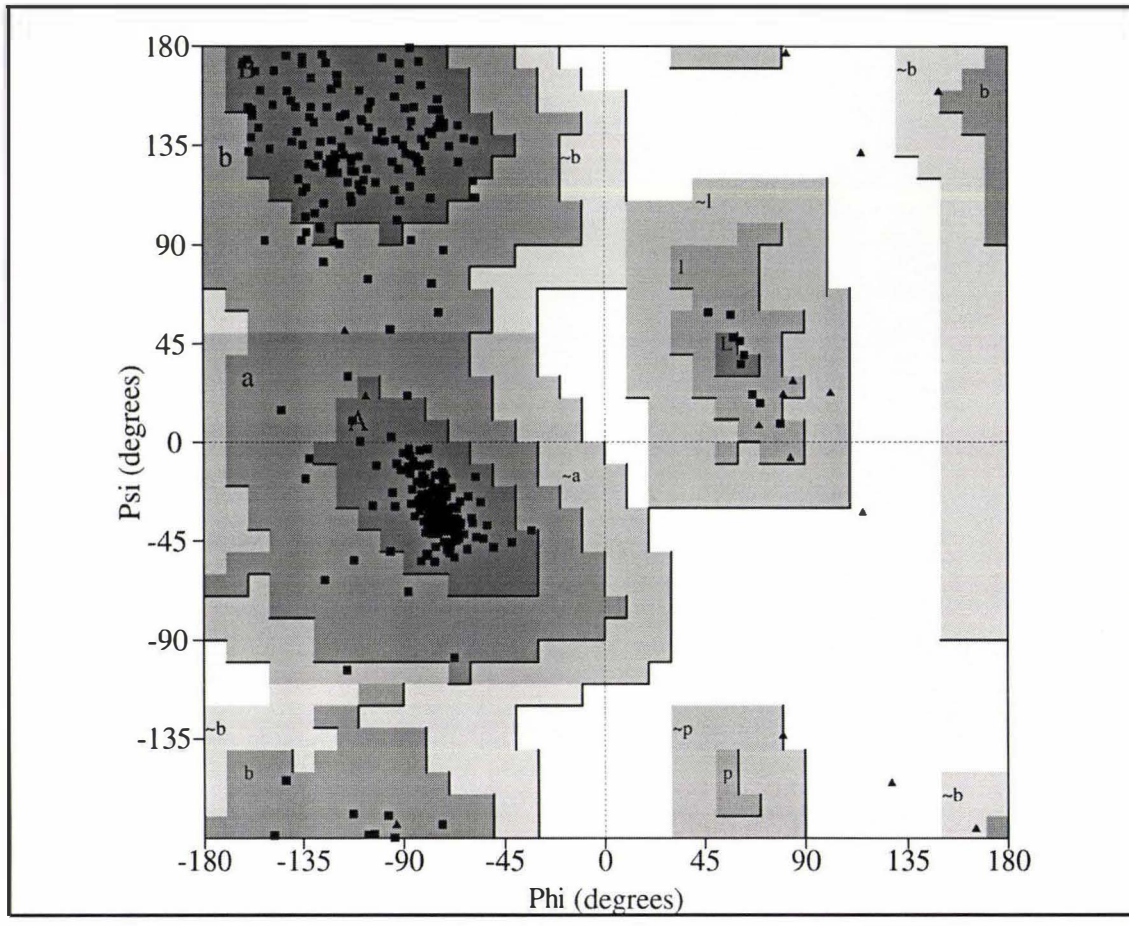
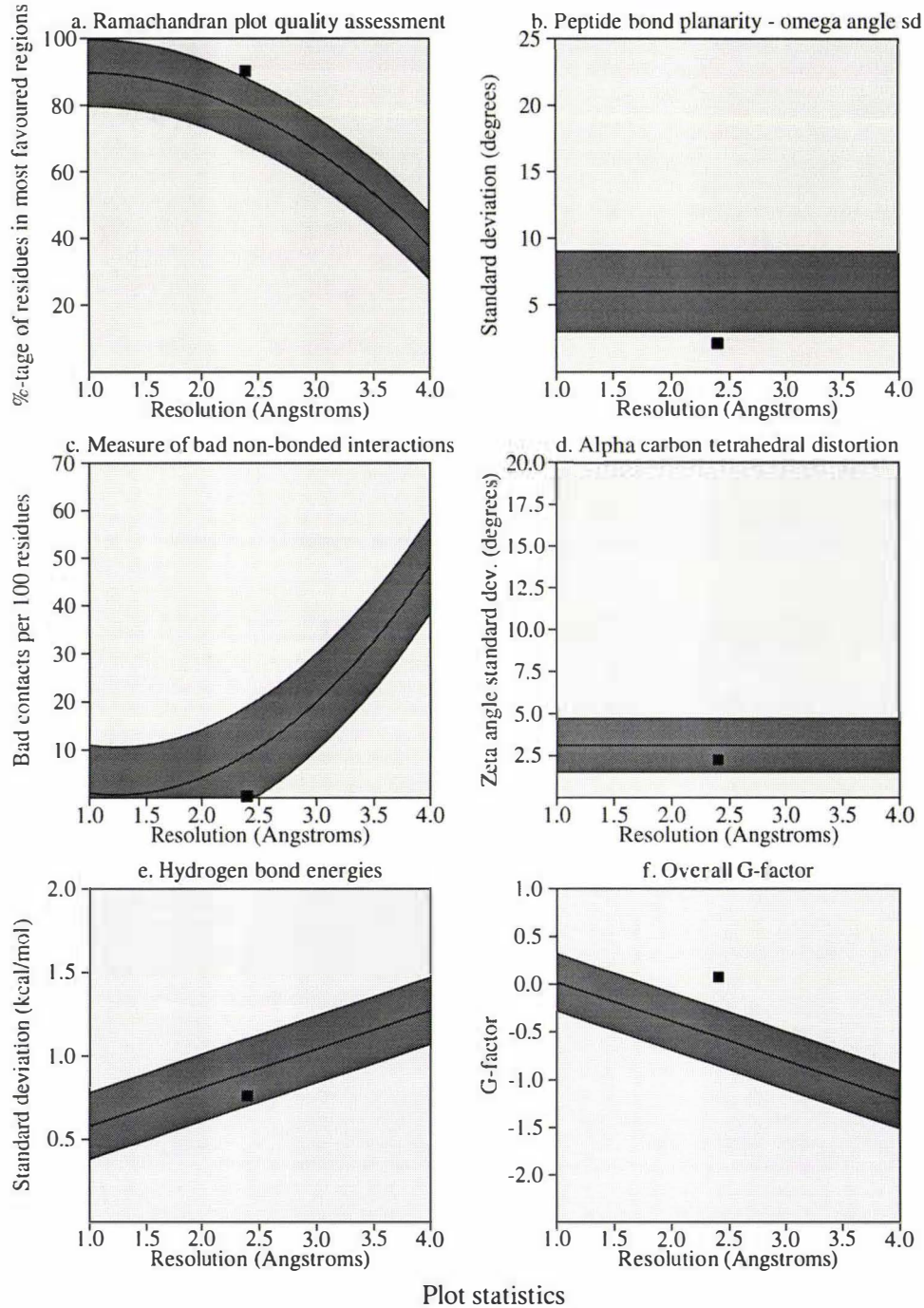


Fig. 2.7.6-II Ramachandran Plot for the refined *L. casei* MgATP-FPGS.
Non-glycine residues are plotted as filled squares, glycine residues as filled triangles. Figure produced with the program PROCHECK (Laskowski et al., 1993).

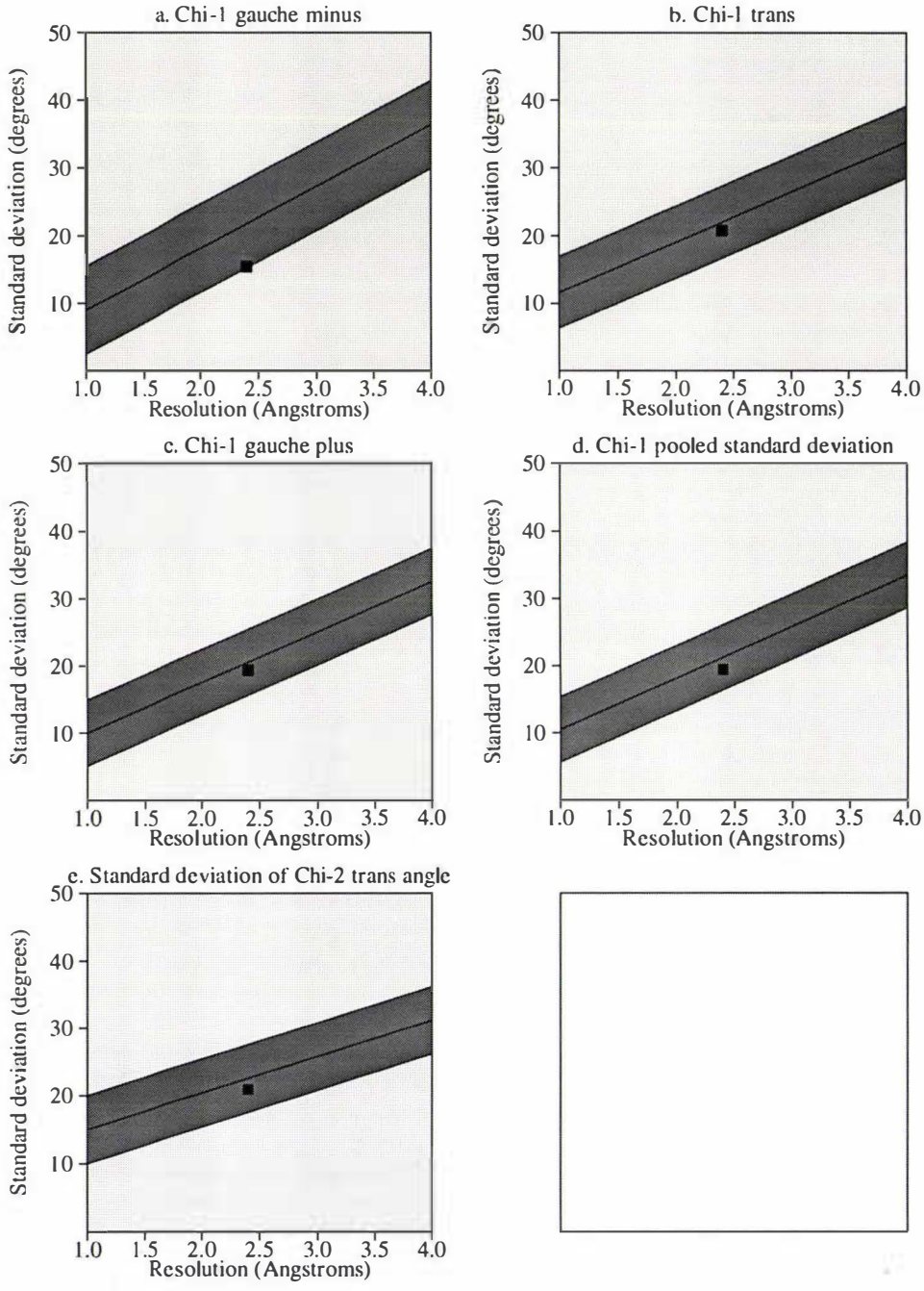
The structure can also be validated by examining the distribution of backbone torsion angles, which were not restrained during the refinement. A Ramachandran plot (Ramakrishnan and Ramachandran, 1965) of the main chain torsion angles was calculated by using the program PROCHECK (Laskowski et al., 1993) and is shown in Fig. 2.7.6-II. All of the nonglycine residues lie within the allowed regions of conformational space, with 90.3% lying within the most favourable regions. Further assessments of the correctness of the model come from the main chain parameters (Fig. 2.7.6-III) and side chain parameters (Fig. 2.7.6-IV) given by PROCHECK, in which main chain and side chain parameters, such as the quality of the Ramachandran plot, α-carbon tetrahedral distortion, and standard deviations of chi-1 and chi-2 dihedral angles, are compared with the average values of known structures at the same resolution. These comparisons show that the final model of MgATP-FPGS is better than average in all respects.



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	339	90.3	78.2	10.0	1.2	BETTER
b. Omega angle st dev	385	2.1	6.0	3.0	-1.3	BETTER
c. Bad contacts / 100 residues	1	0.3	9.0	10.0	-0.9	Inside
d. Zeta angle st dev	367	2.2	3.1	1.6	-0.5	Inside
e. H-bond energy st dev	254	0.8	0.9	0.2	-0.7	Inside
f. Overall G-factor	393	0.1	-0.6	0.3	2.1	BETTER

Fig. 2.7.6-III Quality indicators: main chain of *L. casei* MgATP-FPGS.



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	39	15.5	21.8	6.5	-1.0	Inside
b. Chi-1 trans st dev	101	20.7	22.0	5.3	-0.2	Inside
c. Chi-1 gauche plus st dev	148	19.3	20.5	4.9	-0.2	Inside
d. Chi-1 pooled st dev	288	19.4	21.2	4.8	-0.4	Inside
e. Chi-2 trans st dev	97	20.9	22.6	5.0	-0.3	Inside

Fig. 2.7.6-IV Quality indicators: side chain of *L. casei* MgATP-FPGS.

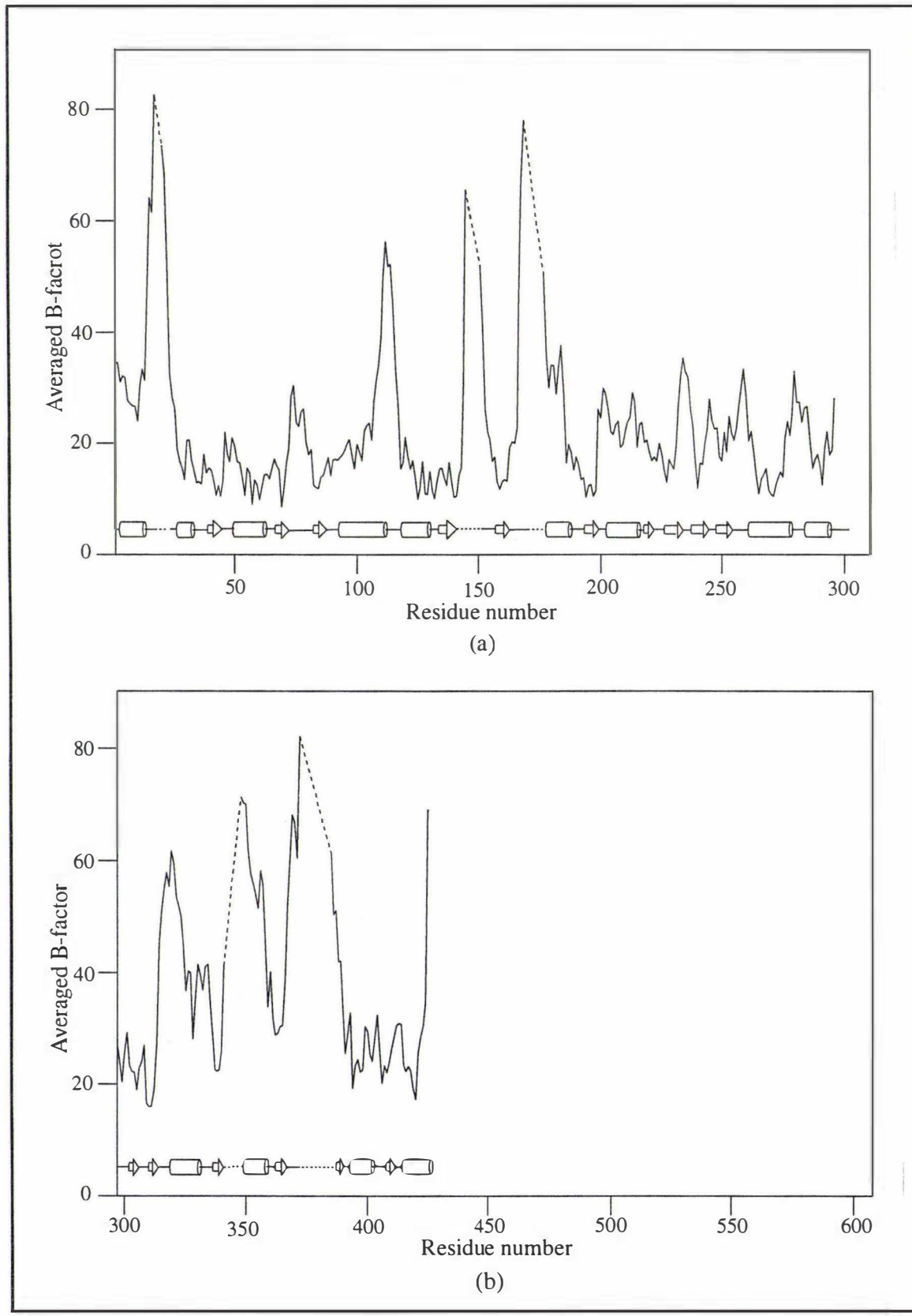


Fig. 2.7.6-V The plot of averaged main chain B factors as a function of residue number. (a) The N-domain residues; (b) the C-domain residues. Secondary structural elements are shown underneath the curve, with α-helix(cylinder), β-strand(arrow), turn(line). The breaks are indicated by the dashed line.

The B factors averaged between main chain atoms as a function of residue number are shown in Fig. 2.7.6-V. The overall mean B value for the main chain atoms is 26.4 $Å^2$, with an rms deviation of 2.7 $Å^2$, close to the overall B (31.2 $Å^2$) estimated from a Wilson plot (Wilson, 1942). The overall mean B value for side chain atoms is 32.4 $Å^2$, with an rms deviation of 5.4 $Å^2$. It is clear from Fig. 2.7.6-V that most of main chain atoms with relatively high B values correspond to the residues at the beginning or end of the five missing loops. These high B values most probably reflect the disorder in these regions of the structure.

The goodness-of-fit of the final model to the electron density is indicated by the real space correlation coefficients (Fig. 2.7.6-VI), which give a residue by residue measure of the agreement between the model and the final $2F_o$-F_c map. The real space correlation coefficients, calculated using the program O (Jones et al., 1991), show that most of the residues have correlation coefficients around 0.9 and that no residues have coefficients below 0.7, indicating a high correlation of the model with the data. A representative section of the final $2F_o$-F_c density map is shown in Fig. 2.7.6-VII.

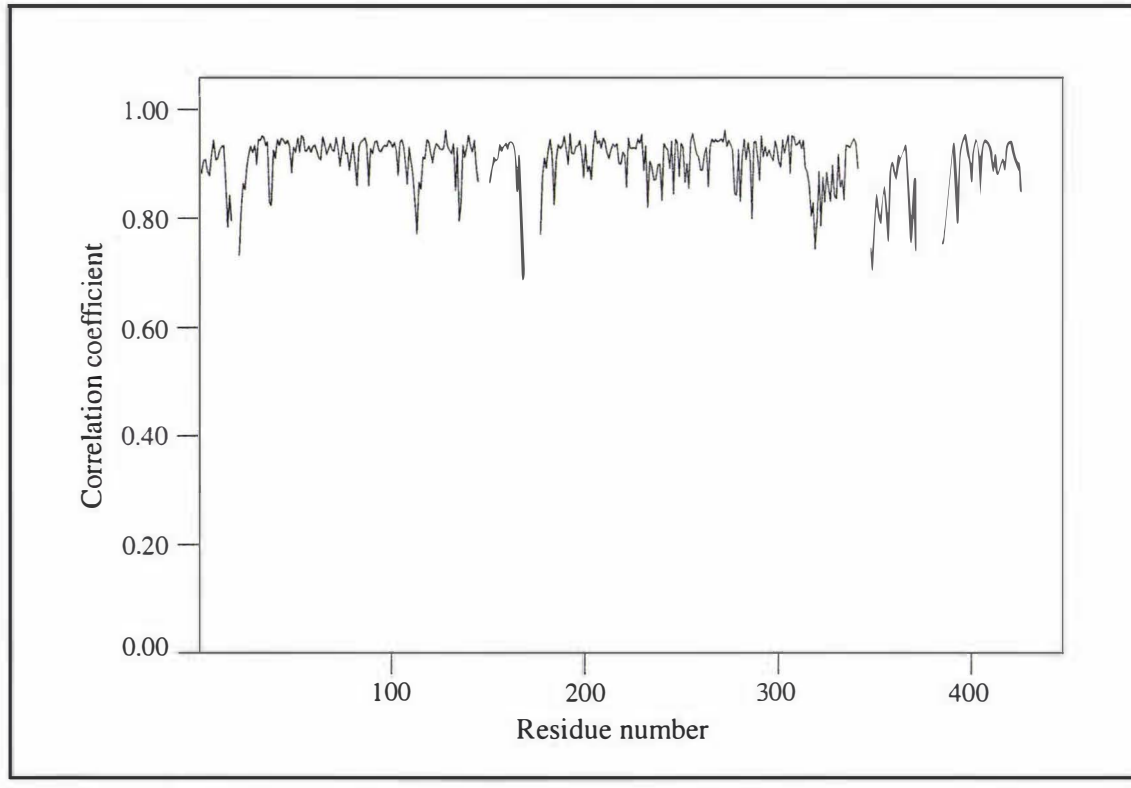


Fig. 2.7.6-VI Real space correlation coefficient for refined *L. casei* MgATP-FPGS.

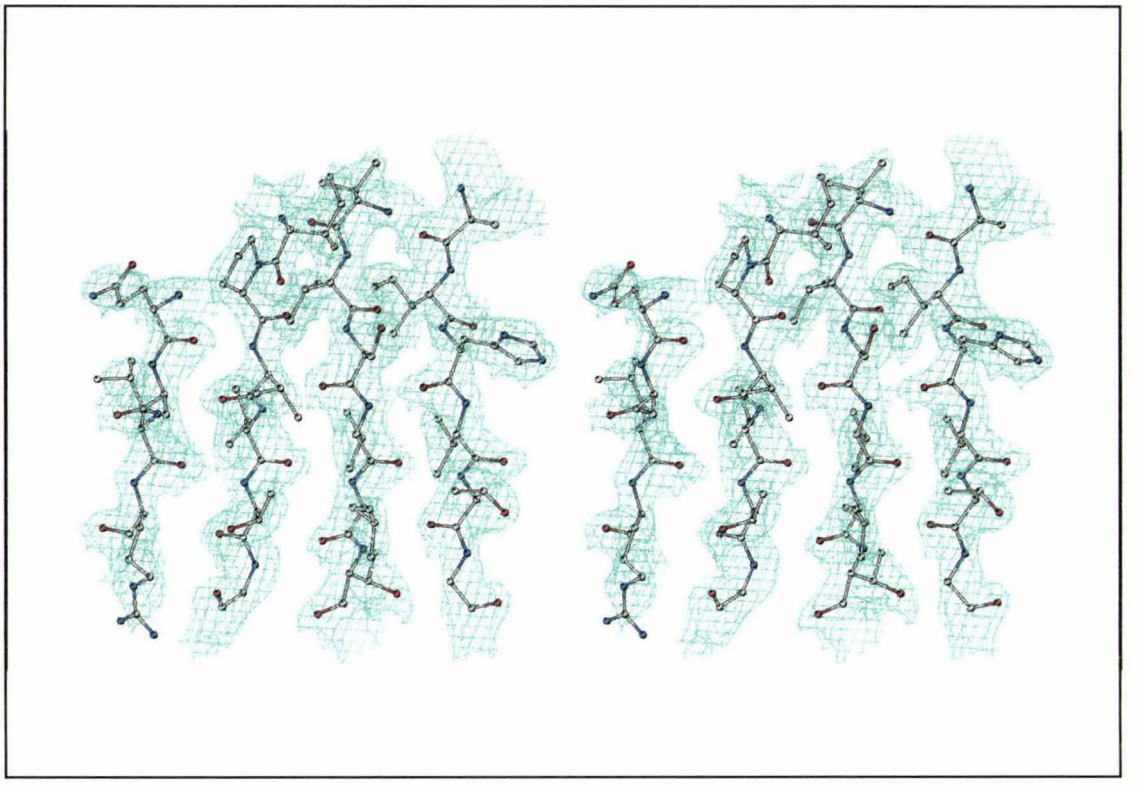


Fig. 2.7.6-VII A representative section of $2F_o$-F_c electron density map (contoured at 1.5 σ) for the part of the N-domain of *L. casei* MgATP-FPGS.

STRUCTURE AND FUNCTION OF FPGS

3.1 Overall polypeptide folding

The FPGS molecule is comprised of two distinct domains. After folding the N-terminal domain (residues 1-293), the polypeptide chain runs into the C-terminal domain (residues 300-428) through a six-residue linker (residues 294-299). A Cα plot of FPGS is shown in Fig. 3.1-I. The large N-terminal domain (residues 1-295) consists of ten β-strands and nine α-helices, in which seven β-strands (B1-B7) form a central β-sheet, flanked on either side by seven α-helices (A2-A8) (see Fig. 3.1-II). The other three β-strands (B8-B10) form an independent small β-sheet, packing against the core of the domain. One helix (A1) starts and another helix (A9) completes the folding of the N-terminal domain. Both occupy positions some distance away from the core of the domain. The small C-terminal domain (residues 300-428) has an α/β structure which is made up of a six-stranded β-sheet and four flanking helices. The nomenclature and location of the structure elements are given in Table 3.1-I.

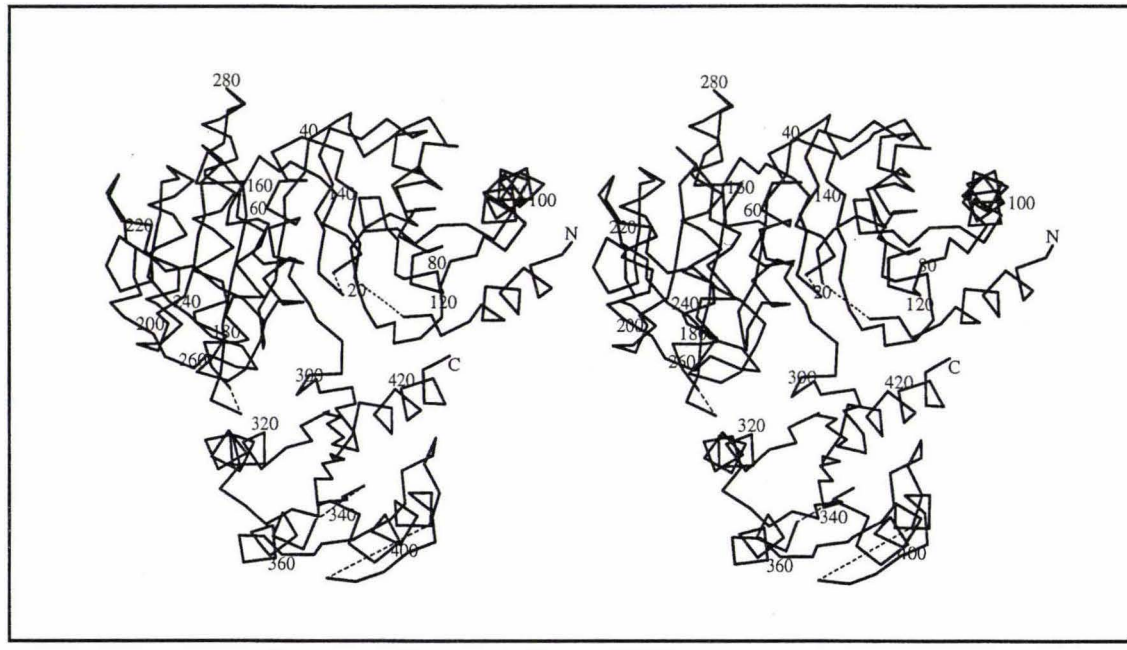


Fig. 3.1-I Stereoview of Cα diagram of *L.casei* FPGS. Every twenty residues is labelled and the breaks are connected with dashed line.

Table 3.1-I The label and location of the secondary structural elements

N-domain				C-domain			
α-Helices		β-strands		α-Helices		β-strands	
Residues	Label	Residues	Label	Residues	Label	Residues	Label
2-11	A1	40-45	B1	318-331	A10	302-305	B11
22-32	A2	67-72	B2	348-358	A11	309-313	B12
51-64	A3	82-86	B3	390-401	A12	336-341	B13
91-112	A4	139-144	B4	412-425	A13	361-365	B14
119-135	A5	160-164	B5			386-388	B15
177-188	A6	194-198	B6			406-410	B16
201-216	A7	218-222	B7				
260-279	A8	226-234	B8				
284-294	A9	237-244	B9				
		247-254	B10				

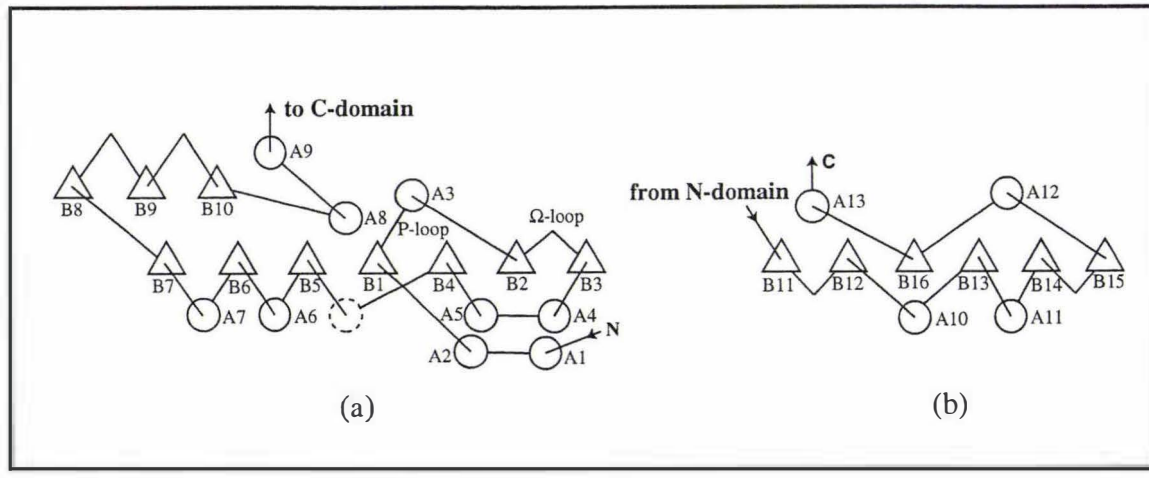


Fig. 3.1-II Topologic diagram of *L. casei* FPGS (a) N-domain; (b) C-domain.

The C-terminal domain appears to be more mobile than the N-terminal domain. The mean B values for the main chain atoms are 21.4 $Å^2$ and 37.1 $Å^2$ for the N- and C-terminal domains, respectively. The mobility of the C-terminal domain is also reflected in the electron density, in which some structure elements such as helix A10 and A11 in the C-terminal domain, have relatively weak density compared to similar features in the N-terminal domain. The overall molecular structure is shown in Fig. 3.1-III, where the molecule is oriented to clearly show the distinct two domains and linking loop.

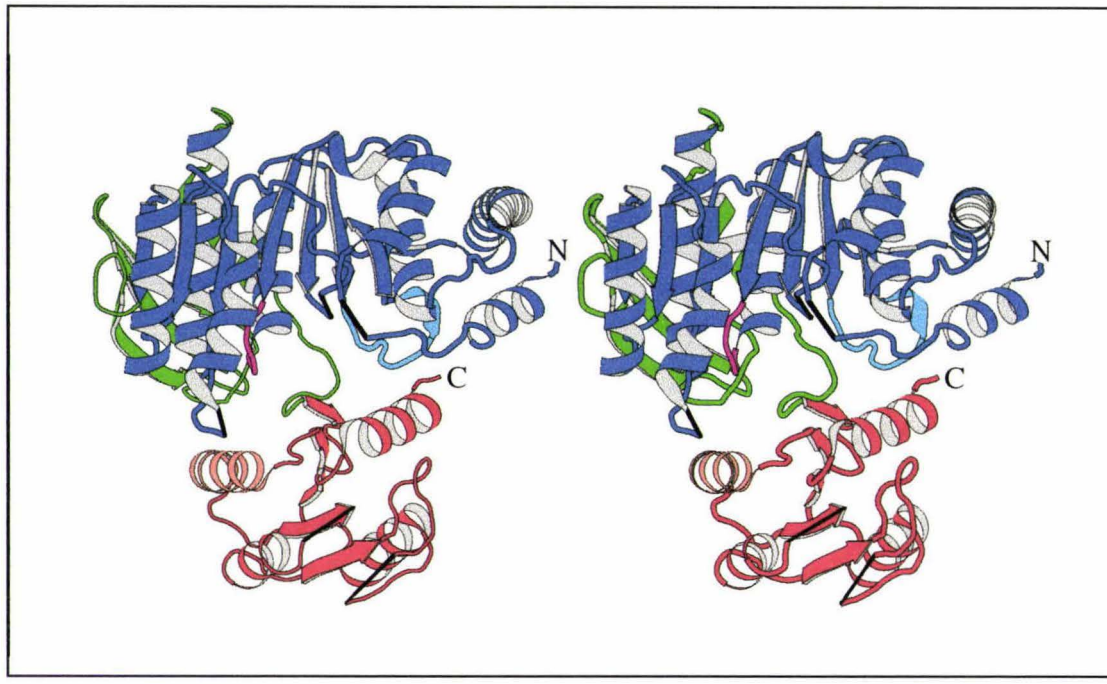


Fig. 3.1-III Stereo ribbon diagram of FPGS. The N-domain is composed of a *ras*-like domain (blue) and a three-stranded β-sheet (green) whereas the C-domain is shown in red. The P-loop is shown in magenta, the Ω-loop is shown in cyan, helix A10 is shown in pink. The breaks in polypeptide chain are joined by thin black lines. The N and C termini are indicated. This figure is generated by using MOLSCRIPT (Kraulis, 1991).

A total of 170 ordered water molecules have been modelled in the current structure. Some water sites are internal, located in discrete sites with space for only one or two waters, and these water molecules make a large number of hydrogen-bonded contacts to the surrounding protein structure. The most common location for such sites is at the points where β-strands diverge away from one another, or in cavities buried deeply within the protein. No water molecules have been included in the vicinity of the pyrophosphate, except for water W466 which bridges between the side chain of Asn264 and one phosphate oxygen (see Fig. 3.2.3-I in next Section), since no clear distinction could be made between the electron density relating to the remainder of the ATP and the ordered solvent. Modelling of water molecules near the ATP-binding site must await further refinement and model building. The majority of the water molecules are located in crevices in the protein surface, attached to main chain or side chain atoms.

3.2 The N-terminal domain

3.2.1 The location of the ras-like domain structure

Since FPGS is a nucleotide-binding protein, it is expected to have some structural homology with other nucleotide-binding proteins. These have been divided into two different families depending on whether they bind dinucleotides or mononucleotides (Schulz, 1992). FPGS falls into the family of the mononucleotide-binding proteins. The core structure of these enzymes consists primarily of a central, parallel four-stranded β-sheet and four invariant α-helices, which has been referred to as the ras-like domain (Coleman et al., 1994) observed in mononucleotide-binding enzymes such as ras-p21 (Pai et al., 1990), elongation factor (EF-Tu) (Kjeldgaard et al., 1993) and adenylate kinase (ADK) (Muller and Schulz, 1992). In addition to the core structure, most of these enzymes have extra secondary structure elements and, in some cases, extra domains inserted at various positions within this core, but they still retain the same structural topology surrounding the nucleotide binding site. Most of these enzymes have a consensus sequence pattern (GXXGXGKT/S, where X denotes amino acids that vary among the different proteins). This glycine-rich segment (P-loop) is located within a large loop between a β-strand from a central parallel pleated sheet and an α-helix. It forms a giant anion hole, which accommodates the phosphates of the mononucleotide (Dreusicke and Schulz, 1986). The glycine residues in the loop clearly play an important conformational role in maintaining the structure of the loop and allowing back-bone hydrogen bonds between adjacent amino acids and the β- and γ-phosphate groups.

Inspection of the topology diagram of the N-domain of FPGS (cf. Fig. 3.1-II) and its putative nucleotide-binding P-loop sequence (residues 46-52) revealed that it is this domain of FPGS that is the nucleotide-binding domain, as it possesses a similar core structure to the ras-like domain. Four parallel β-strands (B1, B4-B6) form the core β-sheet, with three flanking helices (A3, A6, A8). Between β-strands B4 and B5, where there should be another helix similar to the other nucleotide-binding proteins, is a disordered region (residues 145-151), although the following residues (152-155) have a distorted 3_{10} turn conformation. A high mobility for this region (residues 145-151) has

also been observed in the equivalent region in the ras-p21 structure (Pai et al., 1990) (see discussion later).

The nucleotide-binding P-loop (residues 46-52, with the sequence GTNGKGS) is located in a deep cleft formed between N- and C-domains (Fig. 3.1-III), and joins β-strand B1 to α-helix A3. Although its sequence resembles the consensus P-loop sequence found in adenylate kinase (ADK) and uridylate kinase (UK) (GXXGXGKGT), it is two residues shorter, which results in somewhat different conformation.

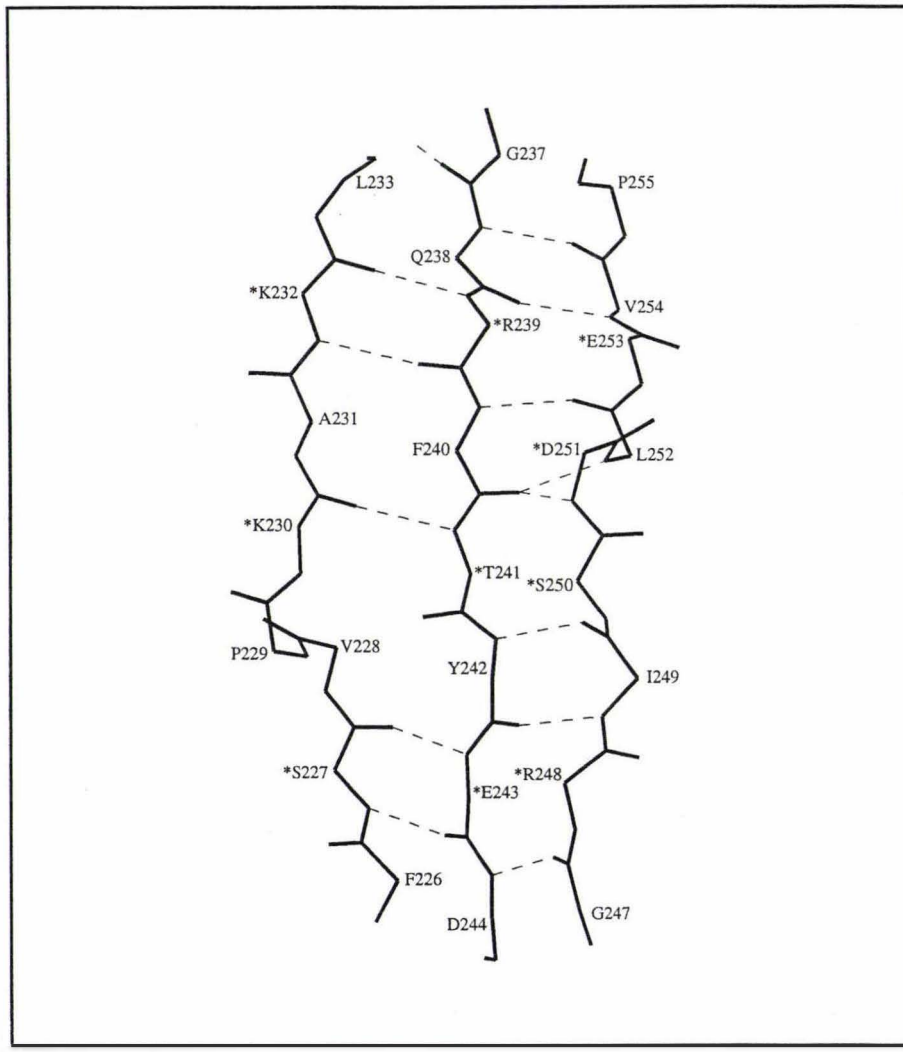


Fig. 3.2.1-I Schematic representation of hydrogen bonding in the small antiparallel β-sheet (B8-B10). The asterisk indicates the residues with their polar side chain pointed outwards from the protein surface.

As in other nucleotide-binding proteins, some extra structural elements exist in the N-domain of FPGS in addition to the core structure described above. They include two N-terminal helices (A1 and A2), one insertion at one end of the core structure (strands B2 and B3; helices A4 and A5), and another insertion at the other end of the core structure (helices A7 and A9, strand B7 and the three-stranded sheet B8-B10). Three of these additional strands (B2, B3 and B7) make up the extended seven-stranded β-sheet along with the four core strands. The independent three-stranded sheet (B8, B9, B10) has a bent antiparallel structure and covers the outer surface of helices A8 and A9 (see Fig. 3.1-III). The hydrogen bonding pattern in this antiparallel sheet is depicted in Fig. 3.2.1-I. Two β-bulges (Pro229 in B8, Asp251 in B10) are located approximately in the middle of the β-sheet, about which the β-sheet has a left-hand twist. The β-bulges are structurally important to orient polar side chains away from the protein and bury the non-polar side chains against the surface of helices A8 and A9. In this way, the interactions between this β-sheet and the rest of the protein are primarily hydrophobic except for one hydrogen bond between the side chain of Gln238 and carbonyl oxygens of Leu256 and Gly258, located at the top of the β-sheet, and another hydrogen bond between the side chains of Asp244 and Lys273 located at the bottom of the β-sheet. The structure and location of this small β-sheet makes it look like an independent central domain attached to the N-terminal domain. This β-sheet seems unlikely to be involved in substrate binding in MgATP-FPGS, because it is far away from the ATP-binding site. An understanding of its function must await further structural analyses.

3.2.2 Comparison with other nucleotide-binding enzymes

The structures chosen for this comparison are listed in Table 3.2.2-I.

3.2.2.1 Topological similarity

As reviewed recently (Schulz, 1992; Smith and Rayment, 1996), the mononucleotide-binding superfamily comprises a large number of proteins of widely different functions. All of these proteins are related by their ability to bind ATP

(adenosine triphosphate) or GTP (guanosine triphosphate) through a similar P-loop motif. They include kinases such as ADK (Muller and Schulz, 1992) and UK (Muller-Dieckmann and Schulz, 1994), G proteins such as H-ras-P21 (Pai et al., 1990) and elongation factor Tu (Kjeldgaard et al., 1993), and motor proteins such as myosin (Smith and Rayment, 1995) and kinesin (Kull et al., 1996).

Table 3.2.2-I The proteins for comparison with the N-domain of FPGS

Protein	PDB	Source	P-loop sequence	Nucleotide bound	Reference
FPGS		*L. casei*	GX—XGKGS	ATP	–
H-ras-P21	121P	Human	GXXXXGKS	GMPPCP	(a)
Elongation factor Tu	1EFT	*T. aquaticus*	GXXXXGKT	GMPPCP	(b)
$G_{i\alpha_1}$	1GIA	*R. norvegicus*	GXXXXGKS	GTPγS	(c)
Adenylate Kinase	1AKE	*E. coli*	GXXGXGKGT	AP_5A	(d)
	1ANK	*E. coli*	GXXGXGKGT	AMPPNP AMP	(e)
Uridylate Kinase	1UKY	*S. cerevisiae*	GXXGXGKGT	ADP ADP	(f)
Myosin	1MNE	*D. discoideum*	GXXGXGKT	PPi	(g)
MurD*	1UAG	*E. coli*	GX—XGKS	SO_4^{2-}	(h)

* MurD: UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase, its domain2 is used for the comparison. Two blank positions are inserted in the sequences of FPGS and MurD in order to align their sequences with those of other enzymes.
References cited: a. Pai et al. (1990); b. Kjeldgaard et al. (1993); c. Coleman et al. (1994); d. Muller and Schulz (1992); e. Berry et al. (1994); f. Muller-Dieckmann and Schulz (1994); g. Smith and Rayment (1995); h. Bertrand et al. (1997).

Although the diversity of proteins related by this fingerprint (i.e., GXXGXGK) is spectacular, the similarity of the respective chain folds together with the similarity of the detailed modes of mononucleotide binding, is indicative of divergence from a common ancestor (Milner-White et al., 1991). This also implies that these proteins share very similar local topologies.

The topology diagrams for the N-terminal domain of FPGS and several other nucleotide-binding proteins are shown in Fig. 3.2.2-I. It is clear by inspection of the diagrams that although the overall topology varies between the proteins, each of them contains a central β-sheet formed from a varying number of strands. In each case, an intact β-strand/P-loop/α-helix motif is conserved, and furthermore a ras-like domain (shaded areas in Fig. 3.2.2-I) in which three other parallel β-strands and three helices are incorporated with the P-loop motif, can be readily seen. In FPGS, as mentioned in the previous section, the position for a helix between strands B4 and B5 (dashed circle) corresponds to a disordered region together with a distorted 3_{10} turn. With such a characteristic core structure, however, the N-terminal domain of FPGS is clearly a new member of this nucleotide-binding superfamily.

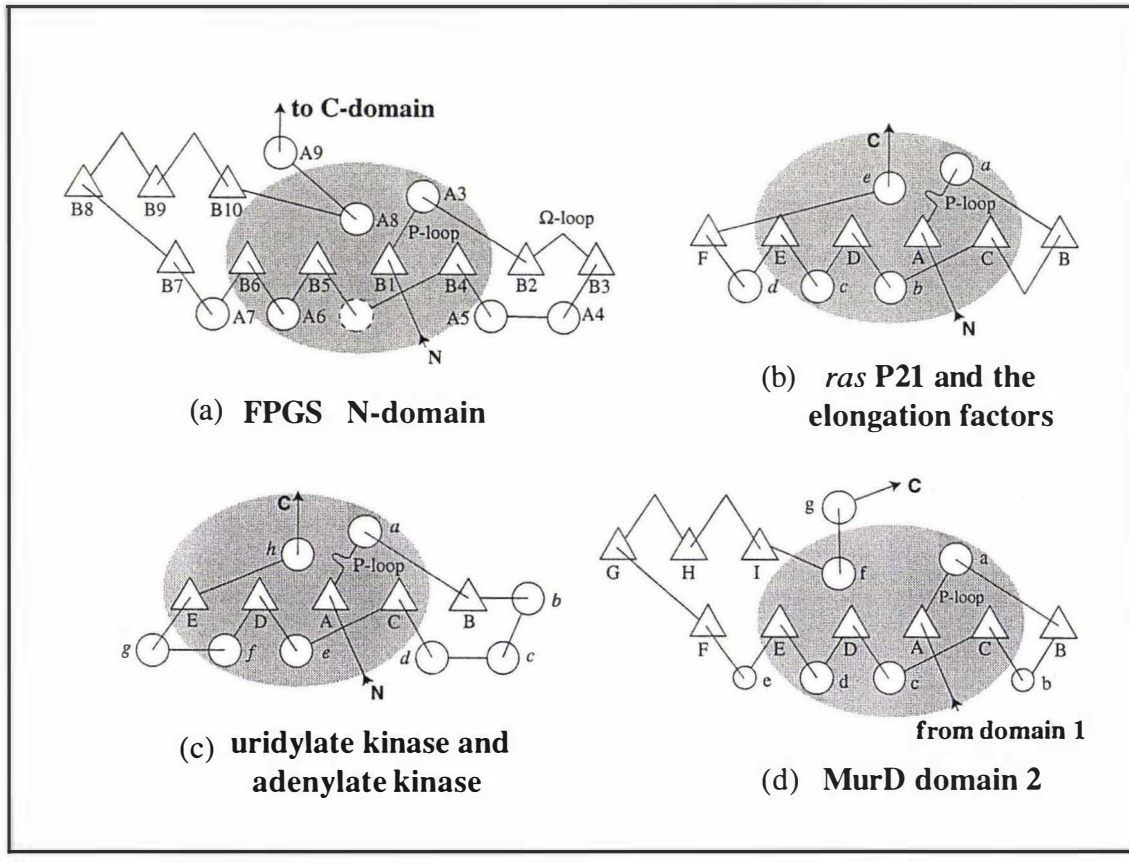


Fig. 3.2.2-I Topology diagrams of (a) the N-domain of FPGS,(b) *ras* and G proteins,(c) ADK and UK, (d) the domain 2 of MurD. The regions of structural similarity among these structures are indicated by the shaded areas.

Surprisingly, the crystal structure of UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase (MurD) reported recently (Bertrand et al., 1997) shows an extremely similar topology, not only for the ras-like domain but also for whole N-domain and C-domain of FPGS. As an ATP-binding domain, domain 2 of MurD has a central ras-like domain (Fig. 3.2.2-I). Moreover, it possesses all of the other secondary structural elements found in FPGS except for one β-strand (B3) and one α-helix (A4). It is also interesting that the small three-stranded sheet (B8-B10) of FPGS reappears in the structure of MurD.

3.2.2.2 Sequence comparison of the P-loop

It has been well established that the P-loop motif is important for mononucleotide binding, and thus is a key structural element of the proteins. Some residues in the P-loop, such as specific glycine and lysine residues, are strictly conserved to maintain the correct conformation of the loop or interactions with the phosphoryl groups of nucleotide, while others seem to vary both in size and character, depending on the particular enzyme (Schulz, 1992; Smith and Rayment, 1996).

Sequence comparisons of the P-loop are given in Table 3.2.2-I. Typical sequence patterns are GXXXXGKS/T for G proteins, GXXGXGKGT for kinases, and GXXGXGKT for myosin, respectively. By comparison, the P-loop sequence of FPGS (GX--XGKGS) is much more similar to that of kinases although it is two residues shorter. Each protein in this comparison has a hydroxyl residue (serine or threonine) directly following the lysine residue, except that one extra glycine is inserted in the case of FPGS and kinases. This serine (or threonine) is important in the majority of these enzymes, since the crystal structures show that it serves as a protein ligand for the divalent cation associated with the bound nucleotide in the case of G proteins and myosin, or hydrogen bonds to the α-phosphoryl group in the case of kinases (Smith and Rayment, 1996). The most important result derived from this comparison and other structural analyses (Pai et al., 1990; Shen et al., 1994) is that the first Gly and the Gly-Lys sequence are invariant in all of these enzymes, whereas there is only a limited requirement for a particular amino acid in the intervening four residues. The well refined structures of ras-p21 (Pai et al., 1990; Tong et al., 1991) and adenylate kinase

(Diederichs and Schulz, 1991; Muller and Schulz, 1992) show that the second and third glycines in the GXXGXGK sequence pattern assume main-chain dihedral angles that would be normally forbidden for residues with side chains, whereas the angles of the first glycine, though forbidden, are near to an allowed region. However, the second glycine is replaced by an alanine in elongation factor Tu, even though its dihedral angles remain the same, and the first glycine is conserved although its angles would permit side chains without too much distortion. The highly conserved lysine in the P-loop is essential as it coordinates to the β- and γ-phosphoryl group and may play a crucial role in phosphoryl transfer (Matte et al., 1998).

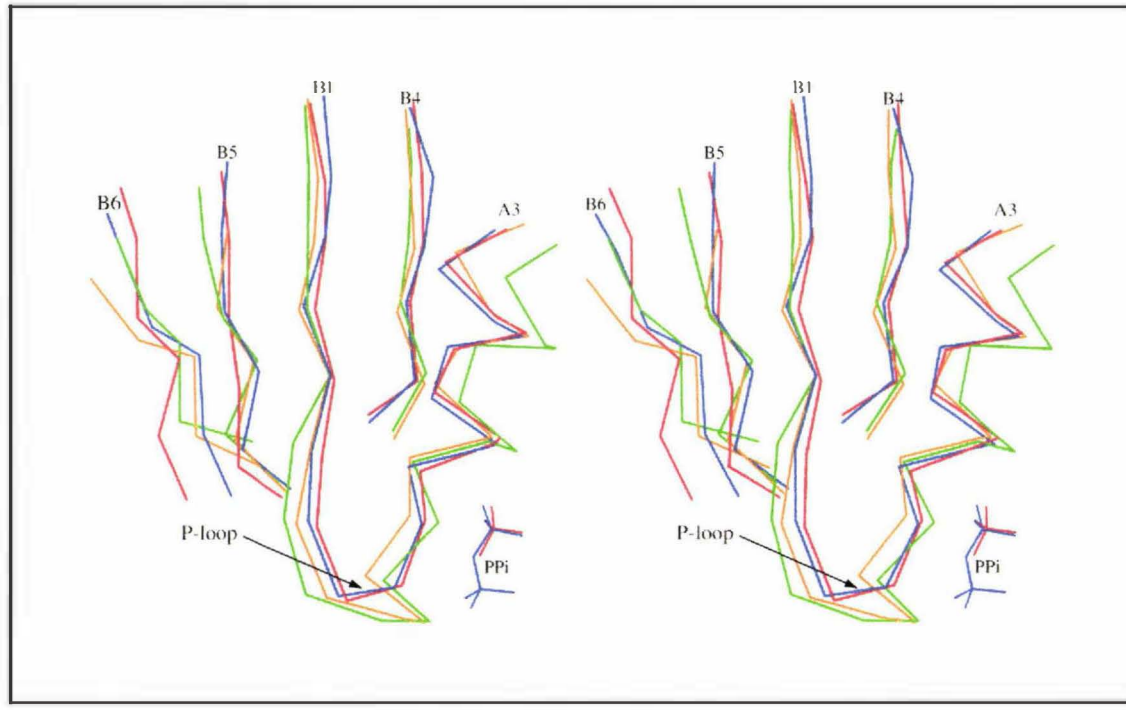


Fig. 3.2.2-II Superposition of some nucleotide-binding proteins on to the N-domain of FPGS, based on four core β-strands (B1,B4-B6), the P-loop, and helix A3. FPGS is shown in blue along with *ras*-p21 (green), myosin (orange), and MurD (red). The PPi from FPGS and sulfate from MurD are also shown.

The shorter P-loop in FPGS, when superimposed on other "normal-length" P-loops, shows that the basic difference between the loops is the "deletion" of two residues between the equivalent of the first Gly and the Gly-Lys in the FPGS sequence (see Fig. 3.2.2-II). The β-strand which precedes the P-loop (B1), and the helix which follows

(A3), both superimpose exactly with similar structural elements in other nucleotide-binding enzymes (e.g. superposition on to the equivalent elements in myosin gives an rms difference of 1.09 Å for 44 Cα atoms). The two-residue "deletion" causes the conformation of the FPGS P-loop to be slightly different from the classical P-loop, in that it forms a rather sharp turn between B1 and A3 compared to the "ledge-like" conformation adopted in the other enzymes. Clearly, however, this "short" loop still retains the primary function attributed to the P-loop; to bind the phosphoryl groups of ATP, as evidenced by the coordination of pyrophosphate in FPGS.

Further evidence for the shorter P-loop comes from MurD. This enzyme contains the same shortened P-loop in domain 2 which has been assigned as an ATP-binding domain (Bertrand et al., 1997). The 3D-structure of this P-loop of MurD is exactly the same as that in FPGS (Fig. 3.2.2-II). Superposition of these two shorter P-loops gives an rms difference of only 0.22 Å (Table 3.2.2-II). Taken together, the shorter P-loop of FPGS and MurD might represent an unique divergence from an ancient nucleotide-binding motif.

Table 3.2.2-II Superposition of nucleotide-binding proteins on to the N-domain of FPGS

Protein	PDB code	rms Difference (Å)[a]						Sequence identity[d]
		P-loop		Central β-sheet + helix[b]		Core structure[c]		
		Ca	main chain	Ca	main chain	Ca	main chain	(%)
ras-P21	121P	0.98(7)	1.08(28)	1.71(38)	1.85(152)	2.45(57)	2.49(228)	14.3
EF-Tu	1EFT	0.89(7)	1.05(28)	1.78(42)	2.03(172)	2.31(61)	2.44(244)	14.3
G_{ial}	1GIA	0.91(7)	1.07(28)	1.86(39)	1.98(156)	2.61(59)	2.62(232)	15.7
AKE	1AKE	0.86(7)	0.98(28)	1.31 (41)	1.42(164)	1.97(59)	1.98(236)	19.6
ANK	1ANK	0.80(7)	0.94(28)	1.37(41)	1.47(164)	2.01(59)	2.03(236)	19.6
UK	1UKY	0.80(7)	0.95(28)	1.71(41)	1.60(160)	2.14(60)	2.17(240)	14.6
Myosin	1MNE	0.78(7)	0.96(28)	1.09(44)	1.30(176)	1.73(57)	1.83(228)	-
MurD	1UAG	0.22(7)	0.22(28)	0.78(52)	0.81(208)	0.89(79)	0.94(316)	15.7

a. The numbers in brackets indicate the number of atoms used for superposition.
b. Structural elements superimposed include four core β-strands (B1, B4, B5, and B6) and the P-loop helix A3.
c. Core structure elements include the whole ras-like domain (β-strands B1, B4, B5 and B6; helices A3, A6 and A8).
d. Sequence identities are based on alignment of the whole N-domain of FPGS and the nucleotide-binding domain of other proteins. They are calculated using the program Pairwise Sequence Alignment (Thompson et al., 1994).

3.2.2.3 Structural similarity

The ras-like domains of the structures listed in Table 3.2.2-I have been superimposed on the N-domain of MgATP-FPGS, based upon the P-loop, core β-sheets and P-loop helix, and core structure of whole ras-like domain (see Fig. 3.2.2-II). The program LSQKAB from the CCP4 program suite (Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography) was used for the structural alignment, and the results are summarized in Table 3.2.2-II. The names for structural elements are referred to FPGS (see topology diagram Fig. 3.2.2-I).

These results clearly indicate that there is a remarkable three-dimensional structural similarity between the N-domain of FPGS and the G proteins, kinases, myosin and MurD, despite little sequence identity (see Table 3.2.2-II).

The rms differences for the superpositions of the P-loop are relatively large despite the small number of atoms superimposed. This stems from the different P-loop conformation resulting from the "deletion" of two residues. By inspection of the rms differences in Table 3.2.2-II, the core β-sheet/P-loop/α-helix motif is shown to be the most highly conserved structural region. In comparison, the other two helices (A6 and A8) conserved in the ras-like domain are less well-aligned. For the G protein structures (ras-P21, elongation factor Tu and $G_{i\alpha 1}$) both helices A6 and A8 are rotated away from the structurally equivalent helices in FPGS by approximately 30° to 40° after the whole ras-like domain was superimposed. In the case of the kinases, helix A8 superimposes quite well, along with the core β-sheet, but helix A6 is rotated about 30° away from the structurally equivalent helix in FPGS. The myosin structure is closer to that of FPGS, with both helices A6 and A8 rotated away from the structurally equivalent helices by only small angles (~15° and 5° respectively) upon superposition of the ras-like domain. This relative variability in position of helices A6 and A8 results in greater rms differences (Table 3.2.2-II) when the whole ras-like domains were superimposed. Nevertheless, the structural similarity between the N-domain of FPGS and these proteins is remarkable, considering the variety of their functions, the diversity of their sources and the low sequence similarity.

As a core structure, the ras-like domain seems to be an essential component of all nucleotide-binding proteins. However, it should be noted that the overlapping secondary structural elements are not limited to the ras-like domain. The β-strand B2 superimposes with a structurally equivalent β-strand in all structures compared. For all kinases, helix A5 also superimposes well, in addition to the superposition of the ras-like domain. Moreover, all seven central β-strands (B1-B7) superimpose with equivalent strands in the myosin structure, in addition to the overlapping ras-like domain. This indicates that the structural similarity between the N-domain of FPGS and other nucleotide-binding proteins is not confined to the ras-like domain.

As a special case, almost all of the secondary structural elements in domain 2 of MurD, including the ras-like domain, superimpose very well on to the structurally equivalent elements in the N-domain of FPGS. The rms differences between these two molecules are considerably less, for more matching atoms, compared to the other structures listed in Table 3.2.2-II, indicating a greater degree of similarity in the spatial arrangement of the secondary structural elements. Moreover, Fig. 3.2.2-II shows a good match between one of the phosphate groups in FPGS and sulphate ion from MurD. The latter is believed to occupy the position of the α-phosphate of ATP (Bertrand et al., 1997). Given the same P-loop sequence and the similarity in 3D-structure of the P-loop, it is likely that FPGS and MurD may have a similar ATP-binding site.

3.2.3 The structure of the putative active site

3.2.3.1 The nucleotide-binding site

Although the P-loop of FPGS has been identified, the limited electron density corresponding to the bound ATP does not allow for complete modelling of the ATP and divalent cation Mg^{2+}. This is presumably a result of the low pH of crystallization and/or the hydrolysis of ATP (see Section 2.7.2). Consequently, only a pyrophosphate (PPi) and magnesium cation (Mg^{2+}) have been modelled into the ATP-binding site of FPGS. While detailed statements concerning Mg-ATP binding geometry must await further structural investigation, some reasonable conclusions concerning ATP binding can be deduced using the knowledge of nucleotide-binding in the other proteins.

Typically, the P-loop in the other proteins listed in Table 3.2.2-I folds to form a small "ledge" perpendicular to the plane of the central β-sheet (Smith and Rayment, 1996). The triphosphate group sits on this ledge and hydrogen bonds to several amide nitrogens from the P-loop and the N-terminus of the helix. The nucleoside projects over the side of this ledge. In FPGS, the shorter P-loop makes the interactions between phosphate and P-loop somewhat different. The putative ATP-binding site in FPGS is shown in Fig. 3.2.3-I. Although the P-loop in FPGS (and MurD) essentially takes a short cut, it is clear from the FPGS structure that the ATP still binds to part of this loop and to the N-terminus of helix A3.

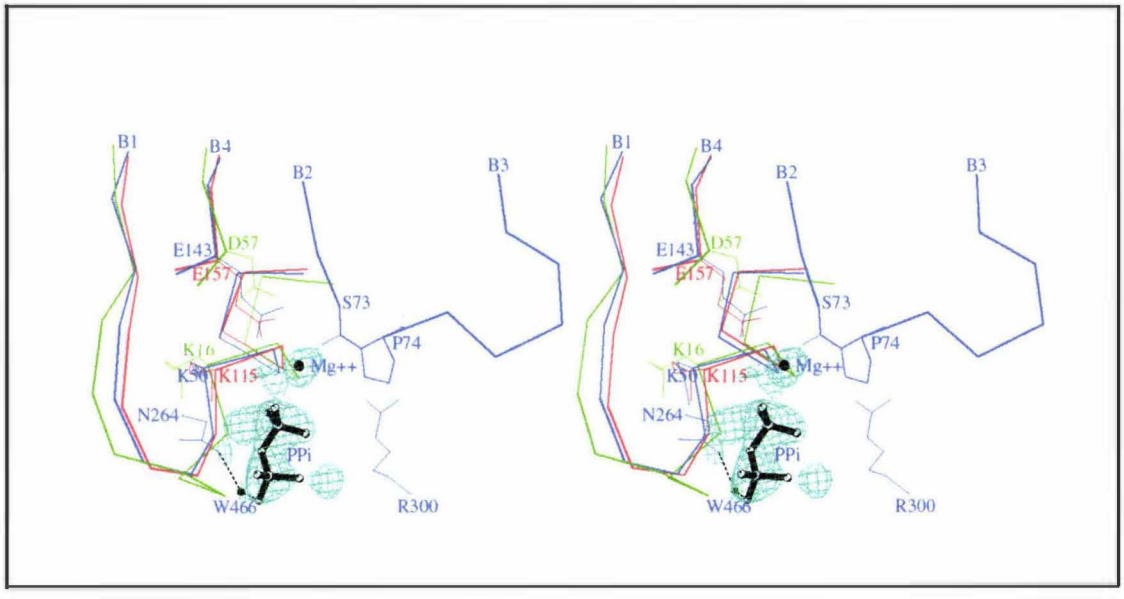


Fig. 3.2.3-I The ATP-binding site in FPGS (blue) overlaid with that from ras-p21 (green) and MurD (red). The β-strands (B1-B4) correspond to labels in FPGS. Side chains from FPGS (E143, K50, N264, R300, S73 and P74), ras-p21 (D57 and K16) and MurD (E157 and K115) are labelled in blue, green and red, respectively. The putative Mg^{2+} and pyrophosphate (PPi) are indicated together with the F_o-F_c omit density (light blue). The *cis*-Pro74 and carbonyl oxygen of S73 are also shown. The water molecule (W466) bridges between N264 Oδ1 and one oxygen of phosphate group is indicated by dashed line.

Assuming that the PPi ion modelled from the omit map represents the α- and β-phosphates of the nucleotide, the γ-phosphate might well occupy the lobe of F_o-F_c density between the PPi and the side chain of Lys50 (see Fig. 3.2.3-I) as it is similar to the location of the γ-phosphate in other proteins. The nucleoside would then project

into the cleft between the N- and C-domains, which is consistent with generally weak binding of nucleoside to protein as observed in adenylate kinase (Muller and Schulz, 1992). In this orientation, there are four hydrogen bonds formed between the phosphate oxygens and the amide nitrogens from the P-loop and the N-terminus of helix A3; they are G49N···OA3 (2.81 Å), G51N···OB1 (3.10 Å), G51N···OB3 (2.91 Å) and S52N···OB3 (2.62 Å). The putative γ-phosphate is adjacent to the side chain of Lys50 which is conserved in all of the nucleotide-binding proteins (Schulz, 1992; Smith and Rayment, 1996). Inspection of the other enzyme structures listed in Table 3.2.2-I reveals that the side chain of Lys50 in FPGS has the same conformation as in all the other proteins. As a representative, the side chain of equivalent lysine of ras-p21 is shown in Fig. 3.2.3-I. This further supports the putative location of the γ-phosphate.

3.2.3.2 The Ω-loop

A special feature of the FPGS active site is an Ω-loop (Fetrow, 1995) of ten amino acids (Ser73-Arg82) that links strands B2 and B3. The sequence alignment of the FPGS family demonstrates that the residues comprising this loop are highly conserved. The hydrogen bonding pattern within and around the Ω-loop is depicted in Fig. 3.2.3-II.

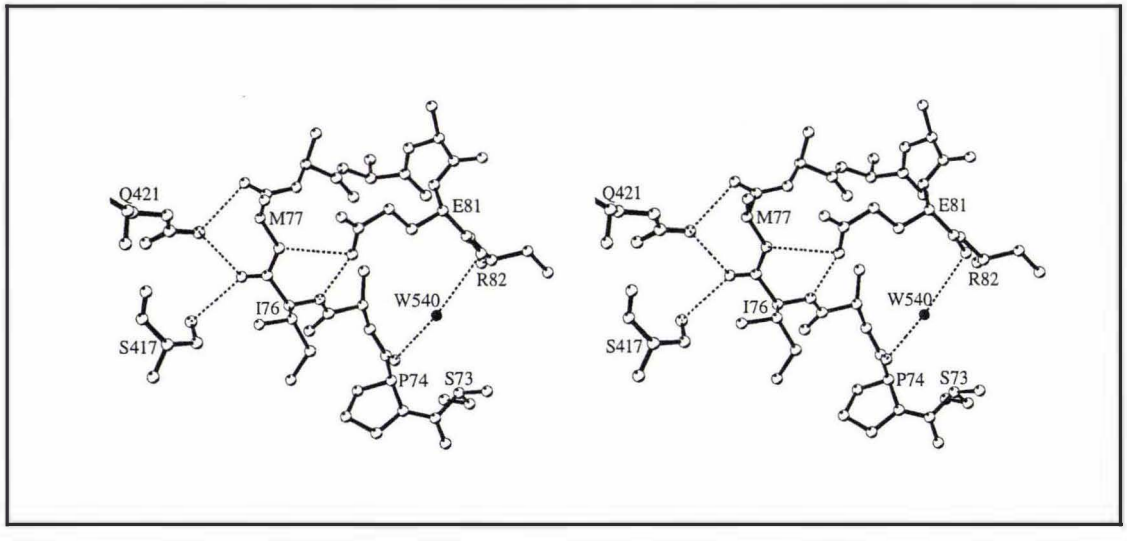


Fig. 3.2.3-II The hydrogen bonding pattern within and around the Ω-loop. For clarity, the side chains of residues 75, 77, 78-80 and 82 are omitted.

A water molecule (W540) hydrogen bonds to the carbonyl oxygens of Glu81 and Pro74 to form a water bridge across the ends of the Ω-loop. Glu81, which is conserved in all FPGS sequences except mouse and *Streptococcus pneumoniae*, (both have an Asp), makes intraloop interactions across to the main chain NH groups of Ile76 and Met77, forming hydrogen bonds. These hydrogen bonds, together with the water bridge, stabilize the loop. Intraloop hydrogen bonds have also been reported in other enzyme structures to maintain loop rigidity during hinge-bending motions (Joseph et al., 1990).

The Ω-loop is located in the region between the ATP-binding site and the C-domain, and it interacts with the C-domain via three hydrogen bonds between the carbonyl oxygen atoms of Ile76 and Met77 and the side chains of Gln421 and Ser417. In addition, there are a number of hydrophobic interactions in this region (see Fig. 3.2.3-II and Table 3.4-I). In the current structure, Pro74 has a cis-conformation and the presence of this cis-proline causes the carbonyl oxygen of the preceding residue (Ser73) to be oriented towards the bound ATP and the putative Mg^{2+} cation (Fig. 3.2.3-I). The Ser73-Pro74 dipeptide is conserved in all known FPGS sequences. Therefore, it is reasonable to assume, given the high level of sequence similarity in this region, that this cis-peptide and whole Ω-loop structure would be retained in the other FPGS molecules.

As reviewed in recent literature (Fetrow, 1995), Ω-loops at substrate-binding sites can be critical for the binding of substrate and subsequent catalysis. Their role may be to help stabilize the bound substrate or enzyme-substrate transition state complex, allowing for efficient catalysis. The Ω-loop in FPGS not only contributes to the ATP-binding site through the cis-peptide, but also mediates interdomain interactions, suggesting that it could play an important role in the activity of FPGS.

3.2.3.3 The divalent cation site

The putative Mg^{2+} cation, modelled in FPGS, is in a position to bond to an oxygen atom from the β-phosphate group of the PPi (2.91 Å), the carbonyl oxygen atom from Ser73 (2.66 Å) and the side chain of Glu143 (2.68 Å) (see Fig. 3.2.3-I). In most nucleotide-binding proteins, the Mg^{2+} coordinates six oxygen ligands in an distorted

octahedral coordination geometry. Generally, two of the ligands are from the β- and γ-phosphates of the nucleotide, two are from water molecules, and two are from the conserved hydroxyl residues (either Thr or Ser). One of the hydroxyl residues is at the position following the lysine residue in the P-loop and another one comes from a region of the polypeptide chain quite removed from the P-loop (Smith and Rayment, 1996). Also, aspartic acid residues have been reported to ligate to Mg^{2+} in other P-loop containing enzymes (Shirakihara and Evans, 1988), such as EF-Tu·GDP complex (an inactive form of the enzyme). In the latter structure, the Mg^{2+} appears to be coordinated to an aspartate residue which is structurally equivalent to the Glu143 of FPGS (Woolley and Clark, 1989). By analogy, the putative Mg^{2+} site in FPGS has a similar coordination environment; potential β- and γ-phosphate oxygens from the ATP, conserved hydroxyl residue (Ser73) and acidic residue (Glu143).

Although the distances between the putative Mg^{2+} site and the ligands in the current structure are longer than typical Mg-O bond lengths (2.2 Å-2.3 Å) (Kjeldgaard et al., 1993; Pai et al., 1990), both the location of the divalent cation binding sites in other proteins and the coordination environment in the ATP-binding site in FPGS support this assignment. Interestingly, MurD has an acidic residue (Glu157) in the same position as Glu143, with the same side chain conformation (Fig. 3.2.3-I). This Glu157 is also believed to bind to Mg^{2+} in the MurD structure (Bertrand et al., 1997). Moreover, all the G proteins listed in Table 3.2.2-I have an aspartate residue in the same position as Glu143, interacting with a Mg^{2+} cation via a water molecule. Glu143 in FPGS and Glu157 in MurD are both conserved in their respective families, perhaps indicative of a similar divalent cation binding role in these two proteins. Three FPGS mutants, E143A, E143D and E143Q, lose enzyme activity (Y. Sheng and A. Bognar, unpublished observation), implying that Glu143 is important for the catalysis, and is probably involved in coordinating the putative Mg^{2+} ion.

Like other nucleotide-binding proteins, FPGS requires Mg^{2+} for its catalytic activity (Bognar and Shane, 1983). The Mg^{2+} is believed to have various possible catalytic functions such as shielding of the negative charge on the attached γ-phosphate or increasing the acid strength of the leaving group (β-phosphate) (Knowles, 1980). Mg^{2+} could also be involved in the stabilization of the transition state of the reaction, and could undergo changes in its coordination (Matte et al., 1998). As discussed previously,

the current structural model for FPGS has a partially bound ATP and perhaps represents an inactive form of the enzyme at low pH (5.3). In this case, another possibility that cannot be ruled out is that the putative Mg^{2+} binding position could either contain loosely bound Mg^{2+} or be occupied by a water molecule.

3.2.3.4 Other residues in vicinity of the active site

There are two potentially significant residues (Arg300 and Asn264) found outside the immediate Mg-ATP coordination sphere (Fig. 3.2.3-I). The residue Arg300 is totally invariant in all FPGS sequences, suggesting a functional role. The side chain of this residue from the C-domain points towards the bound phosphate groups. Although there are no strong hydrogen bonding interactions between its side chain and the phosphate groups in the current structure (the nearest distance is 3.65 Å between Arg300 Nε and an oxygen of β-phosphate), it could approach closely to the bound ATP and hydrogen bond to the phosphate oxygen atoms if the N- and C-domains move closer together during catalysis (see discussion below). There is also a highly conserved arginine residue (Arg302) in the MurD structure, occupying the same position as Arg300 in FPGS and hydrogen bonding to the sulphate, a phosphate analogue. Another residue Asn264, which is conserved in all bacteria and is Gln in both human and mouse FPGS, is located behind the active site and interacts with a phosphate group through a water molecule (Fig. 3.2.3-I). Again, MurD has a conserved residue (Asn271 in MurD) at the same position as Asn264 in FPGS, although there is no corresponding water molecule because the ATP-binding site in MurD structure was modelled with only a single sulphate ion.

3.3 The C-terminal domain

As shown in the topology diagram (Fig. 3.1-II), the small C-domain (residues 300-428) contains a six-stranded β-sheet with five parallel strands (B12, B13, B14, B15 and B16) and one antiparallel strand (B11), together with four surrounding α-helices. There are two surface loops missing in this domain, which correspond to the loop (residues

342-347) connecting strand B13 and helix A11 and the long loop (residues 373-384) connecting strands B14 and B15, respectively.

3.3.1 Comparison with other related proteins

3.3.1.1 Structural similarity to the folate-binding enzymes

It seems probable that the N-domain of FPGS is mainly involved in ATP-binding, presumably functioning in a way similar to other nucleotide-binding proteins. In order to find a possible function for the C-domain of FPGS, a search of the Protein Data Bank was carried out, using DEJAVU (Kleywegt and Jones, 1997), to compare the C-domain structure with structures in the PDB database. The results showed that the C-domain of FPGS bears a remarkable similarity to dihydrofolate reductase (DHFR), strongly implicating this domain as the functional site of folate binding. The DHFR from human (PDB code 1drf) and chicken (PDB code 8dfr) consistently gave rms differences substantially lower than any other molecule, with values of 1.1 Å and 1.3 Å, respectively, for six secondary structure elements. The next best hits were uridine diphosphogalactose 4-epimerase and 3-α,20-β-hydroxysteroid dehydrogenase with rms differences of 3.4 Å and 4.0 Å, respectively. Inspection of the resulting superposition indicated that, although these two structures showed a similar topology to the C-domain of FPGS with regard to the six secondary structure elements used in the search, the connectivity was different.

The topology of the C-domain of FPGS is compared with that of DHFR in Fig. 3.3.1-I. Like the C-domain of FPGS, DHFR also has a central β-sheet with four flanking helices. The structurally equivalent elements include five parallel β-strands (B12-B16) and all four helices (shaded region in Fig. 3.3.1-I).

The folate-binding site in human DHFR is located in a pocket formed from strands A and E, and helix b (see Fig. 3.3.1-Ib for labelling). The only secondary structural differences between DHFR and FPGS are at the N- and C-terminal ends of the molecules (outside the shaded area). The C-domain of FPGS begins with a β-strand (B11) antiparallel to strand B12, whereas DHFR begins with strand A, the equivalent of B12. At the C-terminus of DHFR after helix f, there are three or four extra strands depending on the source of the enzyme (Bolin et al., 1982; Davies et al., 1990). Strand

B11 of FPGS occupies the same position as strand F of DHFR but runs in the opposite direction.

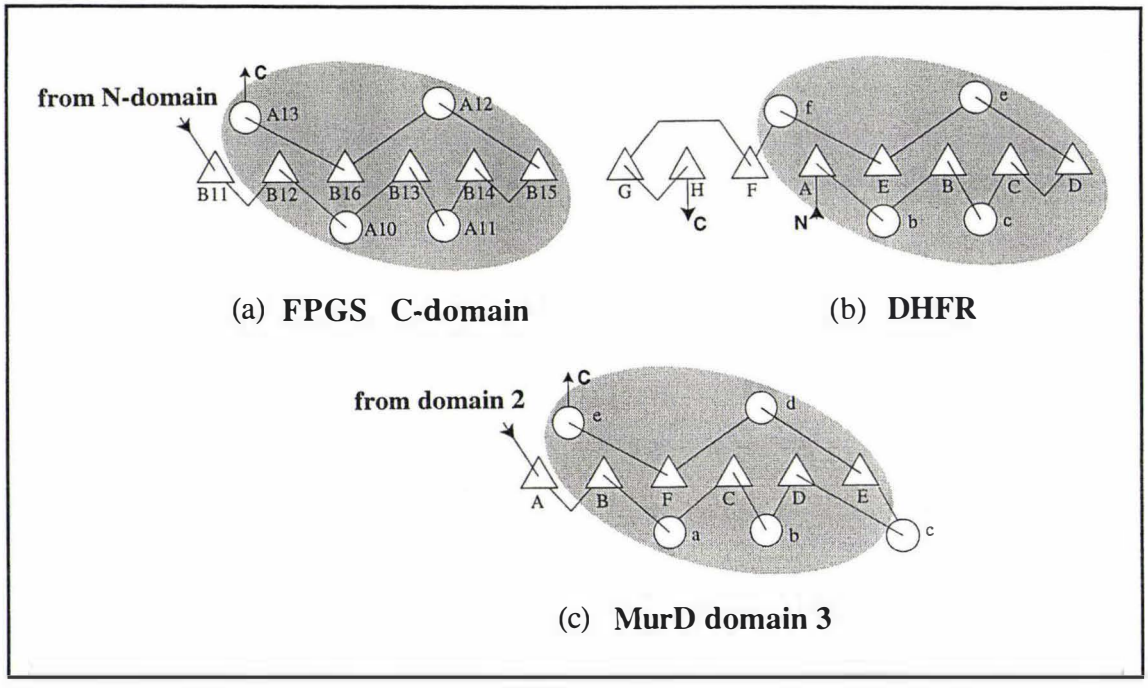


Fig. 3.3.1-I Topology diagrams of (a) the C-domain of FPGS, (b) DHFR, (c) the domain 3 of MurD. The regions of structural similarity among these structures are indicated by the shaded areas.

Although there is some sequence identity (16.4%) between the C-domain of FPGS and DHFR, five β-strands (B12-B16) and two helices (A12 and A13) in the C-domain of FPGS can be superimposed on to equivalent structural elements in human DHFR with an rms difference of only 1.4 Å for 192 main chain atoms, indicating a strong similarity in the spatial arrangement of these secondary structure elements, which constitute most of the domain. The orientation of helix A10 in FPGS is clearly different from the corresponding helix b in DHFR, as helix A10 is rotated in toward the plane of the β-sheet by approximately 11°. The loop connecting strand B12 to helix A10 in FPGS is much shorter (4 residues) than the corresponding region in DHFR (17 residues), and as a result, helix A10 is pulled toward the β-sheet. Nevertheless, the C-domain of FPGS shows a strong resemblance to DHFR, possibly indicative of a functional similarity.

3.3.1.2 Structural similarity to domain 3 of MurD

In addition to the close structural similarity between the N-domain of FPGS and domain 2 of MurD (described in the previous section), domain 3 of MurD also displays a remarkably high degree of structural similarity to the C-domain of FPGS (the MurD structure was not identified by the DEJAVU search because its PDB file was not available from the database at the time). From the topological comparison (Fig. 3.3.1-I), domain 3 has exactly the same topology as that of the C-domain in FPGS, except for one additional helix c (MurD labelling in Fig. 3.3.1-I). All secondary structural elements in the C-domain of FPGS (six strands B11-B16 and four helices A10-A13) can be superimposed on to equivalent elements in domain 3 of MurD with an rms difference of 1.8 Å for 67 Cα positions. Furthermore, the rms difference is only 1.1 Å for 53 Cα atoms when two helices (A11 and A13), which are not perfectly matched, are excluded from the superposition.

The similarities between FPGS and MurD extend even further, to the interdomain linker. As in FPGS, domain 3 of MurD is connected to domain 2 by a single linker (7 residues) which follows the same path as the 6 residue linker in FPGS. Despite limited sequence identity (16.6%) between FPGS and MurD (domain 2 and 3), the similarity of the P-loop sequence and ATP-binding pocket, the striking similarities in topology and spatial arrangement of the secondary structure elements, all suggest that FPGS and the two domains of MurD might carry out their biological functions in a very similar way.

MurD catalyzes the addition of D-glutamate to the nucleotide precursor as shown in the following reaction (Van Heijenoort, 1994):

UDP-MurNAc-L-Ala+D-Glu+MgATP ⇒ UDP-MurNAc-L-Ala-D-Glu+MgADP+Pi

The reaction has been proposed to proceed via phosphorylation of the C-terminal carboxylate of UDP-MurNAc-L-alanine (UMA) by the γ-phosphate of ATP to form an acyl phosphate intermediate, followed by the nucleophilic attack by the amino group of the D-glutamate to produce UDP-MurNAc-L-Ala-D-Glu, ADP and inorganic phosphate (Tanner et al., 1996; Vaganay et al., 1996). This is remarkably similar to the FPGS-catalyzed addition of L-glutamate to folate and follows a similar proposed mechanism

of action (Shane, 1980c; Tang and Coward, 1983). The reaction mechanisms suggested for both enzymes are shown in Fig. 3.3.1-II.

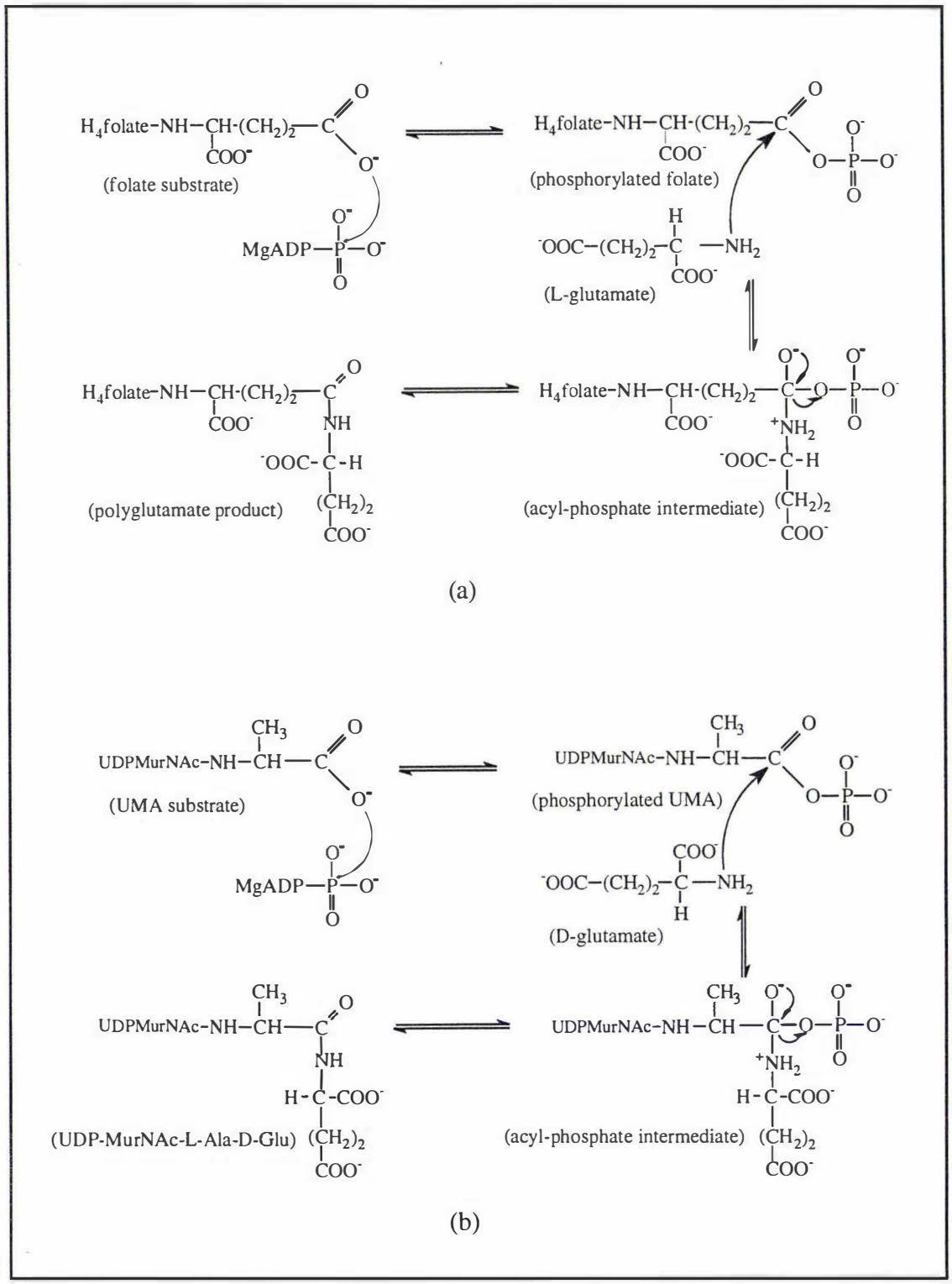


Fig. 3.3.1-II The reaction mechanisms suggested for (a) FPGS; (b) MurD.

In comparison, the substrates for these two enzymes are different; MurD requires UMA and D-glutamate, while FPGS requires tetrahydrofolate and L-glutamate as substrates and the unique negatively charged α-carbonyl group of terminal glutamate of the folate is believed to be involved in the binding of the folate substrate (see Fig. 1.3.4-III). Moreover, MurD has an extra small domain (domain 1) for docking the UDP moiety of UMA. Although the two enzymes possibly bind the UMA and folate in different ways, the domains 2 and 3 of MurD could be expected to have similar catalytic functions to the N- and C-domain of FPGS, considering their highly similar structures, the location of their ATP-binding sites and similar reaction mechanisms proposed (Fig. 3.3.1-II).

3.3.2 Domain movement

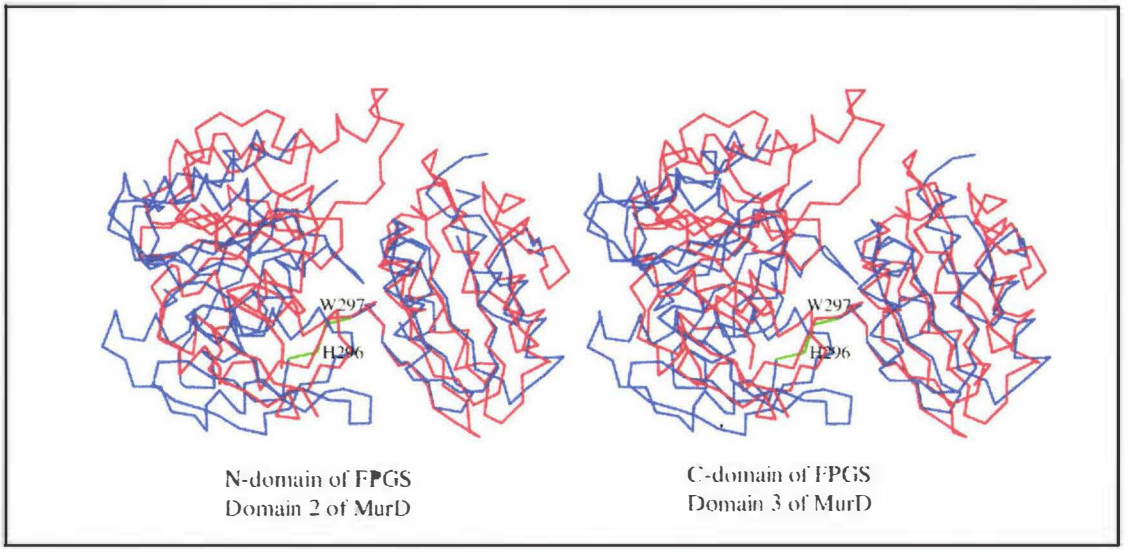


Fig. 3.3.2-I Superposition of domain 3 of MurD (red) on to the C-domain of FPGS (blue). Residues His296 and Trp297 of the interdomain linker of FPGS are indicated (green). For clarity, domain 1 of MurD and helices A1, A2, A4, A5 of FPGS are not shown in this figure.

Superposition of all structurally equivalent elements between FPGS and domains 2 and 3 of MurD (total 132 Cα atoms) gives an rms difference of 2.4 Å, which is significantly greater than the individual differences between the N-domain of FPGS and domain 2 of MurD (rms difference 0.89 Å for 79 Cα atoms), and between the C-

domain of FPGS and domain 3 of MurD (rms difference 1.10 Å for 53 Cα atoms). This is a result of the different relative orientations of the two domains. As shown in Fig. 3.3.2-I, domains 2 and 3 of MurD are closer together than the N- and C-domain of FPGS, by approximately 23°.

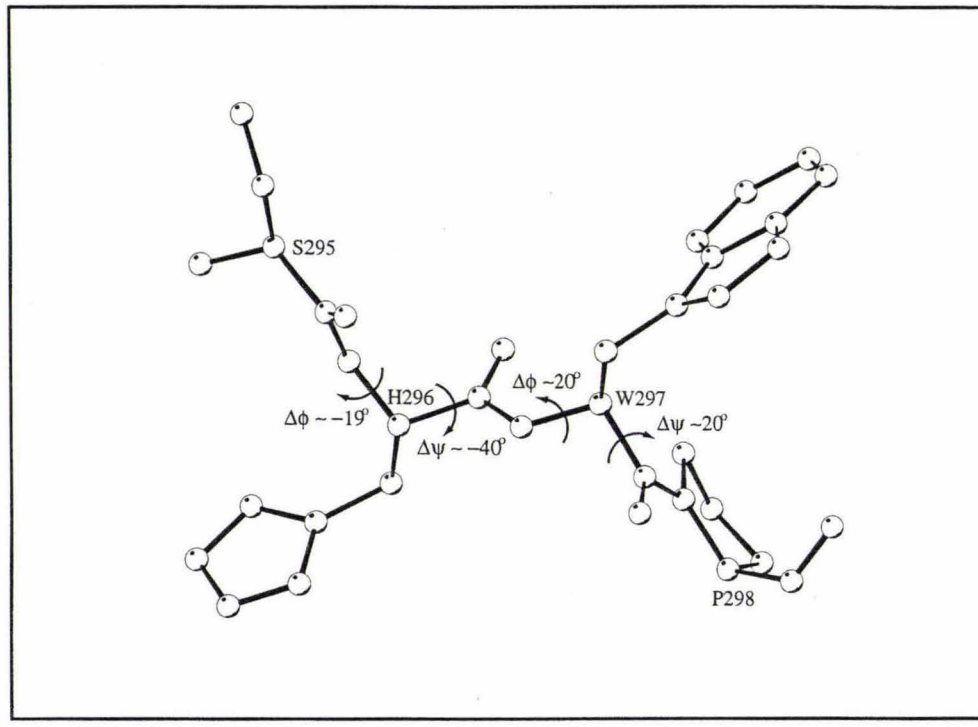


Fig. 3.3.2-II Residues His 296 and Trp297 involved in the possible hinge motion of the N-domain of FPGS. The changes of phi and psi angle responsible for the motion are indicated. The $\Delta\phi$ and $\Delta\psi$ are defined as $\Delta\phi = (\phi_{MurD} - \phi_{FPGS})$; $\Delta\psi = (\psi_{MurD} - \psi_{FPGS})$, respectively.

If the C-domain of FPGS and domain 3 of MurD are initially superimposed as in Fig. 3.3.2-I, the N-domain of FPGS can then be brought into coincidence with domain 2 of MurD simply by changing the main chain torsion angles (ψ, ϕ) of residues His296 and Trp297 (illustrated in Fig. 3.3.2-II). A small additional translation (1.0 Å) of the N-domain of FPGS, away from the C-domain along the interdomain linker, is also required for its complete superposition on the domain 2 of MurD. This may result from the insertion of one additional residue (Thr297) in the interdomain linker of MurD. This domain movement requires torsion angle changes in ϕ (296), ψ (296), ϕ (297) and ψ

(297), as shown in Fig. 3.3.2-II, with one angle (ψ(296)) changing by over 40°. These angles are conjugated across the peptide bonds to produce an overall N-domain rotation of ~23°. This overall domain movement caused by the local torsion angle changes means that the rotation is principally a rigid body motion.

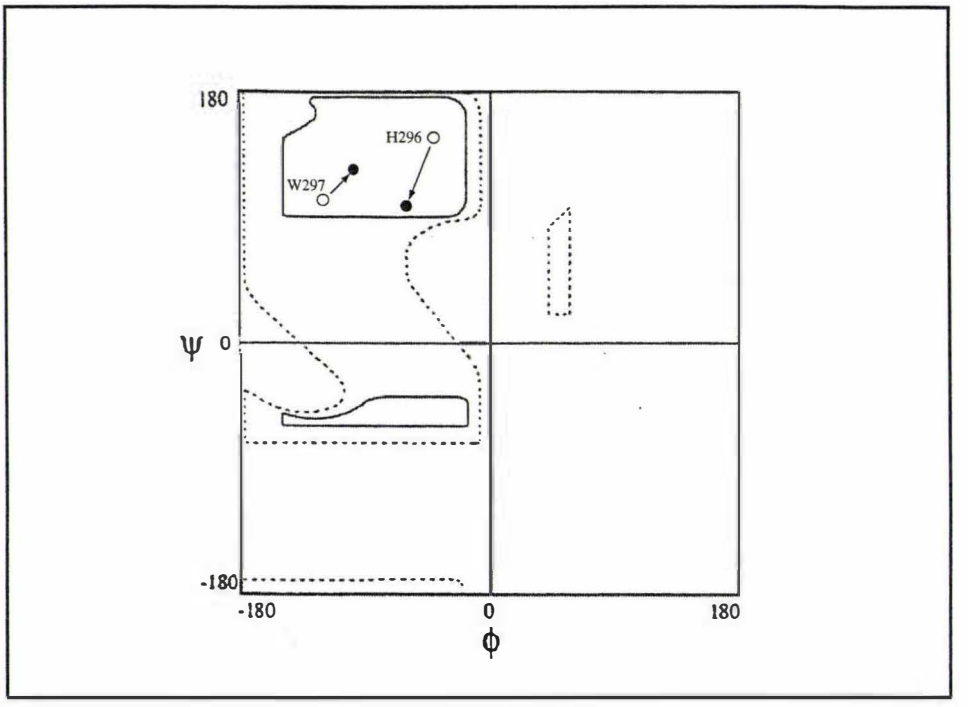


Fig. 3.3.2-III Ramachandran Plot of the presumed main chain torsion angle changes of His296 and Trp297 indicated by arrows. Torsion angles are shown with open circles for the conformation before rotation, and with filled circles for the conformation after rotation.

Like other proteins with a hinge-bending interdomain movement, the main chain in the interdomain linker (or hinge region) of FPGS is free from steric constraint, with the main chain atoms of Ser295, His296 and Trp297 making no interactions with the rest of the protein. This absence of main chain packing constraints, and the absence of internal hydrogen bonds is also found in hinge mechanisms in human lactoferrin (Anderson et al., 1990), lactate dehydrogenase (Gerstein and Chothia, 1991) and adenylate kinase (Gerstein et al., 1993), and appears to be a crucial structural requirement for hinged movement. The torsion angle changes in His296 and Trp297 occur in the most favoured region of the Ramachandran plot (Fig. 3.3.2-III), indicating that the domain motion of

FPGS would involve very low energy transitions and therefore is energetically permissible. The interdomain interactions in FPGS are also few in number in the current crystal structure, with only limited hydrogen bonding and hydrophobic interactions (see next section), and this also gives the domains more freedom to move. Taken together, the presumed rigid body motion between the two domains of FPGS, suggested by structural comparison with the domains of MurD, is structurally reasonable because of the low energy requirement of the principal torsion angle changes and the weak interdomain interactions.

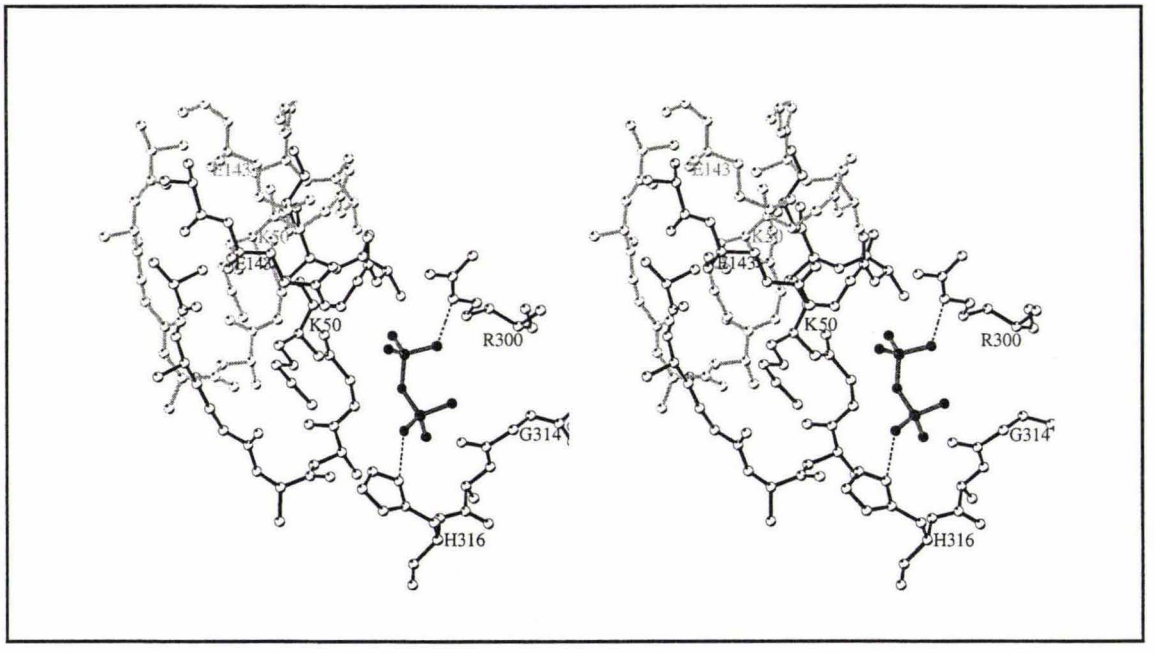


Fig 3.3.2-IV The hydrogen bonding interactions between the pyrophosphate group and residues from C-domain as a result of the presumed domain movement. Some residues, from the β-strand/P-loop/α-helix motif (44-54), β-strand (B4,142-143), loop connecting strand B12 and helix A10 (313-316) and Arg300, are shown. The N-domain residues before domain movement are shown in grey colour. For simplicity, all side chains except for those labelled are omitted.

As a result of such a movement in FPGS, the ATP-binding site could be brought closer to the C-domain, the putative folate-binding domain, as shown in Fig. 3.3.2-IV. The side chain of Arg300 which is three residues away from the proposed hinge region (His296 and Trp297) would approach closer to the enzyme-bound ATP, resulting in hydrogen bonding (2.5 Å) between Arg300 Nε and oxygen of one of the phosphate

groups. At the same time, the His316 Nδ1 can hydrogen bond to an oxygen atom of the adjacent phosphate as well (2.8 Å). These resultant hydrogen bonding interactions caused by the domain movement may be important for stabilization of the reaction intermediate. It has been observed in some kinase structures (Rossemeyer et al., 1993; Schlichting and Reinstein, 1997) that there is a positively-charged basic residue, usually arginine, near the bound ATP. This residue undergoes a structural rearrangement in order to neutralize the strong negative charge of the γ-phosphate for stabilization of the transition state during phosphoryl transfer, and its importance has also been confirmed by mutagenesis (Dahnke et al., 1992). By analogy, the side chain of Arg300 could be expected to play a similar role in the phosphorylation of the γ-carboxyl of the folate substrate. This is supported by its favourable location and complete conservation. The likely importance of this arginine residue lends further support to the idea that interdomain motion could be involved in catalysis.

3.4 Interdomain interactions

The only covalent connection between the two domains is the linker, residues 294-299. As mentioned above, the main chain of the linker, in particular the proposed hinge region (residues 295-297), makes very few interactions with surrounding protein groups. In the current structural model, the pyrophosphate modelled in the ATP-binding pocket binds only to the P-loop and N-terminal residues of following helix A3 in N-domain, and there are no residues from C-domain involved in the interaction with pyrophosphate.

The most striking feature of the interdomain area is the relatively small number of direct protein-protein interactions between the two domains (Table 3.4-I). There are only two small regions of contact which lie on both sides of the linker. One of them comprises residues Ile76 and Met77 from the Ω-loop in the N-domain, interacting with residues Ser417, Arg420, and Gln421 from the last helix (A13) in the C-domain. Another region of contact comprises Trp236 from the loop connecting strands B8 and B9 and Val257 from the loop connecting strand B10 and helix A8 in N-domain, interacting with residues Leu301, Lys303, Val311 and Leu331 from strands B11 and B12 and helix A10 in the C-domain.

Table 3.4-I Interdomain interactions*

Hydrogen bond	Distance (Å)	Non-bonded contact	Distance (Å)
Ile76O···Ser417Oγ	3.02	Ile76O···Gln421Cδ	3.63
Ile76O···Gln421Nε2	2.90	Ile76Cγ2···Ser417Cβ	3.30
Met77O···Gln421Nε2	2.85	Met77O···Gln421Cγ	3.67
Trp236Nε1···Leu301O	2.93	Met77Cε···Arg420O	3.32
		Met77Sδ···Arg420Cδ	3.69
Water bridges		Met77Sδ···Arg420C	3.78
Pro74O···W531···Arg300Nη2	(2.67 Å, 141°, 2.73 Å)†	Trp236Cδ2···Lys303Cβ	3.95
Asp259N···W475···Gln330Nε2	(3.19 Å, 120°, 3.05 Å)	Trp236Cε2···Lys303Cβ	3.72
Gly258N···W477···Gln330Nε2	(2.80 Å, 129°, 3.27 Å)	Trp236Cζ2···Leu301Cδ2	3.49
Val257N···W492···Leu301O	(2.89 Å, 125°, 3.17 Å)	Trp236Cζ2···Val311Cγ1	3.72
Val257O···W522···Arg300O	(2.84 Å, 127°, 3.16 Å)	Trp236Cη2···Val311Cγ2	3.50
		Trp236Cη2···Leu331Cδ1	3.65
		Val257Cγ1···Leu331Cδ1	3.76
		Val257Cγ1···Leu331Cδ2	3.85

* Hydrogen bonds and water bridges listed if the distance is less than 3.5 Å. Non-bonded contacts listed if the distance is less than 4 Å. Water bridges only listed where a single water is involved.

† Figures in parentheses indicate bond length between carbonyl oxygen (or amide nitrogen) and water oxygen, bond angle O (or N)···O (H_2O)···O (or Nε, Nη), and bond length between water oxygen and carbonyl oxygen (or Nε, Nη), respectively.

Three hydrogen bonds are found in the region involving the Ω-loop and only one hydrogen bond in the region involving Trp236. In addition to the hydrogen bonds, a number of non-bonded contacts are present in these two regions. The non-bonded contacts in the region involving Trp236 are mainly hydrophobic in nature. In comparison, non-bonded contacts in the region involving the Ω-loop are only formed between residues that also make hydrogen bonds, suggesting that hydrogen bonds may be the predominant interactions in this region (Table 3.4-I).

A cluster of eight water molecules are found within the interdomain region near Trp236. These water molecules are highly inter-connected to form a network. Four of them form single-water bridges between the two domains (Table 3.4-I), contributing to the interdomain interactions. The other four water molecules are hydrogen bonded to both protein groups and other water molecules. In contrast, the interdomain region involving the Ω-loop has no such cluster of water molecules; there is only one single-water bridge between the two domains in this area.

In summary, the two domains in FPGS interact with each other through two small contact regions. The interactions in the region involving the Ω-loop include three hydrogen bonds and limited hydrophobic contacts. The contact region involving Trp236 contains only one direct protein-protein interdomain hydrogen bond, but a number of hydrophobic contacts and interactions mediated by water bridges. As discussed in the previous section, such few interactions between two domains in FPGS would provide favourable conditions for the proposed hinge-bending motion.

3.5 Putative folate-binding site and implications for catalysis

3.5.1 Folate-binding site in DHFR

In human DHFR (Davies et al., 1990), the polarity of the substrate binding site is complementary to that of folate. The folate molecule consists of polar pteridine and glutamate moieties linked by a relatively nonpolar benzoyl ring of the p-aminobenzoic acid group.

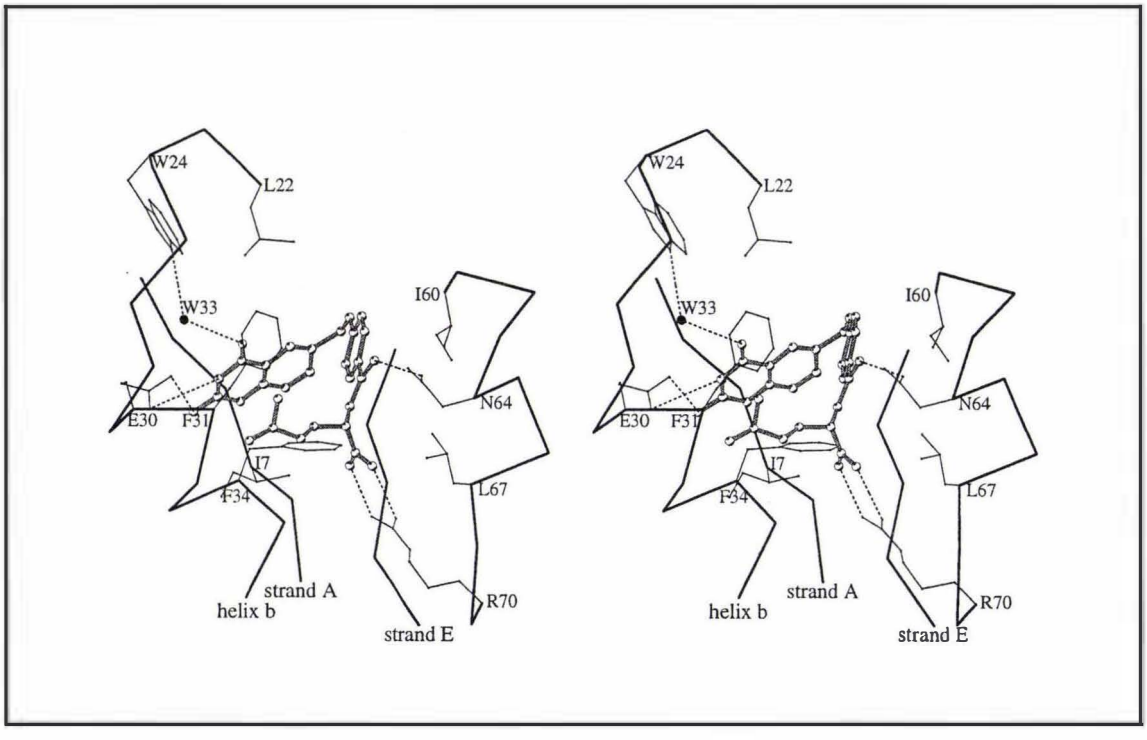


Fig. 3.5.1-I Folate-binding site observed in human DHFR (Davies et al., 1990).

As shown in Fig. 3.5.1-I, the folate substrate is bound in a pocket formed between strands A and E, helix b, and the loop following helix c (see Fig. 3.3.1-Ib for labelling). The pteridine group of the folate is buried against the back surface of the binding pocket and surrounded by backbone carbonyls and polar side chains (strands A and E), while the hydrophobic p-aminobenzoic acid moiety packs against hydrophobic residues from the inner side of the covering helix b and the loop at the C-terminus of helix c. The glutamate group then extends out into solution between helices b and c.

3.5.2 Putative folate-binding site in FPGS

In FPGS, a cavity similar to that in human DHFR exists, which is shown in Fig. 3.5.2-I and is formed between strands B12 and B16 and helix A10 (spatially equivalent to strands A and E and helix b in DHFR). It is reasonable to suggest that this cavity could be the folate-binding site in FPGS, considering the remarkable similarities in both topology and three-dimensional structure between DHFR and FPGS.

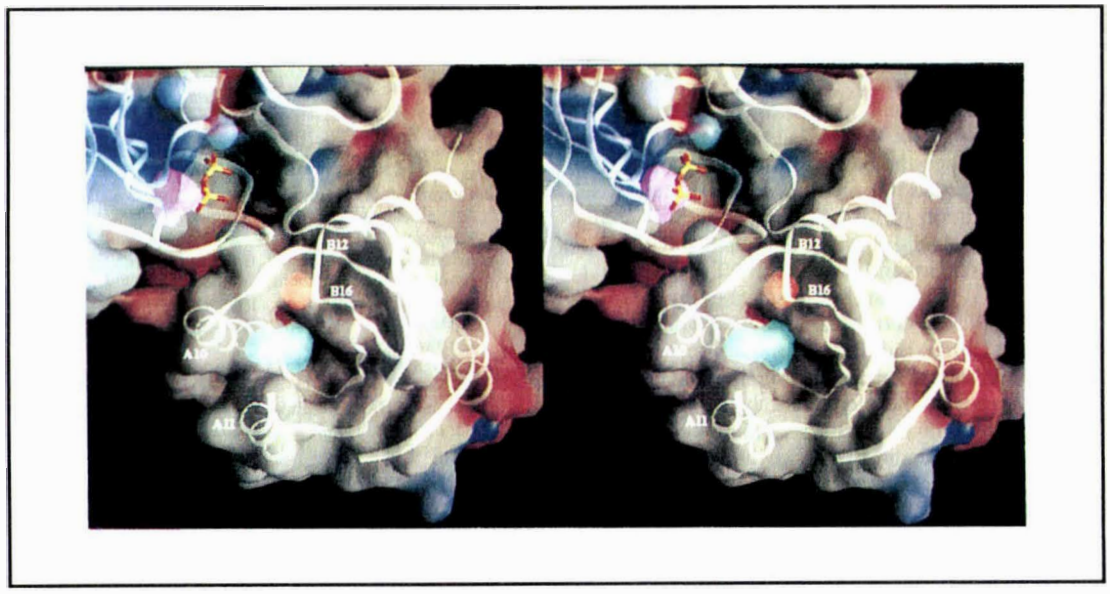


Fig. 3.5.2-I Stereoview of the molecular surface calculated for the N-domain (at top of figure) and the C-domain (lower) of FPGS, coloured according to electrostatic charge (red=negative, blue=positive). A backbone trace of the molecule is shown in white, with the location of the pyrophosphate indicated in red and yellow sticks, lying in a deep interdomain cleft. The putative folate binding site is indicated by the cyan and light red cavities between helices A10 and A11 and strands B12 and B16. This figure was made by using GRASP (Nicholls et al., 1991).

The two enzymes, FPGS and DHFR, catalyze different reactions, with the reaction centre at the pteridine ring of folate for DHFR (Davies et al., 1990; Filman et al., 1982), and at the glutamate tail for FPGS. They also have different folate substrate specificity; *L. casei* FPGS utilizes 5,10-methylene-tetrahydrofolate (Bognar and Shane, 1983) whereas the preferred substrate for DHFR is dihydrofolate. Hence it is reasonable to suppose that the folate binding site in FPGS might not necessarily be exactly the same between the two enzymes.

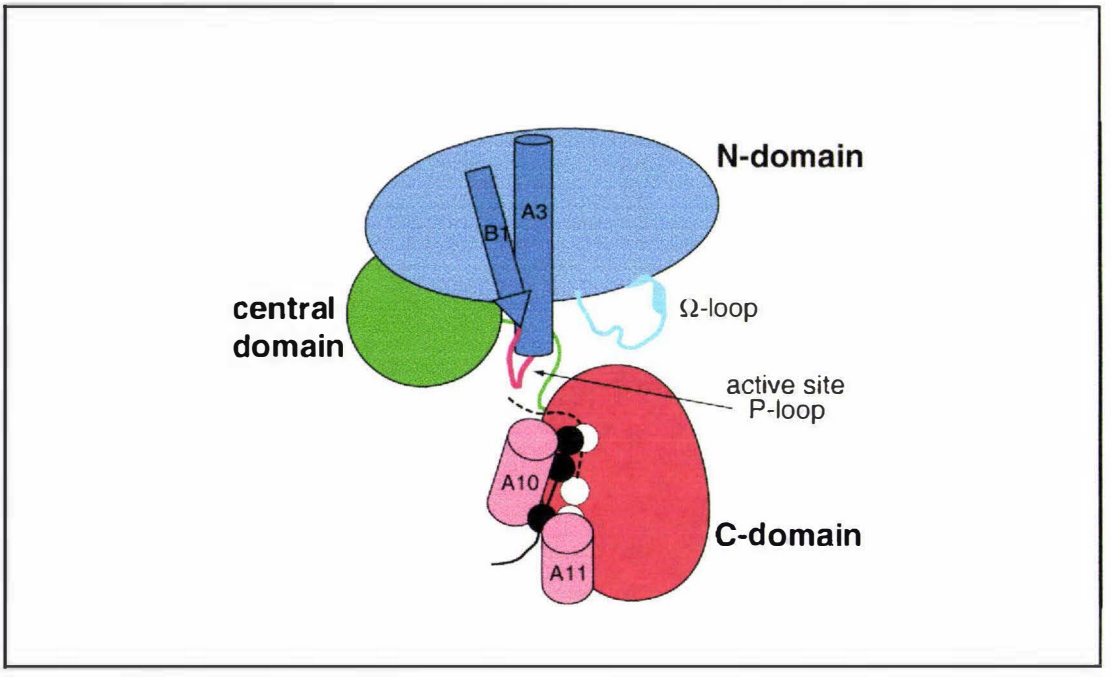


Fig. 3.5.2-II Cartoon representation of FPGS showing the two possible binding modes of folate adjacent to helices A10 and A11. The pteridine is indicated as two fused circles, P-aminobenzoic acid as a single circle, and the polyglutamate tail as a long line. Binding mode for DHFR is presented in black, and mode for FPGS is given in white.

The folate binding mode observed in DHFR is represented in black in Fig. 3.5.2-II, with the glutamate tail projecting out between helices A10 and A11. This binding mode appears to be disallowed in FPGS because of a difference in orientation of helix A10. As noted earlier in Section 3.3.1.1, the helix A10, compared with the corresponding helix b in DHFR, is rotated in toward the plane of the β-sheet due to the shorter loop connecting strand B12 and helix A10. As a result, the cavity in the putative folate binding site in FPGS can be further divided into a larger and a smaller cavity (indicated

in cyan and light red in Fig. 3.5.2-I). This does not allow enough space between A10 and A11 for the polyglutamate tail. The DHFR binding mode would also have a disadvantage in fact that a very large domain reorientation would be needed to bring the glutamate tail close to the ATP-binding site for addition of further glutamate residues. Therefore, we suggest an alternative folate binding mode for FPGS in which the pteridine ring and p-aminobenzoic acid moiety bind the other way around with the glutamate tail projecting into the ATP-binding site from under the loop connecting strand B12 and helix A10. This is shown in white in Fig. 3.5.2-II. In this way the glutamate tail would be placed close to the active site between the domains. Moreover, the elongated shape of the bigger cyan cavity (Fig. 3.5.2-I) suggests that the pteridine ring would be bound more readily in this site than in the smaller site indicated by the red cavity.

3.5.3 Structural implications for catalysis

FPGS catalyzes a complex reaction involving three substrates, ATP, folate and glutamic acid, and a series of steps in which glutamic acid residues are added one at a time to a growing chain. As discussed previously, the structure of FPGS reveals an intriguing modular construction, comprising an ATP-binding N-terminal domain homologous with other nucleotide-binding proteins and a C-terminal domain homologous with DHFR, presumed to be responsible for folate binding. The unexpected and striking structural similarity to the ligase MurD strongly suggests that the domains of FPGS could undergo a possible hinge-bending motion, centred around residues His296 and Trp297 in the interdomain linker. Although further structural analyses are needed to confirm folate binding and find the binding site for glutamate, it seems now reasonable to propose that the catalytic reaction of FPGS is associated with a conformational change involving a hinge-bending motion of the two domains. As a result of this rigid body rotation, the γ-carboxylate group of folate bound in the C-domain would be brought closer to the ATP bound in the P-loop pocket to form an acylphosphate intermediate with the γ-phosphate of ATP, followed by the nucleophilic attack by the amide group of L-glutamate. As discussed in the previous section, the two domains are connected only by a single piece of polypeptide, both areas of interdomain

contact are small and the interdomain interactions are predominately hydrophobic in nature and water mediated. Therefore, such conformational flexibility appears very likely.

3.6 Some structurally important regions common to FPGS family

With the structural model for MgATP-FPGS now available, some structurally important regions common to the whole FPGS family can be seen from the sequence comparisons given in Fig. 1.3.3-I. In addition to the highly conserved P-loop motif which is between Gly46 and Ser52 (all numbering corresponds to *L. casei* FPGS), the Ω-loop region (Ser73-Arg82) described before is also highly conserved. The first two residues, Ser73 and Pro74 which form the cis-peptide and force the carbonyl oxygen of Ser73 to orient towards the putative Mg^{2+}, are invariant in all members of the family, and residues Glu81 and Arg82 in other end of the loop, are almost completely conserved. This means that the Ω-loop motif could be a common structural feature of FPGS.

The next major region of high similarity is located in the loop between Glu143 and Asn154 with sequence pattern EXGXGGXXDXTN, which could correspond to the nucleotide binding B motif (Garrow et al., 1992; Toy and Bognar, 1990). The B motif, DXXG (Walker et al., 1982), follows the β-strand (equivalent to B4 in *L. casei* FPGS) which is parallel and adjacent to the P-loop β-strand (equivalent to B1 in FPGS) and is near the ATP-binding site. It is conserved among the majority of the purine nucleotide-binding proteins, including the ras proteins, EF-Tu, the G-proteins, and myosin. The side chain of the conserved Asp in these proteins adopts an almost identical conformation and is hydrogen bonded to a water molecule bridging to the Mg^{2+}. The B motif, together with the subsequent polypeptide chain in these proteins, exhibits conformational changes as a consequence of nucleotide hydrolysis that are propagated beyond the active site to signal the nature of the nucleotide bound to the enzyme (Smith and Rayment, 1996). During hydrolysis, the amide nitrogen of the conserved Gly moves toward the bound GTP (or ATP) and hydrogen bonds to the γ-phosphate oxygen, altering the conformation of the B motif and changing the orientation of the down stream polypeptide chain (Fisher et al., 1995; Pai et al., 1990).

Glu143 in FPGS occupies the same position as the Asp in the B motif and its side chain has a similar conformation to the Asp, although it directly coordinates to the putative Mg^{2+} (see Fig. 3.2.3-I). Unfortunately, most of residues in this loop are completely disordered in the MgATP-FPGS complex structure, perhaps indicating the inherent flexibility of the glycine residues. This may be an important factor in the conformational change during the catalysis, since these glycines are conserved in all members of FPGS family. The structure of this loop in the *L. casei* apo-protein, in contrast, is well-defined (see discussion in next chapter). In the apo-protein, the carbonyl oxygens of Val144 and Gly145 hydrogen bond to the Nζ atom of Lys50 which is presumably hydrogen bonded to β- and γ-phosphate in the ATP-bound form. The amide nitrogen of Gly147 also points towards the active site and hydrogen bonds to two water molecules near the ATP-binding site. The ras-like domain of the apo-protein (shaded area in Fig. 3.2.2-I) has been superimposed on to the structurally equivalent elements of both MgADP·BeFx·Myosin and MgADP·AlF_4·Myosin complex, where the ADP·BeFx complex mimics ATP in the active site and the ADP·AlF_4 complex is an analogue of the transition state for hydrolysis. In these myosin structures, the conformations of the B motif and the following polypeptide chain were different (Fisher et al., 1995). The superposition shows no correspondence between the loop in apo-FPGS and either of the two B motifs in the myosin structures, suggesting that this loop in FPGS might behave in a somewhat different way from other purine nucleotide-binding proteins. Structural analyses of a complete ATP-bound FPGS and an analogue of the acyl phosphate intermediate for catalysis are needed to observe any possible conformational changes occurring as a result of catalysis. Nevertheless, this loop appears to participate in interactions with the active site and is therefore likely to be involved in catalysis, given its highly conserved sequence and its location near the active site.

Some other significant residues conserved in the N-domain of FPGS are Asp169, His170 and Lys185. Residues Asp169 and His170, in the loop region connecting strand B5 and helix A6, were also disordered in the MgATP-FPGS complex but can be modelled in reasonable density in the apo-protein structure. These two residues are located near the ATP-binding site, with the side chain of His170 pointing towards the active site, although they are not involved in hydrogen bonding interactions with this

part of the active site in the apo-protein structure. Lys185 in helix A6 is located beneath both the P-loop and the conserved loop (Glu143-Asn154), with its side chain pointed towards the active site.

Compared to the N-domain, the C-domain of FPGS has less highly conserved regions but these regions again occupy structurally important positions. Arg300, a potentially important residue for the enzyme catalysis, as described earlier, is conserved in all members of the family. A nearby residue, Glu302 (which is either Glu or Gln in all FPGS enzymes), probably has also a structural function. Its side chain hydrogen bonds (2.6 Å) to the amide nitrogen of Ala299, holding Arg300 in a ring which is connected to strand B11.

The short loop (Asp313-Asn317) connecting strand B12 and helix A10 (two secondary elements possibly involving in the folate-binding) is another conserved region. The side chain of Asp313 hydrogen bonds to the amide nitrogen of Leu301 in strand B11 on one side, and to the Oγ of Thr410 in strand B16 on the other side. The imidazole ring of His316 projects towards the ATP-binding site, and could be involved in hydrogen bonding interactions with the ATP following the presumed domain movement, shown in the Fig. 3.3.2-IV. The high sequence identity in this region, and its location, imply that some of these residues could also be implicated in the enzyme catalysis.

The invariant Gly411 and Ser412 are the two residues in the loop connecting strand B16 and last helix A13. Gly411 has main chain torsion angles ϕ 115° and ψ -32° in a normally disallowed region of the Ramachandran plot. The hydroxyl oxygen of Ser412 hydrogen bonds to two amide nitrogens (Tyr414 and Leu415) of the N-terminus of helix A13. These two conserved residues may be primarily concerned with the proper folding of the polypeptide, correct orientation of helix A13 and the interdomain contacts as discussed earlier.

Interestingly, a comparison between the sequence alignments (a) and (b) in Fig. 1.3.3-I shows distinct differences in sequence between bacterial and eukaryotic FPGS. For example, in bacterial FPGS, the side chain of highly conserved Arg24 hydrogen bonds to the side chain of Glu122 which is also highly conserved in all bacterial FPGS. This hydrogen bond crosses from the N-terminus of helix A2 to that of helix A5, and holds these two adjacent helices together. In contrast, the mammalian FPGS have no

arginine at the position equivalent to the Arg24 and retain arginine at the position equivalent to the Glu122 in *L. casei* FPGS; this may result in the loss of the corresponding hydrogen bonding interaction. As another significant sequence difference, Asn317 is conserved in all bacterial enzymes, however this position has been replaced by threonine in all eukaryotic enzymes, and this residue might be involved in the interactions with the different folate substrates recognized by bacterial and mammalian FPGS.

STRUCTURE OF APO-FPGS

4.1 Structure determination

4.1.1 Initial model and refinement

The starting model for structure determination of apo-FPGS came from the refined structure of MgATP-FPGS at 2.4 Å, with the omission of Mg^{2+}, pyrophosphate, and all solvent molecules. The refined atomic coordinates of MgATP-FPGS were first used as a search model for molecular replacement, using the program AMORE (Navaza, 1994). A cross-rotation function and translation function were evaluated using the data set collected for the apo-FPGS crystal (diffraction terms between 10 and 3.5 Å resolution; inner and outer integration radii for the Patterson functions of 8.0 and 30.0 Å respectively). Following Patterson correlation refinement, a single top orientation of the search model was clearly distinguished from all other peaks of the rotation function, with a correlation coefficient of 25.9%, compared to the second peak with a correlation coefficient of only 6.8%. A translation search was subsequently carried out using the top orientation, which gave an unique solution with correlation coefficient of 54.5% and R-factor of 39.4%. The next best solution gave only 36.2% (correlation coefficient) and 46.6%(R-factor).

Using the model orientated with the unique solution, rigid body fitting was performed, resulting in an initial model with conventional correlation coefficient of 78% and the conventional crystallographic R-factor of 28% for all observed data between 10 and 3.5 Å resolution. The correlation coefficient Cc, is defined as follows:

$$Cc= \Sigma(|F_{obs}|-<|F_{obs}|>)(|F_{calc}|-<|F_{calc}|>)/\Sigma[(|F_{obs}|-<|F_{obs}|>)^2 \bullet (|F_{calc}|-<|F_{calc}|>)^2]$$

where $|F_{obs}|$ and $|F_{calc}|$ are observed and calculated structure factor amplitudes, respectively.

In order to allow for any possible domain movements, rigid body refinement was then employed in the first round of refinement of the initial structural model using the

program TNT (Tronrud et al., 1987). The model was divided into two rigid bodies, the N-domain (residues 1-295) and C-domain (residues 296-425). However, the refinement showed only a small improvement of the overall R-factor, dropping from 27.9% to 26% after 10 cycles of rigid body refinement.

The next rounds of restrained least squares refinement were carried out following the same refinement strategy as that described for the MgATP-FPGS structure in Section 2.6.2, which involved extending the resolution range gradually, adapting the model to the diffraction data first and then bringing the model into agreement with ideal geometry. Individual B-factors were kept constant in the early stages, and then two combined B-factors per residue were used (for the main chain and side chain atoms respectively). Finally individual B-factors were refined in the latter stages of refinement. The free R-factor was calculated using a randomly selected 5% of the data. After each round of refinement, both $2F_o$-F_c and F_o-F_c maps were calculated for model rebuilding. The maps had phases from the model and Fourier coefficients of ($2m|F_o|$-$D|F_c|$) and ($m|F_o|$-$D|F_c|$) as described in Section 2.7.

4.1.2 Structural model rebuilding

The model rebuilding following the first round of refinement mainly involved the repositioning of some long mobile side chains with high B values, such as Gln113, Glu280 and Gln286, and the loop region, residues 314-320 which has weak density in the MgATP-FPGS structure and may be flexible. Unexpectedly, one loop region which was disordered in the MgATP-FPGS structure (residues 146-150), proved to be well-defined in both $2F_o$-F_c (>2σ) and F_o-F_c (>3σ) maps, and all residues of this loop could be convincingly and readily built into the model.

Some well-ordered solvent molecules (regarded as water) were also added into the model if they had potential hydrogen bond partners from the protein and were visible in both $2F_o$-F_c and F_o-F_c maps. Water molecules which refined to high B values (>70 Å^2) were reviewed in each of rebuilding cycle and those with poor density were removed.

Omit maps were calculated to help in the rebuilding of side chains with weak density, and to check regions in doubt. In the latter stages of rebuilding and refinement, more new density was seen for regions in the N-domain that could not be modelled in

the MgATP-FPGS structure. This allowed two loops, 18-20 and 169-176, to be built into the model, although some residues still had poor density and were built as alanine. At the same time, some new density also appeared for the missing loop regions in the C-domain, although it was not sufficient to build all residues into these two missing loops.

After each round of model rebuilding and refinement, difference Fourier maps, with the mask bulk solvent correction described in Section 2.7.4, were calculated as well. These proved useful for building the missing residues, because they often provided some new density in the regions of missing loops, compared to the conventional difference Fourier maps with the exponential bulk solvent correction.

4.2 The quality of final model

The final refined model contains 413 out of 428 residues, a total of 3122 non-hydrogen atoms and 202 solvent molecules, all regarded as water. There are still two loops missing in the C-domain (residues 344-347 and 375-385). The geometry of the model is tightly restrained with an rms deviation of 0.007 Å from standard bond lengths and 1.4° from standard bond angles. The final crystallographic R-factor is 19.4% for all 11431 reflections in resolution range of 20-2.5 Å. The refinement statistics are shown in Table 4.2-I.

Table 4.2-I Refinement statistics for apo-FPGS

Resolution range:	20-2.5 Å
Number of protein atoms:	3122
Number of solvent (water):	202
Crystallographic R-factor:	0.194 (11431 reflections)
free R-factor:	0.293 (570 reflections)
TNT stereochemistry,	rms deviations:
Bond length	0.007 Å
Bond angles	1.4°
Torsion angles	20.8°
Trigonal planes	0.008 Å
General planes	0.012 Å
Bad contacts	0.042 Å
B value correlation	1.8 $Å^2$

The average maximum coordinate error for the final model of the apo-FPGS was estimated between 0.2 and 0.25 Å from a Luzzati plot (Luzzati, 1952), and 0.25 Å from a σ_A plot, which is shown in Fig. 4.2-I. The Ramachandran plot for apo-FPGS shows that ~90% of the residues in the model are in the "most favoured regions" with the rest of the residues in the "allowed regions". None of the backbone torsion angles were restrained during refinement.

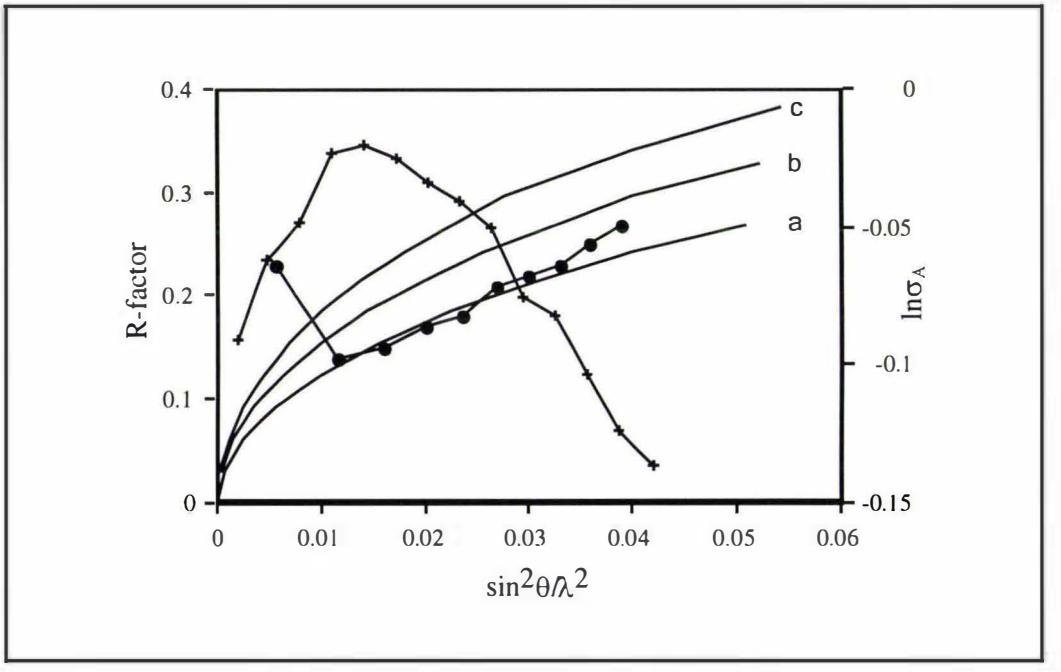


Fig. 4.2-I Plots of the variation of the crystallographic R factor (•) and $\ln\sigma_A$ (+) with resolution for apo-FPGS. The theoretical variation of the R factor for coordinate errors of 0.20 Å, 0.25 Å and 0.3 Å (curves a, b, and c, respectively) is shown.

The mean B value is 33.4 Å^2 (rms deviation 2.4 Å^2) for all main chain atoms and 39.3 Å^2 (rms deviation 5.2 Å^2) for all side chain atoms. Although the patterns of B value distribution are consistent between the apo-FPGS and MgATP-FPGS structures, implying similar dynamic properties, the mean B values for both main chain and side chain atoms are greater in the apo-FPGS structure than in the MgATP-FPGS structure (26.4 Å^2 for main chain, 32.4 Å^2 for side chain). This may correlate with the better quality of the diffraction data of MgATP-FPGS (cf. Fig. 2.3.2-I), compared to the apo-FPGS.

4.3 Structural comparison with MgATP-FPGS

4.3.1 Overall structural similarity

The apo form of FPGS has the same polypeptide folding pattern as MgATP-FPGS. There is no interdomain movement observed when these two structures are superimposed on each other, which suggests that the possible interdomain movement discussed in Section 3.3.2 may only occur when the enzyme binds all three substrates or is in the process of forming the acylphosphate intermediate. It may also be that the MgATP form is effectively in an apo-like state because the ATP is so poorly bound. All 1568 main chain atoms of MgATP-FPGS superimpose upon those of apo-FPGS with an rms difference of only 0.48 Å. The residues with relatively greater differences are mainly those at the beginning and end of the missing loop regions. These residues have high B values and their omission results in an rms difference of 0.34 Å for 1536 main chain atoms.

All main chain atoms around the ATP-binding site of apo-FPGS, such as in the P-loop, remain in the same conformation as in MgATP-FPGS. There is a cluster of four water molecules in the ATP-binding cleft. Two of these, W550 and W582, occupy the positions of the pyrophosphate in MgATP-FPGS, hydrogen bonding to the amide nitrogen of Gly49 and side chain of Asn48, respectively. The other two water molecules have no hydrogen bonding partners from the protein but bind to W550 and W582.

Several side chains in the vicinity of the ATP-binding site adopt a somewhat different orientation. The Nζ of Lys50 shifts away from the pyrophosphate positions to hydrogen bond to the carbonyl oxygens of Val144 and Gly145. The side chain of Glu143 also shifts away from the putative Mg^{2+} site, resulting in the loss of the hydrogen bond between Oε1 of Glu143 and the amide nitrogen of Ser73, which is present in the MgATP-FPGS structure. These minor side chain conformational changes seem to be attributed to absence of the Mg^{2+} and pyrophosphate in apo-FPGS.

4.3.2 Structure of the recovered missing loops

As described in Section 4.1.2, all three loops (residues Lys18-Gly20, Ile146-Thr150 and Asp169-His176) which were missing in the N-domain of MgATP-FPGS are visible in the apo-FPGS structure. The electron density for the loop Ile146-Thr150 is particularly clear and the residues were built unambiguously into the model. The hydrogen bonding pattern around this loop is shown in Fig. 4.3.2-I. A water molecule, W433, stabilizes the loop by making hydrogen bonding interactions with carbonyl oxygen and amide nitrogen atoms from the loop. Further, the side chain of Asp151 hydrogen bonds with the amide nitrogens of Ile146 and Gly145, and may play a role in ordering this flexible loop and maintaining its well-defined structure.

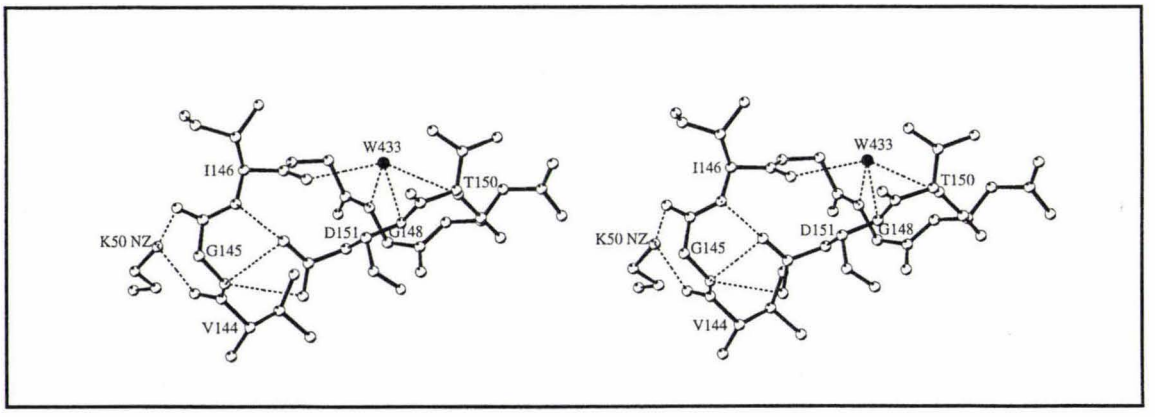


Fig. 4.3.2-I The hydrogen bonding pattern around the loop of Ile146-Thr150.

The refinement of this structural model was based on a diffraction data set collected from an apo-FPGS crystal which was grown in acetate buffer (pH 5.3) and soaked in Hepes buffer (pH 7.5) overnight, prior to data collection. Thus, the current structural model for the apo-FPGS represents a structure of FPGS at a pH which is near the physiological pH value (7.4). The ionizable side chains of the acidic and basic amino acids often play important roles in establishing the structure, stability, and function of proteins (Creighton, 1993); therefore, considerable experimental and theoretical efforts have been directed toward understanding the factors that dictate the pKa values of amino acids in proteins (Yang et al., 1993). In addition to the "normal" value (~3.9),

the pKa of the carboxyl group of aspartic acid residue in proteins can vary dramatically between 2 and 9, depending on its functional and structural environment, such as its degree of exposure to the solvent (Joshi et al., 1997), catalysis (Wilson et al., 1995), and ligand binding (Katz, 1997). Direct measurement of pKa values for aspartic acid residues show that these can have pKa values for the carboxyl as high as 6.5 (Xie et al., 1997) or 7.3 (Jeng and Dyson, 1996). A correlation between the ordering of a flexible loop and the pKa of an aspartic acid residue has been reported in the crystal structure of streptavidin (Katz, 1997). At the current stage, however, it is not possible to say whether the pKa of Asp151 of FPGS also has an unusual value and more FPGS structures under different pH are required to clarify the possible correlation between the ordering of the loop and pKa of the carboxyl group.

As shown in Fig. 4.3.2-I, the Nζ of Lys50 in this apo-FPGS structure moves to hydrogen bond to the carbonyl oxygens of Val144 and Gly145, which holds these two residues more rigid. These hydrogen bonds should also make the amide hydrogens of Gly145 and Ile146 more positive through polarization of the peptide groups; as a result, the hydrogen bonds between the carboxyl group of Asp151 and amide nitrogens of Gly145 and Ile146 could also be strengthened. This might contribute to the ordering of the loop (Ile146-Thr150) as well.

As discussed in Section 3.6, both loops (Ile146-Thr150 and Asp169-His176) show high sequence conservation within the FPGS family. The loop Ile146-Thr150, which is similar to the B motif (Smith and Rayment, 1996) in its location, might participate in interactions with the bound ATP, although it takes a different course from that in other nucleotide-binding proteins. However, structural comparisons show that these two loops follow similar courses to the structurally equivalent loops in MurD when the central β-sheets of domain 2 of MurD and the N-domain of apo-FPGS are superimposed (Fig. 4.3.2-II). Of the two loops, the second (Asp169-His176) is much closer to its counterpart in MurD, each residue in the loop of FPGS matching an equivalent residue in MurD. More significantly, two residues in FPGS, Asp169 and His170 which are located near the ATP-binding site in the apo-FPGS structure, have their counterparts in MurD, Asp182 and His183, interacting with the substrate UDP-N-acetylmuramoyl-L-alanine (Bertrand et al., 1997).

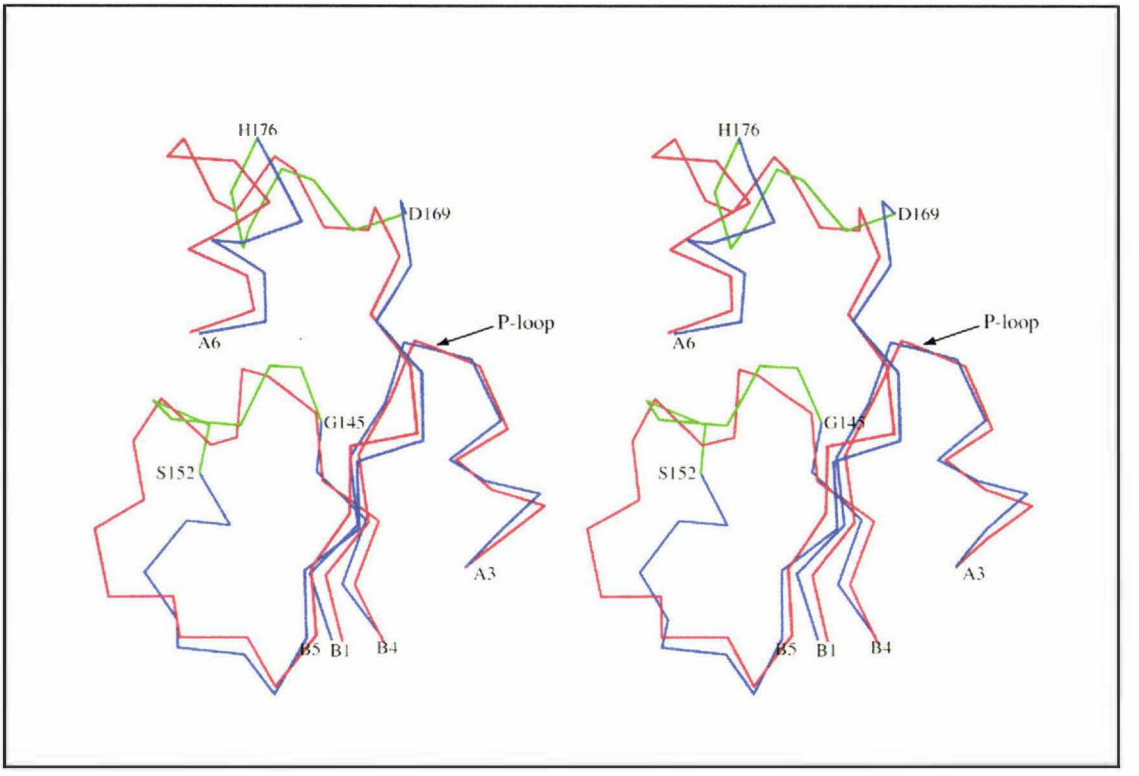


Fig. 4.3.2-II Comparison of loop regions. The Cα superposition between the N-domain of apo-FPGS (blue) and the domain 2 of MurD (red) is based on the central β-sheet. The loop regions of FPGS are indicated in green. The labels for β-strands and α-helices are from FPGS.

As mentioned previously in Section 3.3.1.2, the overall structural similarity between FPGS and MurD, including these two flexible loops, may be related to similar structural and catalytic functions. However, more structural analyses are needed to fully establish this relationship.

PART B

CRYSTALLOGRAPHIC STUDIES OF RECOMBINANT HUMAN LACTOFERRIN

INTRODUCTION AND LITERATURE REVIEW

5.1 Introduction

Iron is extensively used as an essential element for the growth and survival of all living cells. However, Fe(III), the physiologically stable form, is essentially insoluble at concentration above 10^{-18}M (Aisen and Listowsky, 1980). The problem of maintaining iron in soluble non-toxic form has been solved by proteins contained in most organisms. In vertebrates, a family of proteins known as transferrins fill this role.

Proteins of the transferrin family, which include serum transferrin, ovotransferrin, melanotransferrin and lactoferrin, share the common function of controlling the level of free iron in biological fluids through their ability to solubilize, sequester and transport Fe(III) ions. Serum transferrin is the major iron-transport protein in blood, existing in all vertebrates and some invertebrates such as crabs and spiders (Huebers et al., 1982; Lee et al., 1978). Ovotransferrin, found in the egg white, is basically identical to the serum transferrin of birds except in its carbohydrate moiety (Thibodeau et al., 1978). Melanotransferrin is a membrane-bound protein, present at low levels on the surface of normal cells (Brown et al., 1982). Lactoferrin, the target protein of this study, is widespread in the secretory fluids of higher animals, including milk, tears, saliva, mucosal, and genital secretions, also in white blood cells (Baggiolini et al., 1970; Masson et al., 1966).

All the transferrins are monomeric glycoproteins of 670-700 amino acids with molecular weights of approximately 80 kDa. The exception is melanotransferrin which has additional amino acids on the C-terminal end of the molecule. This extension is thought to be involved in anchoring melanotransferrin to the plasma membrane of melanocytes (Rose et al., 1986), and is responsible in part for its higher molecular weight of 97 kDa. There is a high level of similarity between the four classes of transferrins, and transferrins also exhibit an internal amino acid sequence identity of ~ 40% between the N- and C-terminal halves, which is reflected in their bilobal molecular structure. The characteristic property of the transferrins is to bind one Fe^{3+} ion together with one CO_3^{2-} ion into each of two lobes. The metal and anion-binding amino acid residues are identical in both lobes of the transferrins and conserved in all members of

the transferrin family. The relationship between the metal ion and the anion is synergistic, neither ion being bound strongly without the other. This binding is tight, but more importantly reversible, and many of the biological functions of the transferrins depend on this reversibility. Melanotransferrin, however, binds only one ferric ion, in its N-terminal binding site, as a result of specific amino acid changes in the C-terminal binding site (Baker et al., 1992).

Although transferrins generally share common roles such as regulating iron levels in the body (Octave et al., 1983), depressing bacterial growth by iron deprivation (Oram and Reiter, 1968) and protecting against the toxic effects of free iron that might catalyse the formation of cell-damaging free radicals (Winterbourn, 1983), their different biological locations and certain individual molecular properties distinguish lactoferrin from serum transferrin. Two features are particularly notable. First, the affinity for iron of lactoferrin is about 300 times higher than that of transferrin (Aisen and Leibman, 1972). Transferrin begins to release iron when the pH is lowered below about 6.0, whereas lactoferrin retains iron down to about pH 3.5. As a biological consequence, this allows lactoferrin to retain its iron-binding capacity in low-pH regions of the body, such as the upper small intestine, female genital tract, and perhaps also at inflammatory foci. The second feature that distinguishes lactoferrin from transferrin is its much higher isoelectric point, about 8.4-9.0 (Moguilevsky et al., 1985), compared with 5.4-5.9 for transferrin (Hovanessian and Awdeh, 1976). A sequence of four arginine residues close to the N-terminus of human lactoferrin that are absent from transferrin seems to make an important contribution to the cationic nature of lactoferrin and may mediate some of its biological properties. The high pI of lactoferrin also endows it with an ability to react with or bind to a large variety of molecules (Brock, 1997).

5.2 Biological roles of lactoferrin

For many years, a considerable amount of effort has been directed at determining the biological functions of lactoferrin. A variety of roles have been postulated. It seems that the most of these functions are related to the ability of lactoferrin to bind iron reversibly or/and to interact with other molecules. However, the mechanisms underlying some of these effects remain unknown despite our advanced knowledge of

lactoferrin structure. Some of the evidence for the putative functions is also contradictory.

5.2.1 Lactoferrin and iron absorption

The fact that serum transferrin has a clearly defined role as an iron-transport protein and can donate iron to cells through a process of receptor-mediated endocytosis (Octave et al., 1983) suggested that lactoferrin might play a similar role under certain circumstances. It was proposed early in 1960 that lactoferrin acted as a specialised iron-transport protein in the gut and might deliver iron to mucosal cells (Montreuil et al., 1960). This was supported later by the evidence that lactoferrin could deliver iron to these cells (Cox et al., 1979) and the identification of lactoferrin-binding sites on mucosal cells (Mazurier et al., 1985). These experiments showed that lactoferrin may interact with specific binding sites on the brush border membrane and release iron to these cells. Exclusively breast-fed infants maintain adequate levels of iron longer than infants fed on bovine milk formula (Saarinen et al., 1977), which suggests that the iron in human milk is present in a more available form. Furthermore, a 100 kDa protein, identified as a lactoferrin receptor has been found on rabbit brush border cells (Mazurier et al., 1989), as well as rhesus monkey brush border cells (Davidson and Lonnerdal, 1988; Davidson and Lonnerdal, 1989). Another location where specific lactoferrin receptors have been found is on mouse small intestinal brush border membrane sections. This receptor was found to bind several lactoferrins (i.e., human, bovine and mouse) and had a higher affinity for mouse lactoferrin (Fairweather-Tait et al., 1987). A question raised here is how the iron is released from the lactoferrin. Unlike transferrin, from which the iron is released spontaneously at the intracellular pH of 5-6 following receptor mediated internalisation, lactoferrin does not lose iron until pH 3.5 which is much lower than the intracellular pH. However, studies on the receptor-mediated endosomal release of iron from transferrin (Bali and Aisen, 1991; Bali and Aisen, 1992) might suggest a possible mode of iron removal for lactoferrin, which involves not only pH but also the action of the receptor itself.

On the other hand, clinical studies of infants receiving formulas supplemented with bovine lactoferrin failed to show any enhancement of iron absorption (Fairweather-Tait et al., 1987; Schulz-Lell et al., 1991), and removal of lactoferrin from breast milk was

reported actually enhancing the bioavailability of breast milk iron (Davidsson et al., 1994). Therefore, whether lactoferrin has a role in iron absorption is still not very clear at present, although the presence of mucosal lactoferrin-binding sites or receptors in various species is now well documented.

5.2.2 Bacteriostatic and bactericidal properties

Like other members of transferrin family, lactoferrin has been studied extensively as an antimicrobial agent which has both bacteriostatic and bactericidal activities. Most micro-organisms require iron for growth. Therefore, lactoferrin has the potential for depriving bacteria of this essential element through its high affinity for iron and retarding their growth. This is consistent with the fact that almost all of the lactoferrin in milk is iron free (Brock, 1985; Lonnerdal, 1985) and thus capable of binding free iron.

Early in 1968, the ability of lactoferrin to inhibit growth of iron-requiring bacteria was reported and the inhibition was believed to be a result of iron limitation in the milieu caused by lactoferrin (Oram and Reiter, 1968). Another study to show the inhibitory effect of human milk and of lactoferrin was published by Bullen et al. (Bullen et al., 1972). They showed that, *in vitro* , the growth of *E. coli* 0111 was inhibited by milk and the inhibition could be reversed by the addition of iron to the milk. Purified lactoferrin was found to have a similar effect, and this effect could also be abolished by saturation of the protein with exogenous iron. Several independent studies have shown that breast-fed infants who ingest substantial amounts of lactoferrin have a lower incidence of gastrointestinal infections than formula-fed infants (France et al., 1980; Howie et al., 1990). Many investigators have also confirmed that lactoferrin has a bacteriostatic effect on a broad spectrum of gram negative bacteria (Bishop et al., 1976; Reiter, 1983). The bacteriostatic activities of lactoferrin *in vitro* were established beyond doubt by a variety of studies with a common putative mechanism of iron deprivation.

However whether or not these effects could exist *in vivo*, depends on whether lactoferrin can remain biologically active in the intestine. Studies on the ability of gastric juices from infants to degrade lactoferrin showed (Britten and Koldovsky, 1989) that lactoferrin was completely degraded at low pH but at pH 5.8 the degradation was

minimal. Mason (Mason, 1962) also reported that feeding increases the pH of the gut of the infant and the pH possibly remains between 5 and 6.5 for about 2 hours after feeding. These results suggest that it is reasonable to expect at least a fraction of lactoferrin from human milk to survive digestion. This is also supported by the finding (Prentice et al., 1987) that the amount of lactoferrin in the stool of breast-fed infants was greater than that of formula-fed controls. Studies of the gastric digestion of lactoferrin revealed that limited proteolysis probably occurs in addition to the possible gastrointestinal survival of lactoferrin. Trypsin digestion of diferric human lactoferrin produces two major species corresponding to N-lobe (30 kDa) and C-lobe (50 kDa) fragments (Legrand et al., 1986), and the trypsin digestion does not affect the iron-binding or bacteriostatic properties of lactoferrin (Legrand et al., 1984). The N-lobe of the molecule, however, is pepsin-sensitive, whereas the C-lobe remains intact after pepsin digestion (Line et al., 1976). Certain fragments of lactoferrin generated by pepsin hydrolysis could, however, posses antibacterial effects. A recent study on the bacteriostatic effects of bovine lactoferrin demonstrated that bovine lactoferrin can exert bacteriostatic effects in the gut of milk-fed animals. Lactoferrin present in mammalian milk may protect infant animals against gastrointestinal infections even after it has been digested (Teraguchi et al., 1997).

In addition to the bacteriostatic activity of lactoferrin through iron deprivation, which is now well documented and has also been extensively reviewed (Griffiths and Bullen, 1987; Sanchez et al., 1992), bactericidal activity which is associated only with lactoferrin and not transferrin was first reported by Arnold et al. (Arnold et al., 1977). This direct bacterial killing effect was irreversible with the addition of exogenous iron, and strongly dependent on the physical environment (i.e., pH, temperature, osmolarity) and required adsorption of lactoferrin to the target cell surface (Arnold et al., 1982; Arnold et al., 1981), This requirement of directly binding to the microbial surface was supported by experiments with the use of a dialysis membrane (Valenti et al., 1987; Valenti et al., 1985), which showed that lactoferrin needed to be in direct contact with the bacterial surface to exert its bactericidal effect. Furthermore, hydrophobic drugs that cross the bacterial outer membrane become more effective in the presence of lactoferrin (Ellison et al., 1988). These results suggested that lactoferrin may affect the bacterial membrane and may have more potential during active cell division (Naidu and Arnold, 1997).

Research by Bellamy et al. (Bellamy et al., 1992; Yamauchi et al., 1993) cast a new light on the nature of the bactericidal activity of lactoferrin. They isolated microbicidal peptides from both bovine and human lactoferrin through limited proteolysis. These peptides corresponded to basic regions of the N-terminal regions of the lactoferrins (residues 18-40 in human lactoferrin and 17-41 in bovine lactoferrin) and proved to be much more active than intact lactoferrin. The bactericidal activities of the peptides are independent of iron concentration but require contact with bacteria, demonstrating that iron deprivation was not involved in the bactericidal properties of the peptides. These arginine- and lysine-rich peptides, like other antimicrobial cationic basic proteins such as those found in lysosomes, may involve direct interruption of the membrane. Further investigation was carried out (Bellamy et al., 1993) on the cell binding properties of the peptide from bovine lactoferrin with both gram positive (*Bacillus subtilis*) and gram negative (*E. coli*) bacteria. It was found that the peptide bound to the cell surface of both bacteria and that the rate of binding was consistent with the rate of killing observed. Cell-binding was pH dependent, which indicates the importance of the cationic property of the peptide. With each bacterial strain, the killing effect was maximum near the optimum pH for cell binding implying that the bactericidal activity of the peptide is related to cell binding. In a recent study, Tomita et al. (Tomita et al., 1994) found that lactoferrin agglutinated the protoplasts of *Micrococcus lutens* and this agglutination was lost by chemical modification of the basic residues of lactoferrin, indicating that the cationic charge from lysine and arginine residues in lactoferrin is essential for agglutinating the bacterial cells. Moreover, both the apo- and iron loaded forms of lactoferrin inhibited the growth of the bacteria, once again confirming that the bactericidal effect of the lactoferrin is through a specific-binding mechanism other than iron deprivation mechanism.

5.2.3 Lactoferrin and inflammation

The role of lactoferrin in modulation of the inflammatory response has been investigated, with a large and diverse range of findings that support a general role of lactoferrin in inflammation, although the exact mechanism is poorly defined and further studies for crucial missing evidence are needed.

In addition to secretory fluids, lactoferrin is also found in the secondary granules of neutrophils (Baggiolini et al., 1970). Neutrophils, which are phagocytic cells that swallow invading bacteria, respond to stimuli and play an important role in the inflammatory response at the site of damage. During the inflammatory response, plasma lactoferrin levels increase and levels of lactoferrin in the granules decline (Malmquist, 1978), suggesting that the neutrophils release the lactoferrin to play a role in inflammatory response. Further studies by Boxer et al. (Boxer et al., 1982), using immunohistological techniques, showed that lactoferrin released from neutrophils bound to the polymorphonuclear leucocytes (PMN) membrane surface and altered the surface properties of these cells. This binding probably increased the stickiness of the PMN, promoting cell-cell interactions. This work correlates with the observation that during inflammation there is an increased adherence of PMN cells to the endothelial cells which helps to retain them at sites of inflammation and amplify the inflammatory response (Oseas et al., 1981).

The lactoferrin released from neutrophils has also been suggested to have an antibacterial role as well (Van Snick et al., 1974). It was postulated that lactoferrin released from the neutrophils bound free iron, possibly derived from transferrin, and once iron-saturated, the lactoferrin bound to macrophages which were then removed by the reticuloendothelial system. This bacteriostatic effect of lactoferrin in neutrophils is supported by the finding that patients suffering from recurrent infections lack the specific granules in neutrophils (Sanchez et al., 1992).

5.2.4 Lactoferrin and growth factor activity

Transferrin is essential for the growth of many cell lines (Casey et al., 1989). This kind of effect has also been proposed for lactoferrin. The evidence supporting the presence of lactoferrin-specific receptors on both a myeloid leukaemic cell line and a human adenocarcinoma cell line (Hashizume et al., 1983; Roiron et al., 1989) was reported with a suggestion that lactoferrin had an ability to stimulate the growth of human lymphoid cell lines. Studies by Mazurier et al. (Mazurier et al., 1989) demonstrated that phytohaemagglutinin-stimulated human peripheral blood lymphocytes express surface receptors binding human lactoferrin. The lactoferrin binding was found to stimulate growth after the receptors had been synthesised. The

stimulatory effect was dependent on both lactoferrin concentration and levels of iron saturation. These results clearly indicated that lactoferrin possesses growth stimulatory activity. However, there were also some conflicting results reported on the role of lactoferrin in lymphocyte proliferation (Brock and Djeha, 1997). It was reported that neither Fe-lactoferrin nor apo-lactoferrin themselves could exert a proliferative effect on lymphocytes in the absence of transferrin, but that lactoferrin can protect lymphocytes against harmful "free" iron by the sequestration effect. Therefore, different mechanisms need to be sought to explain the ability of lactoferrin to stimulate cell proliferation.

5.2.5 Lactoferrin as oxidant or antioxidant

Iron may act as a catalyst for producing hydroxyl radicals by the Haber-Weiss reaction, and hydroxyl radicals can act to destroy most known biomolecules. Many studies have suggested conflicting roles for lactoferrin in this reaction. Lactoferrin-bound iron has been reported to enhance •OH production (Ambruso and Johnston, 1981; Bannister et al., 1982); consequently this effect may contribute to the bactericidal activity of neutrophils by promoting the formation of the toxic hydroxyl radical. Lactoferrin has also been reported (Zhao and Hutchens, 1994) to mediate the formation of oxygen radicals and may contribute to the hydrolysis of nucleic acids. Also, the copper-binding sites in Cu(II)-bound lactoferrin serve as centres for repeated production of •OH radicals via the Haber-Weiss reaction (Zhao and Hutchens, 1994). This activity is consistent with the hypothesis that lactoferrin's bactericidal effect is mediated by targeted radical generation at a susceptible bacterial site.

On the other hand, several studies have shown that lactoferrin is in fact a very poor catalyst of hydroxyl radical formation (Gutteridge et al., 1983; Winterbourn, 1983), and that it actually complexes free iron, preventing hydroxyl radical formation and lipid peroxidation. Transferrin was found to have a similar role, and lactoferrin may be a more potent antioxidation agent as it retains iron with higher affinity. Interestingly, neutrophil lactoferrin has also been suggested as having a role in trapping iron released from ingested micro-organisms and removing it from sites of inflammation (Molloy and Winterbourn, 1990). This would prevent the free iron from catalysing undesirable oxidation reactions and also prevent growth of any micro-organisms that survived the bactericidal effect.

5.3 Structure of lactoferrin

5.3.1 Primary structure

The first two complete amino acid sequences of the transferrin family, those of human transferrin and chicken ovotransferrin, were reported in 1982 (MacGillivray et al., 1982; Williams et al., 1982). These were then followed by that of human lactoferrin (Metz-Boutigue et al., 1984). The homologous relationship between these proteins established them as transferrins. Since then, many more sequences have become available through cDNA sequencing techniques. At the present time, more than 20 primary structures of the transferrin family have been determined. They are listed in Table 5.3.1-I.

Table 5.3.1-I Primary sequences of the transferrin family known to date

Protein	Ref.	Protein	Ref.
Human serum transferrin	(a, b)	Bovine lactoferrin	(n, o)
Horse serum transferrin	(c)	Goat lactoferrin	(p)
Pig serum transferrin	(d)	Mouse lactoferrin	(r)
Rabbit serum transferrin	(e)	Porcine lactoferrin	(s, t)
Rat serum transferrin	(f, g)	Human neutrophil lactoferrin	(l)
Xenopus serum transferrin	(h)	Human melanotransferrin	(u)
Bovine transferrin	(i)	Chicken ovotransferrin	(v, w)
Rainbow trout transferrin	(j)	Atlantic salmon transferrin	(x)
Atlantic cod transferrin	(k)	Hornworm transferrin	(y)
Human lactoferrin	(l, m)	Cockroach transferrin	(z)

References cited: (a) Park et al. (1985); (b) MacGillivray et al. (1982); (c) Carpenter and Broad (1993); (d) Baldwin and Weinstock (1988); (e) Banfield et al. (1991); (f) Schreiber et al. (1979); (g) Hoshino et al. (1996); (h) Moskaitis et al. (1990); (i) Retzer et al. (1996); (j) Tange et al. (1997); (k) Denovan-Wright et al. (1996); (l) Rado et al. (1987); (m) Metz-Boutigue et al. (1984); (n) Mead and Tweedie (1990); (o) Goodman and Schanbacher (1991); (p) Le Provost et al. (1994); (r) Pentecost and Teng (1987); (s) Alexander et al. (1992); (t) Lydon et al. (1992); (u) Rose et al. (1986); (v) Jeltsch and Chambon (1982); (w) Williams et al. (1982); (x) Kvingedal et al. (1993); (y) Bartfeld and Law (1990); (z) Jamroz et al. (1993).

All members of the transferrin family demonstrate a high degree of similarity which is reflected in 60-70% sequence identity between the lactoferrins and 55-60% identity between the transferrins of high animals and lactoferrin (Baker, 1994). Melanotransferrin and insect transferrins are slightly less similar with ~40% and 30%

identity with lactoferrins and transferrins. Another striking feature of the similarity is that the N-terminal and C-terminal halves of each transferrin molecule are very similar as well, with the level of identity between the two halves ranging from ~26% in the insect proteins to ~40% in higher transferrins. This similarity of the two lobes is also taken as evidence that modern transferrin evolved from a one-site iron-binding ancestral protein of 40 kDa by a gene duplication and fusion. The sequence homology indicates that the transferrin family shares basically the same bilobal three-dimensional structure with a similar iron-binding site in each of the two lobes.

5.3.2 Three-dimensional structure

Since the first crystallographic studies on transferrins more than 20 years ago, many crystals of various transferrins have been reported. A 6 Å low-resolution analysis of rabbit serum transferrin (Gorinsky et al., 1979) clearly showed the bilobal nature of the molecule. However, full details of transferrin structure were not available until the three-dimensional structure of human diferric lactoferrin at 3.2 Å resolution was solved (Anderson et al., 1987). Soon after, the same structure at 2.8 Å resolution and further refined at 2.2 Å resolution were published (Anderson et al., 1989; Haridas et al., 1995). So far, these provide the most detailed description of human diferric lactoferrin. Thereafter, more crystal structures of transferrins were reported. They included rabbit serum transferrin (Sarra et al., 1990), chicken ovotransferrin (Dewan et al., 1993), the recombinant N-terminal half molecule of human lactoferrin (Day et al., 1993), duck ovotransferrin quarter-molecule (Lindley et al., 1993), human copper-lactoferrin (Smith et al., 1992), human oxalate-lactoferrin (Shongwe et al., 1992), bovine lactoferrin (Moore et al., 1997) and many mutant crystal structures of various transferrins. The structural description of diferric human lactoferrin below is based on the model refined at 2.2 Å resolution (Haridas et al., 1995).

5.3.2.1 Structure of diferric human lactoferrin

The single polypeptide chain of 691 amino acid residues is folded into two globular lobes, representing its N-terminal and C-terminal halves. The N-lobe consists of residues 1-333 and the C-lobe, residues 345-691. These two lobes are joined by a short

3-turn α helix corresponding to residues 334-344. The main interactions between the two lobes, apart from the covalent connecting helix, are a number of hydrophobic interactions between N1 and C1 domains as well as some water-mediated contacts through hydrogen bonding. This could allow for movement of one lobe relative to the other, and supports the observation that the relative orientations of the N and C-lobes are different between human lactoferrin and rabbit transferrin (Baker and Lindley, 1992) and between human and bovine lactoferrin (Baker et al., 1991).

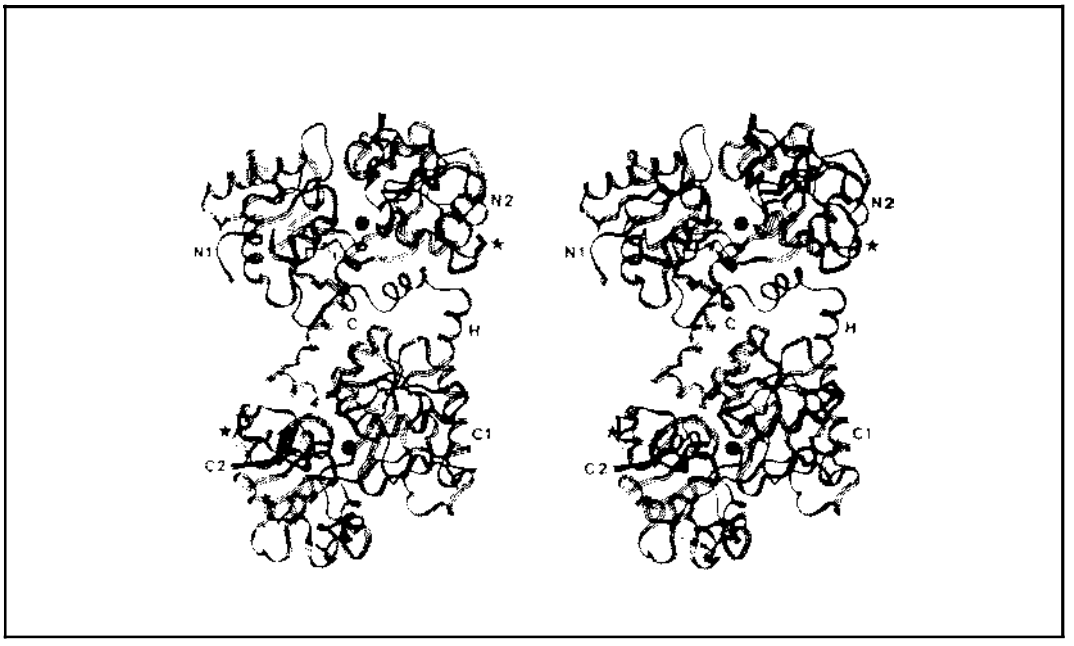


Fig. 5.3.2-I Stereo diagram for diferric-lactoferrin, showing the organization of the molecule (N-lobe above and C-lobe below). The four domains (N1, N2, C1, C2), the connecting helix (H) and the region of hydrophobic contacts between the lobes (C) are also indicated. The iron atoms are shown by filled circles and the glycosylation sites by stars. Taken from Baker et al. (1991) with permission.

Each lobe is further subdivided into two separately folded domains: domains N1 (residues 1-90 and 252-333) and N2 (residues 91-251) in the N-lobe, domains C1 (residues 345-433 and 596-691) and C2 (residues 434-595) in the C-lobe. The iron and anion binding sites, one in each lobe, are located in the cleft formed between the two domains of each lobe. The overall polypeptide folding of the structure is shown in Fig. 5.3.2-I.

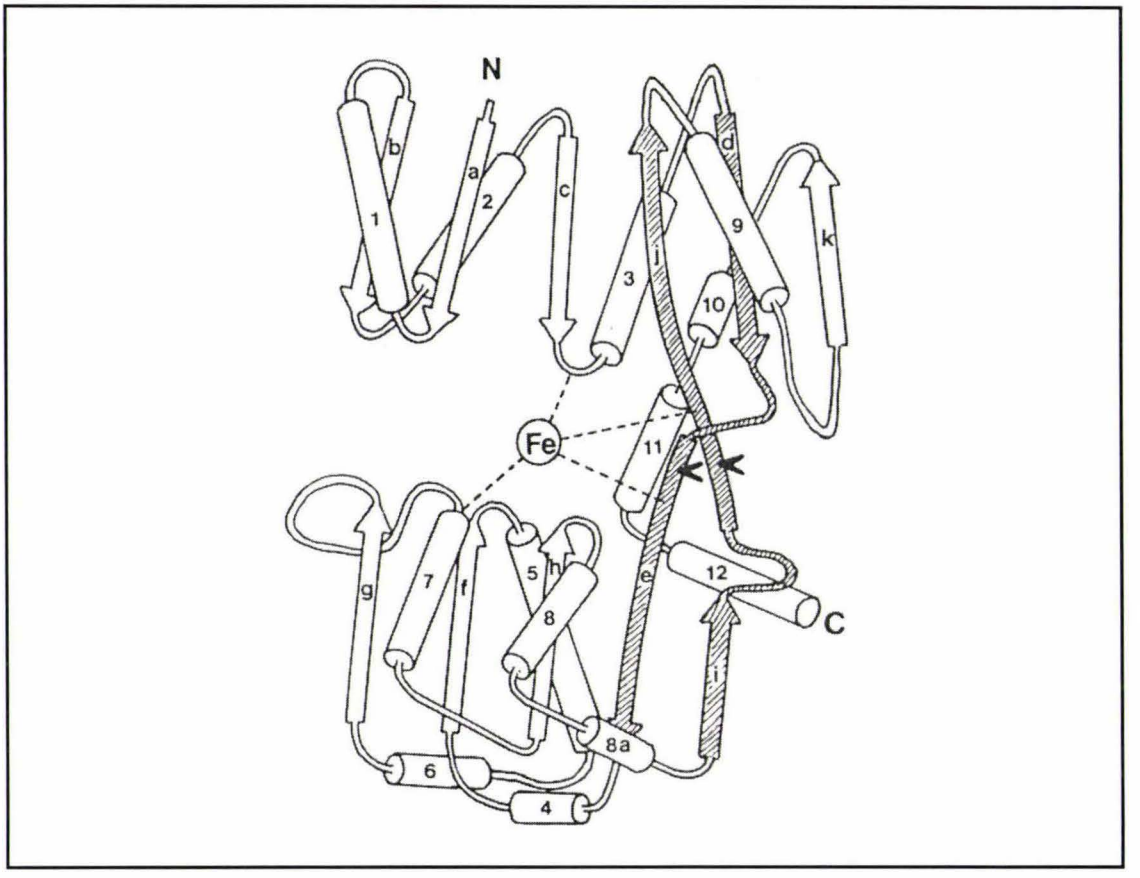


Fig. 5.3.2-II Polypeptide folding pattern for the N-lobe of lactoferrin. Helices (cylinders) are numbered 1 to 12 and β-strands (arrows) are labelled "a" to "k" as in Anderson et al. (1989). The interdomain backbone strands are shaded and the position of the hinge is indicated by arrow. Taken from Baker (1994) with permission.

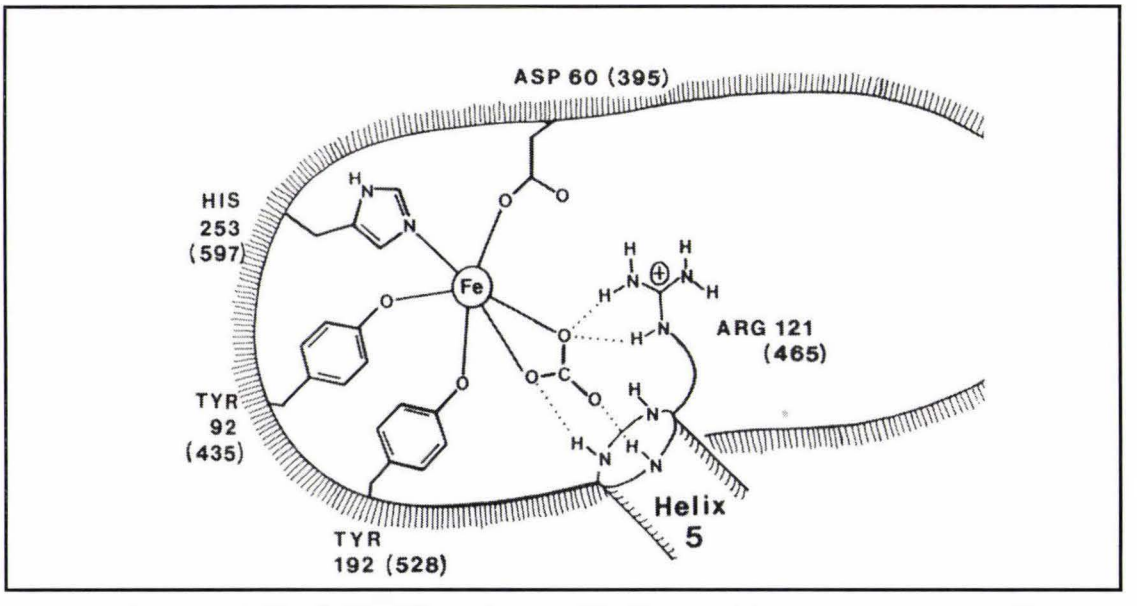


Fig. 5.3.2-III Schematic diagram of the iron and carbonate binding site in lactoferrin. Residue numbers are for the N-lobe and corresponding residues in the C-lobe are given in parentheses. Taken from Baker et al. (1991) with permission.

The two lobes have very similar folding as expected with their high (42%) sequence identity. They are related by an approximate twofold screw axis, the C-lobe being superimposable on the N-lobe by a rotation of 179.5° and a translation of 24.5 Å. When the two lobes are superimposed, the rms difference is only 1.0 Å for 264 Cα positions. The regions not included are primarily the insertions and deletions in a sequence alignment between two lobes, together with the final helix of each lobe (helix 12). Differences also arise from the N-lobe being slightly more open than the C-lobe, by a 6.3° rotation. When the individual domains are superimposed, the rms differences are 0.86 Å for N1 and C1, and 0.75 Å for N2 and C2 domains. The folding pattern for a lobe is shown in Fig.5.3.2-II. Each domain is built around an α/β folding pattern which has a central mixed β-sheet with helices packed against it. After forming about half of the first domain (N1 or C1) with ~70 residues, the polypeptide chain (β-strand d and e) runs behind the iron binding site to fold the whole of the second domain (N2 or C2), then the polypeptide chain (another long extended strand i and j) crosses back behind the iron site to complete the folding of the first domain. These two extended β-strands linking the two domains are known as "backbone" strands in which a hinge (corresponding to Thr90 and Pro251) is very important for the conformational change of the molecule during iron binding and release (Anderson et al., 1990).

The iron and anion binding site is shown schematically in Fig.5.3.2-III. The iron coordination is distorted octahedral, with ligands provided by four amino acid side chains from the protein, together with a bidentate bound carbonate anion. The protein ligands are a carboxylate oxygen of an aspartate (Asp60, or Asp395 in C-lobe), two phenolate oxygens from tyrosine (Tyr92 and Tyr192, or Tyr435 and Tyr528 in C-lobe), and an imidazole nitrogen of histidine (His 253, or His 597 in C-lobe). The binding sites in the two lobes are remarkably similar. The metal-ligand bond lengths are all close to 2.0 Å as expected for Fe^{3+} coordination and consistent with EXAFS studies (Garratt et al., 1986). The bond angles are distorted from true octahedral symmetry, largely because of the small chelate bite angle (~ 63°) of the carbonate anion.

The anion binds to a positively charged pocket on domain 2 (N2 or C2) as shown in Fig.5.3.2-III. The pocket is formed by the N-terminus of helix 5, residues 121-124 (465-468 in C2) and the side chain of Arg121 (Arg465 in C2), as well as side chain of Thr117 (Thr461 in C2) from the loop joining β-strand f to helix 5. The positive charge of the environment is attributed to both the side chain of Arg121 (Arg465) and the N-

terminus of the helix, which has a partial positive charge between +0.5 and +0.75 because of its dipolar nature (Hol et al., 1978). The anion fits between metal ion and protein so perfectly that the full hydrogen-bonding potential of the carbonate ion is utilised. The hydrogen bonds of the anion all have highly favourable geometry with lengths ~2.8 Å, angles at the hydrogens close to 180°, and angles at the oxygens close to ~ 120°.

The anion binding is considered important for the iron binding of the protein. There is a synergistic relationship between iron and carbonate, in that each is required for the tight binding of the other. Both spectroscopic (Harris et al., 1990; Zweier et al., 1981) and crystallographic studies (Anderson et al., 1989) suggested that the anion binds to the apo-protein first so as to prepare the site for iron binding by neutralising the positive charge at the binding site. Other evidence for the crucial role of the anion binding is that all residues involved in anion binding (i.e., 117, 121-125 or 461, 465-469 in C2) are conserved in all transferrins except for melanotransferrin and *Manduca sexta* transferrin which have a reduced iron-binding ability.

Except for certain species of fish, all known vertebrate transferrins are glycoproteins, the absence of glycosylation in these fish species suggesting that the carbohydrate chains of transferrins are not important in the physiological functions of the proteins. This is supported by the finding that the differences in specific carbohydrate residues between the native and recombinant human lactoferrin do not affect the functional activity of the protein (Ward et al., 1997c). The carbohydrate chains are heterogeneous from different species and from different tissues within a species (Spik et al., 1988). Glycosylation sites vary in number, from one in rabbit serum transferrin to two in human lactoferrin, and four in bovine lactoferrin. The sites are located over the surface of the transferrin molecules, indicating no direct functional role in iron-binding and release. However, it is likely that differential glycosylation may contribute to at least the stability and half-life of the protein in certain biological fluids (Prieels et al., 1978). The two carbohydrate chains of human lactoferrin are both N-linked, to Asn137 in the N-lobe and Asn478 in the C-lobe, and the terminal N-acetylglucosamine and its attached fucose in the N-lobe and 6 sugar residues in the C-lobe have been modelled (Haridas et al., 1995). There are very few direct interactions between the carbohydrate chains and the protein, which is consistent with the fact that the carbohydrate clearly

suffers from disorder. Thus, there is no well defined conformation available, as reflected in the poor density.

5.3.2.2 Structure of apo human lactoferrin

It has long been known from biophysical studies (Kilar and Simon, 1985) that transferrins could undergo a substantial conformational change during iron binding and release, and the protein probably becomes more compact after the metal is bound. The nature of this conformational change has now been well defined by crystallographic studies of apo-lactoferrin in two different crystal forms (Anderson et al., 1990; Baker et al., 1997), complemented by solution X-ray scattering measurements (Grossmann et al., 1993; Grossmann et al., 1992).

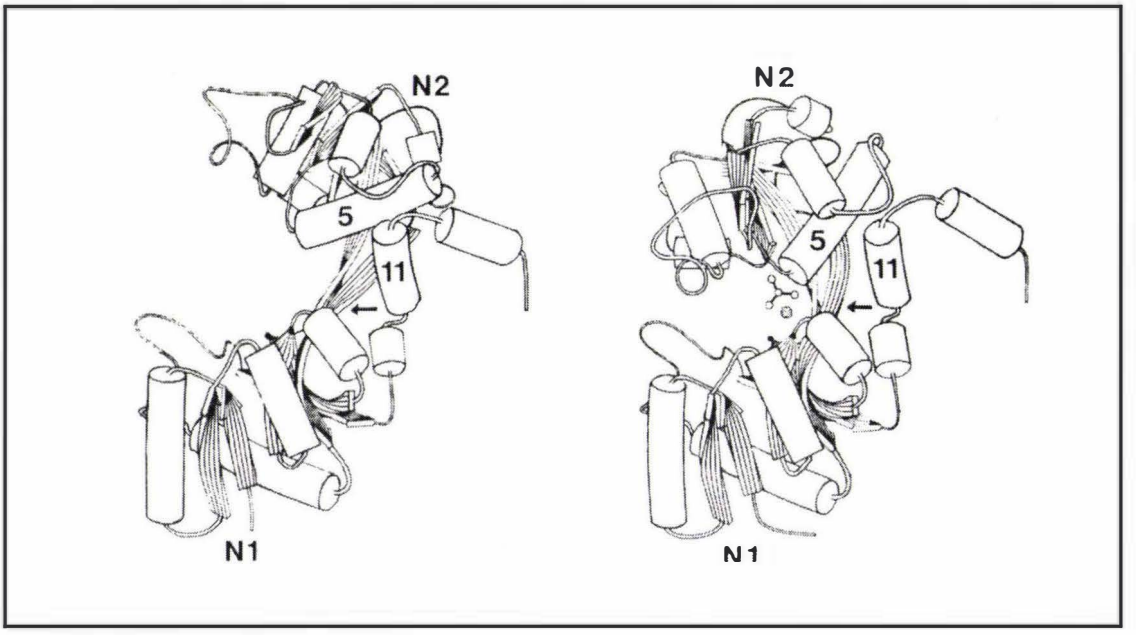


Fig. 5.3.2-IV Schematic diagram of the N-lobe of lactoferrin showing the change from the "open" form of apo-lactoferrin (left) to the "closed" form of iron-lactoferrin (right). Helices are shown as cylinders, β-strands as arrows. The hinge point in the "backbone" β-strands is shown with an arrow. Taken from Baker et al. (1991) with permission.

The first crystal form of apo-lactoferrin, which was from deglycosylated protein, was crystallised in the orthorhombic space group $P2_12_12_1$, with one molecule in the asymmetric unit. The second crystal form of apo-lactoferrin in its native glycosylated

form was grown in the same space group but with two molecules in the asymmetric unit. Two striking results came from the crystal structure of the first crystal form of apo-lactoferrin. The N-lobe was shown to have undergone a large-scale conformational change, whereas the C-lobe remained essentially unchanged even though no metal was bound.

As shown in Fig. 5.3.2-IV, the conformational change in the N-lobe involves an opening of the binding cleft through relative movement of the two domains; the N2 domain rotates 54° relative to N1, about an axis passing through the two backbone strands that run behind the iron site, connecting the two domains. The conformational change is a rigid body movement in which the hinge is located to only a few residues, 90-91 and 250-251. This is shown by superposition of the individual domains of diferric lactoferrin on those of apo-lactoferrin with rms differences: N1, 0.52 Å; N2, 0.65 Å; C1, 0.43 Å; C2, 0.44 Å (Anderson et al., 1990). The rigid body movement causes a helix in the N2 domain (helix 5) to pivot on another one (helix 11) in the N1 domain and has the effect of opening the binding cleft wide.

In the "open" form, two tyrosine ligands (Tyr92 and Tyr192) and the anion ligands remain with the N2 domain, whilst the other two ligands (His253 and Asp60) remain with N1 domain. This has important implications for mechanisms of binding and release. In contrast, the C-lobe in this crystal structure, is closed, as in diferric lactoferrin, even though there is no metal bound. This was unexpected. It seemed certain that the C-lobe must be able to open in a similar way to the N-lobe, although a disulfide bridge, number 7, which links the C1 and C2 domains and which has no counterpart in the N-lobe, might exert extra rigidity to some extent. It was suggested that an equilibrium between "open" and "closed" forms may exist in solution due to the flexibility of the molecule, and that the requirements of crystal packing had preferentially selected out one of several conformations existing in solution (Anderson et al., 1990). If this is so, there must be very little energy difference between the "open" and "closed" forms, given the weakness and small number of intermolecular contacts in the crystals. Solution X-ray scattering measurements are consistent with the opening of the N-lobe of apo-lactoferrin seen in the crystal structure, but these experiments further showed that both lobes are open in solution (Grossmann et al., 1992).

A full understanding of the conformational change is now available from the analysis of a second crystal form of apo-lactoferrin in which both lobes were found to be open,

just as in the solution studies. The opening of the N-lobe is again ~ 50°, whereas the opening of the C-lobe is less, at ~ 18° (Fig. 5.3.2-V). In the molecule as a whole, the N1 and C1 domains form a relatively rigid core, with the N2 and C2 domains rotating away from N1 and C1, respectively, by a rigid-body motion, which opens both of the binding clefts. This study confirmed the conformational flexibility of the protein molecule in solution suggested before.

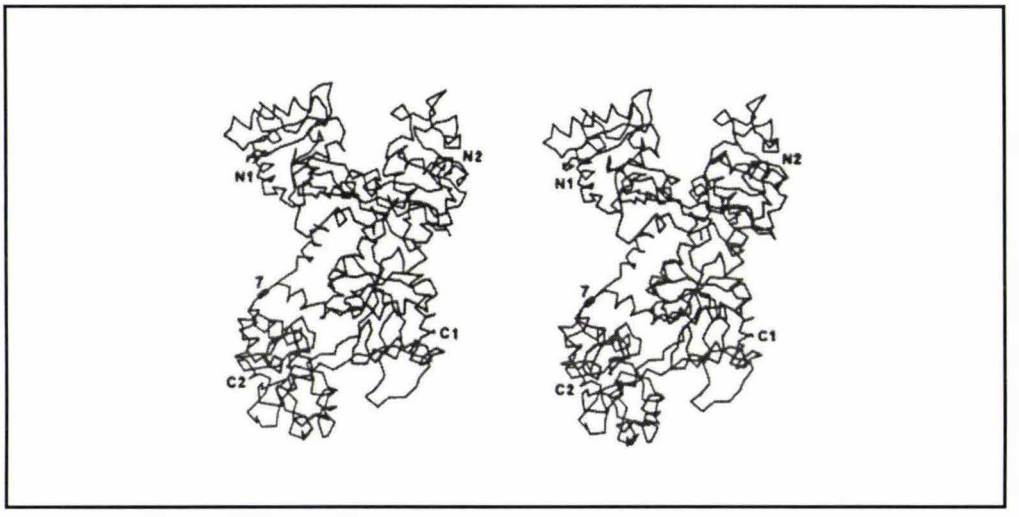


Fig. 5.3.2-V Stereo view of the polypeptide chain conformation in the fully "open" form of apo-lactoferrin. The four domains (N1, N2, C1, C2) are labelled, and the disulfide bond (number 7) that restrains the mobility of the C-lobe is also shown. Taken from Baker et al. (1997) with permission.

From all of the structural information obtained so far, some important conclusions regarding conformational change and iron binding by lactoferrin can now be made (Baker et al., 1997). Both lobes in lactoferrin do open, the importance of this all-open form being that it is almost certainly to this form that iron binding occurs. At the same time, the "open" form is also necessary for iron release, since such a change could pull the iron site apart. The C-lobe opens less than the N-lobe, which is consistent with several observations on other transferrins, including stronger iron binding by the C-lobe, greater acid stability, slower iron release and greater thermodynamic stability (Baker, 1994; Harris and Aisen, 1989). The reason for the lesser flexibility in the C-lobe may be related to a disulfide bridge (number 7, residues 483-677) which has no counterpart in the N-lobe and may exert an extra constraint to movement of the C-lobe domains.

The fact that both "open" and "closed" C-lobe structures have been found in crystals of apo-lactoferrin shows that the energy difference between them must be very little, because the crystal packing forces are weak (Anderson et al., 1990). Both "open" and "closed" forms are accessible and they exist in a dynamic equilibrium.

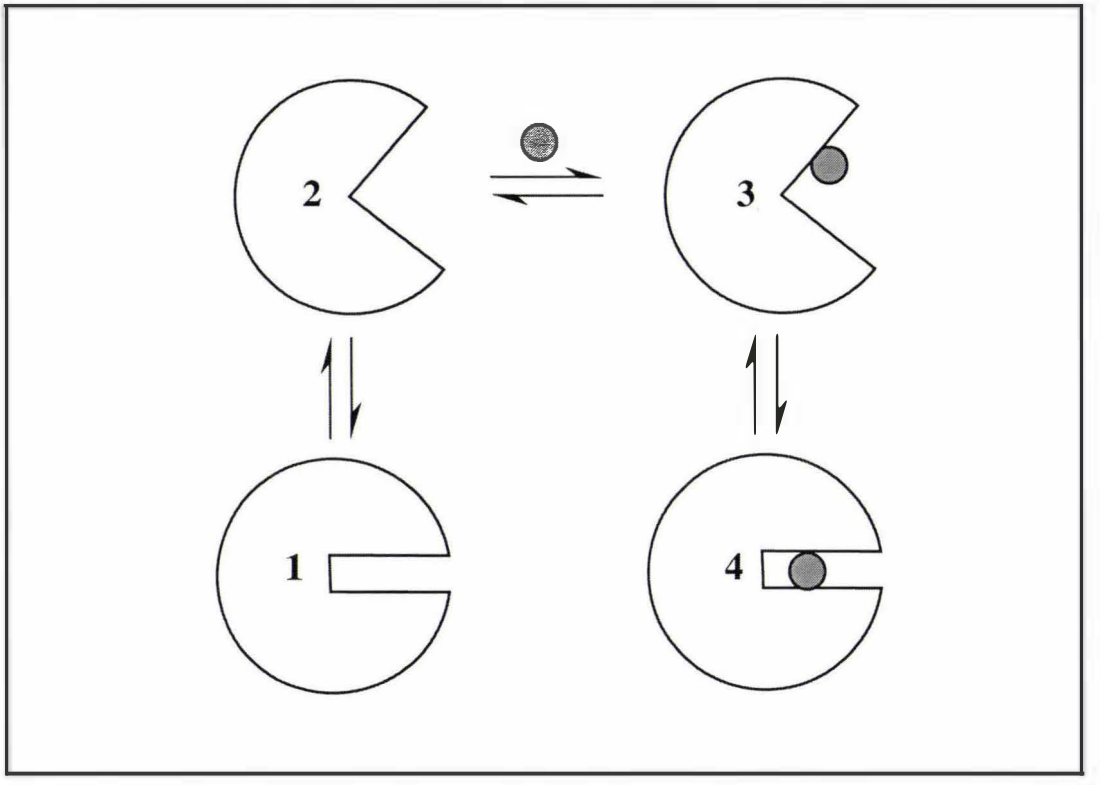


Fig. 5.3.2-VI Representation of dynamics of the lactoferrin molecule (shown for a single lobe only). 1 and 2 represent the "closed" and "open" forms of the apo-protein, in equilibrium. Iron binding to form 2 gives an intermediate 3 in which iron is bound to one domain only. This form is also in equilibrium with a "closed" form, but the latter is locked closed if iron is present. Taken from Baker et al. (1997) with permission.

Therefore, a mechanism can be suggested for iron binding and release by lactoferrin as shown in Fig. 5.3.2-VI. It was proposed that iron binding occurs first on domain 2 (N2 or C2) because the carbonate anion and two tyrosine ligands are associated with this domain. The domains can access both "open" and "closed" states because of the dynamics of the molecule. When the domains close, there are two possibilities. If no iron is present, they simply return again to the "open" form. If iron is bound weakly to the domain 2 (N2 or C2), then it locks on to the other two ligands on domain 1 (N1 or

C1) to complete binding. So it is clear that dynamics are essential for the function, and iron binding can not be completed unless both "open" and "closed" configurations can be accessed. On the other hand, after iron is bound, the molecule remains in its "closed" state until some external factors, such as pH or receptor binding, destabilise it.

5.4 Recombinant DNA studies

5.4.1 Recombinant transferrin for structural and functional studies

While X-ray crystallographic studies enable us to have a full understanding of three-dimensional structures of transferrins, recombinant DNA studies have opened the way to the production of mutant transferrins, in which single or several selected amino acids are changed. Mutagenesis can be carried out either on the whole transferrin molecules or on half-molecules. To date, a variety of transferrin mutants have been designed and constructed to examine the specific properties of each individual site, such as amino acid ligands and other residues involved in forming the iron or anion-binding pocket.

The first successful expression of any transferrin was reported in 1990 (Funk et al., 1990), in which the recombinant human transferrin N-terminal half-molecule was expressed in a cell culture system using baby hamster kidney cells. Similar methods have been used to express both the whole molecule (Stowell et al., 1991) and the N-terminal half-molecule (Day et al., 1992) of human lactoferrin, and intact human serum transferrin (Mason et al., 1993). Since then a lot of attempts have been made to develop appropriate expression systems for the rapid and efficient production of large quantities of transferrins. These include expression of intact and half-molecules (N- and C-terminal) of human serum transferrin in *E. coli* (Ikeda et al., 1992; Steinlein and Ikeda, 1993); use of the baculovirus expression system for the production of human serum transferrin and lactoferrin (Ali et al., 1996; Salmon et al., 1997); and of the Aspergillus expression system for the production of both human and murine lactoferrin (Ward et al., 1997a; Ward et al., 1992). The recombinant intact transferrins are indistinguishable from native transferrins with respect to size, immunoreactivity, and iron-binding capacity. They all possess the same biological functions such as receptor binding and antimicrobial activities. However, different glycosylation patterns between the recombinant and native proteins were commonly observed. The recombinant proteins

generally have different carbohydrate content from their native counterpart, and some recombinant transferrins were even expressed in nonglycosylated form (Ikeda et al., 1992). Many studies demonstrated that the differences in glycosylation do not affect the functional activity of the protein (Anderson et al., 1990; Kawakami and Lonnerdal, 1991; Mason et al., 1993). In addition, heterogeneity exists in the carbohydrate attached to transferrins from different species and from different tissues within a species (Spik et al., 1988), thus also supporting the view that carbohydrate does not appear to affect the biological activity of transferrins.

A number of site-specific mutants have been prepared, both of transferrin and of lactoferrin, in order to investigate the effects of the mutation on structure and function. The targets of mutation have been amino acids in and around the metal binding sites. The three-dimensional structures of some mutants have been determined by X-ray crystallography. For the mutant Asp60Ser of the N-terminal half-molecule of human lactoferrin, two major differences can be seen in the structure, compared with its parent N-terminal half-molecule (Faber et al., 1996b). The side chain of Ser60 neither binds to the iron atom nor makes any interdomain contact as Asp60 does; instead a water molecule fills the iron coordination site and participates in interdomain hydrogen bonding. The loss of the Asp side chain also causes a significant change in domain closure, the domains moving closer together by 7° in the mutant. As a result of the above mutation effects, the iron binding is weakened considerably. The crystal structural analysis of the mutants Arg121Ser and Arg121Glu showed that destabilisation of iron binding of these mutants probably results from disruption of the anion binding site, in which the carbonate ion coordination to the iron atom becomes obviously asymmetric and even monodentate in Arg121Glu mutant, in contrast to the parent N-terminal half-molecule where its coordination is symmetric bidentate (Faber et al., 1996a). Similarly, other crystallographic studies of site-specific mutants of both lactoferrin and transferrin also give new insights into the ways in which particular amino acids influence iron binding and release.

5.4.2 Recombinant lactoferrin for large-scale production

As described in previous sections, the functions proposed for lactoferrin are diverse, including antimicrobial activity against a broad spectrum of bacteria, immuno-

modulatory activity, cellular growth promotion, antioxidant effects and regulation of iron homeostasis in the gastrointestinal tract. All of these functions for lactoferrin highlight the potential of this protein as an important antimicrobial, anti-endotoxin and immune regulatory agent which may have some nutritional and pharmaceutical applications. However, to examine the clinical and nutritional uses of lactoferrin, it is necessary to develop a suitable expression system which would produce large quantities of biologically active protein. So far, great efforts have been made for this purpose. Biologically active recombinant lactoferrin has been expressed in eukaryotic systems, including baby hamster cells (Stowell et al., 1991) and human 293 cells (van Berkel et al., 1995), but these systems can not readily be scaled up to produce large quantities of lactoferrin. Recombinant lactoferrin has also been produced in yeast (Liang and Richardson, 1993); however the secretion of the recombinant lactoferrin in this system was inefficient. High levels of recombinant lactoferrin are being achieved using transgenic cows, where lactoferrin is expressed in the bovine mammary gland (Nuijens et al., 1996). Although high levels of lactoferrin may be attained in transgenic cows, the success of this approach, including the ability to purify the recombinant protein in a viral- and toxin-free form and the cost-effectiveness of a large transgenic animal program, remains to be established. The successful large scale production of biologically active recombinant human lactoferrin has therefore been realised by using an Aspergillus expression system (Ward et al., 1995).

The optimum expression system for human lactoferrin has been designed with three features. The high secretory Aspergillus strain *A. awamori* has been chosen and the cDNA encoding mature human lactoferrin is expressed as a chimera fused to a highly expressed endogenous *Aspergillus* gene. Further increase in the yield of lactoferrin was achieved using a U.V. and chemically mutagenized strain improvement approach. To date, recombinant human lactoferrin has been produced at levels in excess of 2.0 g/l by using this *A. awamori* expression system (Ward et al., 1995). Functional analysis demonstrated that the recombinant protein is indistinguishable from native human lactoferrin, in iron-binding, antimicrobial and human enterocyte receptor binding activity. The attached carbohydrate is similarly N-linked to both the recombinant and native protein and the extent of glycosylation is similar for both proteins too.

However, analysis of the carbohydrate composition showed that the recombinant protein contains a high proportion of mannose-type oligosaccharides which are similar

to the type observed on bovine lactoferrin and may display a simpler carbohydrate structure in contrast to the complex carbohydrate structure of native human lactoferrin. This heterogeneity apparently does not affect the functional activity of the recombinant protein, although its potential influence on specific properties such as protein solubility and stability remains to be further clarified.

In addition to differences in carbohydrate residues, mutagenesis has inadvertently introduced an amino acid mutation (Ala10Thr) in the recombinant lactoferrin, during the strain improvement approach.

5.5 Aims of this project

The purpose of this project was to determine the three-dimensional structure of the recombinant human lactoferrin produced in the *Aspergillus awamori* expression system, by X-ray diffraction methods, and to compare it with that of the native human milk protein.

As described above, the *Aspergillus awamori* expression system has enabled the commercial production of recombinant lactoferrin for therapeutic as well as nutritional applications. Functional identity between the recombinant and native proteins has also been demonstrated by functional analysis experiments (Ward et al., 1997b). Crystallographic studies give detailed structural characterisation at the molecular level for the recombinant protein, and so give an alternative way to determine whether the structure of the recombinant protein is in any way affected by the method of production.

EXPERIMENTAL APPROACH AND METHODS

6.1 Protein purification and crystallization

6.1.1 Iron-saturation for the protein samples

The protein used in this study was provided by Agennix Inc. (Houston, Texas) from large-scale production of recombinant lactoferrin in *Aspergillus awamori* as described before (Ward et al., 1995). The protein samples are freeze-dried, with light brown colour. Iron saturation was carried out to make sure that the protein was fully loaded with iron.

The protein was dissolved in 20 mM Tris/HCl pH 7.0, 0.2 M NaCl, to give 2.5 ml protein solution with concentration ~65 mg/ml. Then 0.6 ml of 10 mM ferric nitrilotriacetate (FeNTA) was added, and the colour of the protein solution changed to brown instantly. The absorbance ratio of A_{280} /A_{466} was 20.2 (A_{280}=0.726, A_{466}=0.036), indicating that the protein had become fully iron-saturated (Aisen and Leibman, 1972).

The iron-saturated protein solution was then dialysed thoroughly against 20 mM Tris/HCl pH 7.0, 0.2 M NaCl, to remove excess FeNTA before further gel filtration.

6.1.2 Protein purification by gel filtration

The protein was purified on a gel filtration column (Pharmacia Superdex 75 HR 10/30) which was mounted on Waters 650E FPLC system. The column was equilibrated with 50 mM Tris/HCl pH 8.0, 0.2 M NaCl. The protein was eluted as a big peak followed by several smaller peaks. The purity of each fraction across the big peak was checked using SDS polyacrylamide gel electrophoresis, and pure fractions combined. The purified protein was dialysed into the 10 mM Hepes buffer, pH 8.0, with 0.5 M NaCl, and concentrated through a Centricon-30 concentrator (Amicon) to 80 mg/ml for crystallization.

6.1.3 Crystallization

Successful crystallizations of various lactoferrin species have involved very similar conditions, i.e., very pure protein, high protein concentration (typically 50-80 mg/ml) and dialysis from low ionic strength buffers, with a pH between 7.6-8.4. A similar crystallization strategy was therefore followed in the present case. The single crystals were grown by microdialysis of the above concentrated protein solution against 10 mM sodium phosphate, pH 8.0, containing 10% (v/v) ethanol. A typical crystal with dimensions 1.5x1.0x0.5 mm was obtained in 3 weeks at 4°C.

6.2 Data Collection and Processing

6.2.1 Data Collection

X-ray diffraction data were collected at room temperature on a Rigaku R-Axis IIC image plate detector with Cu-Kα radiation (λ= 1.5418 Å) generated by an RU200 rotating anode generator. One crystal with approximate dimensions 0.5x0.5x0.4 mm was used to collect 60 oscillations of 1.5° each at a crystal-to-plate distance of 110.0 mm. The crystal diffracted to 2.2 Å and proved to be isomorphous with the native human lactoferrin crystals. The unit cell was orthorhombic, space group $P2_12_12_1$, with dimensions: a=156.25 (156.26), b=97.53 (97.40), c=55.87 Å (55.85 Å) (the numbers in parentheses correspond to native protein crystal). There is one molecule in the asymmetric unit, and the volume per unit mass, V_m, for the crystal is 2.7 $Å^3$/Dalton, indicating a solvent content of ~55% (Matthews, 1968).

6.2.2 Data processing

The diffraction images were indexed and integrated with the program DENZO (Otwinowski, 1993), and the programs ROTAVATA and AGROVATA from the CCP4 program suite (Collaborative Computational Project No.4, 1994) were used to scale and merge all data. The resultant data set has a total of 36329 independent reflections in the resolution range of 36.74 to 2.20 Å. Some details of the data set are given in Table 6.2.2-I. Data analysis as a function of resolution is shown in Fig. 6.2.2-I.

Table 6.2.2-I. Statistics for reflection data collected

Dmin(Å)	Nref	Nmeas	Compl(%)	Redu.	Av-I	I/σ(I)	R_{merg}
6.86	1513	3596	95.7	2.4	10540	20.5	0.019
4.89	2589	6185	97.3	2.4	5261	20.1	0.023
4.00	3209	7429	95.7	2.3	6353	17.1	0.026
3.47	3751	9096	95.5	2.4	3533	18.0	0.031
3.11	3998	10260	90.4	2.6	1817	14.1	0.049
2.84	4156	11386	85.5	2.7	884	9.0	0.084
2.63	4236	12306	80.3	2.9	482	5.4	0.143
2.46	4281	13133	75.9	3.1	269	3.1	0.242
2.32	4302	13684	71.7	3.2	183	2.1	0.357
2.20	4294	13456	67.8	3.1	119	1.4	0.542
Overall	36329	100525	82.3	2.8	1908	11.1	0.046

Nref (Number of reflections); Compl (Completeness); Redu. (Redundancy); Av-I (Average intensity); $R_{merg} = \Sigma|I-\langle I\rangle|/\Sigma I$ (Merging R-factor).

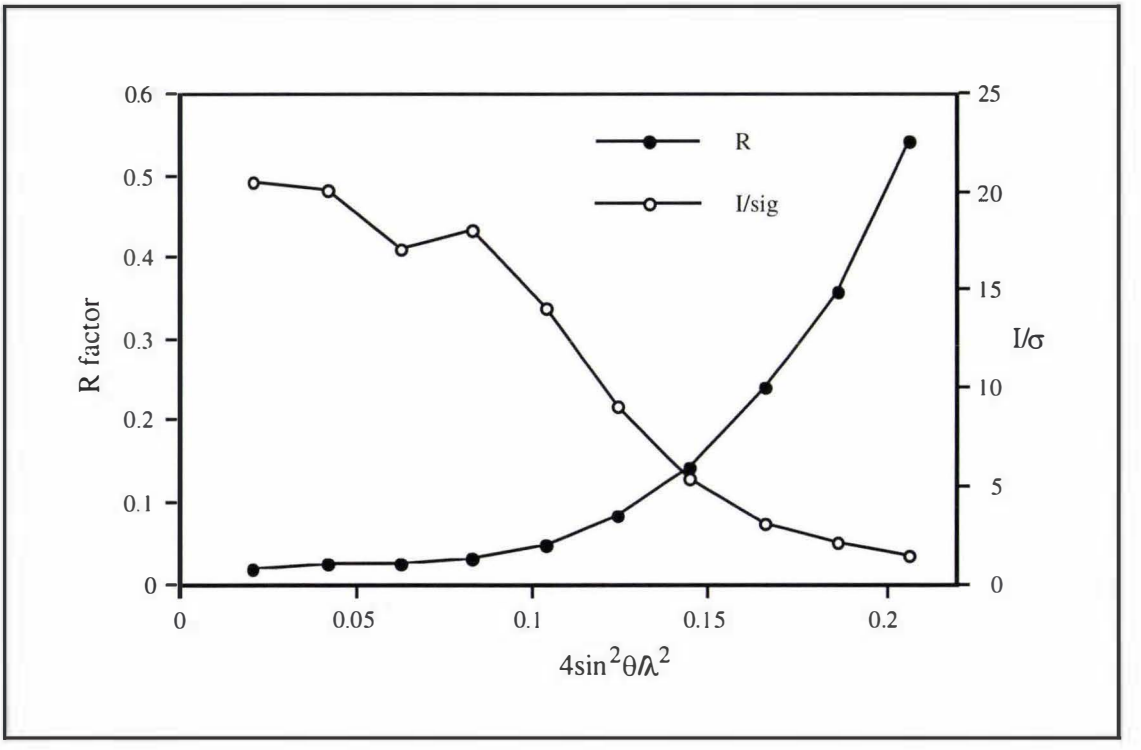


Fig. 6.2.2-I Data analysis against resolution.

The data set was 82.3% complete with a redundancy of 2.8 and overall R_{merge} of 4.6%. It is observed from Fig. 6.2.2-I that the R_{merge} increases rapidly as the intensity of the

diffraction falls off with resolution, a phenomenon which is quite normal for most protein crystals. The proportion of observed data is still reasonable at the cut-off of 2.2 Å resolution (the outermost shell, 2.3-2.2 Å, is 67.8% complete with redundancy of 3.1, average $I/\sigma(I)$ of 1.4). A Wilson plot (Wilson, 1949) for the final 2.2 Å data set gave an estimate of 43 $Å^2$ for the overall temperature factor.

6.3 Structure determination and refinement

The starting model was the native human diferric lactoferrin structure (Haridas et al., 1995) refined at 2.2 Å but with all the solvent molecules omitted. The restrained least-squares refinement program TNT (Tronrud et al., 1987) was used in all refinements. A total of 32667 reflections in the resolution range of 40.0 to 2.2 Å were used for refinement and Fourier syntheses, and another 3616 reflections, ~8% of all data, were used for free R factor calculations in the same resolution range. The main progress of the refinement is summarised in Table 6.3-I.

Table 6.3-I Outline of refinement

Main procedures	B range	Resolution range
1. Rigid body refinement. The molecule was divided into four rigid bodies: N1 (1-90, 252-333), N2 (91-251), C1 (334-433, 596-691), C2 (434-595).	fixed at 30 $Å^2$	10-2.8 Å
2. Repeated model rebuilding mainly of loops, side chains and iron-binding sites. Some water molecules were added.	15-60 $Å^2$	10-2.5 Å ↓ 40-2.2 Å
3. Omit map calculated and model rebuilding: residues 1-4, 220-221, 283-284, 293-294, and 418-424 were omitted.	15-80 $Å^2$	40-2.2 Å
4. Restraints on geometry and B correlation were applied on the carbonate anions and Fe-ligand bonds, but removed in later stages.	10-100 $Å^2$	40-2.2 Å

Rigid body refinement was employed in the first rounds of refinement. The model comprised four rigid bodies, the N1 and N2 domains (residues 1-90, 252-333 and 91-251, respectively) of the N-lobe and the C1 and C2 domains (residues 334-433, 596-691 and 434-595, respectively) of the C-lobe. The overall R factor was reduced from 0.330 to 0.309 after 6 cycles of refinement in which all B values were fixed at 30 $Å^2$ and the resolution was set in the range of 10.0 to 2.8 Å. After a further 6 cycles of refinement, removing the rigid body restrictions and constraining individual B values to a range of 15.0 to 60.0 $Å^2$, the overall R factor had decreased to 0.199. $2F_o$-F_c and F_o-F_c electron-density maps were calculated after the resolution range had been reset to 10.0 to 2.5 Å. Some residues were rebuilt to fit the electron-density and the iron ion in the C-lobe was moved from a site with strong negative electron density to strong positive electron density. Some solvent molecules, regarded as water, that appeared in both $2F_o$-F_c ($>1\sigma$) and F_o-F_c ($>3\sigma$) maps and had suitable hydrogen bond partners were also added to the model. The R factor went from 0.287 to 0.192 following 12 cycles of refinement. The resolution range then was expanded to 40.0-2.5 Å and finally to 40.0-2.2 Å accompanied by rounds of refinement and rebuilding, which mainly involved remodelling of side chains and loop regions. Solvent water molecules were reviewed in each rebuilding procedure and some waters with poor density were removed.

Regions with high B values and negative density (less than -3σ) in F_o-F_c maps, which included residues 1-4, 220-221, 283-284, 293-294 and 418-424, were removed from the model in earlier stages of the refinement. The quality of the $2F_o$-F_c map was much better after this, and correspondingly the percentage of residues in the most favoured region of a Ramachandran plot was also improved greatly. These residues were rebuilt and re-introduced to the model later, as the overall quality of the maps increased. In initial refinements, in the absence of restraints, the iron-carbonate geometry became severely distorted and the iron ions, especially that in the C-lobe, tended to fluctuate considerably. This happened in the refinement of the structure of native human lactoferrin as well (Haridas et al., 1995). Restraints on the geometry of the carbonate anions and iron-ligand bonds, and B correlation parameters were therefore applied. These restraints were cancelled during the later stages of refinement and the geometry of the carbonate and iron-ligand bonds remained reasonable. The bond lengths of the carbonate ions were in the range 1.29 to 1.30 Å, and the bond angles 122.5° to 118.6°.

The iron ions, protein ligands and carbonate ions in both lobes fit the final $2F_o$-F_c map quite well, and as a double check, the F_o-F_c map calculated with no carbonate ions provided a very strong (>10σ) density peak which was fitted well with the model.

Sequencing results for the recombinant protein indicated that the only sequence difference is that Ala10 has been changed to Thr10. The electron density showed that density was present for the two extra atoms and that this residue could be built as threonine without disturbance to the structure.

RESULTS AND DISCUSSION

7.1 The final model

7.1.1 The quality of the final model

The final model contains 691 amino acids with 5307 non-hydrogen atoms, two Fe^{3+} ions, two carbonate ions and 362 solvent molecules all regarded as water. The stereochemistry of the model is very close to ideal with rms deviations of 0.009 Å and 1.4° from standard values of bond lengths and angles. The final crystallographic R factor is 20.0% for all 32667 reflections in resolution range of 40.0 to 2.2 Å, with a free R factor of 30.0% (3616 reflections). Some further cycles of refinement have also been performed on the final model with SHELXL-97 (Sheldrick, G. M. and Schneider, T. R., 1997), in the resolution range 10-2.2 Å, giving a better R factor (18.1%) and free R factor (27.4%) with basically the same structure. Refinement statistics are listed in Table 7.1.1-I.

Table 7.1.1-I TNT refinement statistics

		TNT stereochemistry rms deviations	
Resolution range :	40.0 - 2.20 Å		
Number of reflections :	36284	Bond length :	0.009 Å
Number of protein atoms :	5307	Bond angle :	1.42°
Number of solvent (water) :	363	Torsion angle :	21.4°
Crystallographic R factor :	0.20 (0.181)†	Trigonal planes :	0.013 Å
Free R factor :	0.30 (0.274)†	General planes :	0.016 Å
		Bad contacts :	0.059 Å
		B-value correlations :	2.66 $Å^2$

† Values in parentheses are from subsequent SHELXL-97 refinement.

The average maximum coordinate error in the final model, estimated from a Luzzati plot (Luzzati, 1952) of the R factor as a function of resolution, is shown in Fig. 7.1.1-I and gives a maximum average error in the structure of ~0.30 Å.

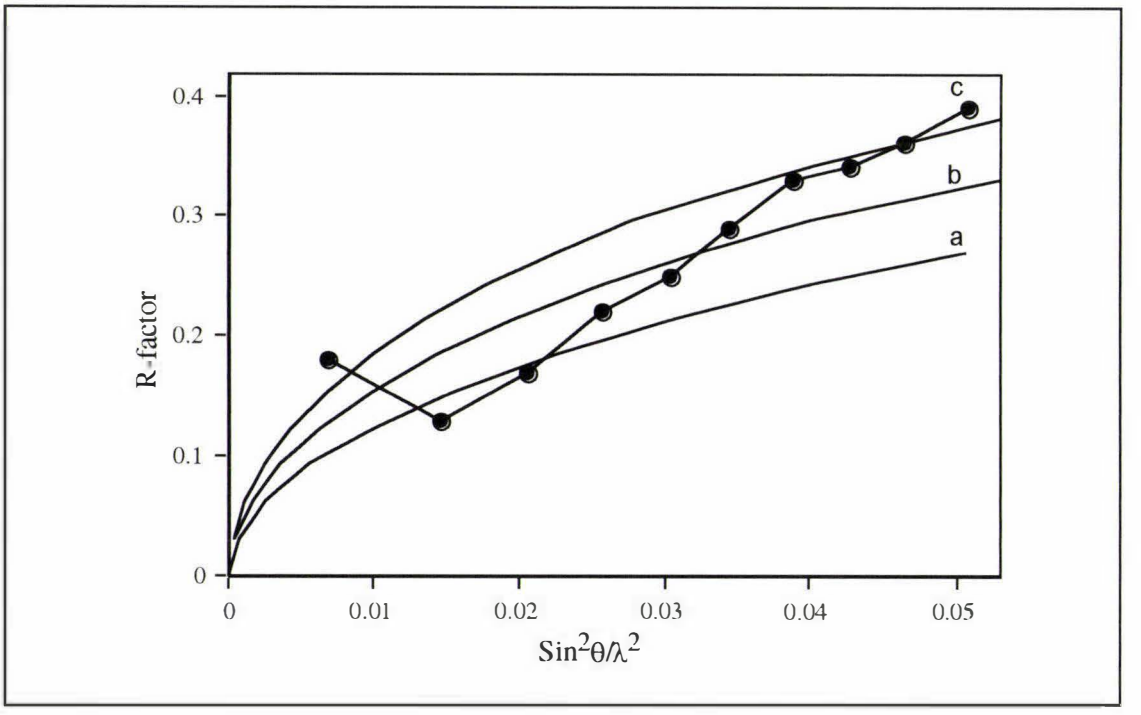


Fig. 7.1.1-I Luzzati plot of the R factor as a function of resolution. The theoretical variation of the R factor for coordinate errors of 0.2, 0.25, and 0.3 Å (curves a, b, and c, respectively) is shown.

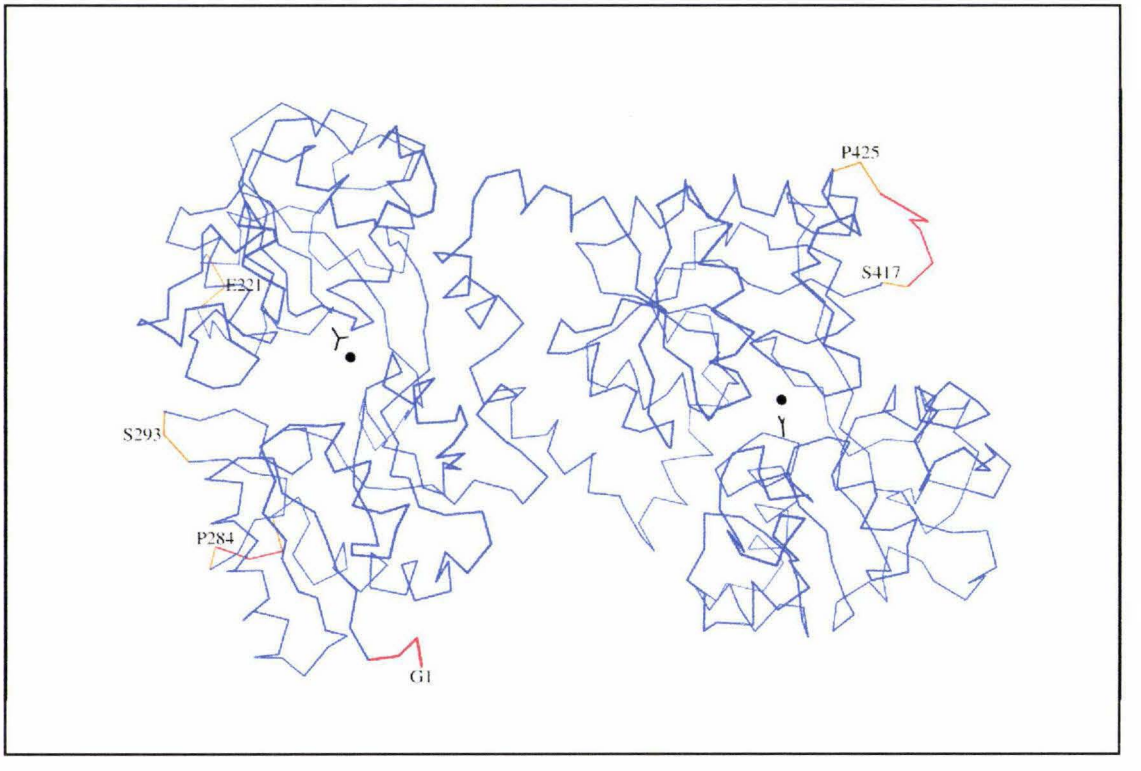


Fig. 7.1.1-II Cα plot of recombinant lactoferrin. The residues with B<60 $Å^2$(blue); 60 $Å^2$<B<80 $Å^2$(orange); B>80 $Å^2$(red) are shown.

Being maximum average value, which assume that the disagreement in observed and calculated structure factors results only from the coordinates, this average coordinate error overestimates the error in the well defined parts of the structure such as the two binding sites in which the error would be much less than average.

The mean B value is 45 $Å^2$ (rms 3.0 $Å^2$) for all protein main chain atoms and 53 $Å^2$ (rms 6.8 $Å^2$) for side chain atoms. The average main chain B value is consistent with the overall B estimated from the Wilson plot which gives an overall B of 43 $Å^2$. The relatively high B value for the main chain atoms also reflects the weakness of the high resolution data and the fact that all data were used in the least-squares refinement. As expected, most of the secondary structures and internal regions have lower than average B values, and some external turns and connecting loops have higher B values. A Cα diagram showing regions with different B values is shown in Fig. 7.1.1-II.

The final model fits with the electron density well except for a few loops and turns. They are residues 1-4, 219-221, 281-285 and 416-423. The N-terminal residues 1-4 project out from the surface of the N1 domain and electron density is very diffuse. As a long surface loop, residues 416-423 have no interactions with neighbours and are severely disordered. As in the native lactoferrin structure (Haridas et al., 1995), we can only locate the general positions of these two regions by using the weak and fragmented density. At the same time, these two flexible regions appear to take the different courses from those in native lactoferrin: rms differences after superimposing the recombinant and the native lactoferrin structures are 4.8 Å for residues 1-4 (averaging among main chain atoms), and 1.0 Å for residues 416-423 (averaging among main chain atoms) respectively. Residues 219-221 and 281-285 contain two ß-turns in the N-lobe that have no counterpart in the C-lobe. All these turns are located on the surface with poor density and resultant high B values. The real space correlation coefficients, which are calculated using the real space fit routine in the program O (Jones et al., 1991) and give a residue by residue measure of the agreement between the model and the final $2F_o$-F_c electron density map, show that in general the correlation of the model with the data is high with most of the residues having correlation coefficients greater than 0.85 (Fig. 7.1.1-III). The residues which have the poorest correlation (<0.7) include the external loops and turns mentioned above. Typical electron density in the well-defined regions is shown in Fig. 7.1.1-IV.

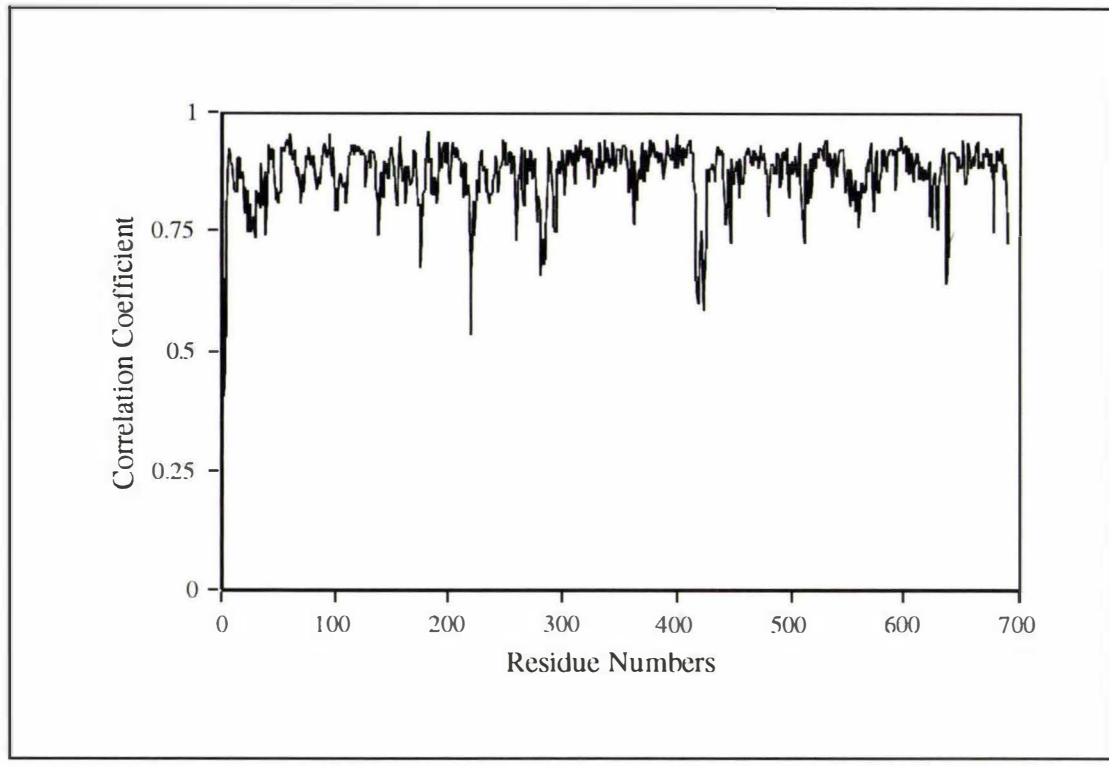


Fig. 7.1.1-III Real space correlation coefficients of the final model.

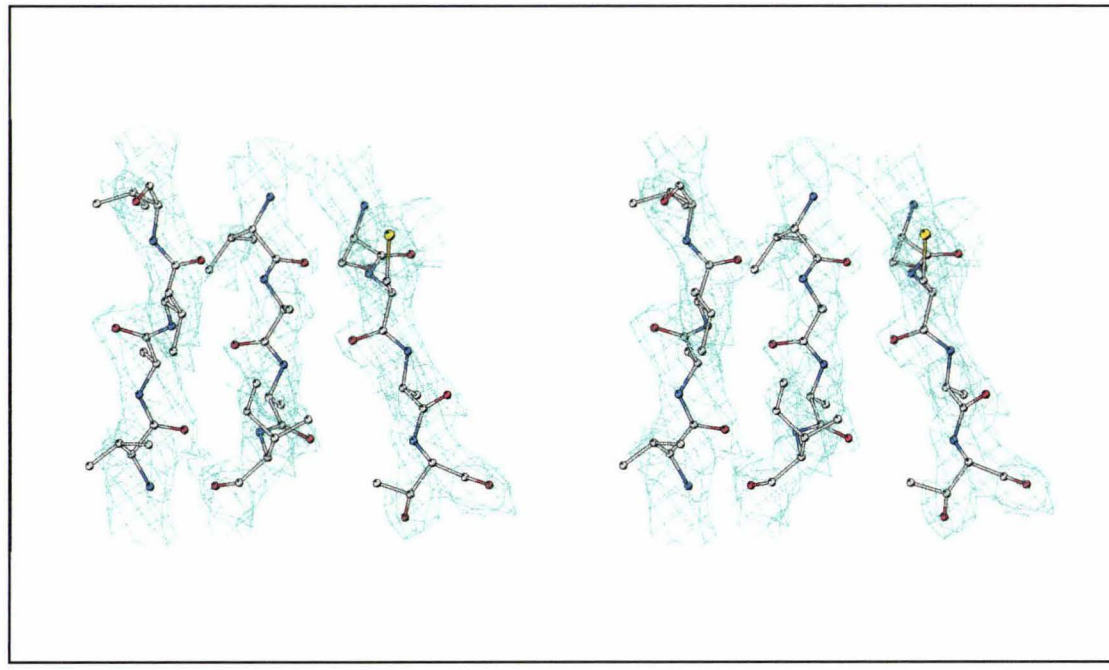


Fig. 7.1.1-IV The final $2F_o$-F_c electron density map for part of N-lobe of the recombinant lactoferrin, contoured at 1.5σ.

The model was also assessed by analysis of a variety of residue properties which are calculated by the program PROCHECK (Laskowski et al., 1993). A Ramachandran plot (Ramakrishnan and Ramachandran, 1965) of the main chain torsion angles shows that 85.5% of the non-glycine residues are in the "most favoured" region, and a further 13.6% of the residues fall in the "allowed region". Residues Ser191, Gln419 and Thr638 occupy the "general allowed region" and two residues, Leu299 and Leu642, are within the "disallowed region" (Fig. 7.1.1-V).

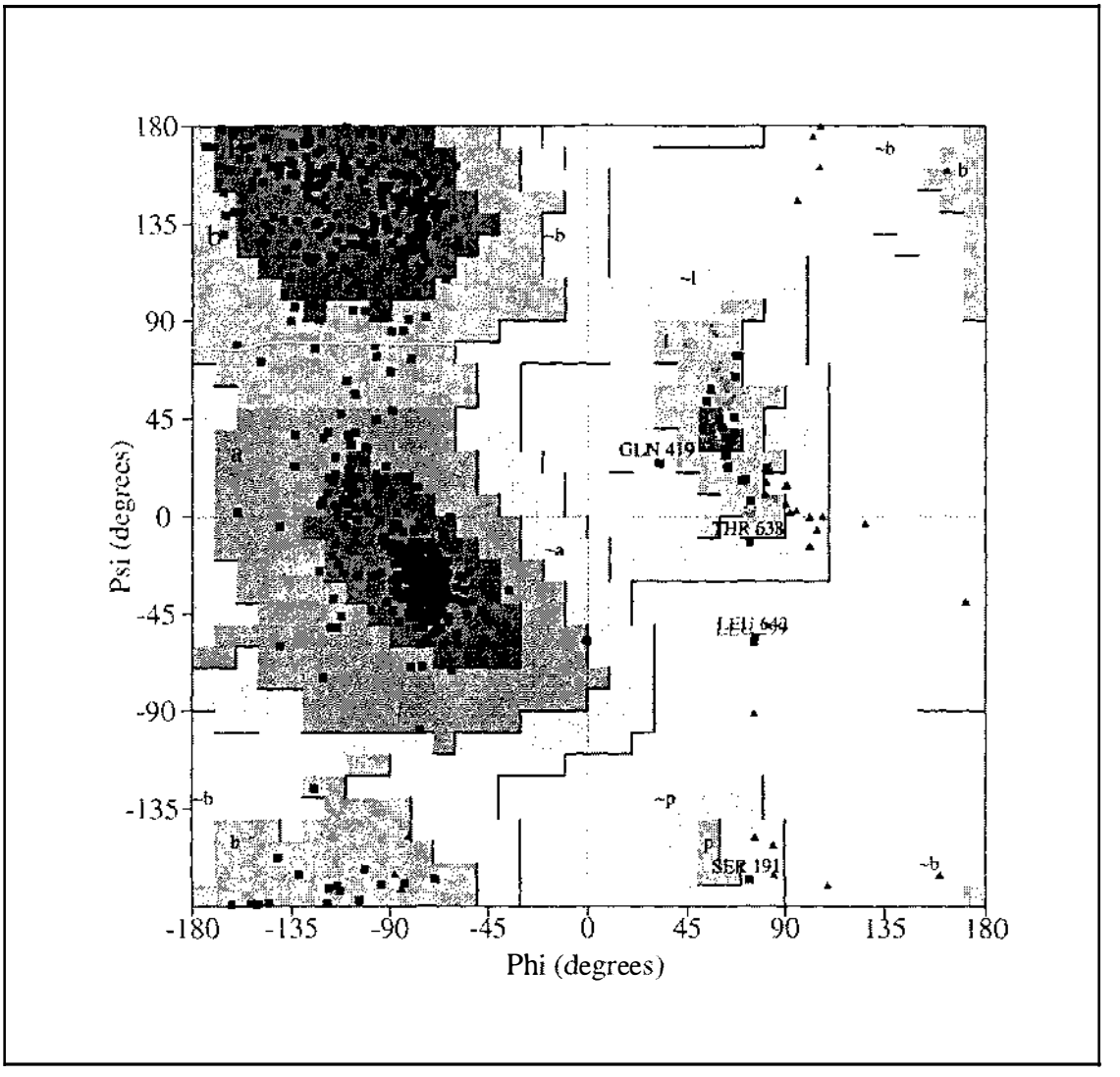
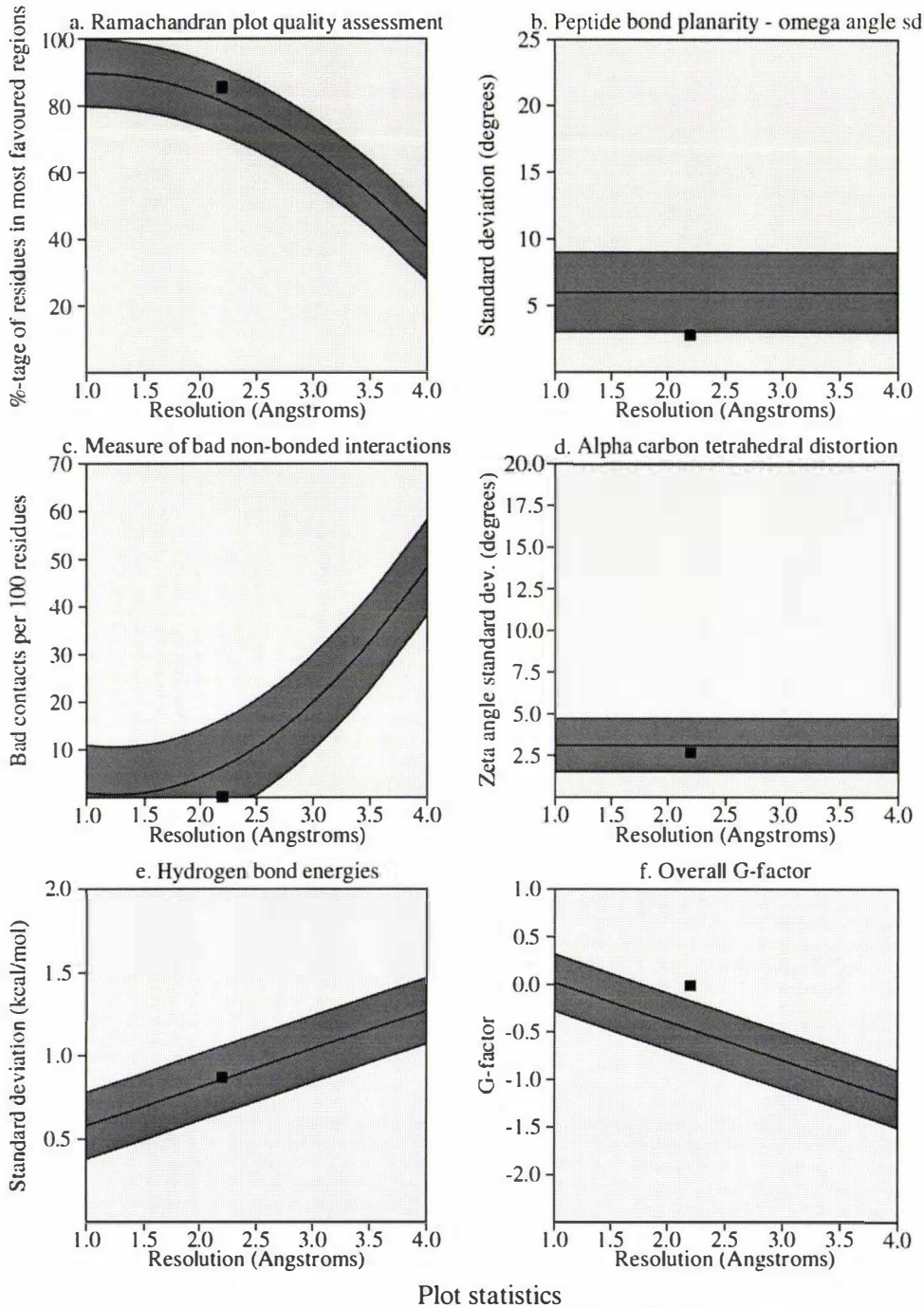


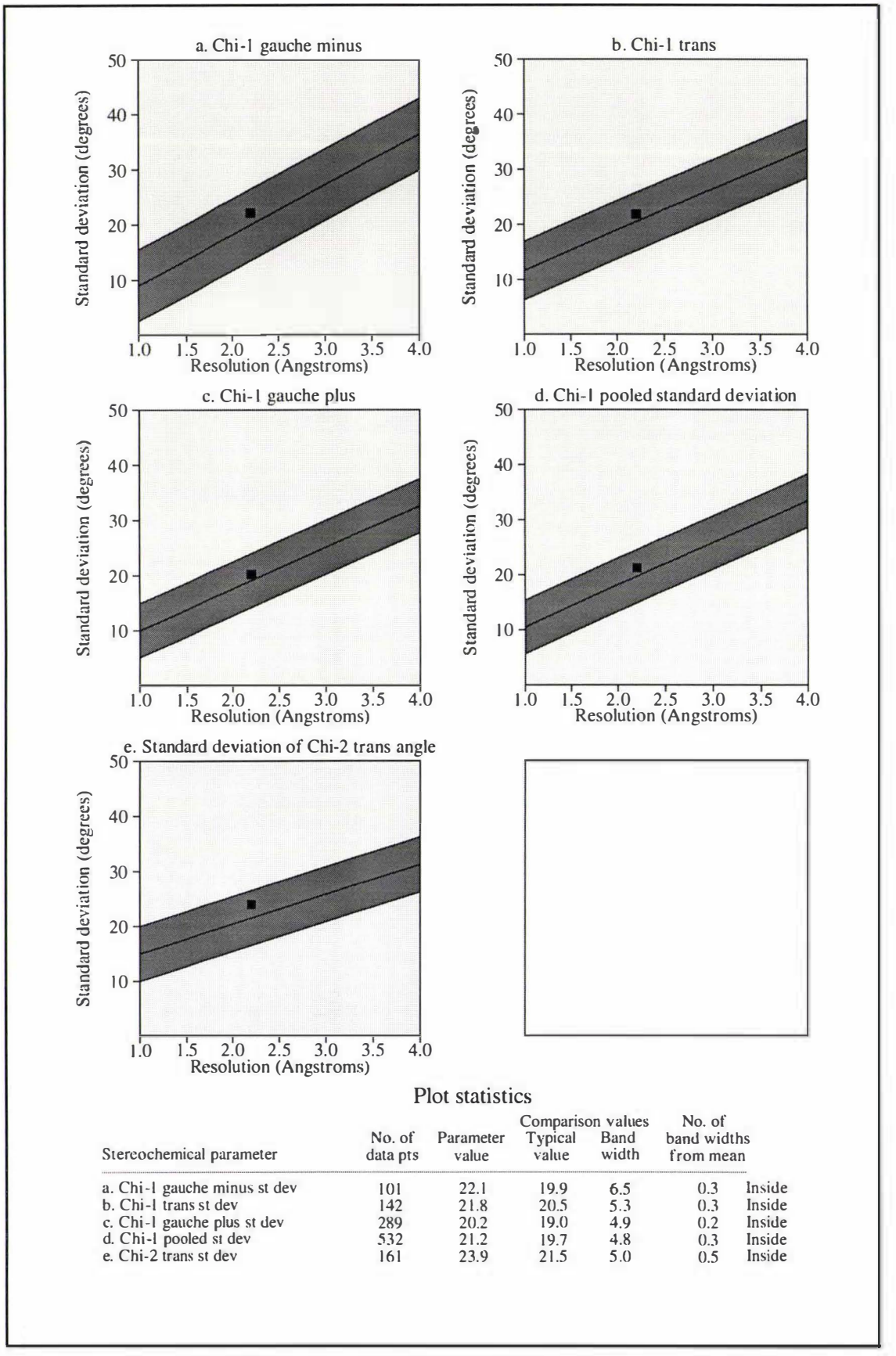
Fig. 7.1.1-V Ramachandran Plot for the final model of recombinant lactoferrin.



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	601	85.5	81.3	10.0	0.4	Inside
b. Omega angle st dev	687	2.8	6.0	3.0	-1.1	BETTER
c. Bad contacts / 100 residues	1	0.1	6.4	10.0	-0.6	Inside
d. Zeta angle st dev	637	2.7	3.1	1.6	-0.3	Inside
e. H-bond energy st dev	417	0.9	0.9	0.2	0.1	Inside
f. Overall G-factor	691	0.0	-0.5	0.3	1.5	BETTER

Fig. 7.1.1-VI Main chain parameters of the final model of recombinant lactoferrin.



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	101	22.1	19.9	6.5	0.3	Inside
b. Chi-1 trans st dev	142	21.8	20.5	5.3	0.3	Inside
c. Chi-1 gauche plus st dev	289	20.2	19.0	4.9	0.2	Inside
d. Chi-1 pooled st dev	532	21.2	19.7	4.8	0.3	Inside
e. Chi-2 trans st dev	161	23.9	21.5	5.0	0.5	Inside

Fig. 7.1.1-VII Side chain parameters of the final model of recombinant lactoferrin.

As in native lactoferrin, the residue Ser191 has the ε configuration (Sibanda et al., 1989) with torsion angles (ψ=72°, ϕ= -166°); this residue makes an important contribution to the N-lobe binding site, hydrogen bonding with anion-binding Arg121. Residue Leu299 and the equivalent Leu642 are located in two γ-turns in the N and C-lobe, respectively: their torsion angles, (75°, -57°) and (74°, -55°), are very close to those for residues at the centre of classic γ-turns, for which torsion angles are around (70°, -50°) (Baker and Hubbard, 1984). These two γ -turns with their highly conserved Leu-Leu-Phe sequence in the transferrin family are located in part of one wall of the binding cleft in each lobe, and seem to be a common feature of transferrin structure (Smith et al., 1994). The overall good quality of the structure is also indicated by the comparison of both main chain and side chain parameters with the average values of well-known structures at the same resolution (Fig. 7.1.1-VI and Fig. 7.1.1-VII). These comparisons show that the structural model is either better than or within the average values.

7.1.2 Glycosylation sites

As mentioned in Section 5.4.2, glycosylation analysis of the recombinant and native lactoferrin demonstrated that although the nature of carbohydrate-protein linkages and the extent of glycosylation are similar for both proteins, the compositions of carbohydrate chains are different; the oligosaccharides in the recombinant protein are high mannose-type compared with the complex carbohydrate structure of the native protein (Ward et al., 1995). Some sugar residues have been modelled in the native lactoferrin structure, N-linked to Asn137 of the N-lobe and Asn478 of the C-lobe. There are very few direct contacts found between the two carbohydrate chains and the protein. The density is generally weak and the carbohydrates clearly suffer from disorder (Haridas et al., 1995). In the present structural analysis, some limited density can be seen extending from the two expected glycosylation sites, Asn137 and Asn478, consistent with the previous glycosylation analysis. However, the density extends only a short distance from the Asn side chains and is only sufficient for a single sugar in each

of the two sites. It is also continuous only at the 0.5σ contour level in the $2F_o$-F_c map. Therefore, no attempt has been made to model the carbohydrate.

7.1.3 Overall polypeptide chain folding

The polypeptide chain folding of recombinant lactoferrin, as shown in Fig. 5.3.2-II, is basically same as that in native lactoferrin which has been described in detail (Haridas et al., 1995). Like the native molecule, the two lobes in the recombinant molecule have a very similar folding pattern. Through a 179.6° rotation (179.5° in native molecule) and a translation of 24.7 Å (24.5 Å in native molecule), the N-lobe can be superimposed on the C-lobe with an rms difference of 1.38 Å for the main chain atoms. Superposition is based on a sequence alignment between the two lobes (Anderson et al., 1989), where a total of 305 residues from each lobe are included in the superposition and some insertions, deletions and the final helix of each lobe (helix 12) are excluded. Helix 12 of the N-lobe joins the two lobes, whereas helix 12 of the C-lobe folds into the interlobe contact area, so they have quite different orientations. After further omitting some residues in loops which do not match in the two lobes and have rms differences greater than 3.0 Å, the structurally equivalent regions of the two lobes superimpose with an rms difference of 0.97 Å for main chain atoms, as compared to 1.0 Å for the same superposition in the native molecule. Furthermore, separately superimposing the N1 domain on to the C1 domain (137 Cα atoms) and N2 domain on to C2 domain (137 Cα atoms) gives even smaller rms differences of 0.85 Å and 0.56 Å respectively. The N-lobe is more open than C-lobe by 6.3°, as in the native molecule. These results show clearly that not only does the recombinant molecule have the same polypeptide chain folding pattern, but it also has the same relative lobe and domain orientations as are found in the native molecule.

All of the 16 disulfide bonds which are present in the native molecule and contribute to the folding and stability of lactoferrin are correctly formed in the recombinant molecule. The conformations of these disulfide bridges are all well conserved between the two structures, demonstrating that the *Aspergillus* expression system provides a suitable medium for correct disulfide bond formation and correct overall folding.

Table 7.2.1-Ia Secondary structures (Helices)‡

N lobe			C lobe		
Residues	Label	Type	Residues	Label	Type
12 - 31	(1)	α_{C2}	351 - 365	(1)	α_{C1}
41 - 53	(2)	α_{C2}	376 - 388	(2)	α_{C2}
60 - 69	(3)	α_{C1}	395 - 407	(3)	α_{C2}
105 - 108	(4)	3_{10}	449 -453	(4)	3_{10}
121 - 136	(5)	π, α_{C1}	465 -481	(5)	π, α
144 - 153	(6)	α_N	484 - 488	(6)	3_{10}
166 - 170	(6a)	3_{10}	501 - 505	(6a)	3_{10}
190 -203	(7)	α_{C2}	526 - 539	(7)	α_{C2}
212 - 218	(8)	α	547 - 554	(8)	α_N , α_{C1}
223 - 227	(8a)	3_{10}	567 - 571	(8a)	3_{10}
241 - 245	(8b)	3_{10}	585 - 589	(8b)	3_{10}
263 - 279	(9)	α	605 - 621	(9)	α
315 - 320	(10)	α	658 - 664	(10)	α
322 - 332	(11)	α_{C1}	664 - 678	(11)	α_{C1}
334 - 344	(12)	α	680 - 690	(12)	α

Table 7.2.1-Ib Secondary structures (ß-Strands)‡

N lobe			C lobe		
Residues	Label	Type	Residues	Label	Type
5 -10	(a)		345 - 350	(a)	
33 - 39	(b)		368 - 374	(b)	
54 - 59	(c)	b 55	389 - 394	(c)	b 390
75 - 82	(d)	b 77	408 -415	(d)	b 410
90 - 100	(e)		435 - 443	(e)	
112 - 116	(f)		456 - 460	(f)	
153 - 158	(g)	b 154	489 - 494	(g)	b 490
204 - 211	(h)	b 205	540 - 547	(h)	b 541
226 - 231	(i)		570 - 575	(i)	
246 -258	(j)	b 247	590 - 602	(j)	b 591
304 - 310	(k)	b 305	647 - 653	(k)	b 648

Table 7.2.1-Ic Secondary structures (turns)‡

N lobe			C lobe		
Residues		Type	Residues		Type
Matching turns (common to both lobes)					
109 - 112		II	453 - 456		II
118 - 121		II	462 - 465		II
158 - 161		II	494 - 497		II
162 - 166		α	498 - 501		I
170 - 173		I	504 - 507		I
179 - 182		II	515 - 518		II
184 - 187		III	520 - 523		III
187 - 190		I	523 - 526		I
210 - 213		III	546 - 549		III
231 - 234		I	575 - 578		III
238 - 241		I	582 - 585		III
287 - 290		I	632 - 635		I
298 - 300		γ (75, -58)	641 - 643		γ (76, -55)
301 - 304		II	644 - 647		I
Non-matching turns					
69 - 72		VIa(*cis*-pro)	362 - 365		I
(84 - 87)*		III	365 - 368		III'
140 - 143		VIa(*cis*-pro)	422 - 424		γ (-81, 90)
177 - 180	(4.02)†	Irregular	424 - 427	(4.15)†	Irregular
220 - 223		I	(442-445)*		III
(279-282)*		Irregular	444 - 446		γ (-89, 84)
283 - 286	(4.63)†	Irregular	(445-448)*		I
292 - 295		I	483 - 485		γ (-84, 85)
311 - 314		III	510 - 513		I
			554 - 557		III'
			560 - 563		III
			563 - 566		I
			602 - 605		III
			621 - 624		I
			624 - 627		I
			626 - 629		VIa(*cis*-pro)
			636 - 639	(4.19)†	Irregular
			653 - 656		I'
			655 - 657		γ (-96, 73)

‡ All secondary structure elements labelled as in Haridas, Anderson & Baker (1995). Helices and γ-turns are defined as in Baker & Hubbard (1984). ß-strands labelled with b means a ß-bulge conformation. ß-turn types are defined according to Wilmot and Thornton (1988).

* 1,4-hydrogen bond distance is within the range of 3.60 - 3.80 Å.

† 1,4-hydrogen bond distance which is substantially different from that in ative lactoferrin.

7.2 Structural comparison with native lactoferrin

The recombinant and the native human lactoferrin have essentially identical overall three-dimensional structure. Superposition of the main chain atoms between the two molecules calculated by program LSQKAB (collaborative computational project, number 4, 1994) gives an rms difference of 0.54 Å in atomic positions. The main regions of structural difference are N-terminal residues 1-4, and loop 416-423 which are poorly defined in both cases. The rms difference in atomic positions for main chain atoms falls to 0.31 Å if these two regions are omitted. This high degree of correspondence can be clearly seen in Fig. 7.2-1.

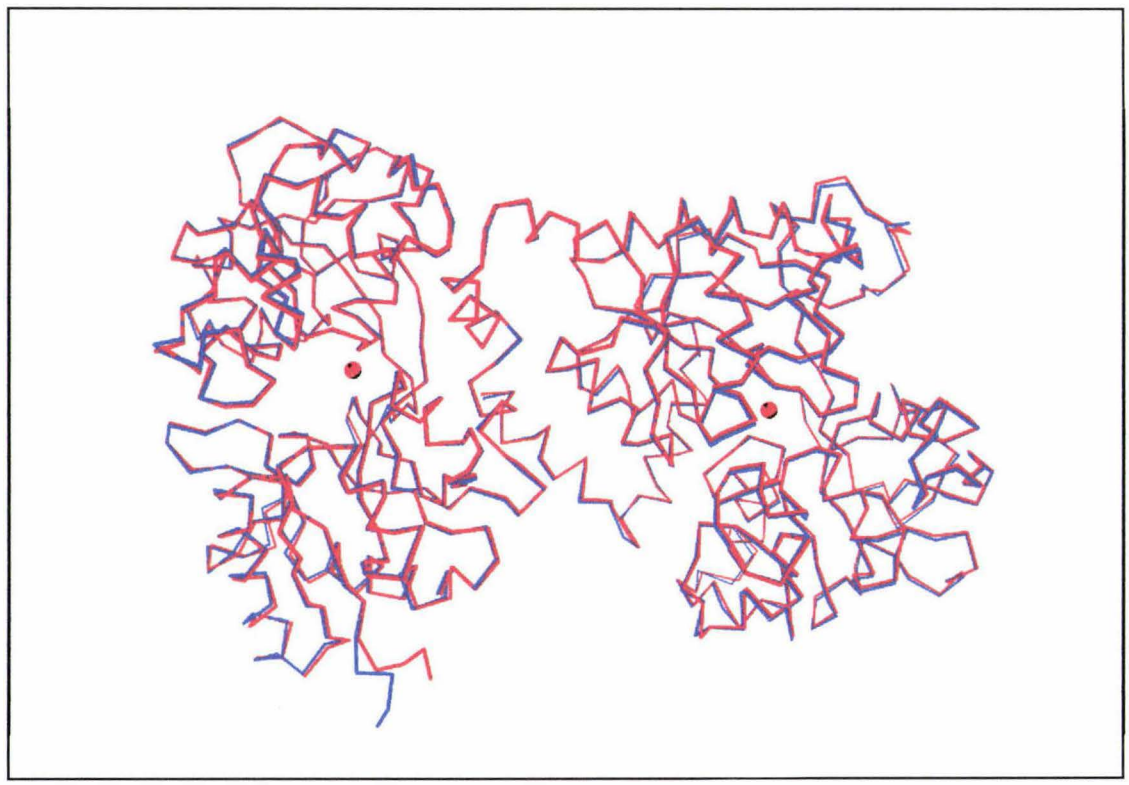


Fig. 7.2-1 Cα superposition between recombinant (red) and native (blue) lactoferrin.

7.2.1 Secondary structure elements

Secondary structural elements are listed in Table 7.2.1-I. Compared with the native lactoferrin structure, all of the β-strands have exactly the same structural details

including β-bulge conformations in some β-strands. All of the helices also have the same hydrogen bonding patterns except that the residue Lys691, which is at the C terminus of helix12 and has an α_{cl} hydrogen bonding pattern in the native molecule (Haridas et al., 1995), appears to be shifted away from the helix, leaving helix12 as an α type in the recombinant molecule. However, the B values for this C-terminal residue are high in both cases (76.5 $Å^2$ for main chain atoms in the recombinant molecule; 80.0 $Å^2$ for main chain atoms in the native molecule) and thus there is considerable positional uncertainty for this residue.

The β-sheets are the most highly conserved structural elements between the two lobes, and the equivalent sheets have identical conformations. As indicated in the native molecule (Haridas et al., 1995), the β-sheets form the core of each lobe. In comparison, most helices match quite well between the two lobes, but some differ where sequence insertions or deletions lengthen or shorten them at the protein surface (see Table 7.2.1-I).

Structural analysis of the native molecule (Haridas et al., 1995) showed that turns are the most variable elements between the two lobes. This is also reflected in the comparison of turns between the recombinant and native molecule. Although all of the turns in the native molecule are retained in the recombinant molecule, some structural details are changed. Three β-turns (575-578, 311-314 and 560-563) have been changed from type I in the native structure to type III in the recombinant structure. Most differences are seen in non-matching β-turns (those which are only present in one lobe or the other). These have relatively large differences in φ-ψ angles compared with those in native lactoferrin. In addition, some non-matching β-turns in the recombinant molecule have larger 1,4 hydrogen bond distances, especially the β-turns 177-180 and 283-286 in the N-lobe and 424-427 and 636-639 in the C-lobe. In fact these are probably no longer hydrogen bonds at all, as judged by the 1,4 hydrogen bond distances (cf. Table7.2.1-Ic), and the torsion angles of the two central residues in these four turns become concomitantly irregular. However, these differences are not surprising in view of the fact that all non-matching turns are located at the surface of the molecule with higher than average B values. The electron density for these four β-turns is also discontinuous at the 1.0σ contour level. A comparison of average B values for main

chain atoms of the recombinant molecule and the positional differences when the recombinant molecule is superimposed on to the native molecule, shown in Fig. 7.2.1-I, gives a high overall correlation. As shown in the Figure, most of the major positional differences originate from the non-matching turns, which also have generally higher B values. Thus, in both native and recombinant lactoferrin structure there are loop regions in common that are very flexible in conformation.

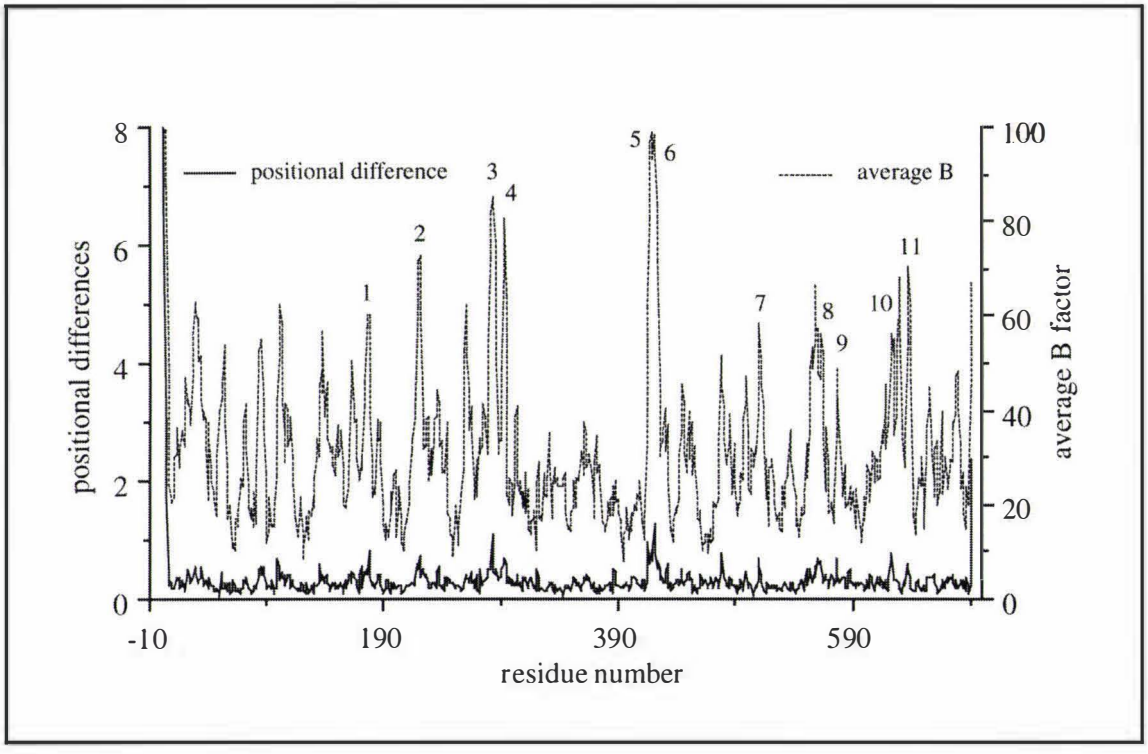


Fig. 7.2.1-I Plot of the positional differences (Å) and average B (Å^2) against residue numbers.
1.177-180(N) 2. 220-223(N) 3. 283-286(N) 4. 292-295(N) 5. 422-424(N) 6. 424-427(N)
7. 510-513(N) 8. 560-563 (III) 9. 575-578(III) 10. 621-624(N) 11. 636-639(N)

7.2.2 Iron and anion binding sites

The geometry of the iron binding sites is given in Table 7.2.2-I, and the hydrogen bonds that hold the carbonate anions are listed in Table 7.2.2-II. Table 7.2.2-I shows the essentially identical geometry and coordination pattern of the iron binding sites between the native and recombinant lactoferrin, at least within the experimental error of the two structural analyses. The maximum difference in the bond angles is only 8° in

N-lobe and 10° in C-lobe, and the average difference is 5.8° for the N-lobe and 3.9° for the C-lobe. When the two structures are superimposed, the atoms of the two binding sites ($2Fe^{3+}$, 2 CO_3^{2+}, Asp60, Tyr92, Tyr192, His253, Arg121, Thr117; Asp395, Tyr435, Tyr528, His597, Arg465, and Thr461; total 130 atoms) have an rms difference of only 0.23 Å. Fig. 7.2.2-I gives a view of the N-lobe and the C-lobe iron-binding sites of recombinant lactoferrin superimposed upon those of native lactoferrin. Most of the anion hydrogen bond lengths are also close to those for the native molecule (cf. Table 7.2.2-II). As in the native molecule, the full hydrogen-bonding potential of the carbonate ion is utilised and the hydrogen bonds all have favourable geometry. The electron density map for the carbonate ion (Fig. 7.2.2-II) also clearly shows its bidentate coordination to the iron ion, as revealed previously in native structure.

Table 7.2.2-I Geometry of the iron binding sites

N lobe		C lobe	
Atoms	Bond length (Å)	Atoms	Bond length (Å)
Fe-O60	2.06 (2.00)*	Fe-O395	2.08 (2.08)*
Fe-O92	2.07 (2.28)	Fe-O435	1.91 (1.92)
Fe-O192	1.92 (1.92)	Fe-O528	2.05 (2.10)
Fe-N253	2.14 (2.21)	Fe-N597	2.22 (2.13)
Fe-O1	2.05 (2.08)	Fe-O1	2.02 (2.13)
Fe-O2	2.27 (2.22)	Fe-O2	2.16 (2.06)

N lobe		C lobe	
Atoms	Bond angle (°)	Atoms	Bond angle (°)
O60-Fe-O92	94 (90)*	O395-Fe-O435	89 (87)*
O60-Fe-O192	165 (161)	O395-Fe-O528	167 (166)
O60-Fe-N253	81 (86)	O395-Fe-N597	80 (80)
O60-Fe-O1	98 (91)	O395-Fe-O1	90 (95)
O60-Fe-O2	79 (84)	O395-Fe-O2	91 (87)
O92-Fe-O192	99 (106)	O435-Fe-O528	98 (102)
O92-Fe-N253	97 (94)	O435-Fe-N597	93 (83)
O92-Fe-O1	87 (95)	O435-Fe-O1	97 (100)
O92-Fe-O2	146 (155)	O435-Fe-O2	161 (161)
O192-Fe-N253	91 (83)	O528-Fe-N597	89 (91)
O192-Fe-O1	89 (97)	O528-Fe-O1	101 (95)
O192-Fe-O2	93 (85)	O528-Fe-O2	86 (88)
N253-Fe-O1	176 (170)	N597-Fe-O1	165 (174)
N253-Fe-O2	114 (109)	N597-Fe-O2	105 (115)
O1-Fe-O2	62 (61)	O1-Fe-O2	64 (64)

* Data in parentheses for the native lactoferrin molecule.

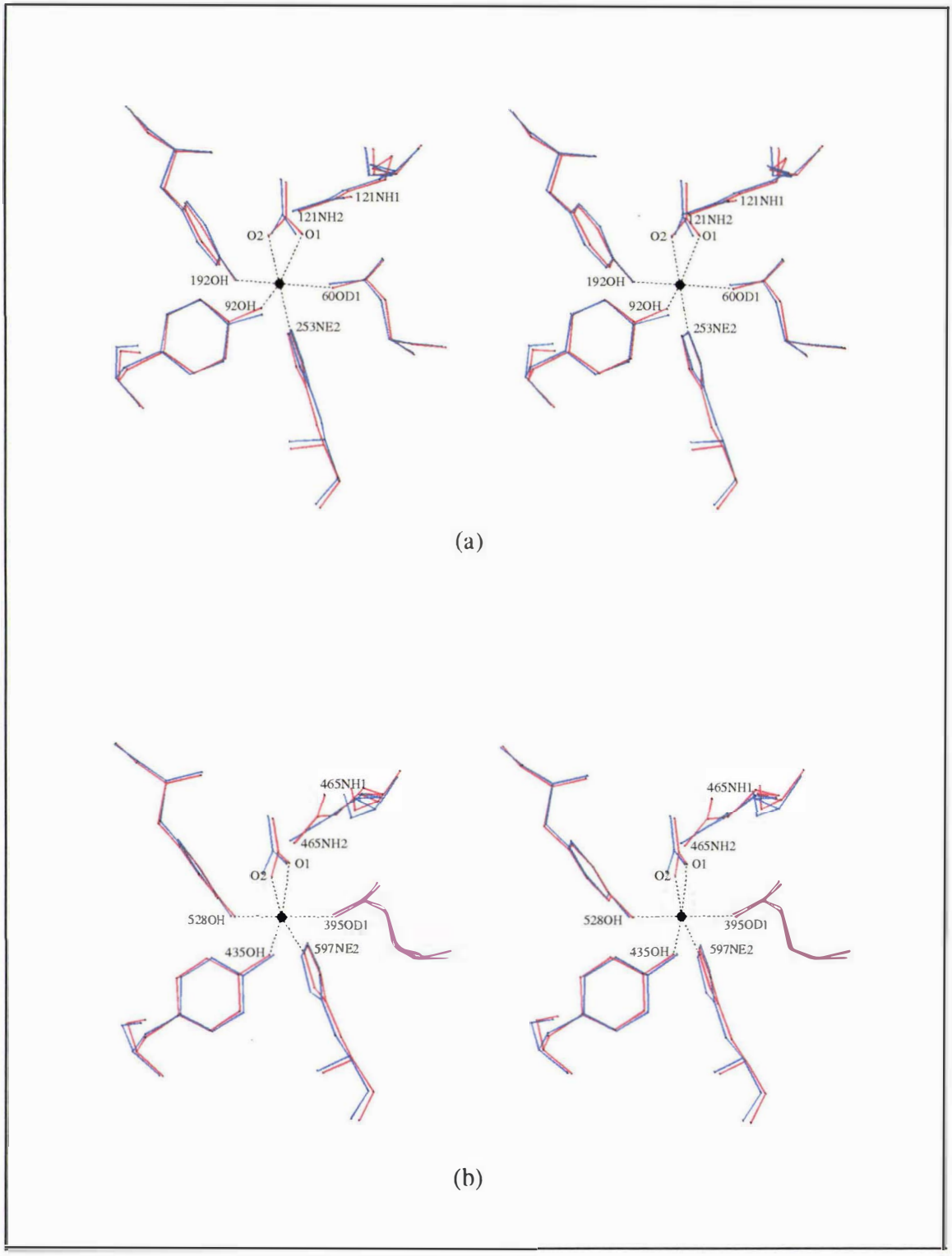


Fig. 7.2.2-I Superposition of the N-lobe (a) and the C-lobe (b) iron-binding sites between recombinant (red) and native (blue) lactoferrin.

Table 7.2.2-II Anion hydrogen bonding distances (Å) and angles (°)

Hydrogen bond	O···N	N-H···O	C-O···H
O1-N(123)	2.82 (2.83)†	166 (160)†	101 (110)†
O2-Ne(121)	2.88 (2.87)	133 (135)	100 (97)
O2-Nh2(121)	2.93 (2.85)	130 (132)	110 (111)
O3-N(124)	3.11 (3.07)	145 (141)	133 (130)
O3-Og1(117)*	2.64 (2.59)	118 (115)	148 (150)
O1-N(467)	2.82 (2.59)	169 (157)	113 (122)
O2-Ne(465)	2.88 (2.65)	139 (139)	110 (97)
O2-Nh2(465)	2.76 (2.88)	138 (128)	132 (126)
O3-N(468)	3.10 (2.94)	153 (153)	133 (125)
O3-Og1(461)*	2.82 (2.53)	108 (112)	147 (149)

* O3-Oγ1 distance is O···O, angles are Cβ-Oγ1···O and C-O···Oγ1

† Native lactoferrin molecule.

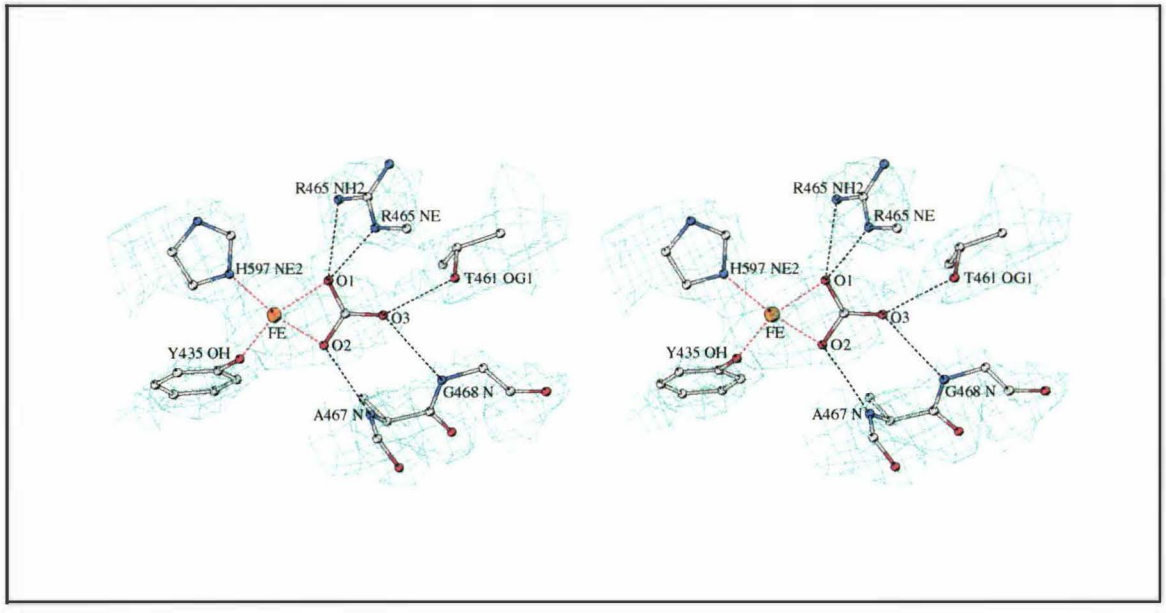


Fig. 7.2.2-II The $2F_o$-F_c electron density (at 1.5σ) for the carbonate ion (N-lobe), showing its bidentate coordination to the metal ion and hydrogen bonding with the surrounding protein structure.

Some important and striking differences between the two lobes, outside the immediate coordination sphere, have been analysed in detail in the native lactoferrin structure (Haridas et al., 1995). They are all retained in the recombinant structure. For example, there are different hydrogen bonding networks at the back of the iron site near

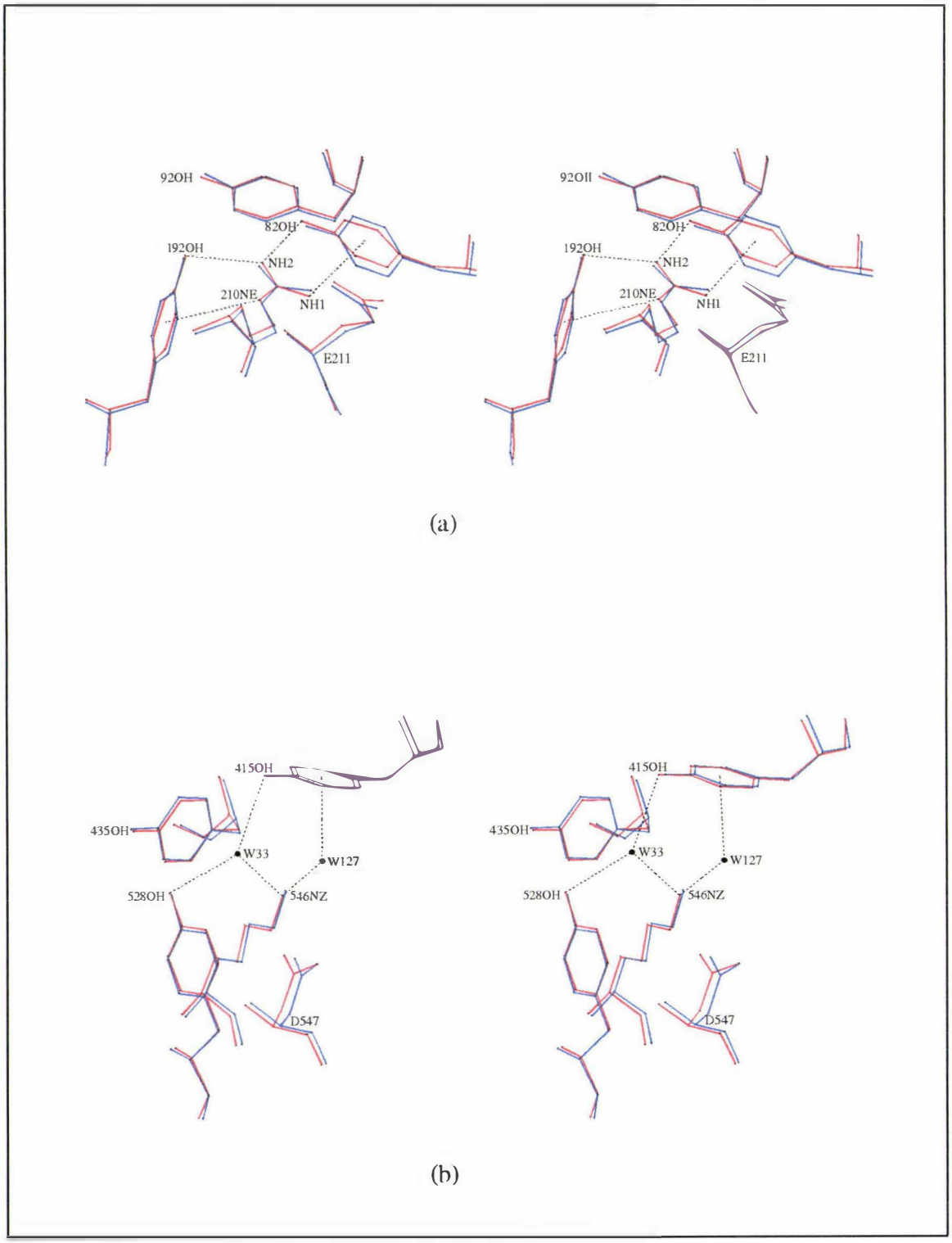


Fig. 7.2.2-III Superposition of hydrogen bonding network behind iron-binding site in N-lobe (a) and C-lobe (b). The recombinant molecule is indicated in red and the native molecule is shown in blue.

the two Tyr ligands, at position Arg210 in the N-lobe and Lys546 in the C-lobe. Arg210 in the N-lobe makes a hydrogen bonding network, including hydrogen bonds to the phenolic O atoms of Tyr192 and Tyr82; two amino-aromatic hydrogen bonds, from Nε to the ring of Tyr192 and from Nη1 to the ring of Tyr82. In the C-lobe, however, the equivalent residue is Lys546, which makes no such interactions, being hydrogen bonded only to water molecules which fill the positions of 210Nη1 and 210Nη2. Thus, Lys546 makes only indirect interactions with protein. Superposition of these two hydrogen bonding networks between the native and recombinant molecule gives an rms difference of only 0.2 Å for 92 atoms of residues Tyr92, Tyr192, Arg210, Tyr82 in N-lobe; Tyr435, Tyr528, Lys546, Tyr415 in C-lobe. As shown in Fig. 7.2.2-III, the recombinant molecule is in very good agreement with the native molecule with regard to these structural features.

7.2.3 Amino acid sequence differences

As mentioned in Section 5.4.2, a mutagenized strain improvement approach had been used to gain high level expression of lactoferrin in the *A. awamori* expression system (Ward et al., 1997b). Increases in yield of recombinant human lactoferrin were achieved using a combination of U.V. and chemical mutagenesis on the *A. awamori* lactoferrin-producing strain, which might introduce mutations within the recombinant lactoferrin gene as well. Sequence analysis indicated that the clone used to produce the recombinant lactoferrin contains one known sequence difference; Ala10 is replaced by Thr10.

The side chain of the threonine residue is accommodated easily in the recombinant protein, without disturbance of the structure. The atoms Thr10Oγ1 and Thr10Cγ2 fit with the final $2F_o$-F_c map (>1σ) well, and the omit maps in which these two side chain atoms had been removed from model showed clearly electron density in the corresponding positions in both $2F_o$-F_c (>1σ) and F_o-F_c (>3σ) maps. The B values for Thr10Oγ1 and Thr10Cγ2 are 63 Å^2 and 54 Å^2, respectively, which are reasonable for residues located near the N-terminal end. This residue is partly buried in one wall of the binding cleft. The methyl group makes no contacts less than 3.6 Å, and the hydroxyl

group makes a hydrogen bond to the nearby carbonyl oxygen of Ile37. Structural superposition between the native and recombinant molecule shows no significant changes caused by the side chain of Thr10.

Inspection of the maps ($2F_o$-F_c and F_o-F_c) for the final model indicates that no other sequence changes can be detected from the electron density.

7.2.4 Solvent structure

In this structure determination, many of the well ordered solvent sites are readily identified as water during refinement by the shape of the peaks and their environment. Although some strong peaks which are on the outside of the molecule and have no apparent hydrogen bond partners from the protein could represent phosphate ions contained in the crystallization solution, the density is not sufficient to build them without doubt. Accordingly, all solvent molecules are treated as water.

The 363 water molecules were located independently of the native structure. Superposition of the recombinant and native structures (based on least-squares fitting of their Cα positions) followed by analysis of the superimposed solvent structure showed that 188 out of 363 recombinant lactoferrin water molecules were in conserved positions (within 1.0 Å). They are mainly distributed in the internal regions and interdomain cavity of each lobe, and some in crevices in the protein surface. All internal water molecules of native structure are retained in recombinant molecule. These water molecules are located in discrete sites and generally have lower B values, they make a large number of hydrogen bonded contacts to the surrounding protein structure and contribute to the stability of the structure. As a typical example of internal water, two water molecules, W15 and W34, are found to be close to the Fe ions in each lobe at the same place as in the native structure: W15, hydrogen bonded to Tyr92O (2.68 Å), Gly61N (3.00 Å), Thr122Oγ1 (2.89 Å) and Pro251O (3.14 Å), is 3.79 Å away from the Fe atom in the N-lobe; W34, hydrogen bonded to Tyr435O (2.68 Å), Gly396N (3.03 Å), Thr466Oγ1 (2.68 Å) and Pro595O (2.63 Å), is 3.83 Å away from the Fe atom in the C-lobe. This confirms NMR evidence suggesting the presence of water molecules near the binding sites (Koenig and Schillinger, 1969).

Table 7.2.4-I Single-water bridges between two domains*

N lobe	C lobe
Arg332Nη1···W8···Asn126Oδ1	Tyr398Oη···W75···Val463O
Ala327O···W302···Asn126O	Glu352Oε1···W121···Asn515Nδ2
Leu331O···W145···Ile145N	Thr377Oγ1···W74···Pro519O
Val11O···W131···Phe183O	Val350O···W28···Ser521N
Asp40O···W248···Phe183O	Glu354N···W140···Ser521Oγ1
Ser12Oγ1···W3···Ser185Oγ1	Gln353N···W120···Asn522Oδ1
Asp297Oδ2···W357···Ser191Oγ1	Asn640Nδ2···W45···Tyr528N
Asp297Oδ2···W86···Ser193N	Asn644Nδ2···W33···Tyr528Oη†
Lys301Nζ···W116···Ser193Oγ1	Asn646Oδ1···W48···Lys546Nζ
Lys301Nζ···W20···Thr213Oγ1	Asn646Nδ2···W127···Lys546Nζ

* In the case of more than one pair of hydrogen bonds for the water bridge, only the strongest hydrogen bond is listed.
† Water molecule occupies the site equivalent to Arg210Nη2 of the N-lobe.

Almost all of the water molecules contained in the interdomain clefts are also retained in the recombinant molecule. The majority of them are anchored to protein atoms around the wall of each cavity but a few, in the centre of cavity, are hydrogen bonded only to neighbouring waters. Compared with the internal waters, these well-ordered waters are highly inter-connected. They stretch between the iron site and the mouth of the cleft and make up most of the interactions between the domains through single-water bridges. Some of the single-water bridges are shown in Table 7.2.4-I. In addition to a large number of conserved waters in the internal sites and interdomain clefts, some water molecules in crevices in the protein surface reappear in the recombinant molecule as well. They attach either to main-chain atoms or to certain side-chain atoms which offer favourable hydrogen bonding sites.

7.2.5 Interdomain and interlobe interaction

Having the identical interdomain and interlobe orientations as in the native molecule, as described in Section 7.1.3, the recombinant molecule retains all of the interdomain hydrogen bonding interactions found in native lactoferrin except that the salt bridge Lys296Nζ···Asp217Oδ1 in the N-lobe, which is relatively long (3.43 Å) in the native

structure, is even longer (3.92 Å) in the recombinant structure. In the meantime, one more hydrogen bond Thr638Oγ1···Glu560Oε1 (3.30 Å) located at the mouth of the cleft appears in the C-lobe. However, electron densities for the atoms 296Nζ, 217Oδ1, 638Oγ1 and 560Oε1 are poor in both native and recombinant structures and the B values are correspondingly high (96.8, 53.5, 90.4, and 88.6 $Å^2$ in the recombinant structure; 92.1, 38.6, 68.4 and 99.9 $Å^2$ in the native structure). Therefore, these four atoms involved in the two hydrogen bonds appear to be not well defined in either case. Direct protein-protein hydrogen bonds between the two domains are given in Table 7.2.5-I. As discussed for the native structure (Haridas et al., 1995), the most striking feature of the binding cleft is the relative lack of direct protein-protein interaction between the two domains. Instead, there are a large number of well-ordered single-water bridges (cf. Table 7.2.4-I). Most of the interactions between the domains are, therefore, mediated by solvent. This is consistent with requirement for reversible binding by transferrins, in which two domains alternate frequently between "open" and "closed" forms.

Table 7.2.5-I Interdomain hydrogen bond interactions*

N lobe	Bond Length(Å)	C lobe	Bond Length(Å)
Asp60Oδ2···Thr122N	2.75	Asp395Oδ2···Thr466N	2.74
Asp60Oδ2···Thr122Oγ1	2.89	Asp395Oδ1···Thr466Oγ1	3.08
Glu66Oε2···Arg332Nη2	3.36	Glu352Oε1···Asn522Nδ2	3.50
Tyr82Oη···Arg210Nη2	3.18	Ser375O···Asn520Nδ2	3.08
Tyr82ring···Arg210Nη1	3.38	Thr638O···Thr529Oγ1	2.80
Arg89Nη2···Glu211Oε2	2.84	Thr638Oγ1···Glu560Oε1	3.30
Lys301Nζ···Glu216Oε2	2.60		

* Only direct protein-protein hydrogen bonds (<3.5 Å) are listed.

Except for the covalent connection provided by the three-turn helix, residues 334-344, there are also very few direct interactions between the two lobes. All the same interactions as in the native structure (Haridas et al., 1995) reappear. They include two salt bridges, one hydrogen bond and a number of hydrophobic interactions, together with a few single-water bridges, (Table 7.2.5-II).

Compared with interdomain interactions, the main feature of the interlobe contact area is hydrophobic interactions of hydrophobic side chains, contributed by both lobes.

Table 7.2.5-II Interlobe interactions*

Hydrogen Bonds	Non-bonded Contacts
Arg313Nη2···Asp379Oδ1	Pro311Cδ···Phe688ring
Asp315Oδ2···Lys386Nζ	Pro311Cγ···Glu684O
Ser322Oγ···Leu385O	Ile314Cδ1···Ala685Cβ
	Ile314Cδ1···Leu689Cδ2
Single-Water Bridges	Gly317Cα···Lys386Cδ
Arg313Nη1···W845···Lys386Nζ	Leu318Cδ2···Leu385Cδ1
Arg313Nη2···W46···Asp379O	Leu318Cδ2···Ala382O
Arg313N···W280···Pro681O	Leu318O···Leu689Cδ2
Thr90Oγ1···W606···Phe688O	Phe325ring···Lys386O

* Hydrogen bonds (< 3.5 Å) and some non-bonded contacts (< 4.0 Å) are listed.

7.2.6 Intermolecular contacts

Direct intermolecular protein-protein contacts are listed in Table 7.2.6-I.

Table 7.2.6-I Intermolecular interactions*

Contacts	Distance(Å)	Contacts	Distance(Å)
Gly101O···Glu51Cβ^{I}	3.26	Ser293Oγ···Ser420O^{I}	2.86
Asp201O···Ser588Oγ1^{I}	2.52	Ser293Cβ···Ser420O^{I}	3.19
Asp201O···Glu585O^{I}	3.33	Arg309N···Asp629Oδ1^{II}	3.39
Gly202O···Ser588Oγ1^{I}	3.36	Arg309Cβ···Asp629Oδ2^{II}	3.35
Gly202O···Lys581Nζ^{I}	2.41	Ala340Cβ···Thr480O^{III}	3.45
Lys277Nζ···Pro423Cγ^{II}	3.35	Lys630Nζ···Glu684Oε2^{II}	2.96

*Hydrogen bonds (< 3.4 Å) and non-bonded contacts (< 3.5 Å) are listed. Symmetry-related positions: (I) 1/2+x, 1/2-y, -z ; (II) x, 1/2+y, 1/2-z.; (III) 1/2-x, -y, 1/2+z.

Many of them are equivalent to those in the native structure, reflecting the isomorphism of the crystal structures. Some intermolecular hydrogen bonds (e.g.

Lys99Nζ···Glu585Oε1^{I}, Arg272Nη1···Cys627O^{II}, Glu276Oε1···Arg622Nη2^{II} and Glu336Oε1···Lys454Nζ^{II}) in the native structure are absent because the atoms involved in such hydrogen bonding interactions are flexible; they are all located at the molecular surface with generally higher B values (>85 Å^2), and some of them are disordered in the recombinant structure. The main intermolecular contact regions involve the C terminus of helix 7 (residues 201-202), an external turn (residue 293) in the N-lobe; the long flexible loop joining β-strand d to e (residues 420-423), and the loop joining β-strand i to j (residues 581-588) in the C-lobe. Overall, only 24 out of 5307 atoms make direct intermolecular contacts for the recombinant lactoferrin and 32 out of 5330 atoms for the native lactoferrin. The small number of interactions indicates that the molecules are not tightly packed, and seems likely that for both native and recombinant lactoferrin the crystal packing does not disturb the protein structure significantly.

CONCLUSIONS

8.1 Structural identity

A high level of structural identity between the native and recombinant lactoferrins has been revealed by this comparison. The superposition of the two lobes of the recombinant molecule itself gives evidence that the interdomain and interlobe orientations are same as in the native molecule. Furthermore, the superposition of the recombinant structure upon the native structure shows that the two have essentially identical overall polypeptide chain folding patterns. Comparisons of detailed structural features such as the iron-binding sites, secondary structure elements and the hydrogen-bonding networks around and behind the iron sites exhibit perfect structural agreement. Most of solvent water molecules, interdomain and interlobe interactions are also retained in the recombinant structure.

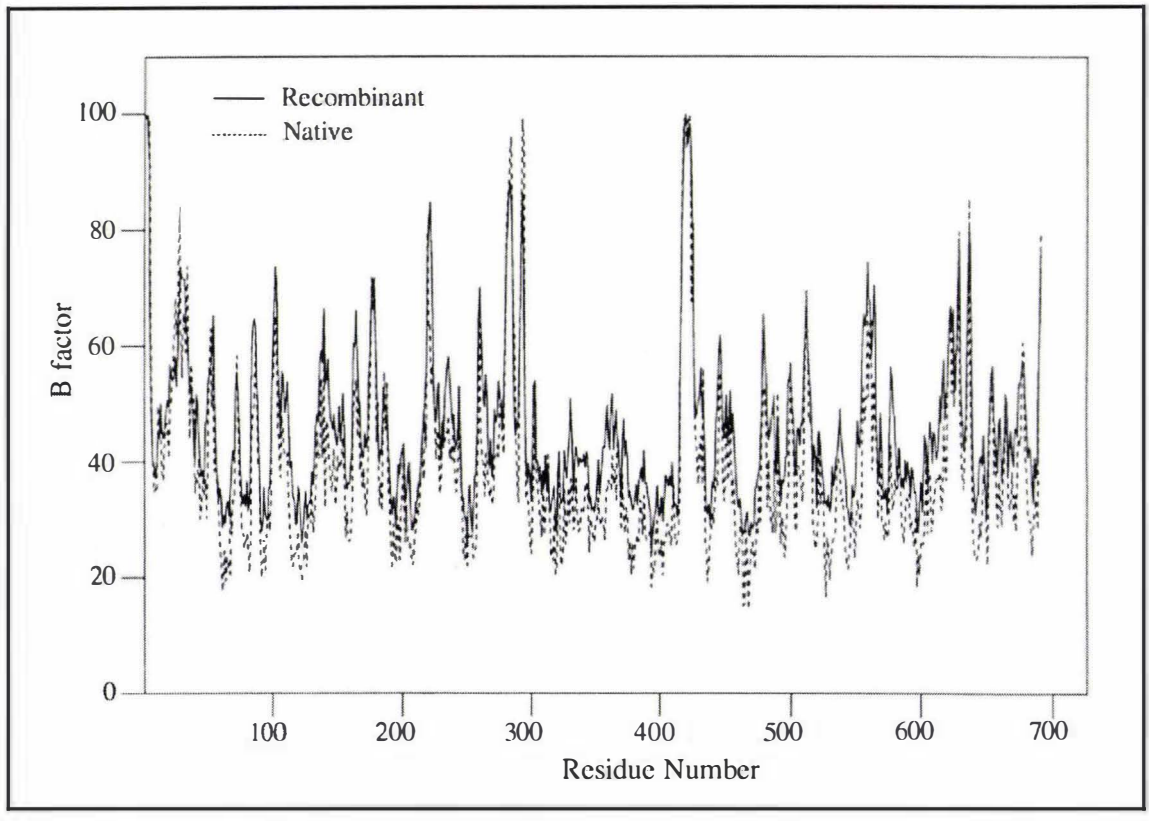


Fig. 8.1-I Comparison of the B factor (averaged among main chain atoms) between the recombinant and the native lactoferrin.

All equivalent β-sheets have exactly identical conformations not only between the two lobes but also between the two molecules from the different sources. This confirms again that the β-sheets are the core of each lobe. The crystallographic B values also show a consistent pattern of agreement between the recombinant and native lactoferrin (Fig. 8.1-I). Although B values can be influenced by model errors or other errors in refinement, they do give some measure of the flexibility of different regions of the molecule. Therefore, the consistency of the pattern of B values between the two structures implies similar dynamic properties. The fact that both structures have very few intermolecular contacts means that the B factors probably represent genuine molecular properties.

Apparent differences between the two structures have been found for some of the non-matching turns which are all located at the surface of the molecule. As depicted in Fig. 7.2.1-I, however, the non-matching turns with significant structural differences have correspondingly high B values. Thus, they are poorly defined in both cases, consistent with their relatively weak density. This is also consistent with similar dynamic properties, since both native and recombinant lactoferrin have the same loop regions which are naturally flexible.

8.2 Functional implications

The results of this structural analysis show that neither the mutagenesis used in strain improvement nor the method of expression have any impact on the three dimensional structure of lactoferrin. Structural identity makes recombinant lactoferrin indistinguishable from native lactoferrin in biological functions. The identity in the iron sites is consistent with the similar iron binding and release characteristics reported for the recombinant lactoferrin, in comparison with native lactoferrin (Ward et al., 1995). Although the altered carbohydrate chains have not been modelled in the recombinant structure, the carbohydrate composition makes no influence on the two protein three-dimensional structures and, consequently, the biological functions.

REFERENCES

Aisen, P., and Leibman, A. (1972). Lactoferrin and transferrin: a comparative study. *Biochim Biophys Acta* **257**, 314-323.

Aisen, P., and Listowsky, I. (1980). Iron transport and storage proteins. *Annu Rev Biochem* **49**, 357-393.

Alexander, L. J., Levine, W. B., Teng, C. T., and Beattie, C. W. (1992). Cloning and sequencing of the porcine lactoferrin cDNA. *Anim Genet* **23**, 251-256.

Ali, S. A., Joao, H. C., Csonga, R., Hammerschmid, F., and Steinkasserer, A. (1996). High-yield production of functionally active human serum transferrin using a baculovirus expression system, and its structural characterization. *Biochem J* **319**, 191-195.

Allegra, C. J., Chabner, B. A., Drake, J. C., Lutz, R., Rodbard, D., and Jolivet, J. (1985). Enhanced inhibition of thymidylate synthase by methotrexate polyglutamates. *J Biol Chem* **260**, 9720-9726.

Ambruso, D. R., and Johnston, R. B., Jr. (1981). Lactoferrin enhances hydroxyl radical production by human neutrophils, neutrophil particulate fractions, and an enzymatic generating system. *J Clin Invest* **67**, 352-360.

Anderson, B. F., Baker, H. M., Dodson, E. J., Norris, G. E., Rumball, S. V., Waters, J. M., and Baker, E. N. (1987). Structure of human lactoferrin at 3.2 Å resolution. *Proc Natl Acad Sci USA* **84**, 1769-1773.

Anderson, B. F., Baker, H. M., Norris, G. E., Rice, D. W., and Baker, E. N. (1989). Structure of human lactoferrin: crystallographic structure analysis and refinement at 2.8 Å resolution. *J Mol Biol* **209**, 711-734.

Anderson, B. F., Baker, H. M., Norris, G. E., Rumball, S. V., and Baker, E. N. (1990). Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins. *Nature* **344**, 784-787.

Arnold, R. R., Cole, M. F., and McGhee, J. R. (1977). A bactericidal effect for human lactoferrin. *Science* **197**, 263-265.

Arnold, R. R., Russell, J. E., Champion, W. J., Brewer, M., and Gauthier, J. J. (1982). Bactericidal activity of human lactoferrin: differentiation from the stasis of iron deprivation. *Infect Immun* **35**, 792-799.

Arnold, R. R., Russell, J. E., Champion, W. J., and Gauthier, J. J. (1981). Bactericidal activity of human lactoferrin: influence of physical conditions and metabolic state of the target microorganism. *Infect Immun* **32**, 655-660.

Atkinson, I., Garrow, T., Brenner, A., and Shane, B. (1997). Human cytosolic folylpoly-gamma-glutamate synthase. *Methods Enzymol* **281**, 134-140.

Baggiolini, M., De Duve, C., Masson, P. L., and Heremans, J. F. (1970). Association of lactoferrin with specific granules in rabbit heterophil leukocytes. *J Exp Med* **131**, 559-570.

Baker, E. N. (1994). Structure and Reactivity of Transferrins. *Advances in Inorganic Chemistry* **41**, 389-463.

Baker, E. N., Anderson, B. F., Baker, H. M., Faber, H. R., Smith, C. A., and Sutherland-Smith, A. J. (1997). Structural and Functional Flexibility of Lactoferrin. In *Lactoferrin: Interactions and Biological Functions*, T. W. Hutchens and B. Lonnerdal, eds. (Totowa, New Jersey: Humana Press Inc.), pp. 177-191.

Baker, E. N., Anderson, B. F., Baker, H. M., Haridas, M., Jameson, G. B., Norris, G. E., Rumball, S. V., and Smith, C. A. (1991). Structure, function and flexibility of human lactoferrin. *Int J Biol Macromol* **13**, 122-129.

Baker, E. N., Baker, H. M., Smith, C. A., Stebbins, M. R., Kahn, M., Hellstrom, K. E., and Hellstrom, I. (1992). Human melanotransferrin (p97) has only one functional iron-binding site. *FEBS Lett* **298**, 215-218.

Baker, E. N., and Hubbard, R. E. (1984). Hydrogen Bonding in Globular Proteins. *Prog Biophys Molec Biol.* **44**, 97-179.

Baker, E. N., and Lindley, P. F. (1992). New perspectives on the structure and function of transferrins. *J Inorg Biochem* **47**, 147-160.

Baldwin, G. S., and Weinstock, J. (1988). Nucleotide sequence of porcine liver transferrin. *Nucleic Acids Res* **16**, 8720.

Baldwin, S. W., Tse, A., Taylor, E. C., Rosowsky, A., Shih, C., and Moran, R. G. (1991). Structural features of 5,10-dideaza-5,6,7,8-tetrahydrofolate that determine inhibition of mammalian glycinamide ribonucleotide formyltransferase. *Biochemistry* **30**, 1997-2006.

Bali, P. K., and Aisen, P. (1991). Receptor-modulated iron release from transferrin: differential effects on N- and C-terminal sites. *Biochemistry* **30**, 9947-9952.

Bali, P. K., and Aisen, P. (1992). Receptor-induced switch in site-site cooperativity during iron release by transferrin. *Biochemistry* **31**, 3963-3967.

Banerjee, R. V., Shane, B., McGuire, J. J., and Coward, J. K. (1988). Dihydrofolate synthetase and folylpolyglutamate synthetase: direct evidence for intervention of acyl phosphate intermediates. *Biochemistry* **27**, 9062-9070.

Banfield, D. K., Chow, B. K., Funk, W. D., Robertson, K. A., Umelas, T. M., Woodworth, R. C., and MacGillivray, R. T. (1991). The nucleotide sequence of rabbit liver transferrin cDNA. *Biochim Biophys Acta* **1089**, 262-265.

Bannister, J. V., Bannister, W. H., Hill, H. A., and Thornalley, P. J. (1982). Enhanced production of hydroxyl radicals by the xanthine-xanthine oxidase reaction in the presence of lactoferrin. *Biochim Biophys Acta* **715**, 116-120.

Barredo, J., and Moran, R. G. (1992). Determinants of antifolate cytotoxicity: folylpolyglutamate synthetase activity during cellular proliferation and development. *Mol Pharmacol* **42**, 687-694.

Bartfeld, N. S., and Law, J. H. (1990). Isolation and molecular cloning of transferrin from the tobacco hornworm, *Manduca sexta*. Sequence similarity to the vertebrate transferrins. *J Biol Chem* **265**, 21684-1691.

Baugh, C. M., Krumdieck, C. L., Baker, H. J., and Butterworth, C. E., Jr. (1971). Studies on the absorption and metabolism of folic acid. I. Folate absorption in the dog after exposure of isolated intestinal segments to synthetic pteroylpolyglutamates of various chain lengths. *J Clin Invest* **50**, 2009-2021.

Baugh, C. M., Stevens, J. C., and Krumdieck, C. L. (1970). Studies on gamma-glutamyl carboxypeptidase. I. The solid phase synthesis of analogs of polyglutamates of folic acid and their effects on human liver gamma-glutamyl carboxypeptidase. *Biochim Biophys Acta* **212**, 116-125.

Beardsley, G. P., Moroson, B. A., Taylor, E. C., and Moran, R. G. (1989). A new folate antimetabolite, 5,10-dideaza-5,6,7,8-tetrahydrofolate is a potent inhibitor of *de novo* purine synthesis. *J Biol Chem* **264**, 328-333.

Bellamy, W., Takase, M., Yamauchi, K., Wakabayashi, H., Kawase, K., and Tomita, M. (1992). Identification of the bactericidal domain of lactoferrin. *Biochim Biophys Acta* **1121**, 130-136.

Bellamy, W. R., Wakabayashi, H., Takase, M., Kawase, K., Shimamura, S., and Tomita, M. (1993). Role of cell-binding in the antibacterial mechanism of lactoferricin B. *J Appl Bacteriol* **75**, 478-484.

Berry, M. B., Meador, B., Bilderback, T., Liang, P., Glaser, M., and Phillips, G. N., Jr. (1994). The closed conformation of a highly flexible protein: the structure of *E. coli* adenylate kinase with bound AMP and AMPPNP. *Proteins* **19**, 183-198.

Bertrand, J. A., Auger, G., Fanchon, E., Martin, L., Blanot, D., van Heijenoort, J., and Dideberg, O. (1997). Crystal structure of UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase from *Escherichia coli*. *EMBO J* **16**, 3416-3425.

Bhat, T. N. (1988). Calculation of an OMIT map. *J Appl Cryst* **21**, 279-281.

Bishop, J. G., Schanbacher, F. L., Ferguson, L. C., and Smith, K. L. (1976). In vitro growth inhibition of mastitis-causing coliform bacteria by bovine apo-lactoferrin and reversal of inhibition by citrate and high concentrations of apo-lactoferin. *Infect Immun* **14**, 911-918.

Blakley, R. L. (1969). The Biochemistry of Folic Acid and Related Pteridines. In *Frontiers of Biology* Vol. 13, A. Neuberger and E. L. Tatum, eds. (Amsterdam: North-Holland Publishing Co.), pp. 219-358.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R., and Klug, A. (1978). Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. *Nature* **276**, 362-368.

Blow, D. M., and Crick, F. H. C. (1959). The treatment of errors in the isomorphous replacement method. *Acta Cryst* **12**, 794-802.

Blundell, T. L., and Johnson, L. N. (1976). *Protein crystallography* (London: Acdemic Press Inc.).

Boffetta, P., and Kaldor, J. M. (1994). Secondary malignancies following cancer chemotherapy. *Acta Oncol* **33**, 591-598.

Bognar, A. L., Cichowicz, D. J., and Shane, B. (1983). Purification and characterization of folylpolyglutamate synthetase from *Lactobacillus casei* and hog liver. In *Chemistry and biology of pteridines*, J. A. Blair, ed. (Berlin: de Gruyter), pp. 627-632.

Bognar, A. L., Osborne, C., and Shane, B. (1987). Primary structure of the *Escherichia coli folC* gene and its folylpolyglutamate synthetase-dihydrofolate synthetase product and regulation of expression by an upstream gene. *J Biol Chem* **262**, 12337-12343.

Bognar, A. L., Osborne, C., Shane, B., Singer, S. C., and Ferone, R. (1985). Folylpoly-γ-glutamate synthetase-dihydrofolate synthetase. Cloning and high expression of the *Escherichia coli folC* gene and purification and properties of the gene product. *J Biol Chem* **260**, 5625-5630.

Bognar, A. L., and Shane, B. (1983). Purification and properties of *Lactobacillus casei* folylpoly-γ-glutamate synthetase. *J Biol Chem* **258**, 12574-12581.

Bognar, A. L., and Shane, B. (1986). Bacterial folylpoly(gamma-glutamate) synthase-dihydrofolate synthase. *Methods Enzymol* **122**, 349-359.

Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., and Kraut, J. (1982). Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7Å resolution. 1. General features and binding of methotrexate. *J Biol Chem* **257**, 13650-13662.

Boxer, L. A., Haak, R. A., Yang, H. H., Wolach, J. B., Whitcomb, J. A., Butterick, C. J., and Baehner, R. L. (1982). Membrane-bound lactoferrin alters the surface properties of polymorphonuclear leukocytes. *J Clin Invest* **70**, 1049-1057.

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.

Britten, J. R., and Koldovsky, O. (1989). *Early Human Development* **19**, 127-135.

Brock, J. H. (1985). Transferrins. In *Metalloproteins* Part II, P. M. Harrison, ed. (London: Macmillan Press), pp. 183-262.

Brock, J. H. (1997). Lactoferrin Structure-Function Relationships. In *Lactoferrin: Interactions and Biological Functions*, T. W. Hutchens and B. Lonnerdal, eds. (Totowa, New Jersey: Humana Press Inc.), pp. 3-23.

Brock, J. H., and Djeha, A. (1997). Regulation of Lymphocyte Proliferation by Lactoferrin. In *Lactoferrin: Interactions and Biological Functions*, T. W. Hutchens and B. Lonnerdal, eds. (Totowa, New Jersey: Humana Press Inc.), pp. 233-243.

Brown, J. P., Hewick, R. M., Hellstrom, I., Hellstrom, K. E., Doolittle, R. F., and Dreyer, W. J. (1982). Human melanoma-associated antigen p97 is structurally and functionally related to transferrin. *Nature* **296**, 171-173.

Brunger, A. T. (1992). *X-PLOR Version 3.1 A system for X-ray crystallography and NMR.*: Yale University, USA.

Buehring, K. U., Tamura, T., and Stokstad, E. L. (1974). Folate coenzymes of *Lactobacillus casei* and *Streptococcus faecalis. J Biol Chem* **249**, 1081-1089.

Bullen, J. J., Rogers, H. J., and Leigh, L. (1972). Iron-binding proteins in milk and resistance to *Escherichia coli* infection in infants. *Br Med J* **1**, 69-75.

Carpenter, M. A., and Broad, T. E. (1993). The cDNA sequence of horse transferrin. *Biochim Biophys Acta* **1173**, 230-232.

Casey, J. L., Koeller, D. M., Ramin, V. C., Klausner, R. D., and Harford, J. B. (1989). Iron regulation of transferrin receptor mRNA levels requires iron- responsive elements and a rapid turnover determinant in the 3' untranslated region of the mRNA. *EMBO J* **8**, 3693-3699.

Cassady, I. A., Budge, M. M., Healy, M. J., and Nixon, P. F. (1980). An inverse relationship of rat liver folate polyglutamate chain length to nutritional folate sufficiency. *Biochim Biophys Acta* **633**, 258-268.

Chen, L., Qi, H., Korenberg, J., Garrow, T. A., Choi, Y. J., and Shane, B. (1996). Purification and properties of human cytosolic folylpoly-gamma- glutamate synthetase and organization, localization, and differential splicing of its gene. *J Biol Chem* **271**, 13077-13087.

Cichowicz, D. J., Foo, S. K., and Shane, B. (1981). Folylpoly-gamma-glutamate synthesis by bacteria and mammalian cells. *Mol Cell Biochem* **39**, 209-228.

Cichowicz, D. J., and Shane, B. (1987a). Mammalian folylpoly-gamma-glutamate synthetase. 1. Purification and general properties of the hog liver enzyme. *Biochemistry* **26**, 504-512.

Cichowicz, D. J., and Shane, B. (1987b). Mammalian folylpoly-gamma-glutamate synthetase. 2. Substrate specificity and kinetic properties. *Biochemistry* **26**, 513-521.

Clarke, L., and Waxman, D. J. (1987). Human liver folylpolyglutamate synthetase: biochemical characterization and interactions with folates and folate antagonists. *Arch Biochem Biophys* **256**, 585-596.

Clarke, S., Ward, W., Planting, A., Spiers, J., Smith, R., Verweij, J., and Judson, I. R. (1992). Phase I trial of ICI D 1694: a novel thymidylate synthase inhibitor. *Proc Am Assoc Cancer Res* **33**, 406.

Clarke, S. J., Jackman, A. L., and Judson, I. R. (1993). The history of the development and clinical use of CB 3717 and ICI D1694. *Adv Exp Med Biol* **339**, 277-287; discussion 289-290.

Cody, V., Luft, J. R., Pangborn, W., Toy, J., and Bognar, A. L. (1992). Purification and crystallization of *Lactobacillus casei* folylpolyglutamate synthetase expressed in *Escherichia coli*. *J Mol Biol* **224**, 1179-1180.

Coleman, D. E., Berghuis, A. M., Lee, E., Linder, M. E., Gilman, A. G., and Sprang, S. R. (1994). Structures of active conformations of $G_{i\alpha 1}$ and the mechanism of GTP hydrolysis. *Science* **265**, 1405-1412.

Cook, J. D., Cichowicz, D. J., George, S., Lawler, A., and Shane, B. (1987). Mammalian folylpoly-gamma-glutamate synthetase. 4. In vitro and in vivo metabolism of folates and analogues and regulation of folate homeostasis. *Biochemistry* **26**, 530-539.

Cowtan, K. (1994). An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter in Protein Crystallography* **31**, 34-38.

Cox, T. M., Mazurier, J., Spik, G., Montreuil, J., and Peters, T. J. (1979). Iron binding proteins and influx of iron across the duodenal brush border. Evidence for specific lactotransferrin receptors in the human intestine. *Biochim Biophys Acta* **588**, 120-128.

Creighton, T. E. (1993). *Proteins: structures and molecular properties*, 2nd edition Edition (New York: W. H. Freeman & Co.).

Dahnke, T., Shi, Z., Yan, H., Jiang, R. T., and Tsai, M. D. (1992). Mechanism of adenylate kinase. Structural and functional roles of the conserved arginine-97 and arginine-132. *Biochemistry* **31**, 6318-6328.

Davidson, L. A., and Lonnerdal, B. (1988). Specific binding of lactoferrin to brush-border membrane: ontogeny and effect of glycan chain. *Am J Physiol* **254**, G580-585.

Davidson, L. A., and Lonnerdal, B. (1989). Specificity of the intestinal lactoferrin receptor. In *Milk Proteins*, C. A. Barth and E. Schlimme, eds. (New York: Springer), pp. 76-82.

Davidson, L., Kastenmayer, P., Yuen, M., Lonnerdal, B., and Hurrell, R. F. (1994). Influence of lactoferrin on iron absorption from human milk in infants. *Pediatr Res* **35**, 117-124.

Davies, J. F., Delcamp, T. J., Prendergast, N. J., Ashford, V. A., Freisheim, J. H., and Kraut, J. (1990). Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafolate. *Biochemistry* **29**, 9467-9479.

Day, C. L., Anderson, B. F., Tweedie, J. W., and Baker, E. N. (1993). Structure of the recombinant N-terminal lobe of human lactoferrin at 2.0 Å resolution. *J Mol Biol* **232**, 1084-1100.

Day, C. L., Stowell, K. M., Baker, E. N., and Tweedie, J. W. (1992). Studies of the N-terminal half of human lactoferrin produced from the cloned cDNA demonstrate that interlobe interactions modulate iron release. *J Biol Chem* **267**, 13857-13862.

DeGraw, J. I., Brown, V. H., Tagawa, H., Kisliuk, R. L., Gaumont, Y., and Sirotnak, F. M. (1982). Synthesis and antitumor activity of 10-alkyl-10-deazaminopterins. A convenient synthesis of 10-deazaminopterin. *J Med Chem* **25**, 1227-1230.

Denovan-Wright, E. M., Ramsey, N. B., McCormick, C. J., Lazier, C. B., and Wright, J. M. (1996). Nucleotide sequence of transferrin cDNAs and tissue-specific of the transferrin gene in Atlantic cod (*Gadus morhua*). *Comp Biochem Physiol B Biochem Mol Biol* **113**, 269-273.

Dewan, J. C., Mikami, B., Hirose, M., and Sacchettini, J. C. (1993). Structural evidence for a pH-sensitive dilysine trigger in the hen ovotransferrin N-lobe: implications for transferrin iron release. *Biochemistry* **32**, 11963-11968.

Dickerson, R. E., Kopka, M., Varnum, J. C., and Weinzierl, J. E. (1967). Bias, feedback and reliability in isomorphous phase analysis. *Acta Cryst* **23**, 511-522.

Diederichs, K., and Schulz, G. E. (1991). The refined structure of the complex between adenylate kinase from beef heart mitochondrial matrix and its substrate AMP at 1.85 Å resolution. *J Mol Biol* **217**, 541-549.

Dolnick, B. J., and Cheng, Y. C. (1978). Human thymidylate synthetase. II. Derivatives of pteroylmono- and polyglutamates as substrates and inhibitors. *J Biol Chem* **253**, 3563-3567.

Drenth, J. (1991). The chemistry of heavy atom attachment. In *Proceedings of the CCP4 study weekend*, W. Wolf, P. R. Evans and A. G. W. Leslie, eds. (Warrington, UK: SERC Daresbury laboratory), pp. 1-8.

Dreusicke, D., and Schulz, G. E. (1986). The glycine-rich loop of adenylate kinase forms a giant anion hole. *FEBS Lett* **208**, 301-304.

Drury, E. J., Bazar, L. S., and MacKenzie, R. E. (1975). Formiminotransferase-cyclodeaminase from procine liver. Purification and physical properties of the enzyme complex. *Arch Biochem Biophys* **169**, 662-668.

Ducruix, A., and Giege, R. (1992). Crystallization of Nucleic Acids and Protein: A Practical Approach (Oxford[England]; New York: IRL Press at Oxford University Press).

Ellison, R. T., Giehl, T. J., and LaForce, F. M. (1988). Damage of the outer membrane of gram-negitive bacteria by lactoferrin and transferrin. *Infect. Immun.* **56**, 2774-2781.

Engh, R. A., and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst* **A47**, 392-400.

Evans, H. J., and Sorger, G. J. (1966). Role of mineral elements with emphasis on the univalent cations. *Annu Rev Plant Physiol* **17**, 47-76.

Faber, H. R., Baker, C. J., Day, C. L., Tweedie, J. W., and Baker, E. N. (1996a). Mutation of arginine 121 in lactoferrin destabilizes iron binding by disruption of anion binding: Crystal structures of R121S and R121E mutants. *Biochemistry* **35**, 14473-14479.

Faber, H. R., Bland, T., Day, C. L., Norris, G. E., Tweedie, J. W., and Baker, E. N. (1996b). Altered domain closure and iron binding in transferrin: the crystal structure of the Asp60Ser mutant of the amino-terminal half-molecule of human lactoferrin. *J Mol Biol* **256**, 352-363.

Fairweather-Tait, S. J., Balmer, S. E., Scott, P. H., and Minski, M. J. (1987). Lactoferrin and iron absorption in newborn infants. *Pediatr Res* **22**, 651-654.

Farber, S., Diamond, L. K., Mercer, R. D., Sylvester, R. F. J., and Wolff, J. A. (1948). Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid (aminopterin). *New Engl J Med* **238**, 787-793.

Fetrow, J. S. (1995). Omega loops: nonregular secondary structures significant in protein function and stability. *FASEB J* **9**, 708-717.

Filman, D. J., Bolin, J. T., Matthews, D. A., and Kraut, J. (1982). Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7Å resolution. 2. Environment of bound NADPH and implications for catalysis. *J Biol Chem* **257**, 13663-13672.

Fisher, A. J., Smith, C. A., Thoden, J., Smith, R., Sutoh, K., Holden, H. M., and Rayment, I. (1995). X-ray Structures of the Myosin Motor Domain of *Dictyostellium discoideum* Complexed with MgADP·BeF_x and MgADP·AlF_4. *Biochemistry* **34**, 8960-8972.

Foo, S. K., and Shane, B. (1982). Regulation of folylpoly-gamma-glutamate synthesis in mammalian cells. In vivo and in vitro synthesis of pteroylpoly-gamma-glutamates by Chinese hamster ovary cells. *J Biol Chem* **257**, 13587-13592.

France, G. L., Marmer, D. J., and Steele, R. W. (1980). Breast-feeding and Salmonella infection. *Am J Dis Child* **134**, 147-152.

Freisheim, J. H., Price, E. M., and Ratnam, M. (1989). Folate coenzyme and antifolate transport proteins in normal and neoplastic cells. *Adv Enzyme Regul* **29**, 13-26.

French, S., and Wilson, K. (1978). On the treatment of negative intensity observations. *Acta Cryst* **A34**, 517-525.

Friedkin, M., Plante, L. T., Crawford, E. J., and Crumm, M. (1975). Inhibition of thymidylate synthetase and dihydrofolate reductase by naturally occurring oligoglutamate derivatives of folic acid. *J Biol Chem* **250**, 5614-5621.

Fry, D. W., Anderson, L. A., Borst, M., and Goldman, I. D. (1983). Analysis of the role of membrane transport and polyglutamation of methotrexate in gut and the Ehrlich tumor in vivo as factors in drug sensitivity and selectivity. *Cancer Res* **43**, 1087-1092.

Funk, W. D., MacGillivray, R. T., Mason, A. B., Brown, S. A., and Woodworth, R. C. (1990). Expression of the amino-terminal half-molecule of human serum transferrin in cultured cells and characterization of the recombinant protein. *Biochemistry* **29**, 1654-1660.

Fussenegger, M., and Meyer, T. F. (1996). Cloning and characterization of the *Neisseria gonorrhoeae* MS11 *folC* gene. *Mol Gen Genet* **250**, 277-285.

Gaertner, F. H. (1978). Unique catalytic properties of enzyme clusters. *Trends Biochem Sci* **3**, 63-65.

Galivan, J. (1979). Transport and metabolism of methotrexate in normal and resistant cultured rat hepatoma cells. *Cancer Res* **39**, 735-743.

Garratt, R. C., Evans, R. W., Hasnain, S. S., and Lindley, P. F. (1986). An extended X-ray absorption fine structure investigation of diferric transferrins and their iron-binding fragments. *Biochem J* **233**, 479-484.

Garrow, T. A., Admon, A., and Shane, B. (1992). Expression cloning of a human cDNA encoding folylpoly(gamma-glutamate) synthetase and determination of its primary structure. *Proc Natl Acad Sci USA* **89**, 9151-9155.

Gawthorne, J. M., and Smith, R. M. (1973). The synthesis of pteroylpolyglutamates by sheep liver enzymes in vitro. *Biochem J* **136**, 295-301.

George, S., Cichowicz, D. J., and Shane, B. (1987). Mammalian folylpoly-gamma-glutamate synthetase. 3. Specificity for folate analogues. *Biochemistry* **26**, 522-529.

Gerstein, M., Anderson, B. F., Norris, G. E., Baker, E. N., Lesk, A. M., and Chothia, C. (1993). Domain closure in lactoferrin. Two hinges produce a see-saw motion between alternative close-packed interfaces. *J Mol Biol* **234**, 357-372.

Gerstein, M., and Chothia, C. (1991). Analysis of protein loop closure. Two types of hinges produce one motion in lactate dehydrogenase. *J Mol Biol* **220**, 133-149.

Godwin, H. A., and Rosenberg, I. H. (1975). Comparative studies of the intestinal absorption of [^{3}H]pteroylmonoglutamate and [^{3}H]pteroylheptaglutamate in man. *Gastroenterology* **69**, 364-373.

Goker, E., Lin, J. T., Trippett, T., Elisseyeff, Y., Tong, W. P., Niedzwiecki, D., Tan, C., Steinherz, P., Schweitzer, B. I., and Bertino, J. R. (1993). Decreased polyglutamylation of methotrexate in acute lymphoblastic leukemia blasts in adults compared to children with this disease. *Leukemia* **7**, 1000-1004.

Goldman, I. D. (1971). The characteristics of the membrane transport of amethopterin and the naturally occurring folates. *Ann N Y Acad Sci* **186**, 400-422.

Goodman, R. E., and Schanbacher, F. L. (1991). Bovine lactoferrin mRNA: sequence, analysis, and expression in the mammary gland. *Biochem Biophys Res Commun* **180**, 75-84.

Gorinsky, B., Horsburgh, C., Lindley, P. F., Moss, D. S., Parkar, M., and Watson, J. L. (1979). Evidence for the bilobal nature of diferric rabbit plasma transferrin. *Nature* **281**, 157-158.

Griffiths, E., and Bullen, J. J. (1987). Iron binding proteins and host defence. In *Iron and Infection*, J. J. Bullen and E. Griffiths, eds. (Chichester, UK: Wiley), pp. 171-209.

Grossmann, J. G., Mason, A. B., Woodworth, R. C., Neu, M., Lindley, P. F., and Hasnain, S. S. (1993). Asp ligand provides the trigger for closure of transferrin molecules. Direct evidence from X-ray scattering studies of site-specific mutants of the N-terminal half-molecule of human transferrin. *J Mol Biol* **231**, 554-558.

Grossmann, J. G., Neu, M., Pantos, E., Schwab, F. J., Evans, R. W., Townes-Andrews, E., Lindley, P. F., Appel, H., Thies, W. G., and Hasnain, S. S. (1992). X-ray solution scattering reveals conformational changes upon iron uptake in lactoferrin, serum and ovo-transferrins. *J Mol Biol* **225**, 811-819.

Gutteridge, J. M., Halliwell, B., Treffry, A., Harrison, P. M., and Blake, D. (1983). Effect of ferritin-containing fractions with different iron loading on lipid peroxidation. *Biochem J* **209**, 557-560.

Habeck, L. L., Mendelsohn, L. G., Shih, C., Taylor, E. C., Colman, P. D., Gossett, L. S., Leitner, T. A., Schultz, R. M., Andis, S. L., and Moran, R. G. (1995). Substrate specificity of mammalian folylpolyglutamate synthetase for 5,10-dideaza-tetrahydrofolate analogs. *Mol Pharmacol* **48**, 326-333.

Haridas, M., Anderson, B. F., and Baker, E. N. (1995). Structure of Human Diferric Lactoferrin Refined at 2.2 Å Resolution. *Acta Cryst* **D51**, 629-646.

Harris, D. C., and Aisen, P. (1989). Physical biochemistry of the transferrins. In *Iron carriers and iron proteins*, T. M. Loehr, ed. (New York: U. C. H. Publishers), pp. 239-351.

Harris, W. R., Nesset-Tollefson, D., Stenback, J. Z., and Mohamed-Hani, N. (1990). Site selectivity in the binding of inorganic anions to serum transferrin. *J Inorg Biochem* **38**, 175-183.

Hashizume, S., Kuroda, K., and Murakami, H. (1983). Identification of lactoferrin as an essential growth factor for human lymphocytic cell lines in serum-free medium. *Biochim Biophys Acta* **763**, 377-382.

Hellman, S., Iannotti, A. T., and Bertino, J. R. (1964). Determination of the levels of serum folate in patients with carcinoma of the head and neck treated with methotrexate. *Cancer Res* **24**, 105-113.

Hendrickson, W. A., and Lattman, E. E. (1970). Representation of phase probability distributions for simplified combination of independent phase information. *Acta Cryst* **B26**, 136-143.

Himes, R. H., and Harmony, J. A. (1973). Formyltetrahydrofolate synthetase. *CRC Crit Rev Biochem* **1**, 501-535.

Hoffbrand, A. V., Tripp, E., Houlihan, C. M., and Scott, J. M. (1973). Studies on the uptake of synthetic conjugated folates by human marrow cells. *Blood* **42**, 141-146.

Hol, W. G., van Duijnen, P. T., and Berendsen, H. J. (1978). The alpha-helix dipole and the properties of proteins. *Nature* **273**, 443-446.

Hoshino, A., Hisayasu, S., and Shimada, T. (1996). Complete sequence analysis of rat transferrin and expression of transferrin but not lactoferrin in the digestive glands. *Comp Biochem Physiol B Biochem Mol Biol* **113**, 491-497.

Hovanessian, A. G., and Awdeh, Z. L. (1976). Gel isoelectric focusing of human-serum transferrin. *Eur J Biochem* **68**, 333-338.

Howie, P. W., Forsyth, J. S., Ogston, S. A., Clark, A., and du V Florey, C. (1990). Protective effect of breast feeding against infection. *Brit Med J* **300**, 11-16.

Huebers, H. A., Huebers, E., Finch, C., and Martin, A. W. (1982). Characterization of an invertebrate transferrin from the crab cancer magister (*Arthropoda*). *J Comp Physiol* **B148**, 101-109.

Ikeda, R. A., Bowman, B. H., Yang, F., and Lokey, L. K. (1992). Production of human serum transferrin in *Escherichia coli*. *Gene* **117**, 265-269.

Imeson, H. C., and Cossins, E. A. (1991). Higher plant folylpolyglutamate synthetase I. Purification, stability and reaction requirement of the enzyme from pea seedlings. *J Plant Physiol* **138**, 476-482.

Jackman, A. L., Marsham, P. R., Moran, R. G., Kimbell, R., O'Connor, B. M., Hughes, L. R., and Calvert, A. H. (1991a). Thymidylate synthase inhibitors: the in vitro activity of a series of heterocyclic benzoyl ring modified 2-desamino-2-methyl-N10-substituted-5,8-dideazafolates. *Adv Enzyme Regul* **31**, 13-27.

Jackman, A. L., Taylor, G. A., Gibson, W., Kimbell, R., Brown, M., Calvert, A. H., Judson, I. R., and Hughes, L. R. (1991b). ICI D1694, a quinazoline antifolate thymidylate synthase inhibitor that is a potent inhibitor of L1210 tumor cell growth in vitro and in vivo: a new agent for clinical study. *Cancer Res* **51**, 5579-5586.

Jamroz, R. C., Gasdaska, J. R., Bradfield, J. Y., and Law, J. H. (1993). Transferrin in a cockroach: molecular cloning, characterization, and suppression by juvenile hormone. *Proc Natl Acad Sci USA* **90**, 1320-1324.

Jancarik, J., and Kim, S.-H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J Appl Cryst* **24**, 409-411.

Jeltsch, J. M., and Chambon, P. (1982). The complete nucleotide sequence of the chicken ovotransferrin mRNA. *Eur J Biochem* **122**, 291-295.

Jeng, M. F., and Dyson, H. J. (1996). Direct measurement of the aspartic acid 26 pKa for reduced Escherichia coli thioredoxin by ^{13}C NMR. *Biochemistry* **35**, 1-6.

Jiang, J. S., and Brunger, A. T. (1994). Protein hydration observed by X-ray diffraction. Solvation properties of penicillopepsin and neuraminidase crystal structures. *J Mol Biol* **243**, 100-115.

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.

Jones, T. R., Calvert, A. H., Jackman, A. L., Brown, S. J., Jones, M., and Harrap, K. R. (1981). A potent antitumour quinazoline inhibitor of thymidylate synthetase: synthesis, biological properties and therapeutic results in mice. *Eur J Cancer* **17**, 11-19.

Joseph, D., Petsko, G. A., and Karplus, M. (1990). Anatomy of a conformational change: hinged "lid" motion of the triose phosphate isomerase loop. *Science* **249**, 1425-1428.

Joshi, M. D., Hedberg, A., and McIntosh, L. P. (1997). Complete measurement of the pKa values of the carboxyl and imidazole groups in *Bacillus circulans* xylanase. *Protein Sci* **6**, 2667-2670.

Kamen, B. A., Wang, M. T., Streckfuss, A. J., Peryea, X., and Anderson, R. G. (1988). Delivery of folates to the cytoplasm of MA104 cells is mediated by a surface membrane receptor that recycles. *J Biol Chem* **263**, 13602-13609.

Karplus, P. A., and Faerman, C. (1994). Ordered water in macromolecular structure. *Curr Opin Struct Biol* **4**, 770-776.

Katz, B. A. (1997). Binding of biotin to streptavidin stabilizes intersubunit salt bridges between Asp61 and His87 at low pH. *J Mol Biol* **274**, 776-800.

Kawakami, H., and Lonnerdal, B. (1991). Isolation and function of a receptor for human lactoferrin in human fetal intestinal brush-border membranes. *Am J Physiol* **261**, G841-846.

Kilar, F., and Simon, I. (1985). The effect of iron binding on the conformation of transferrin. A small angle x-ray scattering study. *Biophys J* **48**, 799-802.

Kim, J. S., Lowe, K. E., and Shane, B. (1993). Regulation of folate and one-carbon metabolism in mammalian cells. IV. Role of folylpoly-gamma-glutamate synthetase in methotrexate metabolism and cytotoxicity. *J Biol Chem* **268**, 21680-21685.

Kisliuk, R. L., Gaumont, Y., Lafer, E., Baugh, C. M., and Montgomery, J. A. (1981). Polyglutamyl derivatives of tetrahydrofolate as substrates for *Lactobacillus casei* thymidylate synthase. *Biochemistry* **20**, 929-934.

Kjeldgaard, M., Nissen, P., Thirup, S., and Nyborg, J. (1993). The Crystal Structure of Elongation Factor EF-Tu from *Thermus aquaticus* in the GTP Conformation. *Structure* **1**, 35-50.

Kleywegt, G. J., and Brunger, A. T. (1996). Checking your imagination: applications of the free R value. *Structure* **4**, 897-904.

Kleywegt, G. J., and Jones, T. A. (1997). Detecting folding motifs and similarities in protein structures. *Methods in Enzymology* **277**, 525-545.

Kleywegt, G. J., and Read, R. J. (1997). Not your average density. *Structure* **5**, 1557-1569.

Kneller, D. G., Cohen, F. E., and Langridge, R. (1990). Improvements in protein secondary structure prediction by an enhanced neural network. *J Mol Biol* **214**, 171-182.

Knowles, J. R. (1980). Enzyme-catalyzed phosphoryl transfer reactions. *Annu Rev Biochem* **49**, 877-919.

Koenig, S. H., and Schillinger, W. E. (1969). Nuclear magnetic relaxation dispersion in protein solutions. II. Transferrin. *J Biol Chem* **244**, 6520-6526.

Kostrewa, D. (1997). Bulk solvent correction: practical application and effects in reciprocal and real space. In *The CCP4 newsletter on protein crystallography*, pp. 9-22.

Kraulis, P. J. (1991). Molscript: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* **24**, 946-950.

Kraut, J., Sieker, L. C., High, D. F., and Freer, S. T. (1962). Chymotrypsinogen: A three dimensional Fourier synthesis at 5 Å resolution. *Proc Natl Acad Sci USA* **48**, 1417-1424.

Krebs, H. A., Hems, R., and Tyler, B. (1976). The regulation of folate and methionine metabolism. *Biochem J* **158**, 341-353.

Krumdieck, C. L., Tamura, T., and Eto, I. (1983). Synthesis and analysis of the pteroylpolyglutamates. *Vitamins and Hormones* **40**, 45-104.

Kull, J. F., Sablin, E. P., Lau, R., Fletterick, R. J., and Vale, R. D. (1996). Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature* **380**, 550-555.

Kvingedal, A. M., Rorvik, K. A., and Alestrom, P. (1993). Cloning and characterization of Atlantic salmon (*Salmo salar*) serum transferrin cDNA. *Mol Mar Biol Biotechnol* **2**, 233-238.

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.

Lawrence, M. C. (1991). The application of the molecular replacement method to the de novo determination of protein structure. *Quart Rev Biophys* **24**, 399-424.

Le Provost, F., Nocart, M., Guerin, G., and Martin, P. (1994). Characterization of the goat lactoferrin cDNA: assignment of the relevant locus to bovine U12 synteny group. *Biochem Biophys Res Commun* **203**, 1324-1332.

Lee, M. Y., Heubers, H., Martin, A. W., and Finch, C. A. (1978). Iron Metabolism in a Spider, *Dugesiella hentzi. J Comp Physiol* **127**, 349-354.

Legrand, D., Mazurier, J., Aubert, J. P., Loucheux-Lefebvre, M. H., Montreuil, J., and Spik, G. (1986). Evidence for interactions between the 30 kDa N- and 50 kDa C-terminal tryptic fragments of human lactotransferrin. *Biochem J* **236**, 839-844.

Legrand, D., Mazurier, J., Metz-Boutigue, M.-H., Jolles, J., Jolles, P., Montreuil, J., and Spik, G. (1984). Characterization and localization of an iron-binding 18 kDa glycopeptide isolated from the N-terminal half of human lactoferrin. *Biochim Biopys.Acta* **787**, 90-96.

Lenz, H. J., Danenberg, K., Schnieders, B., Goeker, E., Peters, G. J., Garrow, T., Shane, B., Bertino, J. R., and Danenberg, P. V. (1994). Quantitative analysis of folylpolyglutamate synthetase gene expression in tumor tissues by the polymerase chain reaction: marked variation of expression among leukemia patients. *Oncol Res* **6**, 329-335.

Levitt, M., and Park, B. H. (1993). Water: now you see it, now you don't. *Structure* **1**, 223-226.

Li, W. W., Lin, J. T., Tong, W. P., Trippett, T. M., Brennan, M. F., and Bertino, J. R. (1992). Mechanisms of natural resistance to antifolates in human soft tissue sarcomas. *Cancer Res* **52**, 1434-1438.

Liang, Q., and Richardson, T. (1993). Expression and characterization of human lactoferrin in yeast *Saccharomyces cerevisiae. J Agric Food Chem* **41**, 1800-1807.

Lin, B. F., Kim, J. S., Hsu, J. C., Osborne, C., Lowe, K., Garrow, T., and Shane, B. (1996). Molecular biology in nutrition research: modeling of folate metabolism. *Adv Food Nutr Res* **40**, 95-106.

Lindley, P. F., Bajaj, M., Evans, R. W., Garrett, R. C., Hasnain, S. S., Jhoti, H., Kuser, P., Neu, M., Patel, K., Sarra, R., Strange, R., and Walton, A. (1993). The mechanism of iron uptake by transferrin: the structure of an 18 kDa NII-domain fragment from duck ovotransferrin at 2.3 Å resolution. *Acta Cryst* **D49**, 292-304.

Line, W. F., Sly, D. A., and Bezkrovainy, A. (1976). Limited cleavage of human lactoferrin with pepsin. *Int J Biochem* **7**, 203-208.

Lonnerdal, B. (1985). Biochemistry and physiological function of human milk proteins. *Am J Clin Nutr* **42**, 1299-1317.

Lu, K., Yin, M. B., McGuire, J. J., Bonmassar, E., and Rustum, Y. M. (1995). Mechanisms of resistance to N-[5-[N-(3,4-dihydro-2-methyl-4- oxoquinazolin-6-methyl)-N-methylamino]-2-thenoyl]-L-glutamic acid (ZD1694), a folate-based thymidylate synthase inhibitor, in the HCT-8 human ileocecal adenocarcinoma cell line. *Biochem Pharmacol* **50**, 391-398.

Lunin, V. Y. (1993). Electron-density histograms and the phase problem. *Acta Cryst* **D49**, 90-99.

Lunin, V. Y., and Skovoroda, T. P. (1995). R-free likelihood-based estimates of errors for phases calculated from atomic models. *Acta Cryst* **A51**, 880-887.

Luzzati, V. (1952). Traitement statistiques des erreurs dans la determination des structures cristallines. *Acta Cryst* **5**, 802-810.

Lydon, J. P., O'Malley, B. R., Saucedo, O., Lee, T., Headon, D. R., and Conneely, O. M. (1992). Nucleotide and primary amino acid sequence of porcine lactoferrin. *Biochim Biophys Acta* **1132**, 97-99.

MacGillivray, R. T., Mendez, E., Sinha, S. K., Sutton, M. R., Lineback-Zins, J., and Brew, K. (1982). The complete amino acid sequence of human serum transferrin. *Proc Natl Acad Sci USA* **79**, 2504-2508.

MacKenzie, R. E., Aldridge, M., and Paquin, J. (1980). The bifunctional enzyme formiminotransferase-cyclodeaminase is a tetramer of dimers. *J Biol Chem* **255**, 9474-9478.

Mackenzie, R. E., and Baugh, C. M. (1980). Tetrahydropterolypolyglutamate derivatives as substrates of two multifunctional proteins with folate-dependent enzyme activities. *Biochim Biophys Acta* **611**, 187-195.

Malmquist, J. (1978). Lactoferrin in haematology. *Scand J Haematol* **21**, 5-8.

Mason, A. B., Miller, M. K., Funk, W. D., Banfield, D. K., Savage, K. J., Oliver, R. W., Green, B. N., MacGillivray, R. T., and Woodworth, R. C. (1993). Expression of glycosylated and nonglycosylated human transferrin in mammalian cells. Characterization of the recombinant proteins with comparison to three commercially available transferrins. *Biochemistry* **32**, 5472-5479.

Mason, S. (1962). Some aspects of gastric function in the newborn. *Arch Dis Child* **37**, 387-391.

Masson, P. L., Heremans, J. F., and Dive, C. (1966). An iron-binding protein common to many external secretions. *Clin Chim Acta* **14**, 735-739.

Masurekar, M., and Brown, G. M. (1975). Partial purification and properties of an enzyme from *Escherichia coli* that catalyzes the conversion of glutamic acid and

10-formyltetrahydropteroylglutamic acid to 10-formyltetrahydropterol-gamma-glutamy-glutamic acid. *Biochemistry* **14**, 2424-2430.

Matte, A., Tari, L. W., and Delbaere, L. T. (1998). How do kinases transfer phosphoryl groups? *Structure* **6**, 413-419.

Matthews, B. W. (1968). Solvent contents of protein crystals. *J Mol Biol* **33**, 491-497.

Matthews, R. G. (1984). Methionine biosynthesis. In *Folates and pterins*, R. L. Blakley and S. J. Benkovic, eds. (New York: Wiley), pp. 497-554.

Matthews, R. G., and Baugh, C. M. (1980). Interactions of pig liver methylenetetrahydrofolate reductase with methylenetetrahydropteroylpolyglutamate substrates and with dihydropteroylpolyglutamate inhibitors. *Biochemistry* **19**, 2040-2045.

Matthews, R. G., Ghose, C., Green, J. M., Matthews, K. D., and Dunlap, R. B. (1987). Folylpolyglutamates as substrates and inhibitors of folate-dependent enzymes. *Adv Enzyme Regul* **26**, 157-171.

Matthews, R. G., and Haywood, B. J. (1979). Inhibition of pig liver methylenetetrahydrofolate reductase by dihydrofolate: Some mechanistic and regulatory implications. *Biochemistry* **18**, 4845-4851.

Matthews, R. G., Ross, J., Baugh, C. M., Cook, J. D., and Davis, L. (1982). Interactions of pig liver serine hydroxymethyltransferase with methyltetrahydropteroyl-polyglutamate inhibitors and with tetrahydropteroylpolyglutamate substrates. *Biochemistry* **21**, 1230-1238.

Mazurier, J., Legrand, D., Hu, W. L., Montreuil, J., and Spik, G. (1989). Expression of human lactotransferrin receptors in phytohemagglutinin-stimulated human peripheral blood lymphocytes. Isolation of the receptors by antiligand-affinity chromatography. *Eur J Biochem* **179**, 481-487.

Mazurier, J., Montreuil, J., and Spik, G. (1985). Visualization of lactotransferrin brush-border receptors by ligand- blotting. *Biochim Biophys Acta* **821**, 453-460.

McBurney, M. W., and Whitmore, G. F. (1974). Isolation and biochemical characterization of folate deficient mutants of Chinese hamster cells. *Cell* **2**, 173-182.

McCloskey, D. E., McGuire, J. J., Russell, C. A., Rowan, B. G., Bertino, J. R., Pizzorno, G., and Mini, E. (1991). Decreased folylpolyglutamate synthetase activity as a mechanism of methotrexate resistance in CCRF-CEM human leukemia sublines. *J Biol Chem* **266**, 6181-6187.

McGuire, J. J., and Bertino, J. R. (1981). Enzymatic synthesis and function of folylpolyglutamates. *Mol Cell Biochem* **38 Spec No**, 19-48.

McGuire, J. J., and Coward, J. K. (1984). Pteroylpolyglutamates: Biosynthesis, degradation and function. In *Folates and Pterins*, R. L. Blakley and S. J. Benkovic, eds. (New York: Wiley Interscience), pp. 135-190.

McGuire, J. J., and Coward, J. K. (1985). DL-threo-4-fluoroglutamic acid. A chain-terminating inhibitor of folylpolyglutamate synthesis. *J Biol Chem* **260**, 6747-6754.

McGuire, J. J., Hart, B. P., Haile, W. H., Rhee, M. S., Galivan, J., and Coward, J. K. (1995). DL-beta,beta-difluoroglutamic acid mediates position-dependent enhancement or termination of pteroylpoly(gamma-glutamate) synthesis catalyzed by folylpoly-glutamate synthetase. *Arch Biochem Biophys* **321**, 319-328.

McGuire, J. J., Hsieh, P., Coward, J. K., and Bertino, J. R. (1980). Enzymatic synthesis of folylpolyglutamates. Characterization of the reaction and its products. *J Biol Chem* **255**, 5776-5788.

McGuire, J. J., Kitamoto, Y., Hsieh, P., Coward, J. K., and Bertino, J. R. (1979). Characterization of mammalian folylpolyglutamate synthetases. *Dev Biochem* **4,** 471-476.

McPherson, A. (1982). *Preparation and analysis of protein crystals* (New York: Wiley).

Mead, P. E., and Tweedie, J. W. (1990). cDNA and protein sequence of bovine lactoferrin. *Nucleic Acids Res* **18**, 7167.

Meister, A. (1978). Inhibition of glutamine synthetase and γ-glutamylcysteine synthetase by methionine sulfoximine and related compounds. In *Enzyme-Activated Irreversible Inhibitiors*, N. Seiler, M. J. Jung and J. Koch-Wester, eds. (Amsterdam: Elsevier), pp. 187-210.

Metz-Boutigue, M. H., Jolles, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G., Montreuil, J., and Jolles, P. (1984). Human lactotransferrin: amino acid sequence and structural comparisons with other transferrins. *Eur J Biochem* **145**, 659-676.

Miller, D. L., and Westheimer, F. H. (1966). The hydrolysis of γ-phenylpropyl di- and triphosphates. *J Am Chem Soc* **88**, 1507-1511.

Mills, J. L., Scott, J. M., Kirke, P. N., McPartlin, J. M., Conley, M. R., Weir, D. G., Molloy, A. M., and Lee, Y. J. (1996). Homocysteine and neural tube defects. *J Nutr* **126**, 756s-760s.

Milner-White, E. J., Coggins, J. R., and Anton, I. A. (1991). Evidence for an ancestral core structure in nucleotide-binding proteins with the type A motif. *J Mol Biol* **221**, 751-754.

Moews, P. C., and Kretsinger, R. H. (1975). Refinement of the structure of carp muscle calcium-binding parvalbumin by model building and difference Fourier analysis. *J Mol Biol* **91**, 201-225.

Moguilevsky, N., Retegui, L. A., and Masson, P. L. (1985). Comparison of human lactoferrins from milk and neutrophilic leucocytes. Relative molecular mass, isoelectric point, iron-binding properties and uptake by the liver. *Biochem J* **229**, 353-359.

Molloy, A. L., and Winterbourn, C. C. (1990). Release of iron from phagocytosed *Escherichia coli* and uptake by neutrophil lactoferrin. *Blood* **75**, 984-989.

Montreuil, J., Tonnelat, J., and Mullet, S. (1960). Preparation et proprietes de la lactosiderophiline (lactotransferrine) du lait de Femme. *Biochim Biophys Acta* **45**, 413-421.

Moore, S. A., Anderson, B. F., Groom, C. R., Haridas, M., and Baker, E. N. (1997). Three-dimensional structure of diferric bovine lactoferrin at 2.8 Å resolution. *J Mol Biol* **274**, 222-236.

Moran, R. G., and Colman, P. D. (1984a). Mammalian folyl polyglutamate synthetase: partial purification and properties of the mouse liver enzyme. *Biochemistry* **23**, 4580-4589.

Moran, R. G., and Colman, P. D. (1984b). Measurement of folylpolyglutamate synthetase in mammalian tissues. *Anal Biochem* **140**, 326-342.

Moran, R. G., Colman, P. D., Forsch, R. A., and Rosowsky, A. (1984). A mechanism for the addition of multiple moles of glutamate by folylpolyglutamate synthetase. *J Med Chem* **27**, 1263-1267.

Moran, R. G., Colman, P. D., and Rosowsky, A. (1987). Structural requirements for the activity of antifolates as substrates for mammalian folylpolyglutamate synthetase. *NCI Monogr*, 133-138.

Moran, R. G., Colman, P. D., Rosowsky, A., Forsch, R. A., and Chan, K. K. (1985). Structural features of 4-amino antifolates required for substrate activity with mammalian folylpolyglutamate synthetase. *Mol Pharmacol* **27**, 156-166.

Moskaitis, J. E., Pastori, R. L., and Schoenberg, D. R. (1990). The nucleotide sequence of *Xenopus laevis* transferrin mRNA. *Nucleic Acids Res* **18**, 6135.

Muller, C. W., and Schulz, G. E. (1992). Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9Å resolution. *J Mol Biol* **224**, 159-177.

Muller-Dieckmann, H.-J., and Schulz, G. E. (1994). The structure of uridylate kinase with its substrates showing the transition-state geometry. *J Mol Biol* **236**, 361-367.

Nahas, A., Nixon, P. F., and Bertino, J. R. (1972). Uptake and metabolism of N 5 - formyltetrahydrofolate by L1210 leukemia cells. *Cancer Res* **32**, 1416-1421.

Naidu, A. S., and Arnold, R. R. (1997). Influence of Lactoferrin on Host-Microbe Interactions. In *Lactoferrin: Interactions and Biological Functions*, T. W. Hutchens and B. Lonnerdal, eds. (Totowa, New Jersey: Humana Press Inc), pp. 259-275.

Nakamura, K., and Kozloff, L. M. (1978). Folate polyglutamates in T4D bacteriophage and T4D-infected Escherichia coli. *Biochim Biophys Acta* **540**, 313-319.

Navaza, J. (1994). AMoRe: An automated package for molecular replacement. *Acta Cryst* **A50**, 157-163.

Nave, C. (1995). Radiation damage in protein crystallography. *Radiation Physics and Chemistry* **45**, 483-490.

Nicholls, A., Sharp, K. A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.

Niemeyer, C. M., Hitchcock-Bryan, S., and Sallan, S. E. (1985). Comparative analysis of treatment programs for childhood acute lymphoblastic leukemia. *Semin Oncol* **12**, 122-130.

Nuijens, J. H., Van Berkel, P. H. C., and Schanbacher, F. L. (1996). *J Mammary Gland Biol Neoplasia* **1**, 285-295.

Octave, J.-N., Schneider, Y.-J., Trout, A., and Crichton, R. R. (1983). Iron uptake and utilization by mammalian cells. I: Cellular uptake of transferrin and iron. *Trends Biochem Sci* **8**, 217-220.

Oram, J. D., and Reiter, B. (1968). Inhibition of bacteria by lactoferrin and other iron-chelating agents. *Biochim Biophys Acta* **170**, 351-365.

Oseas, R., Yang, H. H., Baehner, R. L., and Boxer, L. A. (1981). Lactoferrin: a promoter of polymorphonuclear leukocyte adhesiveness. *Blood* **57**, 939-945.

Otting, G., Liepinsh, E., and Wuthrich, K. (1991). Protein hydration in aqueous solution. *Science* **254**, 974-980.

Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Proceedings of the CCP4 study weekend* (Warrington, UK: SERC Daresbury laboratory), pp. 80-86.

Otwinowski, Z. (1993). Oscillation data reduction program. In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, L. Sawyer, N. Isaacs and S. Bailey, eds. (Daresbury, England: SERC Daresbury Laboratory), pp. 56-62.

Pai, E. F., Krengel, U., Petsko, G. A., Goody, R. S., Kabsch, W., and Wittinghofer, A. (1990). Refined structure of the triphosphate conformation of H-*ras*-p21 at 1.35 Å resolution: Implications for the mechanism of GTP hydrolysis. *EMBO J* **9**, 2351-2359.

Paquin, J., Baugh, C. M., and MacKenzie, R. E. (1985). Channeling between the active sites of formiminotransferase-cyclodeaminase. Binding and kinetic studies. *J Biol Chem* **260**, 14925-14931.

Park, I., Schaeffer, E., Sidoli, A., Baralle, F. E., Cohen, G. N., and Zakin, M. M. (1985). Organization of the human transferrin gene: direct evidence that it originated by gene duplication. *Proc Natl Acad Sci USA* **82**, 3149-3153.

Pentecost, B. T., and Teng, C. T. (1987). Lactotransferrin is the major estrogen inducible protein of mouse uterine secretions. *J Biol Chem* **262**, 10134-10139.

Petsko, G. A. (1985). Preparation of isomorphous heavy-atom derivatives. *Methods Enzymol* **114**, 147-156.

Phillips, D. C. (1966). Advances in protein crystallography. *Advances in Structural Research by Diffraction Methods* **2**, 75-140.

Pizzorno, G., Sokoloski, J. A., Cashmore, A. R., Moroson, B. A., Cross, A. D., and Beardsley, G. P. (1991). Intracellular metabolism of 5,10-dideazatetrahydrofolic acid in human leukemia cell lines. *Mol Pharmacol* **39**, 85-89.

Prentice, A., Ewing, G., Roberts, S. B., Lucas, A., MacCarthy, A., Jarjou, L. M., and Whitehead, R. G. (1987). The nutritional role of breast-milk IgA and lactoferrin. *Acta Paediatr Scand* **76**, 592-598.

Prieels, J. P., Pizzo, S. V., Glasgow, L. R., Paulson, J. C., and Hill, R. L. (1978). Hepatic receptor that specifically binds oligosaccharides containing fucosyl alpha1 leads to 3 N-acetylglucosamine linkages. *Proc Natl Acad Sci USA* **75**, 2215-2219.

Pristupa, Z. B., Vickers, P. J., Sephton, G. B., and Scrimgeour, K. G. (1984). Folypolyglutamate synthetase from beef liver: assay, stabilization, and characterization. *Can J Biochem Cell Biol* **62**, 495-506.

Rado, T. A., Wei, X. P., and Benz, E. J., Jr. (1987). Isolation of lactoferrin cDNA from a human myeloid library and expression of mRNA during normal and leukemic myelopoiesis. *Blood* **70**, 989-993.

Ramakrishnan, C., and Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformation. *Biophys J* **5**, 909-933.

Ray, M. S., Muggia, F. M., Leichman, C. G., Grunberg, S. M., Nelson, R. L., Dyke, R. W., and Moran, R. G. (1993). Phase I study of (6R)-5,10-dideazatetrahydrofolate: a folate antimetabolite inhibitory to de novo purine synthesis. *J Natl Cancer Inst* **85**, 1154-1159.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst* **A42**, 140-149.

Read, R. J. (1994). Model bias and phase combination. In *Proceedings of the CCP4 study weekend*, S. Bailey, R. Hubbard and D. Waller, eds. (Warrington, UK: EPSRC Daresbury laboratory), pp. 31-40.

Reiter, B. (1983). The biological significance of lactoferrin. *Int J Tissue React* **5**, 87-96.

Retzer, M. D., Kabani, A., Button, L. L., Yu, R. H., and Schryvers, A. B. (1996). Production and characterization of chimeric transferrins for the determination of the binding domains for bacterial transferrin receptors. *J Biol Chem* **271**, 1166-1173.

Rice, D. (1981). The use of phase combination in the refinement of phosphoglycerate kinase at 2.5 Å resolution. *Acta Cryst* **A37**, 491-500.

Rice, D., Anderson, B. F., and Baker, E. N. (1988). A practical guide to the use of partial structural phase combination. In *Proceedings of the CCP4 study weekend*, S. Bailey, E. Dodson and S. Phillips, eds. (Warrington, UK: SERC Daresbury laboratory), pp. 113-120.

Roiron, D., Amouric, M., Marvaldi, J., and Figarella, C. (1989). Lactoferrin-binding sites at the surface of HT29-D4 cells. Comparison with transferrin. *Eur J Biochem* **186**, 367-373.

Rose, T. M., Plowman, G. D., Teplow, D. B., Dreyer, W. J., Hellstrom, K. E., and Brown, J. P. (1986). Primary structure of the human melanoma-associated antigen p97 (melanotransferrin) deduced from the mRNA sequence. *Proc Natl Acad Sci USA* **83**, 1261-1265.

Ross, C. A., and Selawry, O. S. (1966). *Proc Am Assoc Cancer Res* **7**, 60.

Rossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H., and Huber, R. (1993). Phosphotransferase and substrate binding mechanism of the cAMP- dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 Å structure of the complex with Mn^{2+} adenylyl imidodiphosphate and inhibitor peptide PKI(5-24). *EMBO J* **12**, 849-859.

Rossmann, M. G. (1961). The position of anomalous scatterers in protein crystals. *Acta Cryst* **14**, 383-388.

Rossmann, M. G., and Blow, D. M. (1961). The refinement of structures partially determined by the isomorphous replacement method. *Acta Cryst* **14**, 641-647.

Rumberger, B. G., Barrueco, J. R., and Sirotnak, F. M. (1990). Differing specificities for 4-aminofolate analogues of folylpolyglutamyl synthetase from tumors and proliferative intestinal epithelium of the mouse with significance for selective antitumor action. *Cancer Res* **50**, 4639-4643.

Saarinen, U. M., Siimes, M. A., and Dallman, P. R. (1977). Iron absorption in infants: high bioavailability of breast milk iron as indicated by the extrinsic tag method of iron absorption and by the concentration of serum ferritin. *J Pediatr* **91**, 36-39.

Sakami, W., Ritari, S. J., Black, C. W., and Rzepka, J. (1973). Polyglutamate synthesis by *Neurospora crassa*. *Fed Proc* **32**, 471.

Salem, A. R., and Foster, M. A. (1972). The microbial biosynthesis of methionine. *Biochem J* **127**, 845-853.

Salem, A. R., Pattison, J. R., and Foster, M. A. (1972). Folic acid and the methylation of homocysteine by Bacillus subtilis. *Biochem J* **126**, 993-1004.

Salmon, V., Legrand, D., Georges, B., Slomianny, M. C., Coddeville, B., and Spik, G. (1997). Characterization of human lactoferrin produced in the baculovirus expression system. *Protein Expr Purif* **9**, 203-210.

Sanchez, L., Calvo, M., and Brock, J. H. (1992). Biological role of lactoferrin. *Arch Dis Child* **67**, 657-661.

Sarra, R., Garratt, R., Gorinsky, B., Jhoti, H., and Lindley, P. F. (1990). High-resolution X-ray studies on rabbit serum transferrin: preliminary structure analysis of the N-terminal half-molecule at 2.3 Å resolution. *Acta Cryst* **B46**, 763-771.

Sato, M., Yamamoto, M., Imada, K., Katsube, Y., Tanaka, N., and Higashi, T. (1992). A high-speed data-collection system for large-unit-cell crystals using an imaging plate as a detector. *J Appl Cryst* **25**, 348-357.

Schirch, V., and Strong, W. B. (1989). Interaction of folylpolyglutamates with enzymes in one-carbon metabolism. *Arch Biochem Biophys* **269**, 371-380.

Schlichting, I., and Reinstein, J. (1997). Structures of active conformations of UMP kinase from *Dictyostelium discoideum* suggest phosphoryl transfer is associative. *Biochemistry* **36**, 9290-9296.

Schreiber, G., Dryburgh, H., Millership, A., Matsuda, Y., Inglis, A., Phillips, J., Edwards, K., and Maggs, J. (1979). The synthesis and secretion of rat transferrin. *J Biol Chem* **254**, 12013-12019.

Schulz, G. E. (1992). Binding of nucleotides by proteins. *Curr Opin Struct Biol* **2**, 61-67.

Schulz-Lell, G., Dorner, K., Oldigs, H. D., Sievers, E., and Schaub, J. (1991). Iron availability from an infant formula supplemented with bovine lactoferrin. *Acta Paediatr Scand* **80**, 155-158.

Shane, B. (1980a). Pteroylpoly(gamma-glutamate) synthesis by Corynebacterium species. In vivo synthesis of folates. *J Biol Chem* **255**, 5649-5654.

Shane, B. (1980b). Pteroylpoly(gamma-glutamate) synthesis by Corynebacterium species. Purification and properties of folypoly(gamma-glutamate) synthetase. *J Biol Chem* **255**, 5655-5662.

Shane, B. (1980c). Pteroylpoly(γ-glutamate) synthesis by Corynebacterium species. Studies on the mechanism of folylpoly(γ-glutamate) synthetase. *J Biol Chem* **255**, 5663-5667.

Shane, B. (1983). Properties of Corynebacterium species dihydrofolate synthetase-folylpolyglutamate synthetase. In *Chemistry and biology of pteridines*, J. A. Blair, ed. (Berlin: de Gruyter), pp. 621-626.

Shane, B. (1989). Folylpolyglutamate synthesis and role in the regulation of one-carbon metabolism. *Vitamins and Hormones* **45**, 263-335.

Shane, B., and Cichowicz, D. J. (1983). Folylpoly-gamma-glutamate synthetases: properties and regulation. *Adv Exp Med Biol* **163**, 149-165.

Shane, B., Garrow, T., Brenner, A., Chen, L., Choi, Y. J., Hsu, J. C., and Stover, P. (1993). Folylpoly-gamma-glutamate synthetase. *Adv Exp Med Biol* **338**, 629-634.

Shane, B., and Stokstad, E. L. (1975). Transport and metabolism of folates in bacteria. *J Biol Chem* **250**, 2243-2253.

Shane, B., and Stokstad, E. L. (1976). Transport and utilization of methyl-tetrahydrofolates by *Lactobacillus casei*. *J Biol Chem* **251**, 3405-3410.

Shane, B., and Stokstad, E. L. (1985). Vitamin B12-folate interrelationships. *Annu Rev Nutr* **5**, 115-141.

Sheldrick, G. M., and Schneider, T. R. (1997). SHELXL: High-resolution refinement. *Methods in Enzymology* **277**, 319-343.

Shen, H., Yao, B. Y., and Mueller, D. M. (1994). Primary structural constraints of P-loop of mitochondrial F1-ATPase from yeast. *J Biol Chem* **269**, 9424-9428.

Shirakihara, Y., and Evans, P. R. (1988). Crystal structure of the complex of phosphofructokinase from *Escherichia coli* with its reaction products. *J Mol Biol* **204**, 973-994.

Shongwe, M. S., Smith, C. A., Ainscough, E. W., Baker, H. M., Brodie, A. M., and Baker, E. N. (1992). Anion binding by human lactoferrin: Results from crystallographic and physicochemical studies. *Biochemistry* **31**, 4451-4458.

Sibanda, B. L., Blundell, T. L., and Thornton, J. M. (1989). Conformation of beta-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J Mol Biol* **206**, 759-777.

Sim, G. A. (1959). The distribution of phase angles for structures containing heavy atoms. II. A modification of the normal heavy-atom method for non-centrosymmetrical structures. *Acta Cryst* **12**, 813-815.

Sim, G. A. (1960). A note on the heavy-atom method. *Acta Cryst* **13**, 511-512.

Sirotnak, F. M., DeGraw, J. I., Moccio, D. M., Samuels, L. L., and Goutas, L. J. (1984). New folate analogs of the 10-deaza-aminopterin series. Basis for structural design and biochemical and pharmacologic properties. *Cancer Chemother Pharmacol* **12**, 18-25.

Smith, C. A., Anderson, B. F., Baker, H. M., and Baker, E. N. (1992). Metal substitution in transferrins: the crystal structure of human copper-lactoferrin at 2.1 Å resolution. *Biochemistry* **31**, 4527-4533.

Smith, C. A., Anderson, B. F., Baker, H. M., and Baker, E. N. (1994). Structure of copper- and oxalate-substituted human lactoferrin at 2.0 Å resolution. *Acta Cryst* **D50**, 302-316.

Smith, C. A., and Rayment, I. (1995). X-ray structure of the magnesium(II)-pyrophosphate complex of the truncated head of *Dictyostelium discoideum* myosin to 2.7 Å resolution. *Biochemistry* **34**, 8973-8981.

Smith, C. A., and Rayment, I. (1996). Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biophys J* **70**, 1590-1602.

Smith, R. F., Wiese, B. A., Wojzynski, M. K., Davison, D. B., and Worley, K. C. (1996). BCM Search Launcher – an integrated interface to molecular biology data base search and analysis services available on the World Wide Web. *Genome Res* **6**, 454-462.

Spik, G., Coddeville, B., and Montreuil, J. (1988). Comparative study of the primary structures of sero-, lacto- and ovotransferrin glycans from different species. *Biochemistry* **70**, 1459-1469.

Spinella, M. J., Brigle, K. E., and GoldmanD (1996). Molecular cloning of murine folylpoly-gamma-glutamate synthetase. *Biochim Biophys Acta* **1305**, 11-14.

Steinlein, L. M., and Ikeda, R. A. (1993). Production of N-terminal and C-terminal human serum transferrin in *Escherichia coli. Enzyme Microb Technol* **15**, 193-199.

Stowell, K. M., Rado, T. A., Funk, W. D., and Tweedie, J. W. (1991). Expression of cloned human lactoferrin in baby-hamster kidney cells. *Biochem J* **276**, 349-355.

Stryer, L., Kendrew, J. C., and Watson, H. C. (1964).The mode of attachment of the azide ion to sperm whale metmyoglobin. *J Mol Biol* **8**, 96-104.

Suelter, C. H. (1970). Enzymes activated by monovalent cations. *Science* **168**, 789-795.

Synold, T. W., Willits, E. M., and Barredo, J. C. (1996). Role of folylpolyglutamate synthetase (FPGS) in antifolate chemotherapy; a biochemical and clinical update. *Leukemia and Lymphoma* **21**, 9-15.

Tang, K.-C., and Coward, J. K. (1983). Synthesis of acyl phosphonate analogues of biologically important acyl phosphates: N-(2-amino-10-methylpteroyl)-5-amino-2-oxopentanephosphonic acid. *J Org Chem* **48**, 5001-5006.

Tange, N., Jong-Young, L., Mikawa, N., Hirono, I., and Aoki, T. (1997). Cloning and characterization of transferrin cDNA and rapid detection of transferrin gene polymorphism in rainbow trout (*Oncorhynchus mykiss*). *Mol Mar Biol Biotechnol* **6**, 351-356.

Tanner, M. E., Vaganay, S., Van Heijenoort, J., and Blanot, D. (1996). Phosphinate inhibitors of the D-glutamic acid-adding enzyme of peptidoglycan biosynthesis. *J Org Chem* **61**, 1756-1760.

Taylor, R. T., and Hanna, M. L. (1975). Folate-dependent enzymes in cultured Chinese hamster ovary cells: induction of 5-methyltetrahydrofolate homocysteine cobalamin methyltransferase by folate and methionine. *Arch Biochem Biophys* **171**, 507-520.

Taylor, R. T., and Hanna, M. L. (1977). Folate-dependent enzymes in cultured Chinese hamster cells: folypolyglutamate synthetase and its absence in mutants auxotrophic for glycine + adenosine + thymidine. *Arch Biochem Biophys* **181**, 331-334.

Taylor, R. T., and Hanna, M. L. (1979). Folate-dependent enzymes in cultured Chinese hamster ovary cells: evidence for mutant forms of folylpolyglutamate synthetase. *Arch Biochem Biophys* **197**, 36-43.

Teraguchi, S., Shin, K., Ozawa, K., Nakamura, S., Fukuwatari, Y., Shimamura, S., and Tomita, M. (1997). Bacteriostatic Effects of Orally Administered Bovine Lactoferrin on Intestinal Bacteria in the Gut of Mice Fed Bovine Milk. In *Lactoferrin: Interactions and Biological Functions*, T. W. Hutchens and B. Lonnerdal, eds. (Totowa, New Jersey: Humana Press Inc.), pp. 303-312.

Thibodeau, S. N., Lee, D. C., and Palmiter, R. D. (1978). Identical precursors for serum transferrin and egg white conalbumin. *J Biol Chem* **253**, 3771-3774.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680.

Tomita, S., Matsue, M., Matsuyama, J., and Kiyosawa, I. (1994). Agglutination of bacterial cell of *Clostridium innocuum, Bifidobacterium longum,* and *Micrococcus luteus* by lactoferrin and ovotransferrin. *Biosci Biotech Biochem* **58**, 722-726.

Tong, L. A., de Vos, A. M., Milburn, M. V., and Kim, S. H. (1991). Crystal structures at 2.2 Å resolution of the catalytic domains of normal ras protein and an oncogenic mutant complexed with GDP. *J Mol Biol* **217**, 503-516.

Toy, J., and Bognar, A. L. (1990). Cloning and expression of the gene encoding *Lactobacillus casei* folylpoly-γ-glutamate synthetase in *Escherichia coli* and determination of its primary structure. *J Biol Chem* **265**, 2492-2499.

Toy, J., and Bognar, A. L. (1994). Mutagenesis of the *Lactobacillus casei* folylpolyglutamate synthetase gene at essential residues resembling an ATP binding site. *Arch Biochem Biophys* **314**, 344-350.

Tronrud, D. E. (1992). Conjugate-direction minimization: an improved method for the refinement of macromolecules. *Acta Cryst* **A48**, 912-916.

Tronrud, D. E. (1994). *The TNT Refinement Package Manual.*

Tronrud, D. E. (1996a). Knowledge-based B-factor restraints for the refinement of proteins. *J Appl Cryst* **29**, 100-104.

Tronrud, D. E. (1996b). The limits of interpretation. In *Proceedings of the CCP4 study weekend*, E. Dodson, M. Moore, A. Ralph and S. Bailey, eds. (Warrington, UK: CCLRC Daresbury laboratory), pp. 1-10.

Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Cryst* **A43**, 489-501.

Urzhumtsev, A. G. (1996). Density growing: a method for local improvement of electron density maps. In *The CCP4 newsletter on protein crystallography*, pp. 37-40.

Urzhumtsev, A. G., Skovoroda, T. P., and Lunin, V. Y. (1996). A procedure compatible with X-PLOR for the calculation of electron-density maps weighted using an R-free-likelihood approach. *J Appl Cryst* **29**, 741-744.

Vaganay, S., Tanner, M. E., van Heijenoort, J., and Blanot, D. (1996). Study of the reaction mechanism of the D-glutamic acid-adding enzyme from *Escherichia coli*. *Microb Drug Resist* **2**, 51-54.

Valenti, P., Visca, P., Antonini, G., Orsi, N., and Antonini, E. (1987). The effect of saturation with Zn^{2+} and other metal ions on the antibacterial activity of ovotransferrin. *Med Microbiol Immunol (Berl)* **176**, 123-130.

Valenti, P., Visea, P., Nicoletti, M., Antonini, G., and Orsi, N. (1985). Sythesis of siderophores by E. coli strains in the presence of lactoferrin-Zn. In *Proteins of iron storage and transport*, G. Spik, J. Montreuil, R. R. Crighton and J. Mazurier, eds. (Amsterdam: Elsevier), pp. 245-249.

van Berkel, P. H., Geerts, M. E., van Veen, H. A., Kooiman, P. M., Pieper, F. R., de Boer, H. A., and Nuijens, J. H. (1995). Glycosylated and unglycosylated human lactoferrins both bind iron and show identical affinities towards human lysozyme and bacterial lipopolysaccharide, but differ in their susceptibilities towards tryptic proteolysis. *Biochem J* **312**, 107-114.

Van Heijenoort, J. (1994). Biosynthesis of the bacterial peptidoglycan unit. In *Bacterial Cell Wall*, J. M. Ghuysen and R. Hakenbeck, eds. (Amsterdam: Elsevier Science B. V.), pp. 39-54.

Van Snick, J. L., Masson, P. L., and Heremans, J. F. (1974). The involvement of lactoferrin in the hyposideremia of acute inflammation. *J Exp Med* **140**, 1068-1084.

Verhoef, P., Stampfer, M. J., and Rimm, E. B. (1998). Folate and Coronary heart disease. *Curr Opin Lipidol* **9**, 17-22.

Vijayan, M. (1980). Phase evaluation and some aspects of the Fourier refinement of macromolecules. In *Computing in Crystallography*, R. Diamond, S. Ramaseshan and K. Venkatesan, eds. (Bangalore: Indian Institute of Science), pp. 19.01-19.25.

Villar, E., Schuster, B., Peterson, D., and Schirch, V. (1985). C1-Tetrahydrofolate synthase from rabbit liver. Structural and kinetic properties of the enzyme and its two domains. *J Biol Chem* **260**, 2245-2252.

Vogler, W. R., Huguley, C. M. J., and Kerr, W. (1965). *Arch Internal Med* **115**, 285.

Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**, 945-951.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol* **115**, 90-112.

Ward, P. P., Chu, H., Zhou, X., and Conneely, O. M. (1997a). Expression and characterization of recombinant murine lactoferrin. *Gene* **204**, 171-176.

Ward, P. P., Cunningham, G. A., and Conneely, O. M. (1997b). Commercial production of lactoferrin, a multifunctional iron-binding glycoprotein. *Biotechnol Genet Eng Rev* **14**, 303-319.

Ward, P. P., Lo, J. Y., Duke, M., May, G. S., Headon, D. R., and Conneely, O. M. (1992). Production of biologically active recombinant human lactoferrin in *Aspergillus oryzae*. *Biotechnology (N Y)* **10**, 784-789.

Ward, P. P., Piddington, C. S., Cunningham, G., Zhou, X., Wyatt, R. D., and Conneely, O. M. (1997c). Expression and functional analysis of recombinant human lactoferrin. in *Lactoferrin: Interactions and biological Functions*, T. W. Hutchens and B. Lonnerdal, eds. (Totowa, New Jersey: Humana Press Inc.), pp. 155-176.

Ward, P. P., Piddington, C. S., Cunningham, G. A., Zhou, X., Wyatt, R. D., and Conneely, O. M. (1995). A system for production of commercial quantities of human lactoferrin: a broad spectrum natural antibiotic. *Biotechnology (N Y)* **13**, 498-503.

Whitehead, V. M., Rosenblatt, D. S., Vuchich, M. J., Shuster, J. J., Witte, A., and Beaulieu, D. (1990). Accumulation of methotrexate and methotrexate polyglutamates in lymphoblasts at diagnosis of childhood acute lymphoblastic leukemia: a pilot prognostic factor analysis. *Blood* **76**, 44-49.

Williams, J., Elleman, T. C., Kingston, I. B., Wilkins, A. G., and Kuhn, K. A. (1982). The primary structure of hen ovotransferrin. *Eur J Biochem* **122**, 297-303.

Wilmot, C. M., and Thornton, J. M. (1988). Analysis and prediction of the different types of beta-turn in proteins. *J Mol Biol* **203**, 221-232.

Wilson, A. J. C. (1942). Determination of absolute from relative X-ray intensity data. *Nature* **150**, 152.

Wilson, A. J. C. (1949). The probability distribution of X-ray intensities. *Acta Cryst* **2**, 318-321.

Wilson, N. A., Barbar, E., Fuchs, J. A., and Woodward, C. (1995). Aspartic acid 26 in reduced *Escherichia coli* thioredoxin has a pKa > 9. *Biochemistry* **34**, 8931-8939.

Winterbourn, C. C. (1983). Lactoferrin-catalysed hydroxyl radical production. Additional requirement for a chelating agent. *Biochem J* **210**, 15-19.

Woolley, P., and Clark, B. F. C. (1989). Homologies in the structures of G-binding proteins: an analysis based on elongation factor EF-Tu. *Biotechnology* **7**, 913-920.

Xie, D., Gulnik, S., Collins, L., Gustchina, E., Suvorov, L., and Erickson, J. W. (1997). Dissection of the pH dependence of inhibitor binding energetics for an aspartic protease: direct measurement of the protonation states of the catalytic aspartic acid residues. *Biochemistry* **36**, 16166-16172.

Yamauchi, K., Tomita, M., Giehl, T. J., and Ellison, R. T. D. (1993). Antibacterial activity of lactoferrin and a pepsin-derived lactoferrin peptide fragment. *Infect Immun* **61**, 719-728.

Yang, A. S., Gunner, M. R., Sampogna, R., Sharp, K., and Honig, B. (1993). On the calculation of pKas in proteins. *Proteins* **15**, 252-265.

Zhang, K. Y. J., and Main, P. (1990). Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. *Acta Cryst* **A46**, 41-46.

Zhao, X., and Hutchens, T. W. (1994). Proposed mechanisms for the involvement of lactoferrin in the hydrolysis of nucleic acids. In *Lactoferrin: structure and function*, T. W. Hutchens, S. Rumball and B. Lonnerdal, eds. (New York: Plenum), pp. 271-278.

Zweier, J. L., Wooten, J. B., and Cohen, J. S. (1981). Studies of anion binding by transferrin using carbon-13 nuclear magnetic resonance spectroscopy. *Biochemistry* **20**, 3505-3510.