National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

THE DOMAIN STRUCTURE OF AN OCTAMERIC BIFUNCTIONAL ENZYME

by

Laura Lea Murley

Department of Biochemistry
McGill University
Montréal, Québec, CANADA
February 1996

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Doctor of Philosophy

ISBN 0-612-12442-8

This thesis is dedicated to my parents,

for their love and encouragement throughout the years.

ABSTRACT

Formiminotetrahydrofolate:glutamate formiminotransferase (EC 2.1.2.5) - formiminotetrahydrofolate cyclodeaminase (EC 4.3.1.4) is a bifunctional enzyme arranged as a circular tetramer of dimers that exhibits the ability to efficiently channel polyglutamylated folate between catalytic sites. A novel, full-length cDNA clone encoding the porcine liver formiminotransferase-cyclodeaminase was isolated and inserted into a prokaryotic expression vector. The recombinant enzyme is expressed as a soluble protein in *Escherichia coli* and was purified to homogeneity using a multistep scheme. Deletion analysis of the cDNA indicated that each subunit consists of an N-terminal transferase-active domain and a C-terminal deaminase-active domain separated by a short linker sequence. These domains were expressed in *Escherichia coli* and demonstrated to exist as monofunctional dimers. This provides direct evidence for the existence of two types of subunit interfaces and suggests that both catalytic activities are dependent on the formation of specific subunit interfaces. Because channelling is not observed between isolated domains, only the octamer appears able to directly transfer pentaglutamylated intermediate between active sites. Thus we have established direct support for Findlay and MacKenzie's model (Findlay, W. A. & MacKenzie, R. E. (1987) *Biochemistry 26*, 1948-1954) that the octamer is the functional unit of the enzyme. The purified dimers show no tendency to associate, suggesting that the linker mediates the only substantial domain-domain interaction. The isolated domains and the full-length enzyme were subjected to urea-induced denaturation in order to characterize the properties of both domains in and outside of the octamer. At low concentrations of urea, both domains undergo a cooperative loss of fluorescence and activity, while maintaining their secondary structure and the majority of their quaternary structure. When the urea concentration is increased, coincident unfolding and dissociation of the isolated domains to monomers is observed. While one of the octameric subunit interfaces is disrupted under the same urea concentrations as the domains, dissociation at the second interface occurs at significantly higher concentrations of urea, indicating that one of the domains exhibits increased stability when part of the full-length enzyme.

RÉSUMÉ

La formiminotétrahydrofolate:glutamate formiminotransferase (EC 2.1.2.5)-formiminotétrahydrofolate cyclodesaminase (EC 4.3.1.4) est une enzyme bifonctionnelle formant un tetramère de dimères circulaire qui peut efficacement transferer le substrat polyglutamylé entre les sites catalytiques. Un clone original d'ADNc de pleine longueur codant l'enzyme d'origine porcine fut isolé et insére dans un vecteur d'expression procaryote. L'enzyme recombinante est exprimée sous forme de proteine soluble chez *Escherichia coli* et fut purifiée jusqu'à homogénéité, en plusieurs étapes. La creation de délétions dans l'ADNc a permi de déterminer que chaque unité de l'enzyme est composée d'un domaine transférase amino-terminal et d'un domaine cyclodésaminase carboxy-terminal, séparés par un court "linker". Chaque domaine, exprimé chez *E. coli*, forme des dimères monofonctionnels. Ceci démontre directement l'existence de deux genres d'interfaces entre les unités chez l'enzyme d'origine et suggère que les deux activités dépendent de la formation d'interfaces spécifiques à chaque activité. On n'observe aucun transfert d'intermédiaire de l'un à l'autre des domaines isolés et donc seul l'octamère semble pouvoir transférer l'intermédiaire pentaglutamylé entre les sites actifs. Ceci appuie directement le modèle de Findlay et MacKenzie (Findlay, W. A. & MacKenzie, R. E. (1987) *Biochemistry 26*, 1948-1954), où l'octamère serait l'unité fonctionnelle de l'enzyme. Les dimères purifiés ne démontrent aucune tendance d'association, ce qui suggère que le "linker" pourvoit à la seule interaction essentielle entre les deux domaines. L'enzyme d'origine et les domaines isolés furent assujettis à la dénaturation par l'urée afin de caractériser les propriétés des domaines lorsqu'ils sont isolés ainsi que dans l'enzyme d'origine. Des faibles concentrations d'urée entraînent une perte coopérative de fluorescence et d'activité chez les deux domaines, en maintenant leur structure secondaire et la majorité de la structure quaternaire. De plus fortes concentrations d'urée entraînent le dépliement et la dissociation simultanés de chaque domaine en monomères. Dans l'enzyme d'origine, une des interfaces se dissocie dans les mêmes conditions que le domaine isolé alors que la deuxième interface de dissocie à des concentrations d'urée plus élevées, ce qui indique que l'un des domaines exhibe une stabilité accrue dans l'enzyme d'origine.

FOREWORD

Parts of this thesis include the text of original papers published or submitted for publication. In compliance with the Faculty of Graduate Studies and Research regulations the following excerpt from the "Guidelines for thesis preparation" is provided:

"Candidates have the option of including, as part of the thesis, the text of a paper(s) submitted or to be submitted for publication, or the clearly-duplicated published text of a published paper(s). These texts must be bound as an integral part of the thesis.

If this option is chosen, connecting texts that provide logical bridges between different papers are mandatory. The thesis must be written in such a way that it is more than a mere collection of manuscripts; in other words, results of a series of papers must be integrated.

The thesis must still conform to all other requirements of the "Guidelines concerning thesis preparation". The thesis must include: A Table of Contents, an abstract in English and French, an introduction which clearly states the rationale and objectives of the study, a comprehensive review of the literature, a final conclusion and summary, and a thorough bibliography or reference list.

Additional material must be provided where appropriate (e.g. in appendices) and in sufficient detail to allow a clear and precise judgement to be made of the importance and originality of the research reported in the thesis.

In the case of manuscripts co-authored by the candidate and others, the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent. Supervisors must attest to the accuracy of such statements at the doctoral oral defense. Since the task of the examiners is made more difficult in these

cases, it is in the candidate's interest to make perfectly clear the responsibilities of all the authors of the co-authored papers. Under no circumstances can a co-author of any component of such a thesis serve as an examiner for that thesis."

PREFACE

The results presented in Chapters 2 and 3 of this thesis have been published in the following journals:

The Nucleotide Sequence of Porcine Formiminotransferase-Cyclodeaminase: Expression and Purification from *Escherichia coli*. L.L. Murley, N.R. Mejia, and R.E. MacKenzie. (1993) *J. Biol. Chem.* **268**, 22820-22824.

The Two Monofunctional Domains of Octameric Formiminotransferase-Cyclodeaminase Exist as Dimers. L.L. Murley and R.E. MacKenzie. (1995) *Biochemistry* **34**, 10358-10364.

The work presented in Chapter 4 has been submitted to *Protein Science* under the title "Urea-induced denaturation of formiminotransferase-cyclodeaminase and its monofunctional domains", by L.L. Murley and R.E. MacKenzie.

All experiments described in this thesis were performed by myself, with the following exceptions. N. Mejia prepared the polyclonal anti-FTCD antibody which was used throughout these studies. He also prepared the cyanogen bromide cleavage fragments of FTCD which provided confirmatory amino acid sequence, as described in Chapter 2. Moreover, he and I worked together to design a scheme to purify recombinant formiminotransferase-cyclodeaminase and he provided Figure 2.5. and Table 2.1 which describe this purification. Further acknowledgements, specific to each chapter, are included at the end of Chapters 2, 3 and 4 and Appendix A.

Dr. R. E. Mackenzie provided the normal advice and supervision appropriate to his role as research director.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ADP	adenosine diphosphate
AICAR	5-aminoimidazole-4-carboxamide ribonucleotide
ANS	8-anilino-1-naphthalene sulphonate
BSA	bovine serum albumin
bp	base pair
C1	one-carbon
CD	circular dichroism
CDH_6	hexahistidine-tagged cyclodeaminase domain
cDNA	complementary deoxyribonucleic acid
CNBR	cyanogen bromide
D\C	dehydrogenase\cyclohydrolase domain
D\C\S	dehydrogenase\cyclohydrolase\synthetase
DNA	deoxyribonucleic acid
ESI-MS	electrospray ionization mass spectrometry.
FAD	flavin adenine dinucleotide
FIGLU	formiminoglutamate
FTCD	formiminotransferase-cyclodeaminase
$FTCDH_6$	hexahistidine-tagged formiminotransferase-cyclodeaminase
FTH_6	hexahistidine-tagged formiminotransferase domain
GAR	glycinamide ribonucleotide
GdnHCl	guanidine hydrochloride
HPLC	high performance liquid chromatography
$H_4PteGlu$	tetrahydrofolate
$H_4PteGlu_n$	tetrahydrofolate, with a total of n glutamates
kb	kilobase(s)
kDa	kiloDalton(s)
LC/MS	liquid chromatography/mass spectrometry
MOPS	4-morpholinepropanesulfonic acid

Mr	molecular mass
mRNA	messenger ribonucleic acid
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
Ni-NTA	nickel-chelated nitrilotriacetic acid matrix
PMSF	phenylmethylsulfonylfluoride
RNA	ribonucleic acid
SDS PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
TBS	tris buffered saline
TFA	trifluoroacetic acid
TS	thymidylate synthase
λ_{max}	wavelength of maximum fluorescence emission

ACKNOWLEDGEMENTS

I am deeply indebted to Dr. Robert MacKenzie for his support and excellent guidance throughout the course of my studies. His encouragement, patience and sound scientific sense have allowed me to develop my own scientific skills. Through his interest in his students, his accessibility and his sense of humour, he provides a superior learning and working environment. I am also grateful for his careful and prompt reading of this thesis.

I am grateful to past and present members of the laboratory who contributed to the pleasant atmosphere within the laboratory. Dr. Dean H., Dr. Carole B., Dr. Xiao Ming Y., Narciso M. and Dr. Krishna P. made me feel welcome when I first came to McGill, and donated much of their time teaching me the ins and outs of molecular biology. Joelle P., Gilles T., Peter P., and most recently Erminia D., provided a lively camaraderie within the laboratory which resulted in stimulating discussions on many topics, both scientific and otherwise. Special thanks are directed to Peter P., Joelle P. and Lionel C. for proof reading sections of this thesis. I would also like to thank Joelle P. for translating my abstract into French. Moreover, I would like to express my gratitude for her friendship and generous spirit which helped to make my years at McGill very special.

I am especially appreciative of the efforts of Narciso Mejia, both for his contributions towards my project and for efficiently maintaining the lab in working order. I would also like to thank Franca D. for providing us with clean glassware every morning.

I have benefitted greatly from the collaborative atmosphere which prevails in the Department of Biochemistry at McGill. I would like to acknowledge the many students, staff and faculty members within the department who have generously helped me on numerous occasions.

The financial support provided by the Faculty of Medicine, and the David Stewart Major Fellowship (through the Faculty of Graduate Studies) at different points during my studies was also much appreciated .

Finally, I would like to express my gratitude to my family for their love, and encouragement, and for providing me with so many opportunities; and to my fiance,

Lionel Chow, who's love and support fuels my accomplishments, and fills my days with laughter

CHAPTER 1

GENERAL INTRODUCTION

The importance of hierarchial structure is evident in almost all aspects of biological organization (as reviewed in Kurganov, 1993). Every level can be viewed as a system containing a set of elements from a lower tier, where the properties of the higher level are greater than the sum of the properties of its components. The multifunctional enzyme formiminotransferase-cyclodeaminase provides an excellent example of hierarchial organization. The covalent association of two sequential enzyme activities results in properties which are specific to the bifunctional enzyme and not observed in the component domains. In the following pages I will discuss the hierarchial structure of proteins, stressing in particular the domain and quaternary structure of enzymes. Examples of enzyme organization and their consequences are considered. A brief review of protein denaturation and folding, with a section concentrating on multidomain and/or oligomeric proteins is offered. A short overview of mammalian folate metabolism is presented and the previously established properties of the folate-dependent formiminotransferase-cyclodeaminase are described.

1.1. PROTEIN STRUCTURE.

Most globular proteins adopt a specific, compact conformation under native conditions. This structure results from the interaction of the amino acid sequence of the protein with the solvent/environment in which it exists. Protein conformation is generally described as a hierarchy of different structural elements, proceeding in order from the primary to the quaternary level.

1.1.1 The primary structure of proteins

Primary structure involves the sequential order of the amino acids, as specified by the mRNA transcript encoding the protein. The residues in amide linkage provide the backbone of the protein (the structure of which is described in Creighton, 1993). As any of 20 different natural amino acids can be specified at any position within the sequence, the primary structure provides an initial level of protein diversity. The amino acid sequence conveys all the information required to fold a protein into its specific 3-dimensional conformation in a particular environment. The acquisition of 3-dimensional structure provides an additional level of structural diversity. Additionally, many proteins are multimeric and must associate with other identical or non identical protein chains and/or prosthetic groups to achieve their native structure.

1.1.2 The secondary structure of proteins

Secondary structure refers to the local conformation of the polypeptide chain, involving the peptide backbone atoms only. It includes structural elements such as α-helices, β-sheets, reverse turns, loops and disordered regions, as well as others. These structures are formed in response to the protein's structural requirement to be compact and to balance buried polar groups by pairing them in hydrogen bonds. Several excellent reviews describe the features and packing of these secondary structure elements (Richardson, 1981; Chothia, 1984; Chothia & Finkelstein, 1990). The interaction of 2 or more secondary structural elements leads to formation of supersecondary structures. This includes the assembly of secondary structure elements and the connecting peptide chain.

Protein structures can be classified on the basis of their secondary structure content (Chothia, 1984). More than 60% of the residues in all-α proteins form α-helices, which can interact by stacking around a central core or forming layer structures. All-β proteins consist primarily of β-sheets and always include at least 2 sheets, usually antiparallel, which pack against each other. α + β proteins contain both helices and sheets, in separate parts of the structure. For example, there may be one antiparallel sheet with helices clustered at one or both ends. In α/β proteins, the helices and sheets can interact, often along the peptide backbone, with β-sheets forming layer structures with helices and/or other β-sheets.

1.1.3 The tertiary structure of globular proteins

Tertiary structure involves the spatial arrangement of all atoms in a single polypeptide chain. The three dimensional arrangement of the polypeptide chain, or the protein fold, (reviewed in Jaenicke, 1987) describes several things: the secondary structures present, their relative arrangement and the chain topology, or path taken by the polypeptide chain, through the structure. It also includes the packing of amino acid side chains, which mediate the long range interactions required to stabilize the fold of a protein. The chain topology is subject to specific limitations. First, chains do not cross each other or make knots. Secondly, pieces of secondary structure which are adjacent in sequence are often adjacent within the fold and generally pack in an anti-parallel fashion. Finally, in β-X-β units (where the β-strands are parallel and X is an α-helix, a loop, or a β-strand from a different sheet) the chain connections formed between secondary structure elements are generally right handed.

Several proteins can be circularly permutated, such that the original N- and C-termini are joined by a linker sequence and new termini are constructed in a surface loop (bovine pancreatic trypsin inhibitor, Goldenberg & Creighton, 1983; dihydrofolate reductase, Buchwalder et al., 1992; aspartate transcarbamoylase, Yang & Schachman, 1993a; T4 lysozyme, Zhang et al., 1993). These circularly permuted proteins often fold and function quite efficiently and have physical properties in common with their wildtype counterparts. Likewise, complementary fragments of proteins such as β-galactosidase,

(Ullman et al, 1967) barnase (Sancho & Fersht, 1992), alanine racemase, (Galaktos & Walsh, 1987) and aspartate transcarbamoylase (Yang & Schachman, 1993b) can fold and associate to form active species. While placement of the chain termini and cleavage sites at nondisruptive positions within surface loops is likely crucial for the proper folding of these proteins, these experiments indicate that chain continuity and the sequential arrangement of secondary structure do not necessarily dictate tertiary structure.

The advent of structural biology has provided an explosion in the number of known protein structures. A surprisingly small number of folds describes the tertiary structure of most proteins, such that similar folds can be encoded by proteins with dissimilar amino acid sequences (Rao & Rossman, 1973; Finkelstein et al., 1993; Laurents et al., 1994). After analysis of the protein structures in the Protein Structure Databank, Thornton et al. (1995) have suggested that there are 71 folds represented by only a single structure and nine superfold structures which account for 46% of all nonhomologous proteins in the databank. The superfolds consist of: the globin; the trefoil; the up-down; the immunoglobulin fold; the alpha-beta sandwich; the jelly roll; the doubly wound; the UB alpha-beta roll; the TIM barrel.

Examination of protein families with similar three dimensional structures but different primary structures, mutagenic analyses of designated positions within a fold, and studies of proteins recovered from random sequence libraries provide generalities concerning the relationship between amino acid sequence and tertiary structure (Knowles, 1987; Lim & Sauer, 1989; Bordo & Argos, 1990; Reidhaar-Olsen & Sauer, 1988; Matthews, 1993). The interior of a protein is mainly occupied by nonpolar residues and occasionally neutral amino acids. Polar atoms within the protein interior form specific hydrogen bonds with other atoms within the same secondary structural unit, such that the surfaces of each secondary structure are also hydrophobic. Highly solvent exposed residues, such as those on the protein surface, are often polar or neutral residues which can interact with the solvent. The sequestering of hydrophobic residues inside the protein, combined with the hydrogen bonding of polar backbone and side chain atoms within secondary structures, results in stably folded globular proteins.

The overall tertiary structure of most small globular proteins is roughly spherical

and close-packed, but with an irregular surface. Typically, the accessible surface area of a small monomeric protein can be approximated by the equation $A_s = 630(Mr)^{0.73}$ nm^2 (Janin et al., 1988), illustrating the relationship between a protein's surface area and its molecular weight. This value is only 23-45% of the surface area that is ascribed to the unfolded polypeptide chain, indicating the degree of compactness observed in globular proteins. Water is generally excluded from protein interiors. When it is present, it may hydrogen bond with the backbone atoms or with the side chain of an internal polar group. In other cases it may function in filling holes. Intramolecular cavities are almost always present in proteins of greater than 100 residues, however, they generally constitute only a small fraction (less than 2.3%) of the total protein volume (Hubbard, Gross & Argos, 1994).

Larger proteins (greater than 200 residues in length) are often subdivided into 2 or more structural units known as domains. While the literature definition of a domain is very subjective (more on this in section 1.1.4), general consensus indicates that domains represent structurally and functionally discrete regions of a single polypeptide chain. Although different domains may interact to various degrees, the interactions of secondary structure elements within the domain are generally stronger than interactions between domains. Domains may be composed of smaller modules of approximately 5 kDa, each with specific functions such as ligand binding or protein interactions (Traut, 1986). Traut suggests that these modules are stable folded structures, corresponding to the amount of polypeptide encoded by the average exon.

1.1.4 The domain structure of proteins

The term "domain" has been used to represent many different concepts of protein organization. As noted by Garel, (1992), the definition of a domain is largely dependent on the method by which the domain has been identified. Limited proteolysis or protein engineering is often used to isolate discrete stable fragments, or domains, of a larger protein. Domains have been defined as compact structural units visible in the crystal structure of a protein, or as functional units responsible for a specific purpose such as cofactor binding. Sometimes domains are identified by amino acid sequence homologies

to other proteins or estimated from exon-intron gene structure. Privalov (1989) has suggested that a domain is a thermodynamic unit, a region of a protein which can fold and unfold in a cooperative manner, existing in one of two macroscopic states, native or denatured. Historically domains have been viewed as independent folding units, as defined by Wetlaufer (1973).

The Rossman fold is a dinucleotide binding domain which displays several of the above properties. This domain can be identified in many dinucleotide binding proteins by its consensus sequence, Gly-X-Gly-X-X-Gly, and characteristic β-α-β supersecondary structure. The dinucleotide binding domain can be stably isolated from many proteins through proteolysis or protein engineering (Jecht et al., 1994), and behaves as an independent folding unit. However, the definitions outlined above do not apply to all domains, as illustrated by structural analysis of the α subunit of tryptophan synthase. Proteolysis of this subunit produces a stable 20 kDa N-terminal fragment and a 9 kDa C-terminal fragment. The cleavage site, however, does not reside within the loops connecting the structural domains identified in the crystal structure of tryptophan synthase (Hyde et al., 1988). In some instances, the existence of extensive interdomain interactions, necessary for stability, may prevent the isolation of a properly folded domain (Rossman & Argos, 1981).

Often domains are identified through inspection of the crystal structure of the protein. Richardson (1981) and others have identified domains which represent structures that might remain stable if independent, that could undergo rigid-body-like movements with respect to the entire protein, or that had structural similarity to known single-domain proteins (reviewed in Islam et al., 1995). Liljas & Rossman (1974) utilized a method which analyzes the distance between inter-residue contacts. Wodak & Janin (1981) have evaluated the minimal interface area between domains versus a cleavage point to determine domain boundaries. Rashin (1981) identified domains on the basis of buried surface area.

Historically, proteolysis has been used to determine the domain structure of proteins whose crystal structure is not known. In fact, the first domains identified were the stable fragments produced by proteolysis of the light and heavy chains of

immunoglobulins (reviewed in Coggins & Hardie, 1986, and Cushley, 1986). This technique has been successful because many domains are formed from contiguous stretches of polypeptide sequence and connected by exposed surface loops, known as hinges or linkers, which are more susceptible to proteolysis than the compact domains themselves. More recently this technique has been combined with deletion mutagenesis and independent expression of individual domains, to further define domain structure. Minard et al. (1989) have shown that the two domains of phosphoglycerate kinase reside within residues 1-184 and 186-415. Hum & MacKenzie (1991) used deletion analysis to localize the dehydrogenase/cyclohydrolase and synthetase domains of human methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate cyclohydrolase-formyltetrahydrofolate synthetase to residues 1-301 and 304-935 respectively. The regulatory and catalytic domains of rat tyrosine hydroxylase can be expressed independently in *E. coli* (Daubner et al., 1993). Likewise, C-terminal truncations indicate that the redox and DNA-repair activities of Ref-1 reside in non-overlapping N- and C-terminal domains (Xanthoudakis et al., 1994).

The combination of protein engineering and knowledge of the crystal structure of the protein can be used to isolate domains which display significant interdomain interaction. The interface domain from dimeric gluthathione reductase makes hydrophobic contacts with the other three domains within the subunit. While this domain can be expressed independently, it is subject to nonspecific aggregation (Lesitler & Perham, 1994). Replacement of 3 amino acids which make contacts with the NADPH binding domain and 4 additional residues which interact with the FAD binding domain produces a soluble stable interface domain which is no longer prone to aggregation.

Domains are not always formed by continuous segments of the polypeptide chain. Several multidomain proteins contain crossover linkages where the C-terminal terminus of the protein interacts with the N-terminal domain (Thornton & Sibanda, 1983; Russel, 1994). Protein engineering can be used to link nonsequential parts of the domain. To express the discontinuous dinucleotide binding domain from the *Thermotoga maritima* glyceraldehyde 3-phosphate dehydrogenase, Jecht et al. (1994) constructed a cDNA with the N-terminal residues 1-148 ligated to the C-terminal α helix (residues 313-333) by a

4 amino acid linker connecting residues 148 and 313. Expression of this cDNA in *E.coli* produces a stable dinucleotide binding domain with native-like structure.

Domain interfaces are generally not as well packed as protein interiors or as optimally matched as some subunit interfaces, however they do retain reasonably good shape complementarity (Lawrence & Colman, 1993; Hubbard & Argos, 1994). Approximately 90% contain cavities or packing defects (Hubbard & Argos, 1994). Clefts between interfaces are often functionally important, and may house ligand binding sites or full catalytic sites involving residues from both domains. Additionally they may provide increased flexibility between domains, perhaps allowing for conformational changes and domain movements resulting from a combination of hinge and/or shear movements between domains (Rashin et al., 1986; Hubbard & Argos, 1994; Gerstein et al., 1994).

1.1.5 The quaternary structure of proteins

Quaternary structure refers to the overall arrangement of subunits within a protein. Most proteins exist as aggregates of identical or nonidentical subunits. The individual subunits are usually distinct globular monomers which then interact with each other. In some cases the subunits are more intimately associated such as the two entwined chains of the dimeric trp repressor (Schevitz et al., 1985) or trimeric dihydrolipoyl transacetylase (Mattevi et al., 1992) which extends a beta strand out of each monomer to interact with a neighbouring subunit.

Most dimers contain isologous interactions which involve the same protein surface on both monomers, thus producing dimers with a two fold axis of symmetry. Even-numbered species are often constructed as n-mers of isologous dimers. For example each monomer of tetrameric lactate dehydrogenase contains 2 surfaces involved in different isologous interactions.

Heterologous interactions involving two different non-overlapping complementary surfaces can form open-ended structures; for example, the polymerizing actin chain, or closed cyclic structures such as the trimeric chloramphenicol acetyltransferase (Leslie, 1990), or tetrameric Mn-dependent superoxide dismutase (Ludwig et al., 1991). A

heterologous dimer would be prone to further aggregation unless dimerization effectively blocked the unpaired contact sites remaining on each monomer.

Recently, the structural and chemical properties of subunit interfaces have come under considerable attention. Cherfils et al. (1991) suggested that the interface region does not significantly differ from the rest of the protein surface in terms of its content of hydrophobic or charged residues. In contrast, Young et al. (1994) suggest that hydrophobicity is very important for protein-protein interactions, since the strongest hydrophobic cluster on the protein surface correlated to the interaction site in more than two thirds of the 38 protein complexes analyzed. Most subunit interfaces have a composition which is intermediate in both its hydrophobic and charge properties (Miller, 1989, Janin et al., 1988). The centre of the interface is often dominated by hydrophobic residues and is similar in composition to the interiors of monomers, while the interface periphery is more reminiscent of the protein exterior and may include loops involved in hydrogen bonding (Miller, 1989). Miller (1989) has identified several structural motifs occurring at subunit interfaces: (i) an extended beta sheet which forms across the interface as observed at the subunit interfaces of several enzymes including prealbumin, Concavalin A, and λcro; (ii) tertiary helix-helix packing which forms most of the interfaces of melitten, citrate synthase and cyctochrome c'; (iii) face to face or aligned packing of beta sheets as observed in the interface of glyceraldehyde-3-phosphate dehydrogenase; (iv) loop interactions which occur to some extent at almost all subunit interfaces. These loops commonly interact by hydrogen bonding to other loops or ends of adjacent secondary structures.

The most consistent aspect of subunit interfaces is that they are complementary in both shape and hydropathy (Korn & Burnett, 1991). Protomers interact by matching hydrophobic centres of adjacent subunits and, similarly, matching hydrophilic centres. The interface formed by wheat germ agglutinin provides a good example of this. It includes both hydrophobic and hydrophilic contacts and has very high hydropathy complementarity, with over one third of the interface interactions occurring between identical residues and even more between conservative residues (matching hydropathies). As well, Korn and Burnett (1991) proposed that the level of complementarity and the

functional role of the interface may be related. Interfaces with high hydropathic complementarity may form inflexible, static interactions, necessary for accurate positioning of subunits. Subunit interfaces which are dominated by hydrophilic forces and have lower hydropathic complementarities are more dynamic and more commonly found in enzymes which undergo allosteric conformational changes, such as haemoglobin.

To determine the degree to which subunit interface residues are conserved, Grishin and Phillips (1994) analyzed five different enzyme families which contain active sites at their subunit interfaces. They concluded that while many of the active site residues are highly conserved, the subunit interface evolves almost as rapidly as the entire sequence and can accommodate amino acid substitutions as long as packing, hydrophobicity and charge are maintained. Hubbard & Argos (1994) have suggested that because they might be more capable of tolerating mutationally prone loops than domain cores, the mutational pressure might be higher at domain and subunit interfaces.

Eisenberg has suggested a method by which oligomeric proteins might evolve from monomers (Bennet et al., 1995). 3D domain swapping occurs when a domain or some part of a domain from one subunit is replaced by the equivalent region of an identical subunit. This can result in intertwined dimers as has been observed for the dimeric form of diphtheria toxin (Bennet et al., 1994). It could also promote formation of higher order cyclic structures or open-ended oligomers which might be prone to aggregation. Domain swapping could occur through a single mutation which would destabilize the monomer relative to an oligomer. For example, a deletion in the loop connecting two interacting domains might stericly prevent domain closure. A second subunit could then interact with the exposed surface of the interdomain interface, producing a dimer which has comparable energies to the original monomer (since the same interfaces are formed by the swapped domain in the monomer and the oligomer). This is a much more efficient method of generating oligomers than the "sticky billiard ball" model, where several rounds of random mutagenesis on the interface surface of a monomer might be required to form a stable dimer.

1.1.6 Consequences of oligomerization

Oligomerization can often influence protein stability. After reviewing the literature concerning the urea-induced denaturation of dimeric proteins which undergo a two-state transition, Neet and Timm (1994) noted that the conformational stabilities of these dimers were significantly greater than the stabilities observed for monomeric proteins (a change in free energy of unfolding of 10-27 kcal/mol for each subunit in the dimer versus 6-14 kcal/mol for the monomer). They estimated that quaternary interactions might provide between 25 and 100% of the conformational stability in such proteins. This stabilization is due primarily to interactions between subunits and is related to the size of the subunit and the structure of the interface. Neet and Timm (1994, and references therein) have compared the free energies of dissociation and unfolding for several of these proteins. For some proteins, such as the arc repressor, the free energy of dissociation is almost identical to the free energy of unfolding, indicating that the monomer is not stable. Other proteins (nerve growth factor, HIV-1 protease, troponin C peptide) form monomers that are stabilized by only 1-2 kcal/mol monomer in the absence of quaternary interactions, also suggesting that the monomer is not stable. A few proteins show considerable stability as a monomer. To illustrate this, the human growth hormone monomer displays a change in free energy upon unfolding of approximately 11.5 kcal/mol. Unlike some of the other examples, this protein is functional as a monomer, since it interacts with a receptor in this form.

The role of oligomerization in providing stability has also been studied by mutating interface residues. Nordhoff et al (1993) have shown that a double mutant G446E/F447P in the interface domain of dimeric human glutathione reductase produces an inactive monomer with a non-native FAD-binding domain, NADPH-binding domain and interface domain. The instability of the dinucleotide binding domains is somewhat of a surprise as these domains are thought to be independent folding units.

In several enzymes, active site residues are donated by both polypeptide chains, thus forming shared catalytic or binding site(s) at the subunit interface (also known as a shared active site). While isologous interactions generally form two active sites per subunit interface, heterologous interactions produce a single site at the interface.

Quaternary structure may play an important role in enzyme regulation through allosteric cooperation between subunits. Aspartate transcarbamoylase (ATCase) and phosphofructokinase are classic examples of oligomeric allosteric enzymes. ATCase is composed of two catalytic trimers linked by three regulatory dimers (Krause et al., 1985). Upon binding carbamyl phosphate and aspartate, this enzyme undergoes a concerted allosteric transition involving changes in quaternary contacts, to reach a more active conformation (Kantrowitz & Lipscomb, 1988). Likewise, binding of phosphoenolpyruvate or ADP to a regulatory site at a subunit interface within the homotetrameric phosphofructokinase produces a change in the relative orientation of subunits, which can result in activation or inhibition of the enzyme (Schirmer & Evans, 1990).

1.2. PROTEIN FOLDING

The field of protein folding includes a tremendous amount of theoretical and experimental data. Out of necessity, this discussion will be limited to a description of factors affecting protein stability and denaturation, and a review of proposed *in vitro* protein folding and assembly pathways. One approach successfully used to elucidate protein folding pathways has been the identification and characterization of intermediates observed in unfolding and/or refolding pathways. While the majority of folding studies have involved single-domain monomeric proteins, most proteins are much larger and their folding and assembly mechanisms are expected to be more complicated.

1.2.1 The thermodynamic hypothesis and the kinetic folding problem

The pioneering work of Anfinsen and colleagues (1966) on the unfolding and refolding of RNaseA indicated that the amino acid sequence of a protein contains all the information necessary for successful folding. Except in cases where irreversible covalent changes occur upon folding, this hypothesis has held true. It has been suggested that a folding code, analogous to the RNA/protein code, could specify the three dimensional conformations of proteins from amino acid sequence alone. Several years of research has indicated that the protein folding question is much more complicated than simply breaking a code. So how does a protein fold? Anfinsen's work suggested that folding is a thermodynamically controlled process, where "the three dimensional structure of a native protein in its normal physiological milieu ... is the one in which the Gibbs free energy of the whole system is lowest; that is, that the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence in a given environment" (Anfinsen, 1973). The evidence for the thermodynamic hypothesis includes the fact that the folding and unfolding of most small proteins is reversible and that the same native conformation is reached through folding *in vivo* and *in vitro* for many proteins. As Levinthal pointed out, a random examination of all possible conformations in order to attain the one most thermodynamically stable is outside of the time frame under which protein folding takes place (Levinthal's paradox, 1968).

Therefore, there must be a mechanism in place to direct folding towards more favourable conformations. Schindler et al. (1995) have summarized some methods of limiting this search. The early formation of partially folded intermediates may restrict the search to a small number of productive conformations. *Cis/trans* isomerization of proline, in particular, and disulphide bond formation and isomerization are some of the slow folding reactions which may synchronize the folding process. Generally the low energy side chain and main chain conformations populated in the unfolded polypeptide are also observed in the native protein. NMR analysis of short peptides has identified locally preferred conformations which are also represented in the folded protein (Dyson & Wright, 1991). Kinetically foldable proteins appear to have evolved to avoid the kinetic traps separating misfolded and native states (Wolynes et al., 1995).

Baker & Agard (1994) have argued that a protein's native conformation occurs at the lowest energy state within a neighbourhood of kinetically accessible states, as opposed to the global free energy minimum. Several examples suggest that folding may be under kinetic control. Refolding of several mature proteases requires the presence of the pro region, either in *cis* as the protease precursor or in *trans* on a separate chain. The pro region may function by increasing the rate of the forward folding reaction or by decreasing the rate of aggregation. Metastable proteins such as plasminogen activator inhibitor, and influenza virus haemagglutonin have been shown to undergo dramatic conformational changes in folding. Upon renaturation, plasminogen activator inhibitor initially adopts a biologically functional conformation, which slowly converts to a latent form. Influenza virus haemagglutonin undergoes a transition at low pH from an initial metastable structure to a more thermostable conformation. This change is irreversible and the low pH conformation cannot be induced back to its original conformation. The implication is that the more stable state is not initially accessible.

1.2.2 Protein stability, folding, cooperativity and denaturation

While the thermodynamic hypothesis suggests that the native conformation occupies the global energy minima, it is only slightly more stable than the unfolded states. In fact, the free energy difference between a folded and unfolded polypeptide has been

estimated to be in the range of only 4-20 kcal/mole, in aqueous solution at room temperature (Creighton, 1993), and this marginal stability is determined by the balance of relatively large opposing energetic forces. When the interaction between the polypeptide chain and its environment is altered, a structural change may occur producing a non-native or denatured conformation. The term "denatured" encompasses all non-native states, and includes intermediate states as well as fully unfolded protein. Denaturation can occur in response to changes in temperature, an increase in pressure, pH changes, addition of chemical denaturants to the solvent (including urea, guanidinium chloride and organic solvents), removal of cofactors required for stability, or changes in the amino acid sequence.

In most single-domain proteins, denaturation occurs as a reversible, cooperative two-state transition between two macroscopic states, the native (N) and an unfolded (U) species. Privalov (reviewed in 1979 and more recently in 1992) provides a thermodynamic analysis of protein unfolding gained from thermal denaturation studies. Prior to unfolding, conformational change is limited to minor changes such as an increase in chain flexibility and local alterations in structure. During this period, the heat capacity of the protein increases in a slow, temperature dependent manner. Unfolding occurs as a cooperative transition over a limited range of experimental conditions. In small proteins (less than 20 kDa) denaturation produces a sharp peak in the heat absorption, which coincides with the midpoint of the transition. Following denaturation, the heat capacity of the protein again increases slowly and in a temperature dependent fashion. Experimentally, this two-state transition is characterized by the coincidence of sigmoidal unfolding curves as measured by different conformational probes, a lack of observable unfolding intermediates, and approximately equivalent calorimetric enthalpy and van'tHoff enthalpy (a measurement of the temperature dependence of the unfolding transition equilibrium constant). Privalov has suggested that the cooperative unfolding transition involves the breaking of intramolecular bonds which stabilize the compact native structure, to form an unstable transition state, followed by hydration of internal groups within the protein.

It was originally thought that the heat capacity increase observed upon protein unfolding was dependent on the method of denaturation employed, and that complete

unfolding occurred only under highly denaturing conditions. However, Privalov has shown that if the effects of preferential binding of protons or of denaturant molecules are taken into account, no difference is observed in the increase in heat capacity as caused by temperature shift, pH change or chemical denaturants (Pfeil & Privalov, 1976, Pfeil et al., 1986). Earlier studies suggested that differences in the conformation of denatured and fully unfolded proteins cannot be distinguished thermodynamically (Privalov, 1979; 1992), and that the denatured state is best represented as an ensemble of interconvertible molecular conformations, exhibiting similar free energy values, within the same macroscopic state. More recently, analyses of proteins which thermally unfold through a molten globule intermediate state have indicated that changes in heat capacity can be measured for the transitions between native and intermediate and between intermediate and unfolded states (Carra, Griko & Privalov, 1995). In the three proteins studied, the first transition was cooperative and involved a significant increase in entropy, however the second transition showed variable cooperativity and heat absorption. Privalov suggests that the sizable increase in heat absorption observed upon unfolding of certain intermediates correlates to independent unfolding of intact domains. Unfolding of a true molten globule involves a small heat absorption and does not constitute a first order reaction (Carra, Griko & Privalov, 1995).

Tanford's classic denaturation experiments suggested that an unfolded protein will adopt the hydrodynamic properties of a random coil (Tanford, 1968). Recent data indicates that unfolded proteins exist in a variety of conformations, some compact, others more expanded (reviewed in Dill & Shortle, 1991). Even under strong denaturing conditions such as 6 M GdnHCl or 9 M urea some proteins may still retain significant residual structure (Shortle & Meeker, 1989; Shortle & Abergunawardana, 1993). As well, mutagenic analysis of proteins has indicated that the primary sequence of a protein can affect the unfolded conformation (Shortle & Meeker, 1986; Shortle et al., 1990; Flanagan et al, 1993). Comparison of the refolding of the β2 subunit of tryptophan synthase from denaturation in 6M GdnHCl or at acidic pH indicates that while identical molecular events occur with the same kinetics, the refolding pathways involve intermediates with different structures (Murray-Brelier & Goldberg, 1989), suggesting that the denatured conformation

will affect the mechanism of refolding.

Dill and Shortle have used heteropolymer theory, where the polypeptide chain is modelled as a polymer linking polar and nonpolar monomers, to predict changes to the conformation of the polypeptide upon denaturation (reviewed in Dill & Shortle, 1991). Any solvent will be a good solvent for some amino acids within the chain but a poor solvent for others, and the free energy of interaction will vary between the solvent and different residues. Water is a poor solvent for proteins which, on average, contain between 25-50% nonpolar residues. Addition of a denaturant such as 6 M GdnHCl transforms water into a good solvent for most proteins. However unfolding transitions usually occur at lower denaturant concentrations, ie, 1-4 M GdnHCl, where the aqueous solvent remains poor. Under these conditions, the unfolding transition is predicted to produce a compact denatured species. Increases in denaturant concentration beyond the unfolding transitional midpoint are predicted to result in gradual expansion of the species, as the solvent improves. Heteropolymer theory predicts that protein volume and behaviour upon denaturation is dependent on the hydrophobicity of the polypeptide chain.

1.2.3 Compact intermediates

While many proteins unfold via a two-state mechanism, other proteins, in particular those containing discrete domains, unfold through one or more intermediate states (reviewed by Kim & Baldwin, 1990; Ptitsyn, 1992; Fink, 1995). In several cases, intermediates observed in denaturation studies are similar to species identified on the folding pathway, fuelling an interest in their structure as possible models of transient folding intermediates.

The molten globule is the name given by Ohgushi and Wada (1983) to the acid-denatured state of cytochrome *c*. This intermediate is compact and retains native-like secondary structure and fluctuating tertiary structure. There is no unique conformation for compact intermediates and this lack of uniformity has contributed to the controversy surrounding the importance of compact intermediates to protein folding. This is compounded by the fact that a protein may form structurally distinct intermediates when exposed to different experimental conditions. As well, an unfolding reaction at equilibrium

may contain more than one species, i.e. both native and intermediate or both native and unfolded, leading one to question what species is being characterized. Others have suggested that compact intermediates may represent misfolded, off-pathway species with questionable relevance.

Despite these uncertainties, a list of key properties characterizing compact intermediates can be assembled (Fink, 1995). First, they retain significant, often native, levels of secondary structure. Second, much of the native side-chain packing has been lost, and in some cases the tertiary structure is disrupted. However, Peng & Kim (1994) have recently shown that the α helical domain of α-lactalbumin retains its tertiary fold even in the absence of tight packing. Third, the protein is in a collapsed conformation relative to the fully unfolded state, although less compact than the native state. Fourth, the hydrophobic core of the protein is retained although there is substantial exposure of hydrophobic surfaces, as measured by binding of hydrophobic dyes. As well, the intermediate is increasingly prone to aggregation. Fifth, the intermediate does not retain biological function. Sixth, there is a variable amount of structure and compactness which is dependent on the protein sequence and the experimental conditions. Seventh, the heat capacity of the intermediate is often similar to that observed for the native state. Finally, the transition from intermediate to an unfolded state is less cooperative than that observed from native to unfolded.

In contrast to Privalov's position that the intermediate and the unfolded protein are part of the same thermodynamic state (outlined in section 2.2), Ptitsyn proposes that the molten globule represents a separate state (Ptitsyn & Uversky, 1994). Thus denaturation produces an all-or-nothing transition between the native state and the molten globule, and unfolding involves another first order transition between the intermediate and the unfolded state. Ptitsyn suggests that the bimodal distribution of protein elution volumes upon GdnHCl-induced unfolding of a molten globule, and the molecular mass dependence of slopes for native-to-molten globule and molten globule-to-unfolded transitions provide evidence that both transitions are first order in nature (reviewed in Ptitsyn et al., 1995).

Fink (1995) has described a model for denaturation via intermediates which is

based on the two following assumptions. First, there are two types of tertiary interactions; one between autonomously folding units (corresponding to structural subdomains) and the second within these same folding units. Second, the initial expansion to a compact intermediate involves significant hydration of the protein interior. Compact intermediates arise when the hydrophobic interactions between subdomains are disrupted, allowing solvent penetration between folding units. Native-like secondary structure and tertiary structure is limited to within the subdomains, and coupled to an expansion in the radius of the protein. Less compact intermediates form when subdomains unfold, as dictated by their intrinsic stabilities. This model is remarkably similar to the modular, cooperative unfolding and refolding reactions described for cytochrome *c* by Englander's group (Bai et al., 1995). In contrast, Ptitsyn, suggests that unfolding involves two separate transitions coinciding with loss of tertiary packing followed by loss of tertiary fold (Ptitsyn et al., 1995). While side-chain packing may loosen upon formation of the intermediate, the solvent remains excluded from the hydrophobic core, and the volume increase arises as a consequence of unfolding in the rest of the protein.

While the properties of compact intermediates are still controversial, these intermediates are found in an increasing number of folding and unfolding pathways.

1.2.4 Protein folding pathways

Anfinsen (1973) suggested that protein folding initiates at nucleation sites on the unfolded polypeptide chain. These nucleation sites transiently adopt their native conformation. Stabilization of these structures could result from formation of specific interactions, formation of secondary structure, hydrophobic collapse, or some combination of the three (reviewed in Matthews, 1993). In order to understand early folding events, investigators have sought to determine the structure and properties of the intermediates which appear in the first milliseconds of folding.

Perhaps the best characterized folding pathway is that of bovine pancreatic trypsin inhibitor (BPTI). Creighton (1988) was able to correlate the order of disulphide bond formation of BPTI with the choice of folding pathway. This suggested that a specific interaction, in this case disulphide bond formation, could initiate and direct protein

folding, but did not rule out the formation of secondary structure or hydrophobic collapse prior to disulphide bond formation.

Stopped-flow CD spectrometry has indicated that several proteins regain significant secondary structure during an initial burst phase occurring within the deadtime of mixing (cytochrome *c* and β-lactalbumin, Kuwajima et al., 1987; dihydrofolate reductase, Kuwajima et al., 1991; staphylococcal nuclease, Sugawara et al., 1991; β_2 subunit of tryptophan synthetase, Chaffotte et al., 1992). Likewise quench-flow NMR studies measuring hydrogen exchange of backbone amides indicate that significant protection of amides within helical segments of several proteins occurs within the first 10 ms of refolding. Such is the case for cytochrome *c* (Roder et al., 1988), barnase (Bycroft et al., 1990; Matouschek et al., 1992), T4 lysozyme (Lu & Dalhquist, 1992) and hen lysozyme (Radford et al., 1992). β-Sheet formation has also been observed in this time range for ribonuclease A (Udgaonkar & Baldwin, 1990), barnase (Bycroft et al., 1990) and T4 lysozyme (Lu & Dalhquist, 1992), as has protection of some amides in the β-sheet domain of hen lysozyme (Radford et al., 1992).

Hydrophobic collapse has been implicated in stabilizing transient elements of secondary structure in the burst phase. Stopped-flow fluorescence studies on the binding of a hydrophobic dye, 8-anilino-1-naphthalene sulphonate (ANS), indicate that non-polar surfaces are present in several burst phase intermediates, including those of carbonic anhydrase and human α lactalbumin (Semisotnov et al., 1991), dihydrofolate reductase (Jones et al., 1994), the α subunit of tryptophan synthase and the trp aporepressor (Matthews, 1993). Shastry & Udgaonkar (1995) have reported that refolding of barnase in 1 M GdnHCl produces a burst phase intermediate which binds ANS but shows no evidence of secondary structure, suggesting that hydrophobic collapse may precede secondary structure formation, at least in this instance. ANS fluorescence increases as hydrophobic regions are sequestered from the solvent (Goldberg et al., 1990; Semisotnov et al., 1991; Ptitsyn et al., 1990) indicating that ANS is binding non-specifically to hydrophobic regions which subsequently become buried as folding proceeds.

Several investigators have reported the formation of partially folded species in intermediary time frames, between 5 and 1000 ms (reviewed in Matthews, 1993). The

available data suggests that these intermediates are more stable than their earlier precursors and can contain native-like secondary and some tertiary structures which are retained in the native protein (Udgakonkar & Baldwin, 1990; Radford et al., 1992). In some instances these intermediates can be recognized by monoclonal antibodies (Goldberg et al., 1990) or can bind ligands (Frieden, 1990; Herold & Leistler, 1992), suggesting the presence of tertiary folds corresponding to structural subdomains and domains. However, these structural elements are not properly docked and the side-chains are more mobile than in their native conformation.

Excluding proline isomerization, disulphide bond formation and association with a required prosthetic group, the rate-limiting step in protein folding is generally the final one and involves the development of native side chain packing and native hydrogen bonds throughout the protein (reviewed in Matthews, 1993; Dobson et al., 1994). It is at this stage that conformational entropy is sacrificed so that the noncovalent forces which stabilize the closely packed native state can form. Investigators have used mutagenic analyses and ligand binding studies to determine the barriers which precede formation of the transition state. Beasty et al. (1986) established that the rate limiting step in formation of the α subunit of tryptophan synthase involves the association of two folding units. To determine the structure of the folding transition state of barnase, Fersht and colleagues have compared the effects of extensive amino acid replacements on the stability of a late-folding intermediate and the native conformation of barnase (reviewed in Fersht, 1993). Kuwajima and colleagues (1989) observed that calcium binds weakly to a molten globule-like intermediate of α lactalbumin, more tightly to the rate limiting transition state and most tightly to the native protein. Meanwhile, staphylococcal nuclease does not bind calcium prior to attaining its native conformation (Sugawara et al., 1991). This difference has been attributed to differences in the structure of the calcium binding site. In α lactalbumin, the calcium binding site is formed from a continuous region of the polypeptide chain, while the site in the nuclease requires the rearrangement of noncontiguous segments (Matthews, 1993).

A detailed study by Dobson and coworkers (1992) of different steps on the folding pathway of hen lysozyme indicate that a burst phase in the appearance of far UV CD

precedes the protection of amide protons from solvent exchange. This suggests that the first step in folding is acquisition of transient secondary structure, within the first 2 ms. Stopped-flow absorption studies (Kuwajima et al., 1985; Ikeguchi et al., 1986) indicate that this is coupled to hydrophobic collapse. The next phase in folding involves protection of the α domain amides in 50% of the protein and protection of the β domain amides in 30% of the protein (Radford et al., 1992). The remainder of the protein folds more slowly. Different parts of hen lysozyme become stabilized at different rates. In particular, folding of the α-helical domain generally precedes development of specific tertiary structure, as measured by appearance of near UV CD signal, and protection of amides in the β-sheet domain. The rate-limiting step in the folding of hen lysozyme is thought to involve improving side chain packing and rearrangement of the α-helical and β-sheet domains (Itzhaki et al., 1994).

The refolding of hen lysozyme demonstrates the tenets of several different folding theories. The appearance of transitory secondary structure and hydrophobic collapse precede formation of specific tertiary structure as predicted by the frame work, diffusion-collision and hydrophobic collapse models of initiation of folding (Lewis et al., 1970; Karplus & Weaver, 1979; Rose & Roy, 1980). The independent folding of the α-helical and β-sheet domains of lysozyme (Miranker et al., 1991) coincides with the modular assembly model which suggests that proteins fold by parts and (sub)domains might serve as folding intermediates. Finally, the observance of different folding pathways indicates that parallel alternative pathways of folding can produce the same native structure, as suggested by the jigsaw model (Harrison & Durbin, 1985), and different from a simpler sequential model. The presence of multiple parallel folding pathways has also been noted in the refolding of dihydrofolate reductase (Jennings et al., 1993) and RNaseA (Li et al., 1995) among others.

To summarize, folding of small proteins in vitro is a relatively quick process, occurring within minutes of removal from denaturants. Factors such as prolyl isomerases, disulphide isomerases and chaperones can increase the rate of folding by catalyzing slow reactions (isomerases) or inhibiting nonspecific aggregation (chaperones). Protein folding pathways likely involve elements found in both sequential and nonsequential models.

Matthews (1993) has described the folding pathway as a funnel where each step is followed by a slower reaction which allows a thermodynamic equilibrium to form between the conformations accessible at that stage. The transitions between different classes of intermediates may involve significant activation barriers such as those involved in disulphide formation, proline isomerization or rearrangement of tertiary structure, and may represent molecules becoming trapped in local energy minima (Wolynes et al., 1995). As folding progresses, the absolute number of available conformations decreases while intermediates with increasing stability are formed and transition reactions become slower.

1.2.5 Folding of multidomain and oligomeric proteins

Folding by parts or modular folding has been proposed to play a role in the rapidity of protein folding. The independent folding of domains in multidomain enzymes should increase the overall rate of protein folding while decreasing the opportunity for nonspecific polypeptide interactions which would result in misfolded or aggregated protein. Folding by parts is therefore a necessary mechanism for the efficient folding of large multidomain proteins (reviewed in Jaenicke, 1987; 1991).

One approach to studying the role of domains in protein folding has been to characterize the isolated component domains and determine their ability to refold and reassociate. The PLP coenzyme binding domain of aspartate aminotransferase can be expressed independently (Herold et al., 1991). While the full length enzyme is dimeric, the domain exists as a monomer but still binds PLP at one site with high affinity. Denaturation of this domain produces a compact equilibrium intermediate, paralleling the unfolding of the dissociated monomer of aspartate aminotransferase. Likewise, a stable dinucleotide binding domain can be isolated from D-glyceraldehyde 3-phosphate dehydrogenase (Jecht et al., 1994). This monomeric domain continues to bind coenzyme strongly and undergoes the same GdnHCl-induced unfolding transition as the native tetrameric dehydrogenase.

Proteolysis of gamma II-crystallin produces an N-terminal domain which is conformationally similar to that of the native protein (Sharma et al., 1990; Rudolph et al., 1990). This domain is more sensitive to denaturation, presumably because it lacks

stabilizing interdomain interactions. Unfolding/refolding kinetics of the isolated N-terminal domain indicate a bimodal equilibrium transition which coincides with that described for the second phase of the three state unfolding/refolding of the intact protein. This supports the authors proposal that folding of gamma II-crystallins proceeds by the sequential folding of the N- and C-terminal domains. The isolated C-terminal domain shows low intrinsic stability (Mayr et al., 1994).

Proteolysis of the β_2 subunit of tryptophan synthase has shown that it is composed of an N- and a C-terminal domain (named F1 and F2, respectively) separated by a hinge region. These fragments can be isolated and will refold independently. Blond & Goldberg (1986) have shown that the N-terminal domain rapidly acquires native-like structure. As well, non-specific hydrophobic collapse promotes the rapid folding of the C-terminal F2 fragment, producing a condensed non-native structure in less than 4 ms (Chaffotte et al., 1991; 1992). Folded F1 and F2 fragments will associate to form a nicked β subunit, analogous to formation of interdomain interactions in the native subunit, prior to subunit association. Association of the domains involves a slow conformational rearrangement of the N-terminal domain in response to interactions with the C-terminal domain within the same chain (Blond & Goldberg, 1986). Two β subunits will rapidly associate and undergo a second slow isomerization step to form the native β_2 dimer.

Yon and colleagues have analyzed the unfolding and refolding of the isolated and covalently linked N- and C- terminal domains of yeast phosphoglycerate kinase (1990). Under most denaturing conditions phosphoglycerate kinase unfolds in a two-state transition. Refolding of the isolated N-terminal domain of this enzyme occurs very quickly. This has been observed for isolated domains from other multidomain proteins (Garel, 1992) and emphasizes that domain pairing is often the rate limiting step in folding. The authors proposed that the refolding of the C-terminal domain includes a slow step, perhaps proline isomerization. In the full length phosphoglycerate kinase this slow step is masked by the even slower rearrangement of domains, as has also been described in the refolding of octopine dehydrogenase (Teschner, Rudolph & Garel, 1987).

Although the domains of phosphoglycerate kinase are known to interact, the isolated domains do not reassociate, even after refolding together. Minard et al (1989)

suggested that the noncovalent forces remaining after separation of the two domains of phosphoglycerate kinase could not counteract the increase in entropy occurring upon cleavage of the hinge, making domain association unfeasible. When the polypeptide is split within a domain, the two resulting fragments will complement each other to form an associated and active phosphoglycerate kinase, presumably as a result of the strength of intradomain interactions (Pecorari et al., 1993). In a review of folding of multidomain proteins, Garel (1992) has noted that the stability of isolated domains is generally not substantially modified by the presence of the rest of the protein chain. He concluded that domain stability results primarily from intradomain interactions.

These and other studies have lead to the following conclusions regarding the folding of multidomain proteins. Early steps in the folding of multidomain proteins are analogous to those observed for folding of single domain proteins. However domains within a protein may attain stable tertiary structure at different rates, depending on their different kinetic parameters, and requirement for stabilizing interactions with other regions of the polypeptide chain. Domain pairing involves minimizing the accessible surface area of the protein through the docking and association of domain surfaces. This generally involves tertiary rearrangements and desolvation to produce a recognition or interaction surface, and often is the rate-limiting step in refolding of multi-domain proteins.

As multidomain proteins are composed of domains which represent independent cooperative units, they often display multiphasic transitions, with stable intermediates observed in their folding and unfolding pathways. These intermediates may contain domains with different levels of folding. Wu, Peng & Kim (1995) have recently used disulphide variants of α-lactalbumin to show that molten globule properties need not encompass the entire polypeptide chain and may be limited to a single domain only. Specifically, the structure of the A-state of α-lactalbumin includes an α-helical domain with native-like secondary structure and tertiary fold, but missing extensive tertiary interactions, while the β-sheet domain is unfolded.

In some multidomain proteins, the domains unfold independently and show little or no evidence of interacting (ovomucoid, Privalov, 1982; domains of plasminogen, Privalov & Potekhin, 1986; cytoplasmic and transmembrane domains of band 3, Brandts

et al., 1989). Mutual stabilization through domain pairing is probably insignificant in these proteins. Other proteins may exhibit very strong domain interactions, where multiple domains unfold co-operatively in a single transition, with no evidence of equilibrium intermediates (heat denaturation of yeast phosphoglycerate kinase in the presence of guanidine hydrochloride, Griko et al., 1989). This level of interaction is no doubt necessary to allow domains to efficiently function together. A third class of proteins displays an intermediate level of domain interaction (serum transferrin, Lin et al., 1994; CD4, Tendian et al, 1995; DnaK, Montgomery et al., 1993).

Presuming that the extent to which partially folded states become stably populated is dependent on the strength of interactions between domains, Freire and colleagues (Freire & Murphy, 1991; Freire et al., 1992) have developed a hierarchial algorithm which makes use of the crystallographic structure of a given protein, as well as thermodynamic measurements, to identify the cooperative folding units within the protein and estimate the relative population of folding intermediates. To this end, Freire and colleagues (1992) have classified some of the structural mechanisms which may result in cooperative unfolding of domains. Unfolding of a domain may result in the exposure of hydrophobic residues, both within the domain that is unfolding and the rest of the protein. While exposure of the apolar residues in the unfolded domain is compensated by the free energy of unfolding, the remainder of the protein is not compensated by an increase in entropy. This constitutes the source of cooperativity in a hydrophobic interaction. Noncovalent interactions including hydrogen bonds, van der Waals contacts, and salt bridges are often found at the interface between domains. The unfolding of one domain is sufficient to break these interface bonds, which may trigger cooperative unfolding of the second domain. Likewise, if the domain interface includes a ligand binding site, and unfolding of one domain results in dissociation of the ligand, this too may destablize the second domain resulting in cooperative unfolding. Brandts et al., (1989) have developed a model for determining the free energy of interaction between domains or subunits from scanning calorimetry data.

The first step in denaturation of an oligomeric protein is often dissociation, as observed in the denaturation of phosphofructokinase (Parr & Hammes, 1975), creatine

kinase (Yao et al., 1984), aminoaspartate transaminase (Herold & Kirschner, 1990), and pyruvate oxidase (Risse et al., 1992). Coincidence of inactivation and dissociation suggests that the oligomer is required for catalytic activity. Sometimes a conformational change precedes dissociation producing an inactive oligomeric species, as seen in the GdnHCl-induced denaturation of glutathione transferase B1 (Sacchetta et al., 1993).

In a few instances, dissociation of oligomeric proteins produces active intermediates. For example, GdnHCl denaturation of tetrameric aspartase produces an active dimeric intermediate (Murase et al., 1993). Dissociated forms of phosphoglycerate mutase (Hermann et al., 1981), alanine racemase (Toyama et al., 1991) and creatine kinase (Grossman et al., 1981) also retain at least partial activity. The dissociated form must contain a complete active site.

Dissociation studies can also provide information about the relative stability of subunit interfaces. Unfolding of aspartase indicates that the native tetramer is better described as a dimer of dimers where the interface between dimers is weaker than the interface within the dimer (Murase et al., 1993). Likewise denaturation of glutamine synthetase, a dodecamer arranged as two stacked hexameric rings, produces even-numbered species which retain the subunit interface between two rings, showing that the inter-ring interface is more stable than that formed between subunits of the hexamer. Sometimes an unfolding intermediate will retain the native quaternary structure. The acid-induced unfolding of creatinase produces a dimeric molten globule (Schumann & Jaenicke, 1993), as does the thermal denaturation of acetylcholinesterase (Kreimer et al., 1995). In both examples, subunit interactions are proposed to play a role in stabilizing the intermediate.

While many oligomeric proteins denature through specific intermediates, others unfold via a two-state transition, as described for the trp repressor (Gittelman & Matthews, 1990), the arc repressor (Bowie & Sauer, 1989), and the λ cro dimer (Pakula & Sauer, 1989). The tetrameric oligomerization domain of p53 undergoes thermal denaturation in one step (N_4 to 4U) (Johnson et al., 1995). 58% of the total hydrophobic surface area of this domain is buried at the subunit interfaces, suggesting that the rapid unfolding occurs with dissociation because the domain is stabilized by intersubunit

interactions as opposed to intrasubunit interactions.

Quaternary interactions between subunits are very specific. Cook and Koshland, (1969) have shown that the presence of other protein species does not interfere with renaturation. While hetero-oligomers have been renatured from mixtures of highly homologous proteins from different species (triosephosphate isomerase, Sun et al., 1992), cross-hybridization appears to be an exception rather than the rule (Cook & Koshland, 1969; Jaenicke, 1987). The denaturation state of the protein and the method of renaturation influence the yield of hybrid species (Lehrer & Quian, 1990; Brown & Scachat, 1985), indicating that the choice between symmetric and asymmetric associations may be a kinetic one (Jaenicke, 1987; Garel, 1992).

In most proteins, extensive folding is required to produce a binding surface prior to subunit association. For example, the appearance of the dimerization site of tetrameric AK-HDH (aspartokinase-homoserine dehydrogenase) occurs at the same time as the kinase active site (Garel & Dautry-Varst, 1980) and the β tryptophan synthase subunit dimerizes after formation of several native epitopes (Blond & Goldberg 1987). Garel (1992) has suggested that formation of a subunit binding site requires the same degree of folding required to form a catalytic or ligand binding site. In contrast, the dimerization site of the tryptophan repressor, an entwined dimer, appears to form early in the refolding pathway (Gittelman & Matthews, 1990; Tasayco & Carey, 1992).

The folding pathway of an oligomeric enzyme is a succession of monomolecular folding steps and bimolecular association steps, yielding a general kinetic scheme where either first or second order processes may be rate limiting (Jaenicke, 1982):

$$n\ M^* \rightarrow n\ M \rightarrow n/2\ D^* \rightarrow n/2\ D \rightarrow n/4\ T^* \rightarrow n/4\ T.....$$

where M, D and T represent monomer, dimer and tetramer respectively, and the asterisk refers to conformationally distinct forms. The overall rate of folding of oligomeric proteins is dependent on protein concentration because, while the folding reactions are not dependent on protein concentration, the association events are. Thus by lowering protein concentration, one can enrich for pathway intermediates with different degrees of association. At high protein concentrations, association is no longer rate-limiting, and the pathway should shift to first order. The experimentally useful range of protein

concentrations is limited by the sensitivity of detection at low concentrations of protein and by aggregation at higher concentrations. This may prevent observation of the shift in the rate-limiting step from first to second order.

This shift can be observed in the refolding of bacterial luciferase, a heterodimeric enzyme consisting of an α and a β subunit. When the protein concentration is above 20 μg/ml, the refolding of luciferase from 5 M urea includes several slow steps involving refolding of each subunit before dimerization, and the isomerization of an inactive heterodimeric intermediate to the active species following dimerization (Ziegler et al., 1993). Below 20 μg/ml, a second order association step is rate limiting.

The refolding and association of luciferase also illustrates how the presence of a polypeptide can modify the folding pathway of a second polypeptide (Waddle et al., 1987). When the α and β subunits are expressed or refolded separately, both subunits refold to form tight binding homodimers which will not associate with one another, suggesting that formation of the heterodimer operates as a kinetic trap on the folding pathway of each subunit.

The early folding pathway of individual subunits of most oligomeric proteins resembles that of any single or multidomain protein, except that folding results in production of a monomer with an exposed surface, competent for association (Jaenicke, 1987). The interface surface and the resulting interactions formed upon subunit association are analogous to those used to secure interacting domains (Jaenicke, 1991). Like multidomain proteins, folding and association of oligomeric proteins must be coordinated to prevent aggregation through an exposed interface surface. Similarly, slow conformational changes which produce the interface surface are often observed prior to association, and following association to stabilize the dimer or allow further oligomerization. Folding of oligomers that are more complex than dimers involves at least two association events, one of which is usually rate-limiting. The β tryptophan synthase dimer undergoes a slow folding step prior to association with the α tryptophan synthase dimer (Blond & Goldberg, 1985). During refolding of tetrameric lactate dehydrogenase, a dimeric intermediate, which precedes the rate-limiting formation of tetramers, accumulates (Jaenicke, 1987). The rate-limiting step in the folding of tetrameric AK-HDH

is the formation of dimers; all subsequent events occur at the same rate (Martel & Garel, 1984). When reactivation of an enzyme and a rate-limiting second order step are observed to coincide, it suggests that the monomer is inactive (Rudolph & Jaenicke, 1976; Hermann et al., 1981). Activity may only be observed following formation of a catalytic site across the dimer interface, or upon adoption of the native conformation as determined by subunit interactions present in the oligomer, as discussed previously in section 1.1.6.

1.3. MULTIFUNCTIONAL ENZYMES IN ENZYME ORGANIZATION

The principle of hierarchial organization continues to be observed at the level of the cell. The structural complexity of the cell necessitates some form of enzyme organization for the efficient functioning of metabolic pathways. Enzyme organization results in the formation of microenvironments, defined by Welch and Marmillot (1991) in the following manner:

"...a subvolume of the cellular space in which the thermodynamic/kinetic character of a metabolic process is different from the hypothetical situation with the reacting components homogeneously 'dissolved' in the bulk cellular space"

The following pages provide a review of some of the mechanisms and consequences of enzyme organization, with emphasis on the formation and structure of multifunctional enzymes.

1.3.1 Mechanisms of enzyme organization

Within the cell, metabolic pathways are compartmentalized into subcellular organelles, separated by membranes. For example, the enzymes involved in the Krebs cycle have been localized to the mitochondrial matrix, while the enzymes of oxidative phosphorylation are found in the inner mitochondrial membrane (reviewed in Srere, 1985). Compartmentalization provides several advantages to the cell. As the membrane limits the entry and exit of molecules, specific chemical environments can be maintained within the compartment. By limiting the diffusion volume available to enzymes and their substrates and cofactors, their effective concentrations are increased with fewer molecules.

Enzymes may localize to specific areas of the cytoplasm by binding to subcellular structures, such as cytoskeletal elements or membranes. Glycolytic enzymes bind to band 3 protein of erythrocytes and to F-actin in the I band of skeletal muscle fibres (Friedrich, 1985). Isoprenylated rhodopsin kinase translocates to the plasma membrane, where it binds to photon-activated transmembrane rhodopsin (Inglese et al., 1992). Like subcellular compartmentation, binding to subcellular structures serves to increase the local concentrations of enzymes and substrates. As well, it limits diffusion of enzymatic

components to only one or two dimensions (Srere, 1987). As proteins may reversibly bind to these elements, localization can provide an effective method of cellular regulation.

Multienzyme complexes, also known as metabolons, are formed by the noncovalent association of distinct enzymes encoded on different polypeptide chains (Welch & Gaertner, 1980; Srere, 1987). Metabolons have been observed in both anabolic and catabolic pathways, including DNA and RNA synthesis, protein synthesis and degradation, and the metabolism of precursor molecules. Some multienzyme complexes are very stable, persisting for the lifetime of the protein components, and may even be purified intact. The high molecular weight pyruvate dehydrogenase is a stable multienzyme complex consisting of multiple copies of at least three different enzymes: pyruvate decarboxylase (E1), dihydrolipoyl transacetylase (E2), and lipoamide dehydrogenase (E3) (Patel & Roche, 1990; Perham, 1991). Complexes from various sources shows significant differences in both structure and subunit composition. For example, the mammalian version includes E1 specific kinase, phospho-E1 phosphatase and protein X as well as the three common components. In all cases, the complex is highly organized and shows considerable symmetry.

Other complexes display weaker interactions and may be more dynamic, existing only transiently. They may form only in the high protein concentration environment of the cell, in the presence of a specific metabolite, or in response to a particular metabolic state. The eukaryotic replitase (Reddy & Pardee, 1980) is a dynamic association of several enzymes involved in DNA replication (including the DNA polymerase complex, dCMP kinase, nucleoside diphosphokinase, ribonucleotide reductase, thymidylate synthase and thymidine kinase), that preferentially channels dNDP's into DNA synthesis. It has only been observed during the S phase of the cell cycle (Reddy & Fager, 1993). Enzymes governing DNA synthesis in prokaryotes may also form complexes (Flanegan & Greenberg, 1977; Allen et al., 1983)

Evidence for the association of glycolytic enzymes into a dynamic metabolon includes studies showing specific interactions between sequential pairs of glycolytic enzymes and between enzymes and subcellular structures, isotope experiments and visualization of complexes by electron microscopy (reviewed in Srere, 1987). Investigators

continue to debate the role of this mechanism in controlling the flux of intermediates through glycolysis, suggesting instead that random, non-specific interactions may be involved. Advocates of random association point to the misinterpretation of kinetic data to support their views, (reviewed in Batke, 1991). On the other hand, Srere (1987) supports the proposed associations, arguing that since glycolysis is less processive than other biosynthetic pathways, and includes intermediates that have multiple metabolic roles, a loose association of glycolytic enzymes may exist. These dynamic complexes would be difficult to detect and isolate but better suited to the needs of the cell.

Multifunctional enzymes (reviewed by Kirschner and Bisswanger, 1976) consist of a single type of polypeptide chain with multiple catalytic functions. In most cases, the catalytic activities are part of the same metabolic pathway although not necessarily sequential reactions. Multifunctional enzymes are more commonly found in (but not limited to) eukaryotes. A trend towards condensation of related activities appears widespread throughout evolution. Prevailing theory suggests that most multifunctional enzymes arose by fusion of the genes encoding pre-existing functional domains (Janin & Wodak, 1983). Fusion products which provided an advantage to the cell were selected for and maintained. In some cases the protein may have undergone several gene fusion events, producing a polypeptide with several enzyme activities, covalently linked by connecting regions. The evolutionary history of a multifunctional enzyme can often be surmised by comparing the distribution of its component activities in nature, phylogenetic analysis of the amino acid sequences encoding these activities, and analysis of the domain structure of the enzyme. Examination of the structural organization of most multifunctional enzymes discloses a modular arrangement.

1.3.2 The domain structure of multifunctional enzymes

The bifunctional enzyme indole phosphate synthase-phosphoribosyl anthranilate isomerase (IGPS:PRAI) catalyzes two reactions in the tryptophan biosynthetic pathway. Where the activities are covalently linked in *E. coli*, PRAI exists as a monofunctional enzyme in yeast. The two domains of IGPS:PRAI were expressed as monofunctional enzymes, in order to determine if ligation of these two activities provides a selective

advantage (Eberhard, et al., 1995). Previously, the crystal structure of IGPS.PRAI (Wilmanns, et al., 1992) had revealed eight putative hydrogen bonds and several hydrophobic contacts between domains. However, the isolated domains no longer interact, indicating that covalent linkage is essential for interdomain interaction (Eberhard et al., 1995). The separated domains and the bifunctional enzyme share similar catalytic efficiencies. As IGPS:PRAI shows no evidence of substrate channelling, Eberhard et al. have suggested that fusion of the PRAI domain may stabilize the more labile IGPS activity.

The pentafunctional AROM protein catalyzes five sequential reactions in the conversion of 3-deoxy-D-arabino-heptulosonic acid-7-phosphate to 5-enol pyruvyl shikimate 3-phosphate in the prechorismate section of the shikimate pathway of fungi and yeast (reviewed by Hawkins et al., 1993). Monofunctional prokaryotic versions exist for each of the activities. Four of the five eukaryotic activities have been expressed independently, with varying results. Both the N-terminal dehydroquinate (DHQ) synthase and the 3-dehydroquinase (fourth domain) can be independently expressed. While the C-terminal shikimate dehydrogenase domain is refractory to independent expression, it can be coupled to the 3-dehydroquinase to produce a stable bifunctional enzyme. The 5-enolpyruvyl 3-phosphate (EPSP) synthase (second domain) cannot be separately expressed as an active enzyme but a bifunctional DHQ synthase-EPSP synthase is active. Thus the AROM protein appears to be divided into two halves, one comprising the adjacent DHQ synthase and EPSP synthase, the second including the shikimate kinase, 3-dehydroquinase, and shikimate dehydrogenase (domains three, four and five). Presumably critical domain interactions occur within each half of the enzyme.

Several steps in nucleotide metabolism are catalyzed by multifunctional enzymes. The first three steps in de novo pyrimidine metabolism are catalyzed by a multifunctional CAD enzyme in mammals (reviewed by Evans et al., 1993). This 243 kDa polypeptide includes glutamine-dependant carbamyl phosphate synthetase (CPSase), aspartate transcarbamylase (ATCase) and dihydroorotase (DHOase) activities, organized into separate domains. In yeast only the CPSase and ATCase activities are covalently linked while the three activities exist separately in prokaryotes. The structural organization of

this enzyme has been determined by sequencing two partial cDNAs which encode the entire protein and by analysis of the fragments produced by limited proteolysis. The CPSase and ATCase domains share a high degree of homology with their prokaryotic counterparts. Whereas the CPSase is composed of two different subunits (a smaller glutamine amidotransferase subunit which transfers an amino group to the active site of the larger CPSase subunit) in prokaryotes, the subunits are covalently linked in yeast and mammals. The dihydroorotase domain is quite dissimilar from monofunctional dihydroorotases, and was postulated to have evolved independently from the linker sequences separating the CPSase and ATCase domains. All three domains can be separately expressed as independent proteins, however the isolated DHOase domain exhibits lower V_{max} and higher K_m values, suggesting that interactions with other CAD activities may optimize its conformation.

Many multifunctional enzymes can be dissected, through proteolysis or protein engineering, into their component domains. Often the isolated domains function quite similarly to those within the intact enzyme, suggesting that most multifunctional proteins are arranged as independent modules linked by short connecting regions.

1.3.3 The role of linker regions in multifunctional enzymes

Connecting regions covalently link the domains within multifunctional enzymes. These regions are often hydrophilic, solvent-exposed sequences which are more susceptible to proteolytic cleavage than sequences within a domain. Linker regions vary in length from only a few amino acids to more than one hundred residues, and may be very flexible or rigid. Argos (1990) has reviewed the structure of linkers from different proteins (mostly between domains within monofunctional enzymes), and has concluded that most linkers are composed of mainly small hydrophilic residues in an extended conformation. The length of the connector sequences may be critical in some instances. In mammalian CAD, the DHOase domain is fused to the C-terminal end of CPSase with no apparent linker, however a hydrophilic 133 residue linker connects the C-terminal end of the DHOase domain to the ATCase domain (Evans et al., 1993). This linker has a phosphorylation site which is proposed to play a regulatory role, in switching between an

open and closed arrangement of CAD (Carrey, 1993). Deletion of this linker region (Guy & Evans, 1994) resulted in a functional CAD which exhibited kinetic parameters similar to that of the wildtype enzyme, however, the CPSase activity became more thermolabile. Guy and Evans suggested that removal of the linker may produce alterations in the juxtaposition of the domains and may prevent interaction between the ATCase and CPSase domains. Interestingly the overall number of residues separating the CPSase and ATCase domains is conserved between yeast (where the activities are part of a bifunctional enzyme) and mammals (Guy & Evans, 1994).

Crawford et al. (1987) used deletion mutagenesis to study the role of the linker in yeast tryptophan synthase. Removal of 18 interdomain residues produced an inactive enzyme, while replacing 14 of the missing residues with nonrelated amino acids restored partial activity, suggesting that length was the important determinant in this linker.

Other linker regions rely on composition to define their biophysical properties. 6-deoxyerythronolide-B synthase 2 and 3 (DEBS2 and DEBS3) are multifunctional enzymes with an internally duplicated modular structure similar to that of the vertebrate fatty acid synthases (Bevitt et al., 1992). Sequence analysis indicates that each half of the enzyme includes a 3-oxoacyl-synthase, acyltransferase, dehydratase (C-terminal half of DEBS2 only), enoylreductase, oxoreductase and acyl-carrier protein domains, connected by potentially flexible linkers, rich in alanine, proline and charged amino acids (Bevitt et al., 1992). ^{1}H-NMR and proteolysis studies of the *E. coli* pyruvate dehydrogenase E2 linkers (Radford et al., 1989; Perham & Packman, 1989, Turner et al., 1993) suggest that the Ala/Pro sequences common in the interdomain regions of the E2 component of multifunctional 2-oxoacid dehydrogenases provide conformational flexibility, facilitating domain movement and promoting interactions. An increase in either alanine or proline content at the expense of the other residue decreases flexibility (Turner et al., 1993). Within linkers in other proteins, the presence of proline-threonine combinations (endoglucanase A, Shen et al., 1991) or charged residues (ton B, Evans et al., 1986; Brewer et al., 1990) has been proposed to constrain the conformation of a linker region, producing a more rigid connector.

While linker regions are critical for the function of many multifunctional proteins,

their sequences are less conserved than the domains which they connect. This has been shown through comparison of homologous yeast and human D\C\S (Hum et al., 1988), tryptophan synthases from different species (Crawford, et al., 1987), and rat and chicken fatty acid synthases (Witkowski et al., 1991).

1.3.4 The oligomeric structure of multifunctional enzymes

Several multifunctional enzymes are oligomeric. Determining the quaternary structure of a multifunctional enzyme may provide further insight into its function and evolution. CAD exists as hexamers and other large oligomers (Coleman et al., 1977; Lee et al., 1985). Strong subunit interactions between the ATCase domains likely form a trimeric interface in CAD, as observed in the solved structure of the catalytic domains of the *E.coli* ATCase (Simmer et al., 1989, Krause et al., 1987). Furthermore, Carrey (1993) has observed that a proteolytic dihydroorotase domain exists as a dimer in solution, and that larger fragments containing both CPSase and DHOase activities crosslink as dimers. Therefore, the DHOase domain might provide an additional interface for further association of two trimers to a hexamer.

The homodimeric mammalian multifunctional fatty acid synthase, contains seven enzyme activities and an acyl carrier domain on a single polypeptide. The two subunits are probably arranged in a cyclic head to tail fashion, such that the oxosynthase activity of one subunit would act on the fatty acid chain bound to the acyl carrier domain of the other subunit (Witkowski, 1991). This arrangement could facilitate transfer of the fatty acid chain between different active sites. Yeast fatty acid synthase has a very different structure, $\alpha_6\beta_6$, in which the α subunit includes the oxosynthase and acyl carrier protein domains while the β subunit contains the remaining activities (Wakil, 1989). Crosslinking and cryo-electron microscopy experiments suggest that the α subunits of the yeast enzyme may also be arranged so that the oxosynthase interacts with the acyl carrier protein from an adjacent subunit (Wakil et al., 1983, Stoops et al., 1992)

Both the bifunctional D\C domain and the trifunctional folate interconversion enzyme exist as dimers in solution (Hum & Mackenzie, 1991), however the arrangement of subunits and the location of subunit interfaces is not known in detail. While the

eukaryotic 10-formyl$H_4PteGlu_n$ synthetase domain may or may not mediate a subunit interface, monofunctional synthetases are generally oligomeric. A dimeric monofunctional synthetase has been identified in spinach (Nour & Rabinowitz, 1991) and *Photobacterium phosphoreum* (Pawelek & MacKenzie, personal communication) and a tetrameric synthetase is found in *Clostridia* (reviewed in MacKenzie, 1984).

1.3.5 Advantages of enzyme organization

The association of components into multienzyme complexes and multifunctional enzymes can result in both catalytic and regulatory advantages. Coggins & Hardie (1986) have suggested several possible advantages, many of which result from efficient compartmentalization. These include catalytic facilitation, protection of unstable intermediates and sequestering of intermediates which might inhibit other reactions, co-ordinate regulation of enzyme activities, and substrate channelling. Several of these advantages were also proposed by Ovadi (1991) as benefits of substrate channelling and will be discussed in this context in section 1.3.6.

One advantage which is independent of channelling is the coordinate regulation of different enzyme activities by an effector molecule, as reported for *E. coli* aspartokinase I-homoserine dehydrogenase I. Threonine binding at a single allosteric site on this enzyme, produces a conformational change which affects both activities (Cohen & Dautry-Varsat, 1980).

Enzyme association may also result in noncatalytic advantages such as increased stability of the complex/enzyme. This may result from stabilizing interactions present in the close packing of the domains within a multifunctional enzyme or between the different subunits of a static multienzyme complex (Hardie & Coogins, 1986).

Multifunctional enzymes also have two advantages over multienzyme complexes. They do not require a genetic mechanism for coordinate expression of enzyme activities and they have no need to evolve interface surfaces for complex formation.

1.3.6 Substrate channelling in multienzyme complexes and multifunctional enzymes

Channelling is defined as the phenomenon where the product of one reaction is

transferred to the active site of the next enzyme without first equilibrating with the bulk solvent. This definition of channelling is deliberately vague, and includes mechanisms which allow direct transfer of intermediate between active sites as well as more general methods of metabolite compartmentation. Ovadi (1991) has provided a review of possible catalytic advantages resulting from metabolite channelling. The proximity of sequential catalytic sites reduces the transit time of intermediates to diffuse between sites, leading to a decrease in the transient time required by the system to reach a new steady state. Steric hindrance or the arrangement of active sites may prevent diffusion of intermediates out into the bulk medium, producing local pools with higher concentrations of intermediates. As more than 80% of metabolites have only one cellular role (Srere, 1987), this would localize the metabolite to the part of the cell where it is required. This allows the pathway to work efficiently at low cellular concentrations of solutes, and conserves the solvent capacity of the cell (Atkinson, 1969). Channelling also allows for efficient removal and cycling of reaction products into the next step of the pathway, preventing or reducing the loss of intermediates by diffusion. In addition, labile intermediates can be protected. The prompt removal of intermediates may prevent the establishment of unfavourable equilibria. Last but not least, channelling could produce separate pools of intermediates for competing reactions.

Sometimes the channelled intermediate is covalently bound to a component of the complex/protein. This is observed in the 2-oxo acid dehydrogenase complexes, where the intermediate is bound to a lipoamide prosthetic group which, in turn, is covalently linked to the E2 component of the complex (Reed, 1974). Likewise, in mammalian and yeast fatty acid synthases the growing fatty acid chain remains bound to the phosphopantetheine prosthetic group of an acyl carrier domain. In both of these examples, the proposed channelling mechanism involves a flexible swinging arm which conveys the intermediate between different catalytic sites in the enzyme complex, while it remains bound to the complex/enzyme. In contrast to such a model, Cohen-Addad et al. (1995) recently demonstrated that the lipoamide arm bound to the H-protein of the glycine decarboxylase complex is unable to move freely in aqueous solvent. The crystal structure of the H-protein reveals that the lipoic acid prosthetic group is attached to a lysine located in the

loop of a hairpin configuration. Upon binding the methylamine intermediate, the cofactor interacts with several conserved residues within a cleft at the surface of the H-protein. The methylamine intermediate is protected from the solvent in a hydrophobic pocket which may explain the reported increased stability of methylamine when bound to the H-protein (Neuberger et al., 1991). The authors suggest that the T protein (which catalyzes the subsequent reaction in the complex) and its folate cofactor must bind close to this surface cleft to allow for the efficient transfer of the methylamine group from the lipoamide to the folate (Cohen-Addad et al., 1995).

Often the channelled intermediate is not bound covalently. The channelling of the indole intermediate between the two sequential activities of tryptophan synthase is the most compelling example of this type of channelling (Hyde et al., 1988). When the two α and two β subunits of this protein combine to form the $\alpha_2\beta_2$ multienzyme complex, the individual reaction rates and the affinities of each subunit for substrate are increased by one to two orders of magnitude. As well, experimental evidence suggested that the indole intermediate remains bound to the complex. The three dimensional structure of this complex revealed that the indole is channelled through a tunnel connecting each pair of α and β catalytic sites. This tunnel prevents hydrophobic indole from escaping into the cytoplasm and potentially out of the cell.

The crystal structure of the bifunctional thymidylate synthase-dihydrofolate reductase from *Leishmania major* has provided a mechanism to explain the observed channelling of dihydrofolate between active sites in this enzyme (Knighton et al., 1994). An electrostatic surface stretching between the synthase and reductase active sites is proposed to promote surface diffusion of the intermediate.

The mammalian sulphate activation pathway includes a bifunctional enzyme with sequential ATP sulfurylase and adenosine 5' phosphosulphate kinase activities (Lyle et al., 1994a). The ability of this enzyme to channel the adenosine monophosphate sulphate (APS) intermediate from the sulfurylase to the kinase active site was shown by evaluating the time course of appearance of intermediate and product, and by isotopic enrichment or dilution studies (Lyle et al., 1994b). The bifunctional enzyme exhibits a channelling efficiency (rate of initial appearance of final product/rate of initial appearance of

intermediate) of 96%, while a mixture of the monofunctional sulfurylase and kinase from *Pennicillium chrysogenum* showed no ability to channel intermediate. As the equilibrium constant for the sulfurylase activity strongly favours the reverse reaction, removal of APS by channelling it to the kinase active site may help drive the pathway in the forward direction. Moreover, the APS intermediate is labile under physiological conditions while the product of the kinase is more stable.

Most of the preceding examples result in perfect channelling, where almost all of the product of the first reaction is channelled to the subsequent active site. However channelling is often less efficient. The bifunctional dehydrogenase/cyclohydrolase domain of human D\C\S is able to channel only 45% of the 5,10-methenylH_4PteGlu$_n$ produced by the dehydrogenase activity (Hum & MacKenzie, 1991). This is surprising because the two activities share a common folate binding site (Pelletier & MacKenzie, 1995). The cyclohydrolase reaction also occurs nonenzymatically under physiological conditions, which may have precluded the need to develop a more channelling efficient mechanism.

Nada et al. (1995) recently described a study of channelling in mitochondrial fatty acid β-oxidation. They developed a tandem mass spectroscopy method to analyze the metabolism of isotopically labelled fatty acid chains in normal and β-oxidation enzyme-deficient human fibroblasts. They observed two distinct types of intermediate transfer: complete channelling by the long-chain specific enzymes bound to the inner mitochondrial membrane and partial channelling by the soluble matrix enzymes responsible for metabolizing short and medium chain fatty acyl-coA thioesters. Incomplete channelling between the 3-ketoacyl-CoA thiolase and subsequent acyl-CoA dehydrogenases may result from the inability to form a complex between the single thiolase and multiple dehydrogenases (Nada et al., 1995). While the arrangement of the matrix enzymes is unknown, it has been proposed that they are organized in a nonrandom fashion which allows for the appropriate dynamic interactions between sequential activities.

The most convincing examples of intermediate channelling are observed in static multienzyme complexes or multifunctional enzymes. They provide a solid example of how enzyme organization can regulate cellular metabolism.

1.4. FOLATE METABOLISM

Tetrahydrofolate (H_4PteGlu) functions as a carrier of one-carbon (C1) units in both prokaryotes and eukaryotes (Figure 1.4.1). This cofactor is involved in a variety of cellular reactions, not all of which are conserved between species. The following review of folate metabolism is limited to one-carbon metabolism in mammalian liver only (the major pathways are outlined in Figure 1.4.2.). Later in this section the emphasis will turn to the role of $H_4PteGlu_n$ in the metabolism of histidine, as mediated by the bifunctional enzyme formiminotransferase-cyclodeaminase.

1.4.1 An overview of folate metabolism in mammalian liver

One-carbon groups at various oxidation states, can be carried at N-5 of tetrahydrofolate to give 5-formyl, 5-formimino or 5-methyl derivatives, at N-10 to produce 10-formyl derivatives, or bridged between N-5 and N-10 to form 5,10- methylene or methenyl derivatives. Within cells, folates are polyglutamylated, containing between 5 and 9 glutamates in gamma linkage. Absorption of dietary folates and hydrolysis of the polyglutamate tail occurs in the small intestine (Halsted, 1989). The most common circulating form of this cofactor is monoglutamylated 5-methylH_4PteGlu (Ratanasthien et al., 1974). Animal studies indicate that the liver is the primary site of reduction and methylation of circulating folate (Kiil et al., 1979) and also serves as the major storage site of dietary folates (reviewed by Cossins, 1984), with the glu_5 and glu_6 derivatives predominant. Within liver, folates are highly compartmentalized, existing in both the cytosol and the mitochondria, and are generally protein bound (Schirch & Strong, 1989). Both the cytosol and the mitochondrial matrix contain serine hydroxymethyltransferases, folylpolyglutamate synthetases and folate interconversion enzymes.

Folates enter the cell through one of two distinct transport systems (reviewed by Antony, 1992, Horne, 1993). The reduced folate carrier is a high affinity/low capacity transporter. This carrier is saturable and actively transports reduced folates and methotrexate preferentially across the membrane. cDNA's encoding the reduced folate carrier, or a component of this carrier, have recently been isolated from mouse (Dixon,

FIGURE 1.4.1. The structure of tetrahydropteroylpolyglutamate. One-carbon substitutions occur at N-5 and/or N-10. The glutamates in the polyglutamate tail are in gamma linkage. (modified from Cossins, 1984).

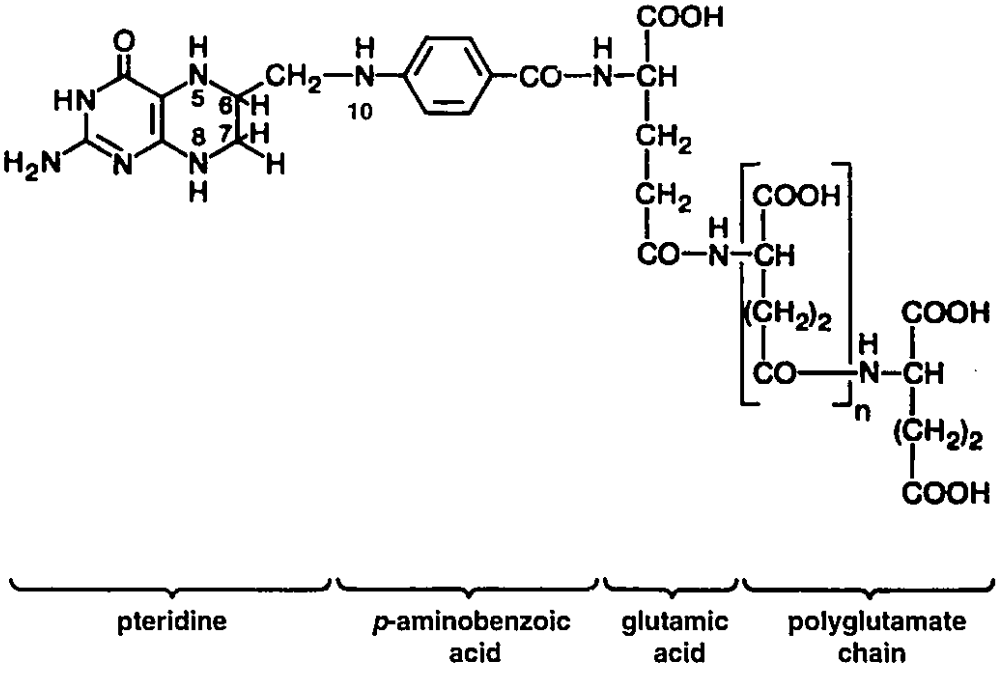
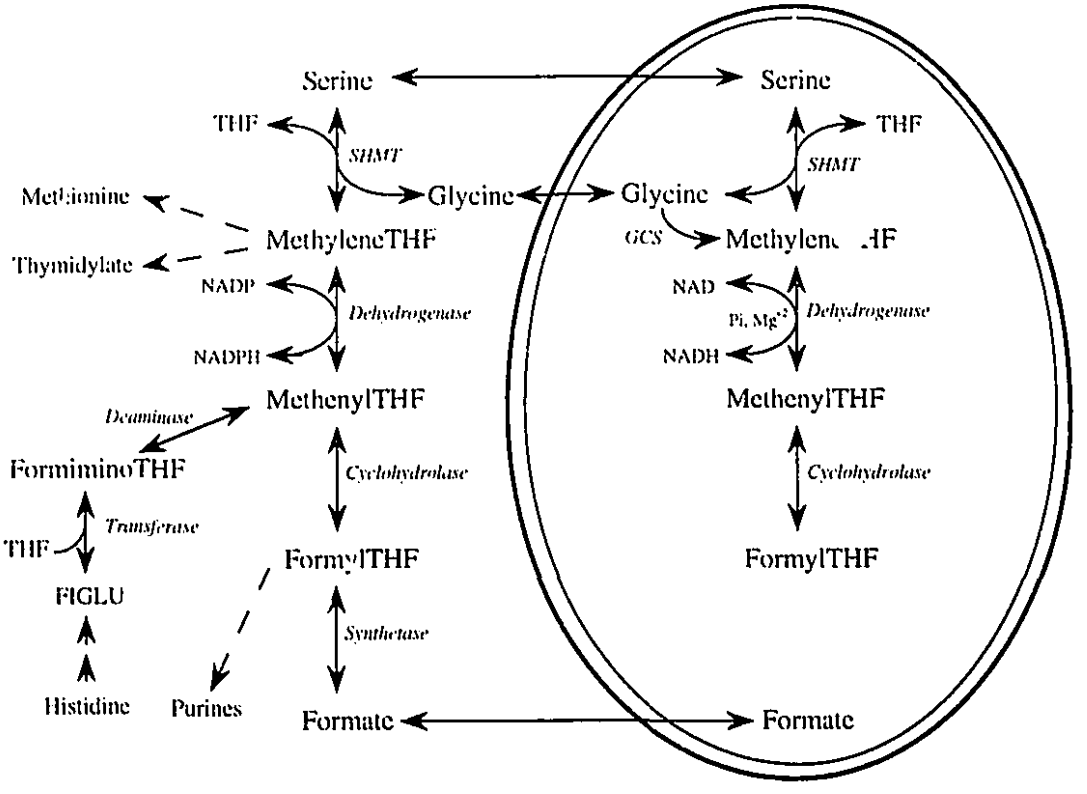

pteridine
p-aminobenzoic acid
glutamic acid
polyglutamate chain

FIGURE 1.4.2. An overview of cellular folate metabolism in mammalian liver. This figure provides an overview of the common cellular sources of one-carbon units, the cellular uses of these one-carbon units, and the pathways for the interconversion of different pools of one-carbon units. (modified from scheme II, which was prepared by G. Tremblay and published in Yang & MacKenzie, 1993)


Serine
Serine
THF
SHMT
THF
SHMT
Methionine
Glycine
Glycine
GCS
MethyleneTHF
Thymidylate
NADP
Dehydrogenase
NAD
Pi, Mg+2
Dehydrogenase
NADPH
NADH
MethenylTHF
MethenylTHF
Deaminase
FormiminoTHF
Cyclohydrolase
Cyclohydrolase
THF
Transferase
FormylTHF
FormylTHF
FIGLU
Synthetase
Histidine
Purines
Formate
Formate
Cytoplasm
Mitochondria

et al., 1994), hamster (Williams et al., 1994) and human (Wong, 1995), and sequence analysis indicates that this carrier is part of a superfamily of transmembrane transporters In contrast, the folate receptor remains bound to the membrane by a glycosylphosphatidylinositol tail and relies on endocytosis to transport folates into the cell This receptor preferentially binds folic acid, 5-formyl and 5-methylH_4PteGlu A reduced folate transporter also promotes transport of folates across the mitochondrial membrane (Horne et al., 1992). There is no evidence for transfer of 5,10-methylene- or 10-formylH_4PteGlu across the mitochondrial membrane (Horne, 1989) and one carbon units are probably metabolized to serine or formate prior to transport out of the mitochondria.

Once inside the cell, poly-γ-glutamate tails must be added to the coenzyme so that it can be retained and used efficiently within the cell (reviewed in Shane, 1989) Folylpolyglutamate synthase (FPGS) catalyzes the stepwise addition of glutamates to the terminal glutamate moiety of folate. In mammalian tissues, poly-γ-glutamate tails are typically between five and nine residues in length. Both cytoplasmic and mitochondrial versions of FPGS exist. The folates that first enter the mitochondria are monoglutamylated, however polyglutamylation is obligatory for their conservation within this organelle. Recently Shane's group have shown that mitochondrial FPGS activity is essential for C1 metabolism within the mitochondria and for normal one carbon flux to the cytosol (Garrow et al., 1992; Lin et al., 1993; Lin & Shane, 1994).

Most of the tetrahydrofolates entering the cell are in the 5-methyl form. Methionine synthase provides the sole means of metabolizing 5-methylH_4PteGlu. This cobalamin-dependent enzyme catalyzes the transfer of the methyl group to homocysteine, producing methionine and $H_4PteGlu_n$ which can re-enter the C1 pool. 5-methyl$H_4PteGlu_n$ is a poorer substrate for FPGS than $H_4PteGlu_n$, 5,10-methylene$H_4PteGlu_n$ or 10-formyl$H_4PteGlu_n$, suggesting that the cofactor must be demethylated prior to polyglutamylation (Cichowicz & Shane, 1987).

1.4.2 Sources of cellular one-carbon units

Carbon-3 of serine is the major source of one-carbon units for cellular metabolism (reviewed in MacKenzie, 1984). Serine hydroxymethyltransferase (SHMT) transfers the

C-3 of serine to $H_4PteGlu_n$, producing glycine and 5,10-methylene$H_4PteGlu_n$. Formate and histidine serve as minor sources of cytosolic C1 units. 10-formyl$H_4PteGlu_n$ synthetase, the third activity of the cytosolic folate interconversion enzyme D\C\S, can couple formate to the N-10 position of tetrahydrofolate in an ATP dependent reaction. Bifunctional formiminotransferase-cyclodeaminase catalyzes the transfer of the formimino group from the histidine catabolite formiminoglutamate to $H_4PteGlu_n$ and the subsequent deamination of this intermediate (described in more detail in section 1.4.7).

Within mitochondria, the glycine cleavage system and catabolism of the N-methylated compounds dimethylglycine and sarcosine by their corresponding dehydrogenases also provide 5,10-methylene$H_4PteGlu_n$. The mitochondrial isoform of SHMT can use 5,10-methylene$H_4PteGlu_n$ and glycine to synthesize serine for transport of C1 units out of the mitochondria.

Barlowe & Appling (1988) have proposed that a mitochondrial pool of C1 units, generated from the SHMT-catalyzed breakdown of serine, is oxidized to formate by a liver mitochondrial D\C\S which is analogous to the yeast mitochondrial D\C\S (reviewed in Appling, 1991). The formate could then be transported out of the mitochondria and incorporated into 10-formyl$H_4PteGlu_n$ by the cytosolic 10-formyl$H_4PteGlu_n$ synthetase. Appling and colleagues have demonstrated that carbon 3 of serine or the N-methyl carbon of sarcosine can be oxidized to formate in rat liver mitochondria (Barlowe & Appling, 1988; Garcia-Martinez & Appling, 1993). This hypothesis provides an explanation for the importance of mitochondrial SHMT and FPGS in cytoplasmic folate metabolism. However, deletion of the yeast mitochondrial D\C\S has no phenotype (Shannon & Rabinowitz, 1988). Moreover, while a mammalian, NADP-dependent, mitochondrial D\C\S has yet to be isolated, an NAD-dependent, mitochondrial D\C has been purified (Mejia et al., 1986) and cloned (Bélanger & MacKenzie, 1989). Yang & MacKenzie (1993) have suggested that this bifunctional enzyme is the homolog of the yeast mitochondrial D\C\S. Certainly mitochondrial folate metabolism has a major, although presently unclear, impact on cytosolic folate metabolism.

1.4.3 Interconversion of one-carbon units

The NADP-dependent trifunctional D\C\S balances the cytosolic pools of 5,10-methylene and 10-formyl derivatives. The NADP dependent dehydrogenase and cyclohydrolase activities reversibly interconvert 5,10-methylene, 5,10-methenyl and 10-formylH_4PteGlu$_n$, such that C1 units are available for synthesis of purines, thymidylate and methionine as required by the cell. Pelletier and MacKenzie (1995) have argued that the cytosolic pools of 5,10-methylene and 10-formylH_4PteGlu$_n$ are kept at or near equilibrium, providing sufficient concentrations of either species as required for biosynthetic purposes, and allowing both serine and formate to efficiently serve as donors to the active C1 pool. This enzyme is expressed in all tissues (Thigpen et al., 1990; Peri & MacKenzie, 1991) and regulated as a house-keeping enzyme (Peri & MacKenzie, 1991). As mentioned earlier, mammalian mitochondria contain an NAD-dependent D\C. Cellular expression of the mitochondrial D\C is highly regulated. While it is present at very low levels in normal cells and tissues, it is overexpressed in fetal tissues and immortalized cell lines, and upregulated in response to mitogens (Peri and MacKenzie, 1993). This enzyme was thought to be involved in supplying formyl groups for initiation of mitochondrial protein synthesis. However, protein synthesis is still observed in an embryonic stem cell line which no longer expresses the D\C protein (Tremblay, 1995). Nevertheless, loss of this D\C does impair cell growth at low concentrations of folate.

Both SHMT (Stover & Schirch, 1990) and the cyclohydrolase activity of D\C\S (Pelletier, 1995) catalyze the hydrolysis of 5,10-methenylH_4PteGlu$_n$ to 5-formylH_4PteGlu$_n$, with low specific activity. The 5-formyl derivative is a possible storage form of folate within the cell (Kruschwitz et al., 1994), and functions as an inhibitor of several folate-dependent enzymes (reviewed in Stover et al., 1993), perhaps playing a regulatory role in C1 metabolism. 5,10-MethenylH_4PteGlu$_n$ synthetase converts 5-formylH_4PteGlu$_n$ back to 5,10-methenylH_4PteGlu$_n$ through an ATP-dependent reaction.

10-formylH_4PteGlu$_n$ dehydrogenase-hydrolase (10-FTHFDH) can regenerate tetrahydrofolate from 10-formylH_4PteGlu$_n$ through either of two separate activities: the NADP dependent dehydrogenase which releases the C1 unit as CO_2, and the hydrolase which produces H_4PteGlu$_n$ and formate. 10-FTHFDH tightly binds pentaglutamylated

tetrahydrofolate, which remains associated with the protein throughout purification. There has been some debate regarding the subcellular distribution of these activites and whether they exist on separate polypeptides (Case et al., 1988). However Cook and colleagues (1991) have cloned a cytosolic 10-FTHFDH from rat liver, which displays both activities. Sequence analysis (Cook et al., 1991), mutational analysis (Krupenko et al., 1995), and proteolysis (Schirch et al., 1994) indicate that this protein contains an N-terminal hydrolase and a C-terminal dehydrogenase domain.

1.4.4 Uses of one-carbon units

One carbon units are required for the *de novo* synthesis of purines, thymidylate synthesis and regeneration of methionine. 10-formylH_4PteGlu$_n$ donates two C1 units to *de novo* purine synthesis, which are incorporated as C-2 and C-8 of the purine ring. The first folate-dependent transfer reaction is catalyzed by GAR (glycinamide ribonucleotide) transformylase. In liver, this transformylase is part of a multifunctional enzyme which also includes the purine synthetic activities GAR synthase and aminoimidizole ribonucleotide synthetase (Daubner et al., 1985; Aimi et al., 1990). Recently a second bacterial GAR transformylase (*purT*) was identified which uses formate instead of 10-formylH_4PteGlu$_n$ as the C1 donor (Nygard & Smith, 1993; Marolewski et al., 1994). However, the formate is thought to be provided by a formylH_4PteGlu$_n$ hydrolase encoded by *purU* (Nagi et al., 1993; 1995). Incorporation of the second C1 unit is catalyzed by AICAR (5-amino-4-imidazole-carboxamide ribonucleotide) transformylase. This activity is part of a bifunctional enzyme which also catalyzes the closure of the purine ring (Mueller & Benkovic, 1981).

After observing that these two folate-dependant transformylases co-purified from chicken liver with D\C\S and SHMT, Caperelli et al. (1980) proposed that these proteins may also associate to form a multi-enzyme complex *in vivo*, which could channel C1 units donated by serine directly into *de novo* purine synthesis.

5,10-methyleneH_4PteGlu$_n$ provides methyl groups for the synthesis of thymidylate. Thymidylate synthase transfers a one carbon unit from 5,10-methyleneH_4PteGlu$_n$ to dUMP to produce thymidylate. During this transfer the pteridine ring becomes oxidized forming

dihydrofolate. Dihydrofolate reductase returns dihydrofolate to the fully reduced state.

5,10-methyleneH_4PteGlu$_n$ also serves as a C1 donor for methionine synthesis. The flavoprotein 5,10-methyleneH_4PteGlu$_n$ reductase (MTFR) catalyzes the reduction of this folate cofactor to 5-methylH_4PteGlu$_n$. This irreversible reaction is the committed step in methionine synthesis. Vitamin B_{12} methionine synthase catalyzes the transfer of the methyl group from 5-methyltetrahydrofolate to homocysteine producing tetrahydrofolate and methionine. The methionine cycle is highly regulated, and only low amounts of these enzyme activities are observed *in vivo*. Methionine is a weak inhibitor of methionine synthase, and when modified to form S-adenosylmethionine (SAM), functions as an allosteric inhibitor of MTFR (Kutzbach & Stokstad, 1971; Jenks & Matthews, 1987). In fact physiological levels of SAM keep the cellular level of MTFR activity very low. When methionine synthase is inactivated and cellular SAM levels drop, MTFR is no longer inhibited. Polyglutamylated cellular folates become trapped as 5-methylH_4PteGlu$_n$ (the methyl trap hypothesis, reviewed in Matthews, 1984; Shane & Stokstad, 1985), and monoglutamylated 5-methylH_4PteGlu$_1$ entering the cell cannot be efficiently polyglutamylated. This diminishes the level of active folates within the tissue, impairing all aspects of C1 metabolism.

1.4.5 The role of polyglutamylation in one-carbon metabolism

The role of polyglutamylation of folates in regulating one-carbon metabolism has received considerable attention (reviewed in Schirch & Strong, 1989; Shane, 1989; Krumdiek et al., 1991). As previously mentioned, folates must be polyglutamylated for retention within the cell or the mitochondria. Many folate-dependent enzymes show increased affinity for folate substrates or enzyme inhibitors with a particular polyglutamate tail length. Polyglutamylation has also been implicated in the channelling of folate intermediates between active sites in multifunctional enzymes.

Several folate-dependent enzymes display preferential binding of polyglutamylated substrates or inhibitors. To illustrate this phenomenon, porcine liver methyleneH_4PteGlu$_n$ reductase shows increased specificity with increasing chain length for both the H_2PteGlu$_n$ inhibitor and 5,10-methyleneH_4PteGlu$_n$ substrate (Matthews & Baugh, 1980), and achieves

maximum affinity with a hexaglutamate tail. Decreasing the folate K_m values results in increased catalytic efficiency, as measured by V_{max}/K_m, and brings the binding affinities of the folate substrate into the range of cellular folate concentrations.

The presence of a polyglutamate tail may also result in a change in the kinetic mechanism. MTFR changes from an apparent ping pong mechanism to a sequential mechanism with polyglutamylated substrates (Matthews & Baugh, 1980). Likewise the order of addition of substrates changes for thymidylate synthase when mono- or pentaglutamylated derivatives are supplied as substrate (Lu et al., 1984). In some instances the presence of a polyglutamylated folate inhibitor or substrate may enhance the binding of a nonfolate substrate (Matthews, 1984; Findlay et al., 1989).

Baggott & Krumdiek (1979) proposed that the relative distribution of folate coenzymes with different polyglutamate tail lengths may play a role in controlling the flux of C1 units down different metabolic pathways. A study of the four cytosolic enzymes which use 5,10-methyleneH_4PteGlu$_n$ as a cofactor suggest that different cellular concentrations of 5,10-methyleneH_4PteGlu$_n$ may regulate the flux of C1 units into nucleotide biosynthesis or methionine regeneration pathways (Matthews et al., 1985; Green et al., 1988).

Polyglutamylation has been implicated in the channelling of a formiminoH_4PteGlu$_5$ intermediate between the sequential transferase and cyclodeaminase activites of FTCD (MacKenzie & Baugh, 1980; Paquin et al., 1985), as discussed in section 1.4.8. Intermediate channelling has also been observed between the dehydrogenase and cyclohydrolase activites in the D\C domain of cytosolic D\C\S and mitochondrial D\C. Pelletier & MacKenzie (1995) have recently shown that the cytoplasmic dehydrogenase and cyclohydrolase activites share a folate binding site, and the mechanism of channelling in this enzyme is likely independent of the polyglutamylation status of the intermediate. Bifunctional thymidylate synthase-dihydrofolate reductase will channel both mono and polyglutamylated derivatives between the two types of active sites. A possible mechanism for channelling of the dihydrofolate intermediate has been discussed previously (section 1.3.6).

1.4.6 The $H_4PteGlu_n$ binding site

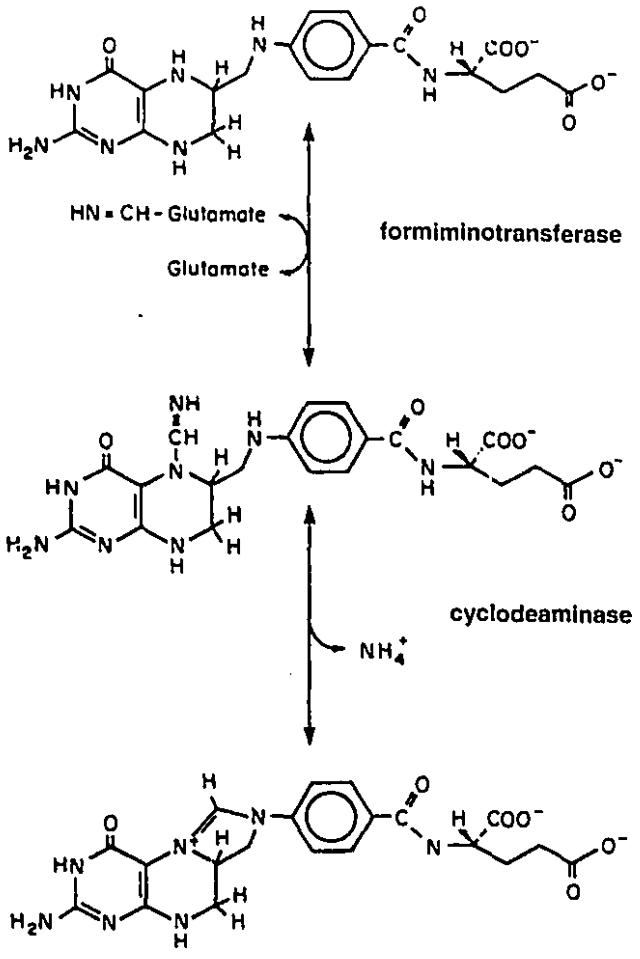
While the amino acid sequences are available for several $H_4PteGlu_n$ binding proteins, sequence analysis has not provided a general consensus sequence for folate binding sites. Cook et al., (1991) however, have identified a putative 10-formyl$H_4PteGlu_n$ binding sequence, $XP\S(X)_2P(X)_{2,3}G$, which has been observed in several different 10-formyl$H_4PteGlu_n$ binding enzymes. While the crystal structures of several folate-dependent enzymes have now been determined, there is very little evidence of a common folate-binding fold (reviewed in Chen et al., 1992). Researchers have been more successful at identifying possible polyglutamate binding sites. Not suprisingly, several studies indicate that basic amino acids are involved in the binding of polyglutamylated substrates to folate-dependent enzymes (Kamb et al., 1992; Usha et al., 1992; Maras et al., 1994; Finer-Moore et al., 1994). These sites will be discussed in more detail in Chapter 5.

1.4.7 Histidine catabolism in mammals

Folates are involved in both the synthesis and the catabolism of histidine (reviewed by Shane & Stokstad, 1984). In mammals, folate-dependent histidine degradation salvages one of the C1 units used in histidine biosynthesis. The preferred route of histidine breakdown in mammals is initiated by histidine-ammonia lyase (known as histidase), which produces urocanate and free ammonia from histidine. Urocanase catalyzes the hydration and rearrangement of urocanate, to form the unstable product imidazolone propionate. This metabolite is then hydrolysed to produce formiminoglutamate (FIGLU). These three reactions are common to eukaryotic and prokaryotic histidine degradation pathways.

The folate-dependent degradation of FIGLU was elucidated in the 1950's by Tabor, Rabinowitz and Wyngarden (1956; 1959). In mammals, FIGLU is the substrate for bifunctional formiminotransferase-cyclodeaminase (FTCD). The formiminotransferase activity catalyzes the transfer of the formimino group from FIGLU to the N5 position of $H_4PteGlu_n$, while the cyclodeaminase catalyzes the deamination of 5-formimino$H_4PteGlu_n$, forming 5,10-methenyl$H_4PteGlu_n$ and releasing ammonia (outlined in Figure 1.4.3). Formiminotransferase-cyclodeaminase serves as a gateway for the entry of one carbon

FIGURE 1.4.3. Bifunctional formiminotransferase-cyclodeaminase catalyzes two sequential reactions in the histidine degradation pathway. (taken from Findlay, 1988)


HN = CH - Glutamate
Glutamate
formiminotransferase
cyclodeaminase
NH_4^+

units from histidine degradation into the folate pathway.

Regulation of this pathway is influenced by a myriad of effectors (reviewed by Schepartz, 1973; Levy, 1989). Hepatic histidase and urocanase levels are subject to several environmental conditions including diet, growth conditions and disease states. Many of the regulatory factors are typically gluconeogenic while others are less obvious. Histidase appears to be developmentally regulated, first appearing at very low levels in rat liver shortly after birth. This level is augmented during puberty, however the increase is greater in the female animal such that histidase levels are doubled. Less is known about the regulation of the third and fourth enzymes in this pathway. It has been reported that FTCD enzyme activity is downregulated in rat hepatocarcinomas (Jackson & Niethammer, 1979). Like histidase and urocanase, rat hepatic transferase activity may be induced by glucagon and decreased by insulin. Stifel et al. (1974) suggested that the transferase may also be stimulated by epinephrine and intravenous cAMP through post translational modification. Developmentally, the transferase levels are low in newborn rats and mice and increase rapidly during the first three weeks of life, plateauing at week five (Rabinowitz & Tabor, 1958). Our lab has observed that full-grown sows are the best source of porcine liver FTCD (MacKenzie, personal communication) suggesting that estragen may be a regulator of this enzyme as has been observed with histidase.

Histidine catabolism is clearly affected by the level of available folate, and vitamin B_{12}. Folate deficiency results in inactivation of urocanase and FTCD, with increased excretion of urocanate and FIGLU in the urine. Vitamin B_{12} deficiency also results in increased excretion of FIGLU. The excretion of this metabolite is used clinically to diagnose folate deficiency and to differentiate between folate and vitamin B_{12} deficiencies (Chanarin, 1969). Addition of methionine will decrease FIGLU excretion if Vitamin B_{12} deficiency is the cause, as outlined in the methyl trap hypothesis (Shane and Stokstad, 1984).

The FTCD activities are primarily found in the liver, but lower levels have also been identified in kidney and jejunum in some mammalian species (Tabor & Wyngarden, 1959; McLain et al., 1975). Formiminotransferase activity has not been detected in insects, bacteria and yeast, however both activities have been identified in filaria (Jaffe

et al., 1980) and an avian homolog has been observed in chicken (Hennig, personal communication, this thesis, Chapter 5). A *Clostridial* cyclodeaminase is responsible for the deamination of formiminotetrahydrofolate produced by catabolism of purines (Rabinowitz & Pricer, 1956a,b). This activity is associated with a 5,10-methenylH$_4$PteGlu cyclohydrolase (Uyeda & Rabinowitz, 1967). Also, an open reading frame from the *Methanobacterium extorquens* shows amino acid sequence homology to the deaminase domain of FTCD (this thesis, Chapter 5).

1.4.8 Porcine liver Formiminotransferase-Cyclodeaminase

Tabor and Wyngarden (1959) first showed that the transferase and deaminase activities co-purified from hog liver acetone powder. Ultracentrifugation of the purified enzyme gave a single peak indicating that the activities were associated. Each activity could be inactivated separately. The transferase activity was preferentially destroyed by treatment with NH_4OH at pH 10.5, while the deaminase activity was lost after proteolysis with chymotrypsin. Thus, FTCD appeared to be either a multienzyme complex or a multifunctional enzyme.

Drury et al. (1975) developed a protocol to isolate the enzyme directly from pig liver. The purified enzyme had a monomer size of approximately 62 kDa estimated from SDS PAGE, and a native molecular weight of 540 kDa as determined by equilibrium sedimentation. This suggested that FTCD was an oligomer of between 7 and 9 subunits. Isoelectric focusing and cyanogen bromide cleavage studies confirmed that the subunits were identical and that FTCD was a multifunctional protein (Beaudet & MacKenzie, 1976). Electron microscopy with rotational reinforcement of negatively stained molecules of FTCD indicated that the enzyme is octameric and arranged such that the eight subunits form a planar ring. Cross-linking with dithiobis(succinidyl propionate) verified that native FTCD is an octamer (MacKenzie et al., 1980)

Several lines of evidence indicated that the transferase and deaminase activities could function independently, suggesting that they were located at separate sites on the polypeptide. Treatment with dithionitrobenzene preferentially inactivated the deaminase while the transferase activity was susceptible to inactivation with diethylpyrocarbonate

(Drury & MacKenzie, 1977). As well, a 39 kDa transferase-active fragment was released upon proteolysis of FTCD with chymotrypsin in the presence of folic acid (MacKenzie et al., 1980).

Beaudet and MacKenzie (1975) proposed that the formiminotransferase employs a rapid equilibrium random kinetic mechanism. This implies that the folate and FIGLU substrates bind to distinct parts of the transferase active site and that either can bind first. Kinetic analyses indicated that both the transferase and the deaminase activities display higher catalytic efficiencies when polyglutamylated substrates are used (MacKenzie & Baugh, 1980; Paquin et al., 1985). The transferase-active proteolytic fragment, however, does not retain specificity for folylpolyglutamates. The presence of a polyglutamate tail also decreases the K_m value for formiminoglutamate approximately ten-fold, bringing it closer to the physiological range (Findlay et al., 1989).

FTCD channels pentaglutamylated forminoH$_4$PteGlu$_n$ between the transferase and deaminase active sites with 100% efficiency, and other polyglutamates (with 4, 6 or 7 glutamates) to a lesser extent (MacKenzie & Baugh, 1980). Channelling is indicated by an increase in the rate of production of 5,10-methenylH$_4$PteGlu$_n$ (the product of the deaminase reaction), such that the formimino intermediate no longer accumulates in the medium. MacKenzie and Baugh, (1980) used chemical modification to produce monofunctional transferase or deaminase. As a mixture of the modified monofunctional transferase and deaminase does not channel the formimino intermediate, substrate channelling appears to only occur between sites within the same octamer. MacKenzie and Baugh (1980) proposed that the polyglutamate tail might function as an anchor, fastening the intermediate to the octamer while the pteroyl moiety swings between active sites.

Paquin et al. (1985), used binding and kinetic studies to further analyze the mechanism of channelling. These studies showed that while FTCD binds hexaglutamates with the greatest affinity and displays similar catalytic efficiencies when H$_4$PteGlu$_n$ (n = 4, 5, 6 or 7) are used as substrate, complete channelling only occurs with the pentaglutamylated substrate. Thus the channelling mechanism can distinguish between different polyglutamate chain lengths, and the pentaglutamate tail may be preferred for steric reasons. The distance between the α-carboxyl of the fourth glutamate (the first

glutamate to be tightly bound to FTCD) and the N-5 position on the pteridine ring, was estimated to be 20-25°A in length. This could provide the pteroyl moiety with a range of 40-50°A while the polyglutamate tail remains attached to the enzyme. While this distance is greater than the diameter of the subunit as estimated from electron microscopy, the relative positions of both types of active sites within the dimer must be determined before one can resolve whether this model provides a feasible explanation of channelling in this system.

Equilibrium binding studies (Paquin et al., 1985) indicate that each FTCD octamer contains four high affinity polyglutamate binding sites. Using combinations of mono and pentaglutamylated $H_4PteGlu_n$ and formimino $H_4PteGlu_n$, Paquin et al. (1985) demonstrated that FTCD will only use exogenous formimino $H_4PteGlu_n$ when one of the substrates is monoglutamylated. This suggests that one polyglutamate binding site exists per pair of transferase/deaminase active sites.

Crosslinking of FTCD with the short bifunctional reagent difluorodinitrobenzene yielded predominantly even-numbered species, indicating that two types of subunit interactions are present within the octamer (MacKenzie et al., 1980). As well, the transferase-active proteolytic fragment exists as a dimer. This implied that the FTCD octamer was in fact a ring-shaped tetramer of dimers.

What would be the smallest functional unit in such an arrangement, able to catalyze both activities as well as channel intermediate between active sites? Findlay and Mackenzie (1987) performed a series of denaturation and renaturation studies to answer this question. Urea induced-dissociation of FTCD was analyzed by monitoring catalytic activity, intrinsic tryptophan fluorescence and subunit association under increasing concentrations of denaturant. In potassium phosphate buffer, FTCD sequentially dissociated, proceeding from octamer to dimers to monomers. Between 2 and 3 M urea octameric FTCD underwent its first dissociation to dimers, accompanied by a simultaneous loss of both catalytic activities and a large decrease in the fluorescence intensity. The second transition, occurring between 3 and 4 M urea, and represented by a redshift in the wavelength of maximum fluorescence emission, was interpreted as a conformational change in the dimer. Above 4 M urea, the dimers dissociated to

monomers. Two different types of monofunctional dimers can be isolated by varying experimental conditions. At 3 M urea in potassium phosphate buffer containing folic acid, FTCD forms deaminase-active dimers. Transferase-active dimers are produced upon incubation of FTCD in triethanolamine hydrochloride buffer containing 1 M urea. These dimers remain stable at 3 M urea when glutamate is present. Proteolysis of the transferase and deaminase-active dimers produced different fragmentation patterns, indicating that they are structurally distinct. This suggested that they might isolate different subunit interfaces.

Renaturation of FTCD from 6 M GdnHCl indicates that the enzyme recovers the native quaternary structure, over 90% of both catalytic activities, and the ability to channel intermediate within 48 hours following dilution (Findlay & MacKenzie, 1988). In the presence of 1.5 M urea, reassembly of FTCD is arrested at the level of a transferase-active dimer. Dialysis leads to recovery of both the deaminase activity and the polyglutamate binding specificity. This also supports the proposal that the transferase activity is associated with one type of subunit interface and the deaminase activity and polyglutamate binding site with a second type of interface. The transferase-active proteolytic fragment could also be renatured from 6 M GdnHCl, indicating that it can function as an independent folding unit.

STATEMENT OF THE PROBLEM

The integrity of alternating subunit interfaces appears essential for the coincident expression of the transferase and deaminase activities and the substrate channelling behaviour. Consequently Findlay and Mackenzie have proposed that the octamer rather than a dimer is the functional unit of FTCD. Assuming that each subunit contains both a transferase and a deaminase domain, an analysis of the domain structure of FTCD and characterization of the isolated domains should finally resolve this issue. As well, it should clarify the role of the two catalytic domains within the FTCD octamer

Prior to such an analysis I first had to generate a system where we could isolate an independent transferase and deaminase domain. A novel full-length cDNA encoding porcine liver FTCD was isolated and used to express the recombinant enzyme in *E. coli*. The protocol described in Findlay et al. (1989) was modified to purify the recombinant enzyme. Deletion analysis of the cDNA encoding FTCD was used to demonstrate the domain structure of each FTCD subunit. Both catalytic domains were independently expressed and characterized in terms of their quaternary structure and ability to bind and channel polyglutamates. The isolated domains and the full-length enzyme were subjected to urea-induced denaturation in order to further characterize the properties of each domain in and outside of the octamer.

CHAPTER 2

THE NUCLEOTIDE SEQUENCE OF PORCINE FORMIMINOTRANSFERASE-CYCLODEAMINASE: EXPRESSION AND PURIFICATION FROM *ESCHERICHIA COLI*

ABSTRACT

We have isolated and characterized cDNA clones encoding the porcine liver octameric enzyme, formiminotetrahydrofolate:glutamate formiminotransferase (EC 2.1.2.5) - formiminotetrahydrofolate cyclodeaminase (EC 4.3.1.4). The cDNA encodes a novel amino acid sequence of 541 residues which contains exact matches to two sequences derived by automated sequence analysis of CNBr cleavage fragments isolated from the porcine enzyme. The recombinant enzyme has been expressed as a soluble protein in *Escherichia coli* at levels 4-fold higher than those observed in liver, and is bifunctional, displaying both transferase and deaminase activities. With a calculated subunit molecular mass of 58926 Da, it is similar in size to the enzyme isolated from porcine liver. Purification of the enzyme from *Escherichia coli* involves chromatography on a novel polyglutamate column which might interact with the folylpolyglutamate binding site of the protein. The purified recombinant enzyme has a transferase specific activity of 39-41 units/mg/min.

INTRODUCTION

Formiminotransferase-cyclodeaminase FTCD is a large octameric protein with two independent catalytic activities. This folate-dependent enzyme serves to channel one-carbon units from formiminoglutamate, a metabolite in the histidine degradation pathway, to the folate pool. After identifying 5-formiminotetrahydrofolate as the product of a formiminoglycine transferase from *Clostridia* (Rabinowitz & Pricer, 1956), Rabinowitz and associates defined the enzymatic steps involved in the folate-dependent catabolism of formiminoglutamate in mammals (Tabor & Rabinowitz, 1956). Specifically, the transferase activity transfers the formimino group of formiminoglutamate to the N5 position of tetrahydrofolate, producing 5-formiminotetrahydrofolate and glutamate. The cyclodeaminase activity catalyzes the cyclization of the folate intermediate, forming 5,10-methenyltetrahydrofolate and releasing ammonia. This liver specific enzyme was first purified and characterized by Tabor and Wyngarden (1959). The presence of FTCD in other organisms is, for the most part, uncertain. The enzyme has not been detected in yeast or bacteria (Shane & Stokstad, 1984) but both enzyme activities have been reported in filaria (Jaffe et al., 1980).

Previous studies on FTCD indicate that its unusual quaternary structure is essential for the full expression of both catalytic activities. The native porcine enzyme has been shown to be composed of eight identical subunits of approximately 62 kDa each, which associate to form a planar, ring-shaped structure (Drury et al., 1975; Beaudet & MacKenzie, 1976). The presence of only four polyglutamate binding sites per octamer, the prevalence of dimers and tetramers in cross-linking studies, and the generation of a dimeric transferase-active proteolytic fragment suggest that the subunits are arranged as a tetramer of dimers (Paquin et al., 1985; MacKenzie et al., 1980). This model has received further support from a series of dissociation and renaturation experiments which indicated that two different types of dimers can be isolated, each containing a different subunit interface and displaying either the transferase or the deaminase activity (Findlay & MacKenzie, 1987; 1988). In order to further explore the molecular structure necessary for expression of each activity, we have isolated the cDNA encoding the porcine liver

FTCD and expressed it in *E. coli*

MATERIALS AND METHODS

Materials Restriction Enzymes and DNA modification enzymes were obtained from Bethesda Research Laboratories, Boehringer Mannheim, Pharmacia or New England Biolabs. All reagents and enzymes for sequencing DNA were from United States Biochemical Corporation or Pharmacia. Nitrocellulose membranes used for Western blotting were from Schleicher and Schuell. Nitrocellulose and nylon filters for screening and nylon membranes for Southern/Northern blot analysis were from Amersham Corporation as were radioisotopically labelled nucleotides and ^{125}I-labelled protein A. Diagnostic X-OMAT AR film was purchased from Kodak. Oligonucleotides were synthesized and purified by the Sheldon Biotechnology Centre, McGill University.

Oligo dT cellulose, DEAE Sepharose and Heparin Sepharose CL-6B were from Pharmacia, Matrex Blue A was from Amicon and Affigel 15 was purchased from BioRad. Poly-L-glutamic acid sodium salt (MW > 8000 Da) was supplied by TaKaRa. Folic Acid and formimino-L-glutamic acid were from Sigma. Filter units (Centricon) were purchased from Amicon. All other chemicals used were of reagent grade.

All cDNA probes were labelled by the random primer method using [α-^{32}P]dCTP as the labelled nucleotide (Feinberg & Vogelstein, 1984). Unless otherwise noted, protocols used were as outlined in Sambrook et al. (1989).

Coupling of polyglutamic acid to activated agarose. Polyglutamate was coupled to affigel 15 as outlined in the general instructions for aqueous coupling supplied by the manufacturer (BioRad, Bulletin 1085). Specifically, 120 µmoles of polyglutamate (assuming M_r of 8000) was coupled to 10 ml of resin (12 µmoles ligand/ml resin) in a final reaction volume adjusted to 20 - 25 ml with 0.1 M $NaHC0_3$ (pH 8.5).

Amino acid sequencing. Purified porcine protein was subjected to cyanogen bromide treatment (Beaudet & MacKenzie, 1976). The resulting fragments were separated by electrophoresis and transferred onto PVDV membranes (Immobilon) (Matsudaira, 1987). Specific protein bands were cut from the membrane and sequenced on an Applied

Biosystems, Inc model 473A protein sequencer.

Isolation of cDNA clones. A polyclonal antibody was prepared in our laboratory against the purified porcine liver enzyme. This was used to screen a λgt11 porcine liver cDNA library (Clontech) as described by Huynh et al. (1985). The antiserum was diluted 1:500 and preabsorbed with total *E. coli* cell lysate (10 mg/ml). After incubation with antiserum, the membranes were treated with ^{125}I-protein A (80-90 μCi/ug) using 10^7 cpm/ml. The largest clone, designated FT2e, contained a 1 kb insert which was labelled and used to screen $7x10^5$ recombinants from a λgt10 porcine liver cDNA library constructed in this laboratory, as described below. Nylon filters were incubated overnight at 37°C in hybridization buffer [40% formamide, 5 X SSPE (1 x SSPE = 0.15 M NaCl, 10 mM NaH_2PO_4, 1.3M EDTA, pH 7.4), 5X Denhardt's solution, 1% dextran sulfate, 50 mM sodium phosphate, 0.1% SDS, 100 ug/ml denatured salmon sperm DNA] and labelled probe at 2.5 $x10^5$ cpm/ml. Filters were washed with 1 X SSC (1 X SSC = 0.15 M NaCl, 15 mM sodium citrate) at 42° for 15 minutes, with 0.1 X SSC, at 42° for 15 minutes, and with 0.1 X SSC at 55° for 15 minutes. Seven positive clones were isolated and plaque purified, 4 of which proved to be full-length.

RNA isolation. Yucatan pig liver, kindly provided by Dr. Gamal Selmy of the Royal Victoria Hospital, Department of Urology, was removed during surgical procedures and immediately frozen in liquid nitrogen. Total RNA was isolated using the urea/LiCl method (Auffray & Rougeon, 1980). Poly A+ mRNA was isolated by two passages over oligo dT cellulose, essentially as described (Aviv & Leder, 1972).

λgt10 porcine liver cDNA library construction. PolyA+ mRNA was converted to double stranded cDNA according to the λgt10 cDNA library construction kit from Bethesda Research Laboratories using Superscript Reverse Transcriptase and oligo dT as primer for first strand synthesis. The cDNA was ligated to EcoRI/NotI adaptors (Pharmacia) and cloned into the EcoRI site of λgt10. Recombinant phage were packaged (Bethesda Research Laboratories) and amplified one time.

cDNA sequencing The full-length cDNAs Cm1ea, Cm5e and Cm7e were subcloned into the EcoR1 site of Bluescript SK⁺. The sequencing strategy used is shown in Figure 2.1. To sequence the noncoding strand, progressive unidirectional deletions of Cm1ea were performed using Exonuclease III and S1 nuclease, as described by Henikoff (1987). Double stranded DNA of appropriate deletion mutants was prepared and sequenced using the dideoxy method of Sanger et al. (1977) and [α-^{35}S]dATP. To sequence the coding strand, Cm1ea, Cm5e and Cm7e were subjected to restriction digestion using sites shown in Figure 2.1. Restriction fragments were subcloned into Bluescript SK and double stranded DNA was prepared and sequenced. Synthetic oligomers, Sp1, Sp2, Sp3, were used to prime double stranded sequencing reactions to verify missing sequences. The complete sequence was obtained from each strand of cDNA.

Expression in E. coli. pBke-Cm1 was constructed, as described in Figure 2.2, to express the cDNA encoding FTCD in *E.coli*. The expression construct pBKe-HB1 was a generous gift from X-M. Yang. This plasmid contains a T7 RNA polymerase promoter, a translational enhancer sequence, a ribosomal binding site and an initiator ATG codon. It also includes a cDNA encoding the human bifunctional NAD dependent dehydrogenase/cyclohydrolase, which we replaced with Cm1ea, the cDNA encoding FTCD. The HB1 insert was removed from pKBe-HB1 by restricting with NcoI and filling in the overhang with Klenow polymerase, before digesting with ClaI. Prior to inserting our cDNA into the vector, Cm1ea was blunt-ended at the 5' end such that the insert begins with the first codon following the putative initiator ATG. To this end, Cm1ea was subcloned into the EcoR1 site of Bluescript KS+ and an NsiI site was introduced into the 5' end by oligonucleotide directed mutagenesis using the mutagenic oligomer 5'-GCCATGCATCCCAGCTG-3', essentially as described by Kunkel et al. (1987). The mutated insert was isolated from the vector pKS-NsiCm1 by digesting with NsiI, removing the resulting 3' overhang with Mung Bean Nuclease and then restricting with ClaI. This insert was then directionally ligated to the vector isolated from pBke-HB1 using T4 ligase. Double stranded sequencing of pBke-Cm1 verified that the resulting protein would include an amino acid sequence identical to that deduced from the Cm1ea

FIGURE 2.1. Physical map and sequencing strategy for FTCD cDNA. A. Restriction sites were determined both from sequence and by digestion with appropriate restriction enzymes, and are indicated along the coding (open bar) and untranslated (closed bar) regions of the cDNA. Abbreviations for restriction endonuclease sites: A, ApaI; BH, BamHI; BX, BstXI, K, KpnI; N, NotI; P, PstI; Sc, SacII; Sm, SmaI. **B.** Arrows indicate the direction and extent of cDNA sequence determined. **C.** FT2e is the partial cDNA isolated from the λgt11 porcine liver cDNA library; the cDNA clones Cm1ea - Cm7e inclusive were isolated from the λgt10 porcine liver cDNA library.

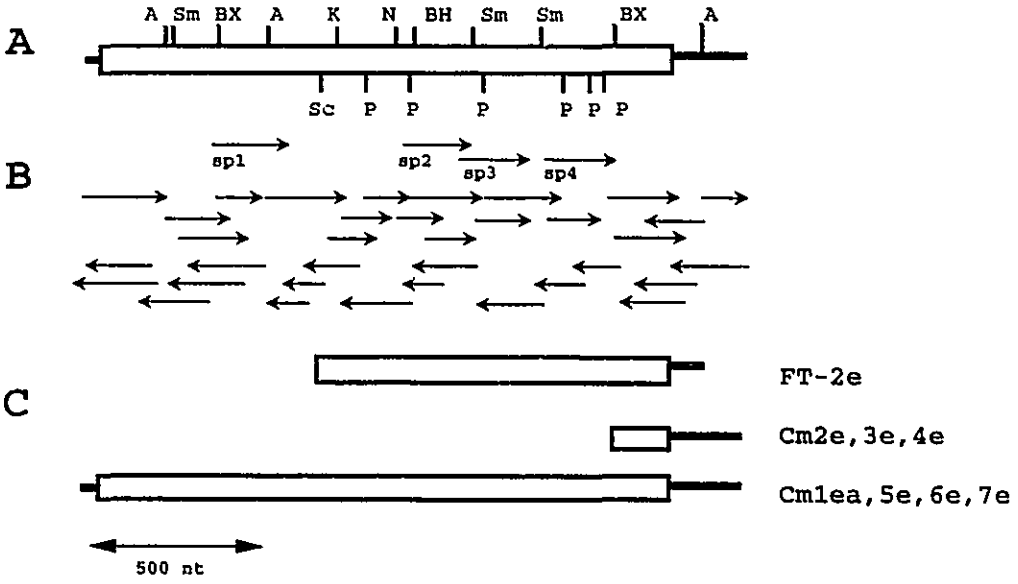
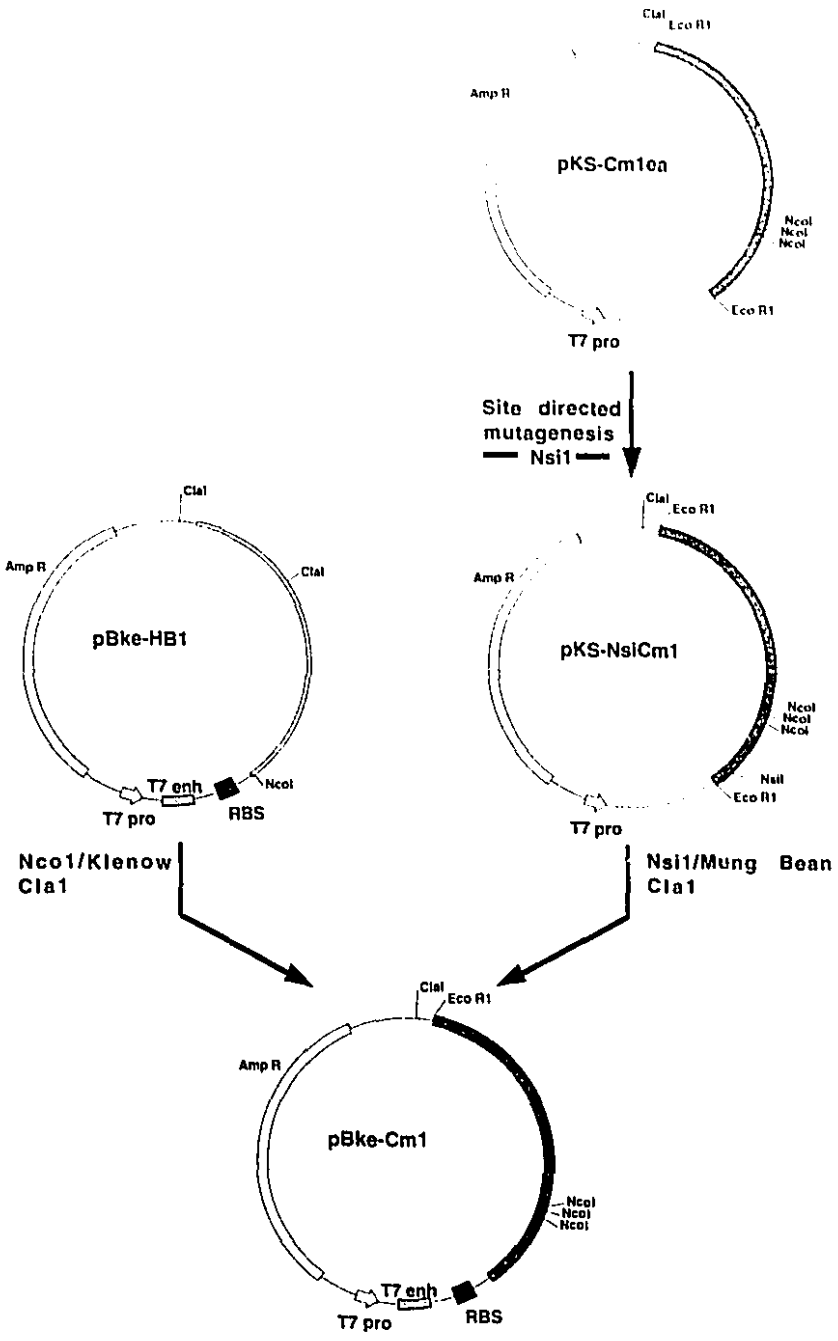

A
A Sm BX A K N BH Sm Sm BX A
Sc P P P P P P
B
sp1
sp2
sp3
sp4
C
FT-2e
Cm2e,3e,4e
Cm1ea,5e,6e,7e
500 nt

FIGURE 2.2. Construction of the expression plasmid pBKe-Cm1. The oligomer used to insert an NsiI site at the 5'coding region of Cm1ea is described in *Materials and Methods*. T7 pro, T7 enh and RBS represent, respectively, the T7 promoter sequence, the T7 enhancer sequence and the ribosome binding site.


ClaI
Eco R1
Amp R
pKS-Cm1ea
NcoI
NcoI
NcoI
Eco R1
T7 pro
Site directed
mutagenesis
Nsi1
ClaI
Amp R
ClaI
pBke-HB1
T7 enh
NcoI
T7 pro
RBS
ClaI
Eco R1
Amp R
pKS-NsiCm1
NcoI
NcoI
NcoI
NsiI
Eco R1
T7 pro
Nco1/Klenow
Cla1
Nsi1/Mung Bean
Cla1
ClaI
Eco R1
Amp R
pBke-Cm1
NcoI
NcoI
NcoI
T7 enh
T7 pro
RBS

cDNA. This construct was transfected into the K38 and the BL21(DE3) strains of *E.coli*. Western analysis and transferase activity assays were used to determine the level of FTCD in these lysates.

Purification of recombinant formiminotransferase-cyclodeaminase. Transformed BL21(DE3) cells were incubated at 37°C in 5 liters of Terrific Broth + 200 µg/ml ampicillin in a New Brunswick Scientific BIOFLO IIc fermenter. Cells were induced at OD_{600} 1.0-1.5 by the addition of IPTG to a final concentration of 0.4 mM. After 30 minutes, rifampicin was added to a final concentration of 15 µg/ml. After 2 more hours at 37°C, cells were harvested by centrifugation at 4500g for 30 minutes. The pellet was washed once with 0.1 M potassium phosphate, pH 7.3, 1 mM benzamidine, 1 mM PMSF. Cells were frozen and stored at -80°C.

Frozen cells (18-20 g) were thawed on ice and lysed in 3 volumes of sonication buffer (0.1 M potassium phosphate, pH 7.3, 35 mM 2-mercaptoethanol, 1 mM benzamidine, 1 mM PMSF). Aliquots (20 ml) of resuspended cells were sonicated for 5 minutes at 15 second intervals using a Vibra-Cell model cv17 sonicator (Sonics and Materials, Inc), and soluble and insoluble fractions were separated by centrifugation at 25000g for 30 minutes. To the supernatant fraction was added 0.1 volume of 250 mM MOPS, pH 7.3 and 0.24 volume of glycerol. Ammonium sulfate was slowly added, while stirring on ice, to 35% saturation. After 30 minutes, the suspension was centrifuged at 25000g for 30 minutes. The pellet was resuspended in one tenth of the original volume with Buffer A (25 mM MOPS, pH 7.3, 5 mM potassium phosphate, pH 7.3, 1 mM benzamidine, 20 % glycerol, 35 mM 2-mercaptoethanol, 0.02 % Triton X 100) and 1 mM PMSF. This was dialysed overnight against 2 X 1.5 l of Buffer B (Buffer A, with glycerol increased to 30 %) + 0.5 mM PMSF. The dialysed solution was brought to 50 ml by adding Buffer A + 1 mM PMSF and then centrifuged for 30 minutes at 25000g before applying at a rate of 130 ml/hr to a DEAE Sepharose column (4.5 x 10 cm) equilibrated in Buffer A. After washing with 3 column volumes of Buffer A, the protein was eluted using a linear gradient of 300 ml Buffer A and 300 ml 0.3 M KCl in Buffer A. The fractions containing transferase activity were pooled and dialysed overnight against 2 X

1.5 l of Buffer B + 0.5 mM PMSF. The enzyme was then applied at a rate of 110 ml/hr to a Dye Matrex Blue A column (5 x 4 cm) equilibrated in Buffer A. The column was washed with approximately 3 column volumes of Buffer A before eluting with a linear gradient of 250 ml Buffer A and 250 ml 0.8 M KCl in Buffer A. The transferase-active fractions were pooled and dialysed overnight against 2 X 1.8 l of Buffer B. The dialysate was then applied at a rate of 45 ml/hr to a Heparin Sepharose column (3.5 x 3 cm) previously equilibrated in Buffer A. This column was washed with 3 column volumes of Buffer A before eluting with a linear gradient of 50 ml Buffer A and 50 ml 0.3 M KCl in Buffer A. The fractions containing transferase activity were pooled and dialysed overnight against 2 X 1.6 l of Buffer B before applying the sample at a rate of 20 ml/hr onto an Affigel 15 polyglutamate column (5 x 1.7 ml) pre-equilibrated in Buffer A. The column was washed with 3 volumes of Buffer A before the purified enzyme was eluted using a linear gradient of 50 ml Buffer A and 50 ml 0.8 M potassium chloride in Buffer A. The transferase-active fractions were pooled and purity was assessed by SDS PAGE (Laemmli, 1970) using 9% gels.

Enzyme assays. Formiminotransferase assays were performed as previously described (Drury et al., 1975; MacKenzie, 1980). Protein determinations were performed using the method of Bradford (1976), or if detergent was present, protein was precipitated and concentrations were determined by the modified Lowry method of Bensadoun and Weinstein (1976).

Western analysis. Samples of protein extracts of *E.coli* were electrophoresed on SDS PAGE using 9% gels (Laemmli, 1970). Separated proteins were electroblotted onto nitrocellulose membranes. The membranes were blocked for a minimum of 2 hours in 5% skim milk containing 0.1% normal goat serum, and incubated for one hour in a 1/1000 dilution of polyclonal anti-FTCD antiserum in TBS. Detection was performed using an alkaline phosphatase conjugated anti-rabbit second antibody, as described for the Promega Protoblot system.

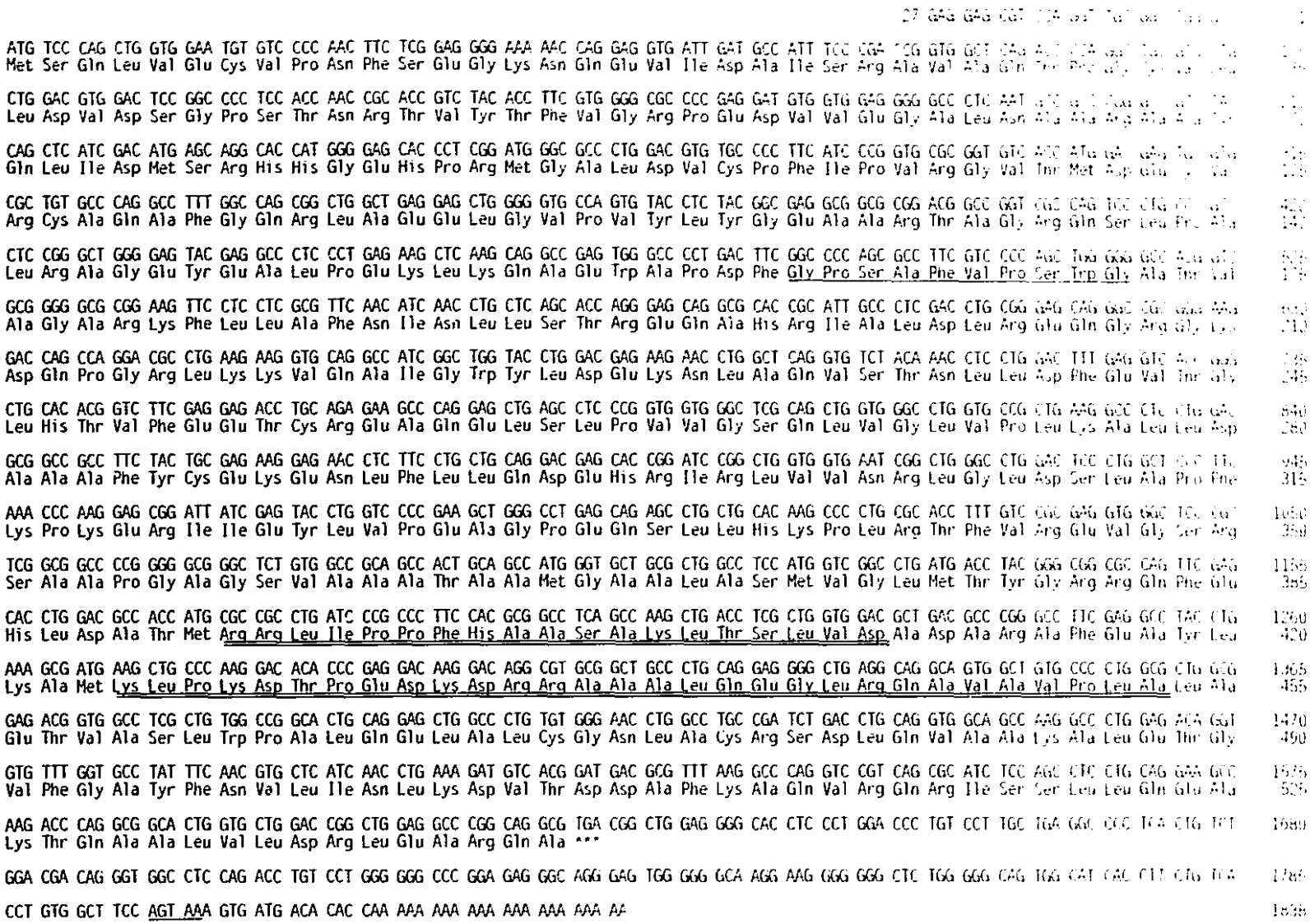
Isolation and Characterization of cDNA clones To isolate the cDNA encoding formiminotransferase-cyclodeaminase, we screened a Clontech λgt11 porcine liver cDNA library with a rabbit polyclonal antibody raised against the porcine enzyme. A total of 6 x 10^6 plaques were screened and 55 putative positives were identified, 35 of which were plaque purified. EcoR1 inserts from these clones, ranging in size from 500 - 1100 bp, cross hybridize on Southern Blot analysis. The largest partial cDNA, FT2e, was missing sequences from both the 5' and 3' ends, but encoded amino acid sequences identical to those in the porcine enzyme as determined by N-terminal sequencing of cyanogen bromide fragments of FTCD.

Because a full-length cDNA could not be isolated from this library, even after extensive screening, we constructed a λgt10 pig liver cDNA library using polyA+ mRNA isolated from Yucatan pig liver. After screening 7 X 10^5 plaques with the partial clone FT2e, seven positives were identified (CM1e to 7e, refer to Figure 2.1). Four clones contained a 1.9 kb EcoR1 insert (CM1e,5e,6e,7e), and the remaining 3 contained a 500 bp EcoR1 insert (CM2e,3e,4e) which proved to be partial sequences of the larger cDNA. CM1e also contained an unrelated 200 bp fragment. Northern analysis suggested that the FTCD message is liver specific and present as a single species of approximately 1.9-2 kb (unpublished observations).

Nucleotide Sequence and Deduced Amino Acid Sequence. The longest clone has a nucleotide sequence of 1865 bp (Figure 2.3). A 27 nucleotide 5' untranslated region precedes an open reading frame of 1623 bp which encodes a 59 kDa protein, followed by a 215 bp 3' noncoding region. The putative initiation codon, located at nucleotides 28-30, is the most plausible site for initiation of translation as it is preceded by a GCC sequence and is the first ATG from the 5'end (Kozak, 1987). The 3' noncoding sequence includes a possible polyadenylation signal, AGTAA (Birnstiel, 1985), located 14 nucleotides upstream from the start of the polyadenylated tail.

The deduced primary sequence of 541 amino acids is described in Figure 2.3. The

FIGURE 2.3. DNA sequence of the FTCD cDNA and the deduced amino acid sequence. Numbering for the cDNA starts at the putative initiation codon. The amino acid sequence is numbered sequentially from the N-terminus of the predicted protein sequence. The stop codon is indicated by asterix. Sequences identical to those determined from the porcine liver enzyme are doubly underlined. A sequence possibly involved in folate binding as described in the results and discussion is singly underlined, as is the putative polyadenylation signal in the 3' untranslated region.



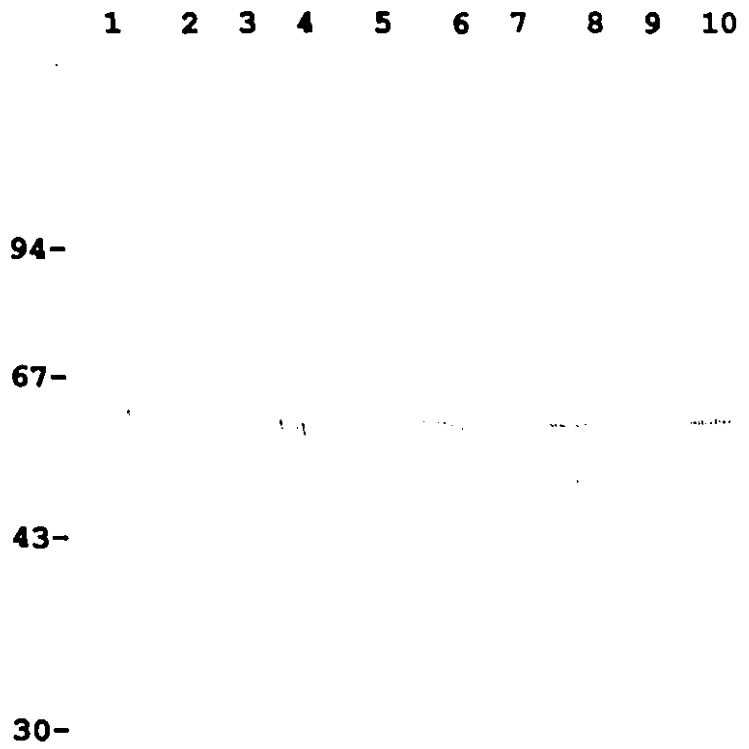
doubly underlined sequences are identical to peptide sequences determined by amino acid sequencing of two CNBR cleavage fragments of the porcine liver enzyme, confirming that this cDNA does indeed encode FTCD.

A FastA search of the SWISS-PROT data base and a TFastA search (Pearson & Lipman, 1988) of the GenBank/EMBL data bases (Genetic Computers Group) found no significant homology between FTCD and other known protein sequences. The amino acid sequence includes a high proportion of basic and hydrophobic residues which correlates well with two previously described amino acid composition analyses of the porcine enzyme (Beaudet & MacKenzie, 1976; Findlay et al., 1989).

No general folate binding sequence has been described to date. However, Cook et al. (1991) have identified a putative 10-formylH_4PteGlu binding consensus sequence, **XPS**$(X)_2$**P**$(X)_{2,3}$**G**, which has been observed in 10-formylH_4PteGlu binding enzymes from several different sources (Nour & Rabinowitz, 1992; Rankin et al., 1993). A similar sequence, **GPSAFVPSWG**, is found in FTCD at residues 163-172 and may play a role in H_4PteGlu recognition. This, or a similar sequence, is not repeated elsewhere in FTCD's coding region.

Expression in E.coli. To confirm that the isolated cDNA can direct expression of functional FTCD we expressed Cm1ea in *E.coli* using procedures based on the Tabor and Richardson's T7 Expression System (Tabor & Richardson, 1985). The expression vector, pKBe-HB1, which had been engineered and successfully utilized in this laboratory by X-M. Yang, provided the T7 promoter, ribosome binding site and initiation codon. It also contained a T7 translational enhancer element previously described by Olins and Rangwala (1989). The coding region following the putative initiator methionine was subcloned into this plasmid as described in Figure 2.2. This construct was transfected into K38 and BL21(DE3) cells. Western analysis of induced extracts indicate that the construct produced a soluble 59 kDa protein, approximately the same size as the enzyme from porcine liver (Figure 2.4). Despite the size of the octameric enzyme (approximately 480 kDa) more than 90% of the enzyme is found in the soluble fraction.

FIGURE 2.4. Comparison of the amounts of FTCD in the soluble and the insoluble fraction of *E. coli* by Western blot analysis. Soluble and insoluble fractions of K38 cell extracts were separated by SDS PAGE and electroblotted onto nitrocellulose. The blots were probed with polyclonal antiserum as described in the *Materials and Methods*. Lane 1, 100 ng porcine FTCD; lane 2, 20 µg K38 whole cell extract (without pBKe-Cm1); lanes 3 to 10 contain insoluble and soluble fractions of cell extracts prepared from induced K38 cells harbouring pBKe-Cm1. Lanes 4,6,8 and 10 contain 30,20,10 and 5 µg of soluble protein, respectively. Insoluble fractions were dissolved in SDS sample buffer and readjusted to original volumes. Lanes 3,5,7 and 9 contain volumes equivalent to those used in lanes 4,6,8 and 10. The numbers refer to the molecular mass of the reference proteins (kiloDaltons).


1 2 3 4 5 6 7 8 9 10
94-
67-
43-
30-

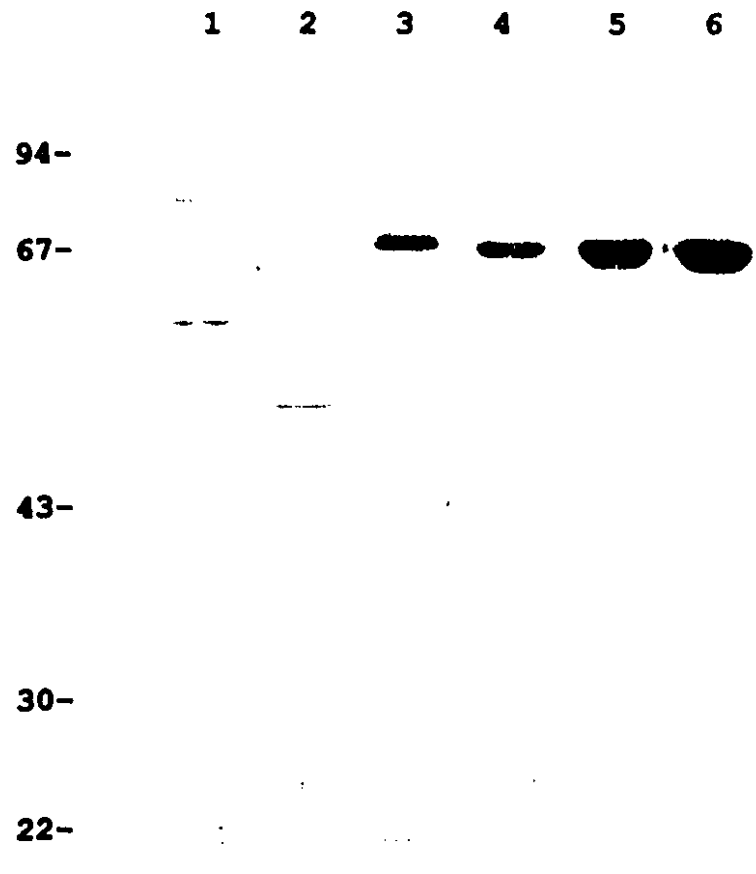
Purification of Formiminotransferase-Cyclodeaminase from E. coli extracts The purification of the recombinant enzyme, as outlined in Table 2.1, relies on an ammonium sulfate precipitation and several chromatographic procedures including a novel step, chromatography on Affigel 15 polyglutamate, which might make use of the polyglutamate binding site on this enzyme. The purified recombinant enzyme has a specific activity of 39-41 units/mg which is in the range observed for the purified porcine liver enzyme (Drury et al., 1975; Findlay et al., 1989). SDS PAGE of samples from different stages of purification (shown in Figure 2.5) confirms that the enzyme is ≥ 95 % pure after the polyglutamate column.

The enzyme also displays cyclodeaminase activity and elutes in the void volume upon gel filtration on Ultrogel ACA 34 (exclusion limit of 350 kDa), indicating that the subunits associate to form the native octameric structure of approximately 480 kDa (data not shown). With an effective expression system in hand, we can direct our attention towards elucidating the domain structure of this bifunctional enzyme and determining the nature of its subunit interactions.

TABLE 2.1. Purification of recombinant formiminotransferase-cyclodeaminase

Fraction	Volume ml	Protein mg	Transferase activity (μmol/min)	Specific activity (μmol/min/mg)	Yield %
Crude extract	98	749	176	0.235	100
Ammonium sulfate	49	111	123	1.11	70
DEAE Sepharose	122	14.0	77	5.5	44
Matrex Blue A	132	4.7	71	15.1	40
Heparin Sepharose	36	1.9	51	26.8	29
Affigel 15 poly-glutamate	35	1.1	45	40.9	26

FIGURE 2.5. SDS PAGE analysis of fractions from various steps in the purification of FTCD from *E. coli*. Electrophoresis was carried out as described in the text and the gel was stained with Coomassie Blue. Lane 1, crude extract, 10 µg; lane 2, 35% ammonium sulfate fraction, 10 µg; lane 3, DEAE Sepharose, 10 µg; lane 4, Matrex Blue A, 7.5 µg; lane 5, Heparin Sepharose, 7.5 µg; lane 6, Affigel 15 polyglutamate, 7.5 µg. The numbers refer to the molecular mass of the reference proteins (kiloDaltons).


1
2
3
4
5
6
94-
67-
43-
30-
22-

ACKNOWLEDGEMENTS

We thank Dr Peter Roughley for peptide sequences, Dr Gamal Selmy for providing fresh porcine liver, Kathy Teng for photography and Xiao Ming Yang for the expression vector pBKe-HB1

CHAPTER 3

THE TWO MONOFUNCTIONAL DOMAINS OF OCTAMERIC FORMIMINOTRANSFERASE-CYCLODEAMINASE EXIST AS DIMERS

ABSTRACT

Formiminotransferase-cyclodeaminase is a bifunctional enzyme arranged as a circular tetramer of dimers that exhibits the ability to efficiently channel polyglutamylated folate between catalytic sites. Through deletion mutagenesis we demonstrate that each subunit consists of an N-terminal transferase active domain and a C-terminal deaminase active domain separated by a linker sequence of minimally 8 residues. The full-length enzyme and both isolated domains have been expressed as C-terminally histidine-tagged proteins. Both domains self dimerize providing direct evidence for the existence of two types of subunit interfaces. The results suggest that both the transferase and the deaminase activities are dependent on the formation of specific subunit interfaces. Because channelling is not observed between isolated domains, only the octamer appears able to directly transfer pentaglutamylated intermediate between active sites.

INTRODUCTION

The bifunctional enzyme formiminotransferase-cyclodeaminase (FTCD) catalyzes two sequential reactions in the histidine degradation pathway. This enzyme transfers a one carbon unit from formiminoglutamate to tetrahydrofolate, thus serving as an additional entry point to the folate pool in liver [for a review see Shane and Stokstad (1984)] Although electron microscopy has shown that porcine liver FTCD consists of 8 identical subunits arranged to form a circular octamer (Beaudet & MacKenzie, 1976), the enzyme is more aptly described as a tetramer of dimers (MacKenzie et al., 1980). A series of denaturation and renaturation experiments indicated that octameric FTCD can dissociate to form two distinct types of monofunctional dimers (Findlay & MacKenzie, 1987; 1988). These dimers display different catalytic activities and affinities for substrate and were proposed to isolate different subunit interfaces. According to this hypothesis, the tetramer of dimers includes two types of subunit interaction which must be maintained for concurrent expression of both the transferase and deaminase activities.

FTCD can directly transfer polyglutamylated formiminotetrahydrofolate between the transferase and deaminase active sites (MacKenzie, 1979; MacKenzie & Baugh, 1980). As first demonstrated by MacKenzie (1979), the efficiency of channelling is dependent on the length of the polyglutamate tail attached to the folate. The specificity of the enzyme for pentaglutamate led MacKenzie and Baugh (1980) to suggest that the polyglutamate chain may act to anchor the substrate to the octamer, while allowing the pteroyl moiety to move between the two types of catalytic sites. Further support for this model was realized when Paquin et al. (1985) observed only 4 high affinity polyglutamate binding sites per octamer. This, in combination with other kinetic experiments, indicated the existence of only one polyglutamate binding site per pair of transferase/deaminase active sites. Findlay and MacKenzie (1988) observed that after dissociation of the FTCD octamer, the polyglutamate specificity was retained by the deaminase active dimer

Multifunctional enzymes are thought to be composed of different modules responsible for specific binding and/or catalytic functions (Wetlaufer, 1973; Rossman & Argos, 1981). As the transferase and deaminase activities are kinetically independent and

a transferase active fragment can be isolated after limited proteolysis (Tabor and Wyngarden, 1959; MacKenzie, 1979; MacKenzie et al., 1980), it seemed possible that the two activities reside within separable domains. Therefore deletion mutagenesis of the FTCD cDNA (Murley et al., 1993) was used to isolate separate transferase and deaminase active domains and to delineate the interdomain region. These domains were characterized in terms of their quaternary structure and ability to channel the product of the transferase activity to the deaminase active site.

MATERIALS AND METHODS

Materials. Restriction and DNA modifying enzymes were obtained from Bethesda Research Laboratories, and New England Biolabs. All reagents for sequencing DNA were supplied by USB Corp. Nitrocellulose membranes were from Schleicher and Schuell. ^{125}I-labelled Protein A and ^{35}S-dATP were purchased from Amersham Corporation. Oligonucleotides were synthesized and purified by the Sheldon Biotechnology Centre, McGill University, or General Synthesis and Diagnostics.

DEAE-Sepharose and the Superose 6 HR 10/30 column were purchased from Pharmacia. Ni-NTA matrix came from Quiagen and DEAE53 from Whatman. Folic acid and formimino-L-glutamic acid were from Sigma. Pteroylpenta-γ-glutamic acid was from Dr. B. Schircks Laboratories (Jona, Switzerland). All other chemicals were of reagent grade.

Folic acid and pteroylpenta-γ-glutamic acid were converted to the corresponding (6*S*)-tetrahydro-derivatives and purified on DEAE-cellulose as described by MacKenzie and Baugh (1980). 5-formiminoH_4PteGlu was prepared enzymatically as described previously (Paquin et al., 1985) except that purified recombinant transferase domain was used to catalyze its synthesis. Unless otherwise noted, molecular biology techniques were performed as outlined in Sambrook et al. (1989).

Bacterial Strains and Plasmids. pBke-Cm1 is a previously described FTCD expression vector (Murley et al., 1993). pET23d (Novagen) was the source of the C-terminal hexahistidine tag. *E. coli* strain BL21/DE3 was employed for expression of recombinant proteins, DH5α was used for cloning purposes and CJ236 was used to produce uracil-containing single strand DNA templates for site-directed mutagenesis (Kunkel et al., 1987).

Preparation of N-terminal and C-terminal Deletion Constructs. The FTCD expression plasmid pBke-Cm1 was used to construct vectors that would express separately the transferase and deaminase domains. The terminology used to describe these vectors

incorporates the amino acid numbering of the full-length FTCD. The N-terminal fragment represented by FT327-1 is a protein starting at the FTCD initiator methionine and terminating after residue 327 plus one non-related amino acid. CD333 represents a protein beginning at FTCD residue 333 and ending at the normal stop codon after residue 541. The plasmids expressing these fragments are pBke-FT327-1 and pBke-CD333.

FTCD proteins truncated at the C-terminus were generated by one of two strategies. FT318-1, FT321-3, FT327-1, FT331, FT335-2 and FT340-2 were produced through Exo/S1 deletion mutagenesis (Henikoff, 1987) and insertion of translational stop codons in all 3 reading frames. Dideoxy sequencing was used to identify constructs which terminated at the desired positions (Sanger et al., 1977). To construct FT322 and FT325, stop codons were introduced immediately following the codons for residue 322 or 325 by oligonucleotide-directed mutagenesis (Kunkel et al., 1987) of pBke-Cm1. Clones containing the desired mutations were identified by dideoxy sequencing.

To create constructs expressing the C-terminal fragments (CD), pBke-Cm1 was re-engineered to remove the intervening coding sequence between the initiating ATG and the desired initiating residue, using a strategy previously described for the construction of pBke-Cm1 (Murley et al., 1993). The new translation initiation site was confirmed by dideoxy sequencing.

Construction of Plasmids Expressing Histidine Tagged FTCD and Domains. Hexahistidine tags (H_6) were engineered onto the C-terminal ends of FTCD and the N- and C-terminal domains. To produce $FTCDH_6$, we first replaced the FTCD stop codon in pBke-Cm1 with an *Xho*I site, creating pBke-Cm1*Xho*I-1. Then, using standard molecular biology techniques, an *Xho*I/*Esp*I fragment encoding the hexahistidine tag from pET23d was inserted in frame immediately after the new *Xho*I site. Thus pBke-$FTCDH_6$ expresses an FTCD with 8 additional amino acids attached to its C-terminus: leucine, glutamate, and six additional histidines. To construct pBke-CDH_6, which expresses a histidine-tagged C-terminal domain (CDH_6), a fragment encoding the N-terminus of pBke-$FTCDH_6$ was replaced with the corresponding fragment from pBke-CD333. pBke-$CD339H_6$ was constructed in a similar manner.

In order to construct pBke-FTH$_6$, a vector expressing the histidine-tagged N-terminal domain (FTH$_6$), we first introduced an *Xho*I site into pBke-Cm1 immediately after codon 328. The cDNA following this *Xho*I site was replaced with an in frame *Xho*I/*Bfa*I fragment from pET23d, which provided the same 8 amino acids and stop codon described above.

Western Analysis. BL21/DE3 expressing the N- and C- terminal deletion mutants were grown and harvested as previously described (Murley et al., 1993). Frozen *E. coli* cell pellets were suspended in approximately 3 volumes of sonication buffer I [0.1 M potassium phosphate (pH 7.3), 25 mM 2-mercaptoethanol, 1 mM benzamidine, 1 mM PMSF] and sonicated for a total of 10 x 15 seconds. Following centrifugation, the insoluble fraction was resuspended in a volume of SDS sample buffer equivalent to the volume of the soluble fraction and aliquots of both fractions were analyzed by Western analysis. Soluble protein concentrations were determined by Bradford analysis (1976) and samples were precipitated (Bensadoun and Weinstein, 1976) before separation on SDS-PAGE slab gels (Laemmli, 1970). Proteins were transferred onto nitrocellulose using a Tyler Research Instruments semi-dry apparatus (Khyse-Anderson, 1984). After blocking, blots were incubated with a polyclonal anti-FTCD antibody and detection was performed with ^{125}I-Protein A as previously described (Murley et al., 1993).

Purification of Histidine-tagged Proteins. Wildtype FTCD was purified as described previously (Murley et al., 1993). Histidine-tagged proteins were purified from frozen cell pellets as outlined below. 17 g of BL21/DE3 cells expressing FTCDH$_6$ were thawed on ice in 3 volumes of sonication buffer II (sonication buffer I with pH increased to 7.8 and 2-mercaptoethanol decreased to 10 mM) and lysed by sonication on ice for 3 minutes (12 x 15 seconds). After centrifugation (25000 g for 30 minutes), the composition of the supernatant was altered to include the following: 0.1 M potassium phosphate (pH 7.8), 0.5 M NaCl, 5 mM histidine, 5 mM glutamate, 0.1% Triton X-100, 20% glycerol, 10 mM 2-mercaptoethanol. This is also the composition of the Binding Buffer. The supernatant (80 ml) was added to 20 ml of a 50% slurry of Ni-NTA resin in Binding Buffer, and

mixed end over end for 1 hour at 4 °C. This mixture was packed into a column under gravity. The column was washed with 3 column volumes of Binding Buffer and Wash Buffer (Binding Buffer with histidine increased to 50 mM). The protein was eluted at a rate of 15 ml/hr using a linear gradient of 30 ml of Wash Buffer (pH 7.8) and 30 ml of 380 mM histidine in Binding Buffer (pH 7.3). Fractions containing activity were pooled and dialysed overnight against 2 x 2 litres of Buffer B (25 mM MOPS (pH 7.3), 5 mM potassium phosphate (pH 7.3), 30% glycerol, 0.02% Triton X-100, 35 mM 2-mercaptoethanol, 1 mM benzamidine). The dialysed protein was further purified by Affigel 15-polyglutamate chromatography as reported for the purification of FTCD (Murley et al., 1993). CDH_6 and $CD339H_6$ were purified as described for $FTCDH_6$.

As FTH_6 does not bind as tightly to the Ni-NTA resin, the Ni-NTA column was washed with 3 column volumes of Binding Buffer before eluting with a linear gradient of 5 mM to 150 mM histidine in Binding Buffer at pH 7.8. Fractions containing transferase active protein were pooled and dialysed as above. The dialysed protein was loaded onto a DEAE Sepharose column equilibrated in Buffer A (Buffer B with glycerol decreased to 20%) and washed with 3 column volumes of the same buffer. FTH_6 was eluted with a linear gradient of 0 to 0.3 M KCl in Buffer A.

N-Terminal Sequencing of the deaminase domain. Purified CDH_6 was bound to Ni-NTA resin and washed with 10 volumes of 0.1 M potassium phosphate, pH 7.8, 0.5 M NaCl, 20% glycerol, 10 mM 2-mercaptoethanol and 20 mM imidazole, to remove traces of Triton X-100. The enzyme was eluted using the same buffer with imidazole increased to 0.3 M. The protein was extensively dialysed against H_2O to remove glycerol, prior to lyophilization. $CD339H_6$ was treated in the same manner and then further purified by chromatography on a reverse phase C18 column (Vydac), using a gradient of 0 to 70% acetonitrile in 0.1% TFA, on a Hewlett Packard 1090M HPLC equipped with a diode array detector. The N-terminal amino acid sequences of CDH_6 and $CD339H_6$ were determined at the Biotechnology Research Institute, National Research Council of Canada (Montreal, Quebec), using an Applied Biosystems 470-A gas phase sequencer coupled to a 120-A PTH-amino acid separation system.

Tryptic digestion and Mass Spectral Analysis of the transferase domain. Purified FTH_6 was bound to Ni-NTA resin and extensively washed with 0.1 M potassium phosphate, pH 7.8, 0.5 M NaCl and 10 mM imidazole to remove glycerol and detergent, and then eluted using the same buffer with imidazole increased to 0.3 M. Eluted FTH_6 was dialysed against 2 x 1 litre of 50 mM ammonium bicarbonate. Dialysed FTH_6 (24 µg) was digested with 0.48 µg of TPCK-treated trypsin (Worthington Enzymes) in 50 mM ammonium bicarbonate (total volume of 85 µl) for 4 hours at 37°C. LC\MS analysis of 20 µl of the tryptic digest was performed by C. Fenwick and G. Tsaprailis, Concordia University. The peptides were separated using a Hewlett Packard 1090 HPLC with a Reverse Phase C18 column (Vydac) and a gradient of 0 to 80% acetonitrile in 0.05% TFA, at a flow rate of 40 µl/min. The HPLC was directly coupled to a Finnigan-MAT SSQ 7000 mass spectrometer equipped with an electrospray ionization source.

Enzyme Assays and Channelling Experiments. Routine assays of transferase and deaminase activity were performed as previously described (MacKenzie, 1980; Drury & MacKenzie, 1975). The K_m value for formiminoglutamate (FIGLU) was determined at 30°C in assay mix containing 0.1 M potassium phosphate (pH 7.3), 35 mM 2-mercaptoethanol, 1 mM H_4PteGlu and various concentrations of FIGLU. The H_4PteGlu K_m value was determined using 5 mM FIGLU and various concentrations of (6*R*,*S*)-H_4PteGlu. To determine the H_4PteGlu$_5$ K_m value the assay mix included 5 mM FIGLU and various concentrations of (6*R*,*S*)-H_4PteGlu$_5$, with 118 mM NaCl. The kinetic constants for the deaminase activity were determined at 30°C using the assay conditions described by Paquin et al. (1985), with 2-mercaptoethanol decreased to 35 mM. The data from these experiments were fit to the Michaelis-Menten equation using the non-linear regression analysis program "Enzfitter" (Leatherbarrow, 1987).

Channelling of the formimino intermediate between active sites was monitored by following the time course of appearance of the products of the transferase and the deaminase reactions, as described previously (Paquin et al., 1985) with 2-mercaptoethanol decreased to 35 mM and 40 ng of FTCD or equivalent units of either $FTCDH_6$, or FTH_6 and CDH_6 added to each assay.

Gel filtration Purified samples of FTCD, $FTCDH_n$, FTH_n and CDH_n were analyzed on a Superose 6 HR 10/30 column equilibrated in 0.1 M potassium phosphate (pH 7.3), 20% glycerol, 0.02% Triton X-100 and 35 mM 2-mercaptoethanol. Aliquots containing 200 µl of approximately 2 mg/ml protein were injected onto the column and chromatographed at a flow rate of 0.2 ml/min. Absorbance of the eluate was monitored at OD_{280}. Ferritin (440 kDa), aldolase (158 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa) and chymotrypsinogen (25 kDa) were used as molecular weight standards. The void volume was determined using blue dextran.

Deletion analysis of FTCD. The plasmids expressing N- and C-terminal fragments of FTCD contained the same translational enhancer sequence, ribosome binding site and intervening sequence up to and including the initiation codon. The plasmids were transformed into BL21/DE3 cells and the resulting protein products were analyzed by enzyme assay and Western blotting of cell extracts. Figures 3.1, 3.2 and 3.3 summarize the results from these analyses.

The FT mutants (Figure 3.1) begin at the same initiator methionine as the wildtype FTCD and end between residues 318 and 335. As shown in Figure 3.2A, mutants ending at or before residue 322 are completely insoluble while larger, transferase-active fragments are at least partially soluble. FT325 partitions between both the soluble and insoluble fractions. When aliquots of soluble extract containing equivalent units of transferase activity were analyzed by Western blotting (Figure 3.2B), each lane contained approximately equal amounts of immunoreactive protein. This indicates that FT325, FT327-1, FT331 and FT335-2 have similar intrinsic transferase activity.

To isolate a C-terminal fragment which did not contain sequence overlap with the smallest transferase active domain, CD327 was constructed. As shown in Figure 3.1, CD327 expresses cyclodeaminase activity. To delineate the N-terminal boundary of the deaminase active domain, mutants beginning at positions further 3' were constructed and the expressed proteins were analyzed in terms of their solubility and deaminase activity. Truncated proteins initiating at or before residue 341 expressed some level of deaminase activity while smaller fragments were completely inactive. All constructs produced soluble protein (Figure 3.3A). However, CD341 was susceptible to proteolysis and CD344 and CD350 were not as highly expressed as their larger counterparts. When aliquots containing equivalent units of deaminase activity are compared (Figure 3.3B), it is clear that the intrinsic deaminase activity is highest for CD333 and CD334 and decreases dramatically as the domain is shortened.

FIGURE 3.1. Analysis of the interdomain region of FTCD. The sites of termination of translation of the N-terminal FT domains and initiation of translation of the C-terminal CD domains are indicated in bold (initiator methionines are not shown). Non-related residues attached to the N-terminal domains are written in italics. The minimum linker region of 8 residues is underlined. The level of enzyme activity in *E. coli* extracts is indicated by: +, >0.001 and <0.1 μmoles/min/mg; ++, >0.1 μmoles/min/mg; -, not detectable.

Transferase Activity		
-	**FT318-1**	PK*D*
-	**FT321-3**	PKERI*MIN*
-	**FT322**	PKERII
++	**FT325**	PKERIIEYL
++	**FT327-1**	PKERIIEYLVP*D*
++	**FT331**	PKERIIEYLVPEAGP
++	**FT335-2**	PKERIIEYLVPEAGPEQSL*IN*
++	**FT340-2**	PKERIIEYLVPEAGPEQSLLHKPL*IN*

317 352

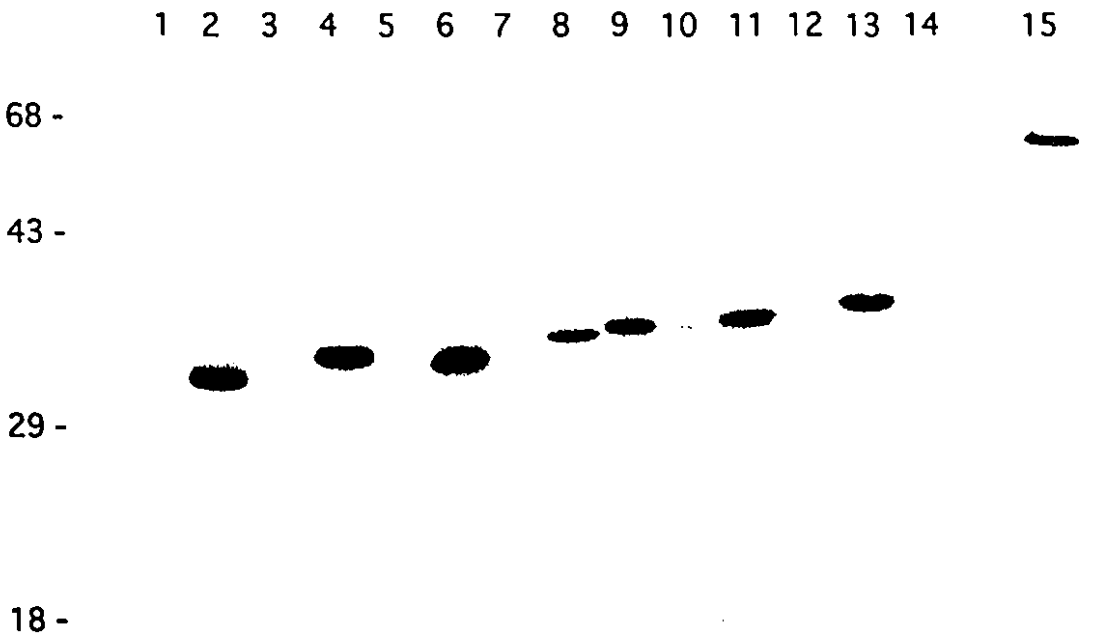
.....PKERIIEYLVPEAGPEQSLLHKPLRTFVREVGSRSA.....

PEAGPEQSLLHKPLRTFVREVGSRSA	**CD327**	++
QSLLHKPLRTFVREVGSRSA	**CD333**	++
SLLHKPLRTFVREVGSRSA	**CD334**	++
HKPLRTFVREVGSRSA	**CD337**	+
PLRTFVREVGSRSA	**CD339**	+
RTFVREVGSRSA	**CD341**	+
VREVGSRSA	**CD344**	-
RSA	**CD350**	-

Deaminase Activity

FIGURE 3.2. Western analysis of the N-terminal fragments. A. Lanes 1 to 14 contain insoluble and soluble fractions of *E. coli* extracts expressing the following N-terminal fragments: (1 & 2) FT318-1; (3 & 4) FT321-3; (5 & 6) FT322; (7 & 8) FT325; (9 & 10) FT327-1; (11 & 12) FT331; (13 & 14) FT335-2. Odd numbered lanes contain 25 µg (1, 3 & 5) or 5 µg (7, 9, 11 & 13) of soluble protein. Even numbered lanes contain volume equivalent amounts of insoluble fractions. Lane 15, 10 ng purified FTCD. **B.** Lanes with aliquots of soluble extract containing 5 nmoles/min of transferase activity: (1) FT325; (2) FT327-1; (3) FT331; (4) FT335-2.

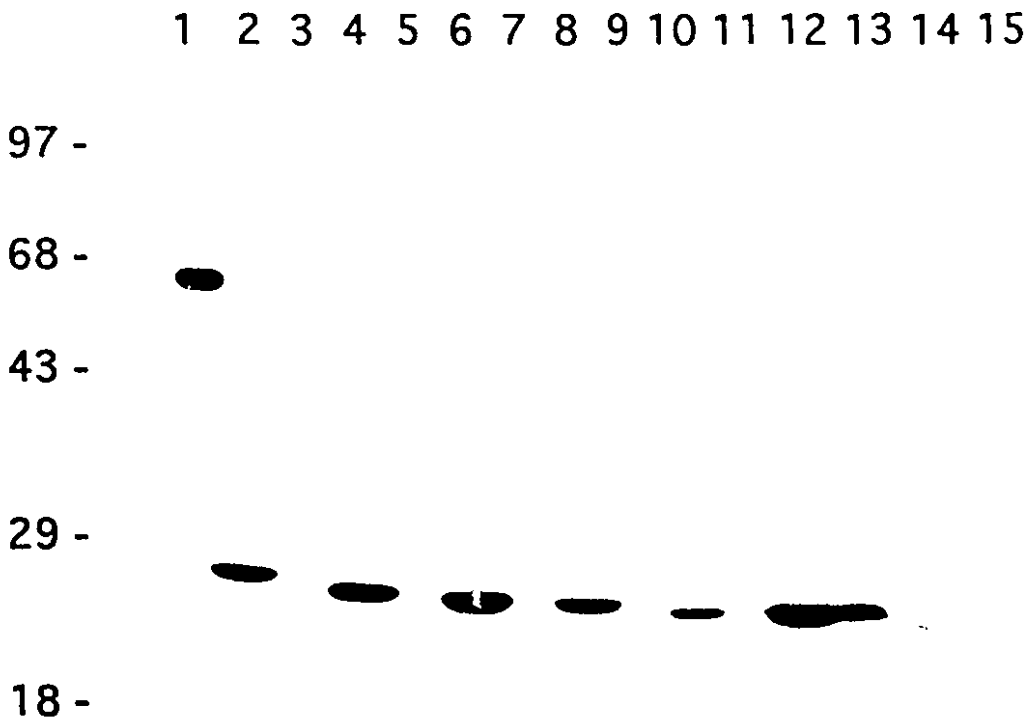
A Equal Protein Loaded



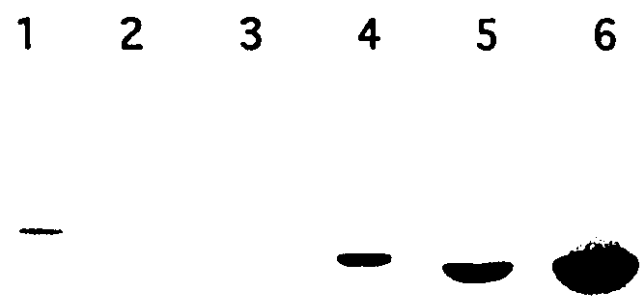
B Equal Activity Loaded

FIGURE 3.3. Western analysis of the C-terminal fragments. A. Lane 1, 10 ng purified FTCD. Lanes 2 to 15 contain insoluble and soluble fractions of *E. coli* extracts expressing the following C-terminal fragments: (2 & 3) CD327; (4 & 5) CD333; (6 & 7) CD334; (8 & 9) CD337; (10 & 11) CD339; (12 & 13) CD341; (14 & 15) CD344. Even numbered lanes contain 5 μg of soluble protein. Odd numbered lanes contain volume equivalent amounts of insoluble fraction. **B.** Lanes with aliquots of soluble extract containing 1 nmole/min deaminase activity: (1) CD327; (2) CD333; (3) CD334; (4) CD337; (5) CD339; (6) CD341.

A Equal Protein Loaded



B Equal Activity Loaded



Expression and Purification of the Histidine-tagged FTCD and Domains. Because the isolated domains were very difficult to purify by conventional methods, hexahistidine tags were appended to the C-terminal ends of the full-length FTCD ($FTCDH_6$), the transferase-active domain (FTH_6), the deaminase-active domain (CDH_6), and a shorter deaminase-active fragment CD339 ($CD339H_6$). Ni-NTA and polyglutamate affinity chromatography effected a 150 fold purification of $FTCDH_6$ and a 100 fold purification of CDH_6 while Ni-NTA followed by DEAE chromatography produced a 60 fold purification of FTH_6. SDS PAGE of the purified histidine-tagged proteins is shown in Figure 3.4.

Sequencing of CDH_6 and $CD339H_6$ yielded the expected 10 amino terminal residues and confirmed that these proteins initiated at amino acids 333 and 339 respectively. While CDH_6 retained the initiator methionine, the majority of $CD339H_6$ did not.

The combination of reverse phase HPLC and ESI-MS was used to separate and identify tryptic fragments of the N-terminal domain FTH_6. Fragments covering greater than 90% of this domain were positively identified by mass spectrometry. A fragment of 2040.7 Da was of particular interest since it corresponded exactly to the predicted mass of the C-terminal tryptic fragment (containing residues 321 to 328 plus the histidine tag).

The quaternary structure of FTCD and isolated domains. Gel filtration was used to determine the association state of FTCD and the isolated domains. Both FTCD and $FTCDH_6$ eluted as high molecular weight complexes (Table 3.1), suggesting that they exist as the previously described octamer. FTH_6 and CDH_6 eluted at apparent molecular weights of twice their subunit sizes, indicating that both are dimers. Cross-linking of the full-length enzyme and of the isolated domains with bis[sulfosuccinimidyl]suberate (Pierce, Rockford, Ill.) supported these results (data not shown).

Kinetic Characterization of the recombinant enzymes. Table 3.2 contains kinetic parameters for the wildtype and histidine-tagged proteins. The transferase activities displayed K_m values for FIGLU in the millimolar range, as previously observed with enzyme isolated from pig liver (Beaudet & MacKenzie, 1975). K_m values for (6*S*,*R*)-

FIGURE 3.4. SDS-PAGE analysis of purified proteins. Proteins were electrophoresed on a 12% acrylamide gel and stained with Coomassie Brilliant Blue. Each lane contains 10 µg of purified: (1) FTCD; (2) $FTCDH_n$; (3) CDH_n; (4) FTH_n.

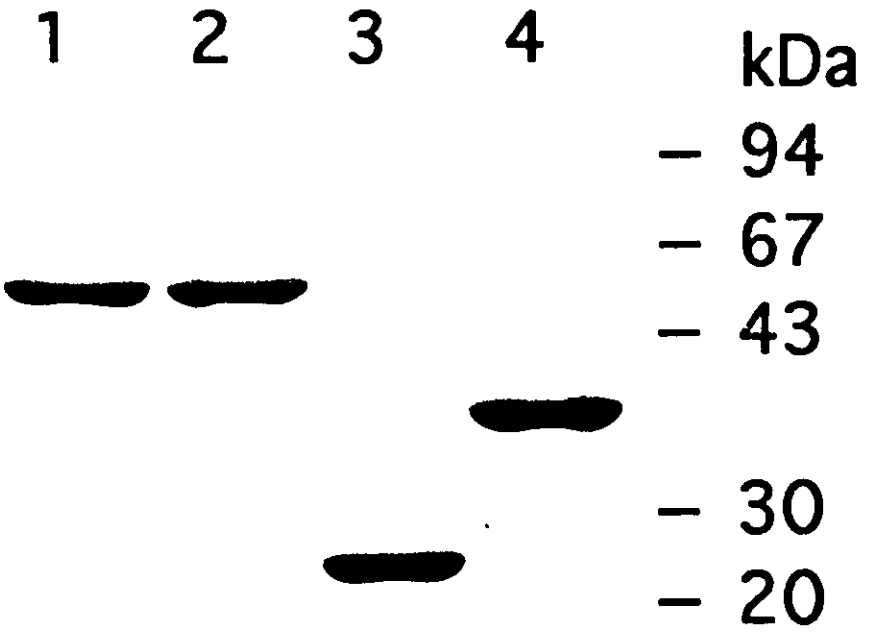

1
2
3
4
kDa
94
67
43
30
20

TABLE 3.1. Comparison of subunit and native molecular weights

Protein	Molecular Weight (kDa)	
	Subunit[a]	Native[b]
FTCD	59	438
$FTCDH_6$	60	380
FTH_6	37	83
CDH_6	24	52

[a] subunit size taken from amino acid sequence

[b] native size determined by gel filtration

TABLE 3.2. Kinetic properties of the transferase/deaminase enzymes

	K_m					k_{cat}	
enzyme	figlu	(6*R*,*S*)-H_4PteGlu	(6*S*)-H_4PteGlu$_5$	(6*S*)-formimino H_4PteGlu	(6*S*)-formimino H_4PteGlu$_5$	transferase	deaminase
	mM	μM	μM	μM	μM	sec^{-1}	
FTCD	5.8 ± 0.3	141 ± 7		66 ± 7		58 ± 3	394 ± 2
FTCDH$_6$	6.7 ± 1.1	148 ± 13	1.7 ± 0.4	70 ± 4	1.4 ± 1.3	77 ± 7	379 ± 25
FTH$_6$	8.3 ± 0.9	111 ± 5	> 75			32 ± 1	
CDH$_6$				70 ± 5	19 ± 2		213 ± 8

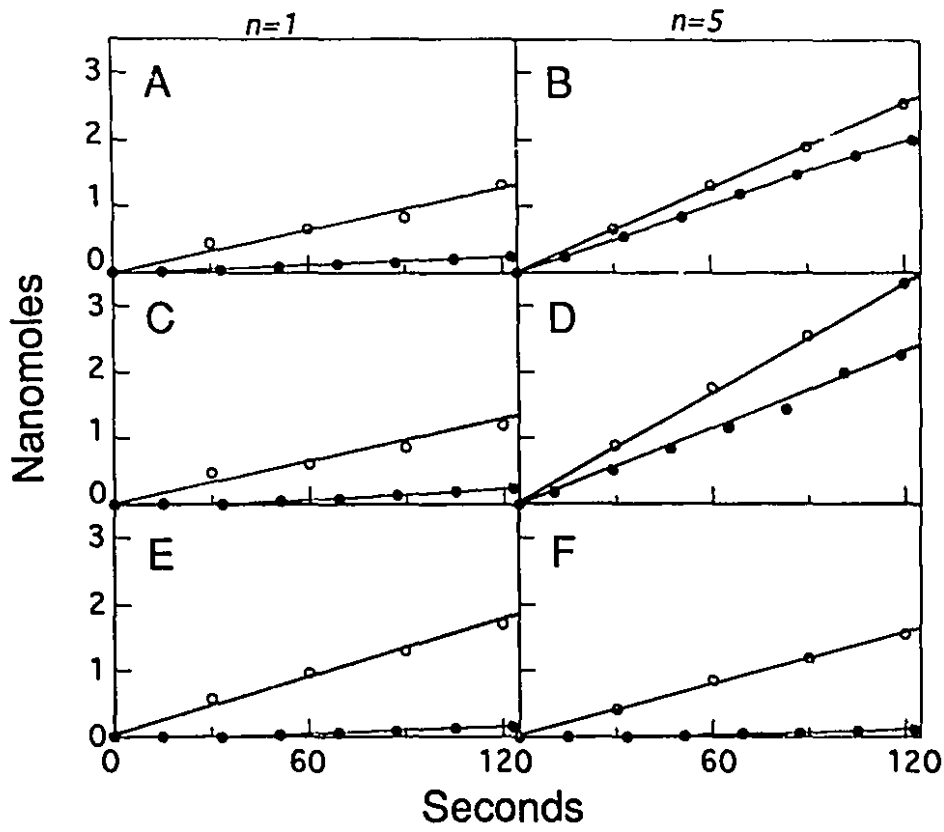
[a] values are expressed as averages ± standard deviations for 3-6 separate determinations

H_4PteGlu are also similar between the transferase enzymes. However, the K_m value of FTH_n for the pentaglutamylated substrate (6*S*)-H_4PteGlu$_5$, at greater than 75 μM, is over 50 times higher than that of $FTCDH_n$ under identical conditions. This observation is in agreement with the K_m value previously described for a transferase-active dimer isolated by denaturation of the octamer (Findlay and MacKenzie, 1988).

The K_m values for (6*S*)-formiminoH$_4$PteGlu were similar for the three deaminase enzymes assayed, but approximately one half to two thirds the value described elsewhere (Paquin et al., 1985; Findlay and MacKenzie, 1988). The K_m values for the pentaglutamate derivative were 1.4 μM for the deaminase activity of $FTCDH_n$ and 19 μM for CDH_6. CDH_6 and a deaminase active dimer produced by denaturation (Findlay and MacKenzie, 1988) display essentially identical K_m values for the pentaglutamylated substrate. Thus the deaminase domain retains significant polyglutamate specificity while the transferase domain does not. The k_{cat} values for each isolated domain are approximately half that of the corresponding activity in FTCD.

Channelling of the formimino intermediate between active sites. Channelling experiments (Figure 3.5) were performed using 50 μM H_4PteGlu$_1$ or H_4PteGlu$_5$, and equivalent units of enzyme activity. When FTCD is used with the monoglutamylated substrate (Figure 3.5A), we observe that the rate of formation of methenylH_4PteGlu$_1$ is much slower than the rate of the transferase (formimino + methenyl), indicating that the formimino intermediate is accumulating in the medium. However, when H_4PteGlu$_5$ is used as substrate (Figure 3.5B), this intermediate is channelled to the deaminase active site and the rate of formation of the second product is similar to that of total products. The same phenomenon is observed using $FTCDH_6$, as shown in panels C and D, although the efficiency of channelling is decreased. When FTH_6 and CDH_6 are mixed, the results (panels E + F) indicate that the isolated domains cannot channel the pentaglutamylated intermediate. In contrast to FTCD(H_6), an increase in transferase activity is not observed when the pentaglutamylated substrate substituted for the monoglutamate with the mixed domains.

FIGURE 3.5. Time course of appearance of products. The products represented are: (○), transferase activity, the sum of 5-formiminoH_4PteGlu$_n$ and 5,10-methenylH_4PteGlu$_n$; (●), deaminase activity, 5,10-methenylH_4PteGlu$_n$ alone. Panels (**A & B**), FTCD; (**C & D**), $FTCDH_6$; (**E & F**), a matched mixture of FTH_6 and CDH_6.


n=1
n=5
A
B
C
D
E
F
Nanomoles
Seconds
0
1
2
3
60
120

A short linker sequence separates the N-terminal transferase domain and the C-terminal deaminase domain. Most multifunctional enzymes are comprised of distinct modules or domains linked together by short flexible linker sequences (Coggins & Hardie, 1986). Previous studies including limited proteolysis of FTCD and denaturation of the octamer indicated that the transferase and deaminase activities might be located on different domains. As expression of functionally active fragments demonstrates the modular composition of a protein, we attempted to experimentally define the domain structure of FTCD through deletion mutagenesis.

As FTCD does not show significant homology to other known proteins, sequence similarity could not be used to estimate its domain structure. However, Findlay et al. (1989) had previously observed that both a 39 kDa transferase active proteolytic fragment and the full length FTCD appear to be N-terminally blocked. We interpreted this to mean that a protease sensitive sequence connecting the transferase and deaminase domains might lie within amino acids 320-360. Using deletion mutagenesis techniques, we produced constructs which would express N-terminal proteins ending at different positions between amino acids 318 and 340, and C-terminal fragments which initiated at different residues within the same region.

The N-terminal boundary of the linker sequence resides within residues 324 to 326. Proteins ending at or before residue 322 apparently do not fold readily and are completely insoluble, while larger proteins ending at or after residue 325 are soluble and equally active. Although all C-terminal peptides were at least partially soluble, CD327, CD333 and CD334 have substantially higher turnover numbers than CD337, CD339 and CD341. The smallest C-terminal fragments, CD344 and CD350, are completely inactive. Therefore, the C-terminal boundary of the linker region lies within residues 333 and 335, and distinguishes the fully active CD fragments from those which are less active.

Thus each FTCD subunit includes an N-terminal transferase domain and a C-terminal deaminase active domain, separated by a short linker. This linker sequence is between 8 (as underlined in Figure 3.1) and 12 residues long and is similar in composition

to other linkers described by Argos (1990). We can remove 6 of these proposed linker residues (327 to 332, inclusive) to produce a protein which is less soluble than the wildtype FTCD yet still expresses both activities, albeit at reduced levels. Upon extending this deletion to include residues 327 to 338, the protein becomes even less soluble and can no longer catalyze either reaction (data not shown).

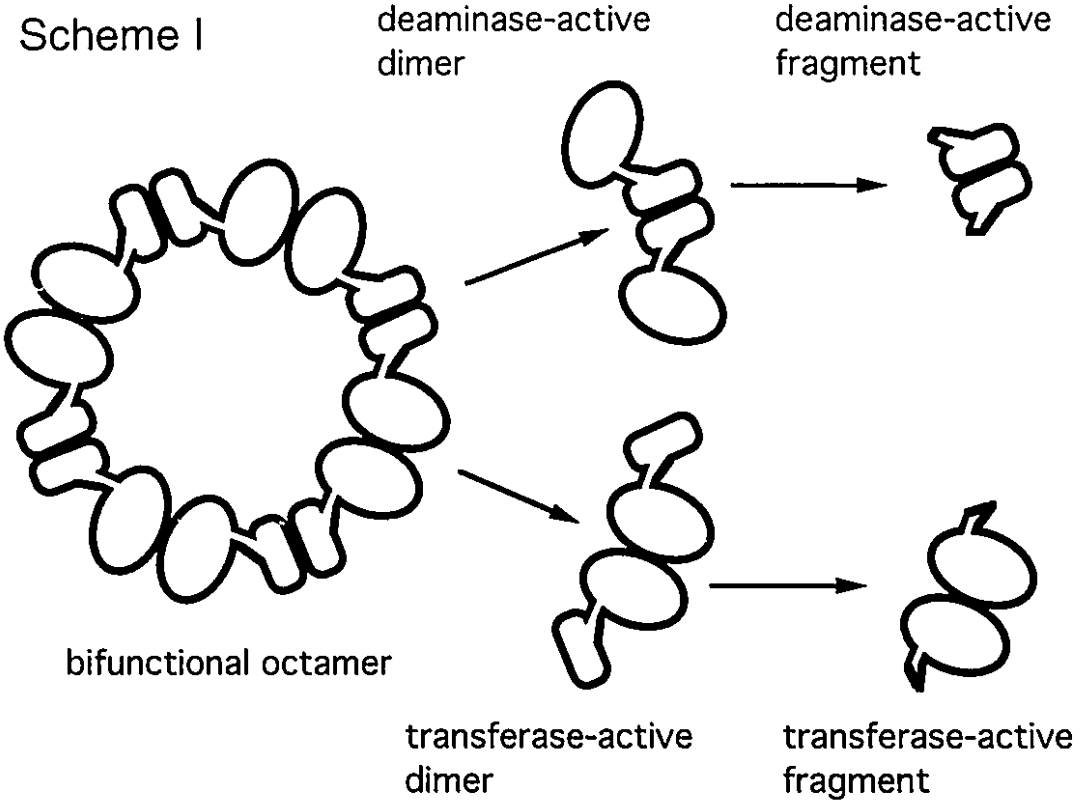
For the remainder of these studies we chose to produce and purify the full-length enzyme and the isolated domains CD333 and FT328 as histidine-tagged fusion proteins to simplify their purification.

The octameric FTCD contains two types of subunit interfaces. Previous work on FTCD suggested that formation of the octamer is required for co-expression of both catalytic activities and for channelling of substrate between active sites. In the proposed model the eight subunits are arranged to form a circular tetramer of dimers, with two types of subunit interfaces (MacKenzie et al., 1980; Findlay & MacKenzie, 1987; 1988). Therefore, each domain might include sequences governing subunit dimerization. Analytical gel filtration indicated that both isolated domains exist as dimers, confirming that octameric FTCD includes two different types of subunit interfaces as represented by the tetramer of dimers shown in Scheme I.

Indirect evidence suggests that both the transferase and deaminase activities are dependent on the retention of specific subunit interfaces (Findlay & Mackenzie, 1987; 1988). If it is a prerequisite for either activity, dimerization may be required to achieve an active conformation of the subunit, or to form a catalytic site(s) at a subunit interface using residues derived from each monomer. Determining the number and position of catalytic sites within the octamer is required to resolve these questions.

As both domains reside within one subunit it seemed reasonable that the isolated domains might continue to associate noncovalently. Several attempts were made to demonstrate such an interaction, without success. The domains could not be crosslinked using a bifunctional reagent, and neither domain could be retarded on a Ni-NTA column previously loaded with the H_6 version of the other domain. As well, the fluorescence spectrum of an equimolar mixture of CDH_6 and FTH_6, initially incubated at high protein

Scheme I


deaminase-active dimer
deaminase-active fragment
bifunctional octamer
transferase-active dimer
transferase-active fragment

concentration (1 mg ml), is identical to the spectrum obtained by mathematically adding the individual spectra of each domain. Finally, lack of channelling by domains incubated at high protein concentration before assay indicates the absence of a specific hetero-domain interaction in this system. The apparent lack of association between the isolated domains emphasizes the importance of the linker region in maintaining domain interactions within each subunit of the octamer

The isolated domains retain kinetic characteristics similar to that of the native enzyme with monoglutamate substrates. The major difference is a decrease in their k_{cat} values to about 50%. Previously, a chymotryptic transferase active dimeric fragment was also shown to have only 67% of the expected activity (MacKenzie et al., 1980). While the domains retain a great deal of integrity, it is likely that their removal from the restraints imposed by the octameric structure results in some conformational changes.

The role of the polyglutamate tail in substrate binding and channelling. Like many other folate-dependent enzymes, FTCD displays a preference for polyglutamylated substrates. Polyglutamylation improves the binding of both the folate substrates and FIGLU (Paquin et al., 1985; Findlay et al., 1989). As well, a polyglutamate tail, 4 or more polyglutamates in length, is required for the direct transfer of formiminoH$_4$PteGlu$_n$ between active sites. Paquin et al. (1985) demonstrated the presence of only 4 high affinity polyglutamate binding sites per octamer, or 1 per dimer. Our results indicate that this polyglutamate binding site resides within the deaminase domain. The existence of only 1 polyglutamate binding site per pair of deaminase domains, suggests that it (and possibly the deaminase active site) may be located at the subunit interface formed between deaminase domains.

Basic amino acids, in particular arginine residues, have been implicated in the binding of polyglutamylated substrates to other folate-dependent enzymes (Kamb et al., 1992; Maras et al., 1994; Finer-Moore et al., 1994). Rabinowitz's group suggested that a 53 residue sequence, which is specific to 10-formyltetrahydrofolate synthetases which bind longer polyglutamates and includes doublets of basic amino acids, may constitute part of that enzyme's polyglutamate binding site (Whitehead & Rabinowitz, 1988; Nour & Rabinowitz, 1992). Within FTCD, arginine doublets (R381,R382; R392,R393; R435,R436)

are only located in the deaminase domain; some of these doublets may aid the binding of polyglutamate to the octamer.

Since pentaglutamylated substrate is transferred with the highest efficiency while longer polyglutamates are more tightly bound (Paquin et al., 1985), substrate channelling within FTCD is thought to involve a steric component. MacKenzie and Baugh suggested a "swinging arm" mechanism whereby the polyglutamate binds to a site on the octamer while the more mobile pterin moiety can interact with either type of catalytic site (1980). As we have shown that the activities reside in different domains, the substrate must be channelled between domains, if not subunits.

As predicted by this model, the separated domains are not able to channel pentaglutamylated intermediate. While some method of direct transfer between sites within the octamer is likely, the slightly altered properties of the CDH_6 domain prevent us from entirely ruling out a "release and rebinding" mechanism. Similar limitations were encountered in testing this model by chemical modification of the native proteins (MacKenzie & Baugh, 1980).

A swinging arm is not the only feasible method of direct transfer. Recently Knighton et al. (1994) described a possible mechanism for channelling of polyglutamylated dihydrofolate within bifunctional thymidylate synthase-dihydrofolate reductase of *Leishmania major*. They observed an unusual distribution of charged residues across the surface of the protein which may serve to guide the intermediate between sites. A similar "electrostatic highway" (Stroud, 1994), the proposed model of a polyglutamate "anchor" or a third unknown mechanism may mediate the channelling described for FTCD. Crystallographic analyses of FTCD and its isolated domains should clarify this issue and, as well, answer questions regarding the number and location of active sites in relation to the subunit interfaces formed by each domain.

ACKNOWLEDGEMENTS

We wish to thank Dr. A. English, C. Fenwick, and G. Tsaprailis, of the Department of Chemistry, Concordia University, and B. Gibbs, Biotechnology Research Institute, National Research Council of Canada, for mass spectral analysis of the transferase domain.

CHAPTER 4

UREA-INDUCED DENATURATION OF FORMIMINOTRANSFERASE-CYCLODEAMINASE AND ITS MONOFUNCTIONAL DOMAINS

ABSTRACT

Each subunit of octameric formiminotransferase-cyclodeaminase consists of a transferase and a deaminase domain connected by a short linker sequence. These domains can be independently expressed in *E. coli* as monofunctional dimers and show no indication of associating, suggesting that the linker mediates the only substantial domain-domain interaction. To better understand the benefits arising from octamer formation, we have used equilibrium unfolding methods to analyze the transferase and deaminase domains independently and within the octamer. The domains and the full-length enzyme exhibit multistate denaturation profiles. Both isolated domains undergo a concurrent loss of intrinsic fluorescence and catalytic activity at low concentrations of urea to form inactive dimers. Unfolding and dissociation of the transferase domain are observed at slightly higher urea concentrations and circular dichroism suggests that a partly unfolded transferase domain, containing 40% of the native molar elipticity, is stable between 3.5 and 4 M urea. At 2.5 M urea, inactive deaminase dimer, folded monomer and unfolded monomer are observed to coexist, suggesting that dissociation and unfolding of this domain occur at similar urea concentrations. While the isolated transferase and deaminase domains undergo dissociation to monomer between 2 and 2.5 M urea, only one type of subunit interface in the octamer is disrupted at this urea concentration. Dissociation of the second interface occurs between 3.5 and 5 M urea, indicating that one domain achieves increased stability within the full length enzyme.

INTRODUCTION

The bifunctional enzyme formiminotransferase-cyclodeaminase (FTCD) catalyzes two sequential reactions in the histidine degradation pathway. The transferase activity transfers a formimino group from the histidine catabolite formiminoglutamate to $H_4PteGlu_n$. The formimino intermediate is deaminated by the cyclodeaminase producing 5,10-methenyl$H_4PteGlu_n$ and NH_4^+. With monoglutamylated folates (n=1), the two activities function independently, however polyglutamylated intermediate ($n \geq 4$) can be efficiently channelled from the transferase to the deaminase active site (MacKenzie & Baugh, 1980; Paquin et al., 1985). FTCD preferentially channels pentaglutamates and MacKenzie and coworkers have proposed that the polyglutamate tail acts as a noncovalently bound swinging arm serving to anchor the intermediate to the enzyme while the pteroyl moiety moves between the two active sites.

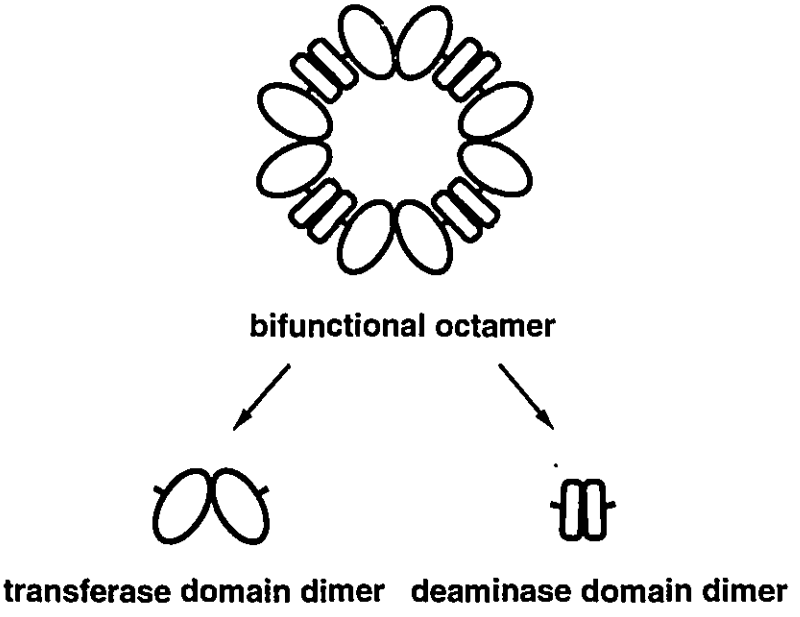
FTCD is composed of 8 identical subunits arranged to form a planar, ring-shaped tetramer of dimers (Beaudet & MacKenzie, 1976; MacKenzie et al., 1980). Denaturation/renaturation (Findlay & MacKenzie, 1987; 1988) and deletion analysis studies (Murley & MacKenzie, 1995) indicate that the octamer is the smallest functional unit of FTCD which retains the transferase and deaminase activities and the ability to channel substrates. Paquin et al. (1985) demonstrated that 4 high affinity polyglutamate binding sites exist per molecule of FTCD and Findlay and MacKenzie (1987, 1988) have suggested that the integrity of alternating subunit interfaces in the native octamer is a prerequisite for maintaining both catalytic activities. While this suggests that active sites may be situated at subunit interfaces, exact numbers of catalytic sites and their relative location within the octamer are not known.

We have recently determined that each subunit of FTCD consists of an N-terminal transferase domain and a C-terminal deaminase domain separated by a short linker sequence (Murley & MacKenzie, 1995). The polyglutamate binding site maps to the deaminase domain. These domains can be independently expressed in *E. coli* and show no indication of interacting with each other once purified. This suggests that the linker mediates the only substantial domain-domain interaction in FTCD. Both domains contain

sequences governing subunit association, and exist as monofunctional dimers in solution (Murley & MacKenzie, 1995), as illustrated in Scheme 1. Apparently the octameric structure results from the fusion of two different dimeric monofunctional proteins. The subunit interfaces formed between identical monomers are required to maintain domain structure, but the domain interface within the subunit is relatively unimportant. If this model were accurate, the two types of domains should behave similarly both within and outside of the octamer. The ability of FTCD to channel substrate and the 50% decrease in k_{cat} values observed for each isolated domain (Murley & MacKenzie, 1995) argue that the transferase and deaminase domains do not function completely autonomously within the octamer.

Substrate channelling is the most obvious benefit of octamer formation. We would like to establish if fusion of the transferase and deaminase domains might result in other advantages, such as increased stability of the domains or subunit interfaces. In this paper we examine the urea-induced denaturation of the transferase and deaminase domains independently and as part of the FTCD octamer.

Scheme I


bifunctional octamer
transferase domain dimer
deaminase domain dimer

MATERIALS AND METHODS

Materials. Folic acid and formimino-L-glutamic acid were from Sigma and ultrapure urea was from ICN. Ni-NTA resin was purchased from Qiagen. The Superose 6 HR 10/30 column was from Pharmacia. 5-formiminoH$_4$PteGlu was prepared enzymatically as described previously (Paquin et al., 1985) except that purified recombinant transferase domain (Murley & MacKenzie, 1995) was used to catalyze its synthesis. Acidified 5-formiminoH$_4$PteGlu was centrifuged in a microfuge for 10 minutes at 4 °C to pellet precipitated enzyme, instead of passing the solution through Centriflo cones (Amicon).

Enzyme Purification. Recombinant histidine-tagged FTCDH$_6$, FTH$_6$ and CH$_6$ were expressed in *E. coli* and purified as previously described (Murley & MacKenzie, 1995). Enzyme preparations had specific transferase activities of approximately 30 μmoles/min/mg for FTCDH$_6$, and 22-24 μmoles/min/mg for FTH$_6$, and specific deaminase activites of 18-21 μmoles/min/mg for FTCDH$_6$ and 16-20 μmoles/min/mg for CDH$_6$ when assayed with 0.1 mM 5-formiminoH$_4$PteGlu synthesized from a racemic mixture of H$_4$PteGlu. Single bands were observed upon SDS PAGE. Triton X 100 was replaced with Tween 20 by rebinding the enzyme to Ni-NTA resin and washing with buffer containing 0.05 % Tween 20, prior to elution with elution buffer with 0.05 % Tween 20 replacing the Triton X 100. Enzyme was concentrated to approximately 4 mg/ml (Centriprep, Amicon) and stored at -20 °C in buffer containing 0.1 M potassium phosphate (pH 7.8), 0.05 % Tween 20, 40 % glycerol and 1 mM DTT. Protein concentrations (per monomer) were determined using the calculated extinction coefficients (Gill & von Hippel, 1989) E_{280} = 37920 $M^{-1}cm^{-1}$ for FTCDH$_6$, 28150 $M^{-1}cm^{-1}$ for FTH$_6$ and 9770 $M^{-1}cm^{-1}$ for CDH$_6$.

Enzyme Activity. Proteins (1 μM) were incubated for 150 minutes at room temperature in Buffer A, containing 0.1 M potassium phosphate (pH 7.3), 0.05 % Tween 20 and 1 mM DTT, containing the appropriate concentration of urea (0 to 8 M). Transferase activity was assayed (MacKenzie, 1980; Drury & MacKenzie, 1975) in 0.5 ml incubation volumes for 3 minutes at 30 °C. The deaminase activity was determined by time drive

at 30°C in assay mix containing 0.1 M potassium phosphate (pH 7.3), 0.1 mM 5-formiminoH$_4$PteGlu and 35 mM 2-mercaptoethanol, and measurement of $A_{\lambda\lambda}$ on a Beckman DU640 spectrophotometer. All activity measurements are the average of assays done in duplicate or triplicate.

Fluorescence Emission. The intrinsic fluorescence of FTCDH$_n$, FTH$_n$ and CDH$_n$ was measured with a Hitachi F3010 fluorescence spectrophotometer, equipped with a thermostatically controlled sample holder at 25 °C. Samples (10 nM to 2μM) were incubated for 150 minutes at room temperature in Buffer A containing the appropriate amount of urea (0 to 8 M), and the emission spectra (excitation at 290 nm) was recorded between 320 and 380 nm. The excitation slit was set at 3 nm and the emission slit at 5 nm. The spectrum of a buffer blank was subtracted from each sample spectrum.

Data Analysis. Data were analyzed using the non-linear regression analysis program "Enzfitter" (Biosoft, Cambridge, UK). Two indicators of protein conformation (activity and fluorescence intensity) showed that both domains undergo a coincident, cooperative transition at low urea concentrations, prior to significant dissociation and unfolding. This suggested that both domains might undergo a two state change in tertiary structure at low concentrations of urea: F ⇌ F*. To characterize this transition, the equilibrium constant of this conformational change, K_{F*}, and the free energy, ΔG_{F*}, were calculated at different urea concentrations within the transition region according to the following relationship:

$$K_{F*} = F^*/F = f_F/(1-f_F) = exp(-\Delta G_{F*}/RT) \quad (1)$$

where f_{F*} is the fraction of protein which has undergone the conformational change. In order to estimate changes in tertiary stability inside and outside of the octamer, the free energy of this conformational change in the absence of denaturant, was estimated from a linear extrapolation of the values of ΔG_{F*} versus the denaturant concentration to 0 M urea according to Pace (1986),

$$\Delta G_{F*} = \Delta G_{H2OF*} + m_{F*}[\text{urea}] \quad (2)$$

where the intercept ΔG_{H2OF*} corresponds to the free energy of the conformational change at 0 M urea and the slope m_{F*} reflects the cooperativity of the urea-induced transition.

Circular Dichroism. CD spectra of samples containing 10 μM protein, incubated for 150 minutes at room temperature in Buffer A containing urea (0 to 7.5 M), were recorded in a Jasco J-710 spectropolarimeter with a cylindrical cuvette of 0.05 cm pathlength. Spectra of blanks containing only buffer were subtracted from sample spectra and molar ellipticity was determined. Scan speed was set at 100 nm/min and measurements are reported from 5 accumulations.

Size Exclusion Chromatography. $FTCDH_n$, FTH_n and CDH_n were incubated at room temperature for 150 minutes in Buffer A containing the appropriate concentration of urea (0 to 5 M). Aliquots of 200 μl of approximately 1 mg/ml protein, were injected onto a Superose 6 HR 10/30 column equilibrated in Buffer A containing the appropriate concentration of urea. Samples were chromatographed at a flow rate of 0.31 ml/min, at 4 °C and the absorbance of the eluate was monitored at OD_{280}. Ferritin (440 kDa), catalase (232 kDa), aldolase (158 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), chymotrypsinogen (25 kDa) and Ribonuclease A (13.7 kDa) were used as molecular weight standards to construct a standard curve to determine the sizes of applied proteins. The void volume was determined using Blue Dextran. BSA and Blue Dextran were chromatographed at each urea concentration and no changes in elution time were observed up to 5 M urea. Each sample was chromatographed 2-3 times.

Dissociation and unfolding of larger proteins can be relatively simple, for example the two-state denaturation of dimeric trp repressor (Gittelman & Matthews, 1990), the arc repressor (Bowie & Sauer, 1989), or the multidomain λ cro dimer (Pakula & Sauer, 1989). In other cases denaturation is considerably more complicated, and may involve intermediates such as inactive oligomeric species (glutamine synthase, Maurizi & Ginsburg, 1982), structured monomers (aspartate aminotransferase, Herold & Kirschner, 1990; Pi-class glutathione transferase, Aceto et al., 1992) or compact unfolding intermediates (reviewed by Fink, 1995).

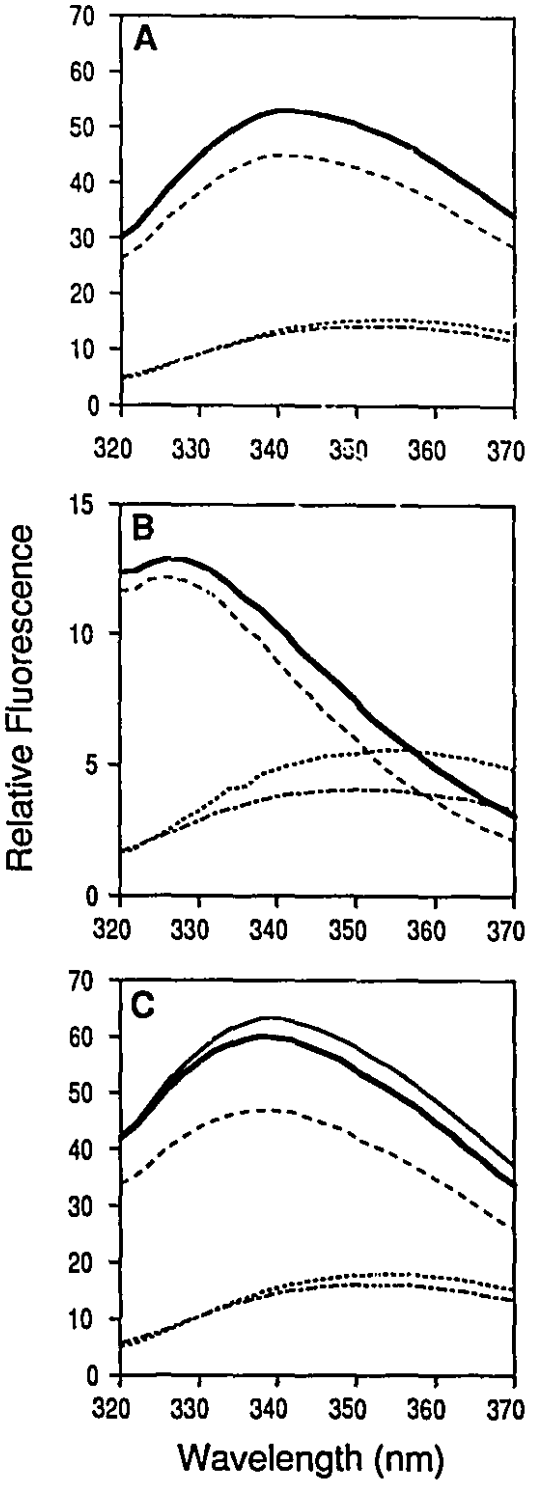
Bifunctional FTCD is both oligomeric and multidomain. Each monomer consists of an N-terminal transferase domain and a C-terminal deaminase domain separated by a short linker peptide. Both domains include sequences mediating subunit association and can be independently expressed as dimeric proteins. Previously, Findlay and MacKenzie (1987) used urea-induced dissociation to analyze the subunit arrangement of FTCD, and proposed a model for the dissociation of this enzyme. They suggested that the octamer first dissociates to form inactive dimers between 2 and 3 M urea (transition I). This is followed by a conformational change in the dimer between 3 and 4 M urea (transition II) which precedes the third transition, dissociation of the dimers to monomers above 4 M urea. Their results suggested that unfolding of wildtype FTCD is multiphasic and involves several intermediates.

Some multidomain proteins exhibit very strong domain interactions where multiple domains unfold co-operatively in a single transition, with no evidence of equilibrium intermediates (λ cro dimer, Pakula & Sauer, 1989; thermal denaturation of yeast phosphoglycerate kinase in the presence of guanidine hydrochloride, Griko et al., 1989). Domains in other proteins unfold independently and show little or no evidence of interacting (ovomucoid, Privalov, 1982). A third class of proteins displays an intermediate level of domain interaction (CD4, Tendian et al., 1995; DnaK, Montgomery et al., 1993). Our previous results indicate that once the linker connecting the two catalytic domains of FTCD is severed, functional and structural association of FTH_6 and CDH_6 is not

observed, suggesting that interdomain interactions are minimal at best (Murley & MacKenzie, 1995). In a review of folding of multidomain proteins, Garel (1992) has noted that often the stability of isolated domains is not substantially modified by the presence of the rest of the protein chain. He concluded that domain stability results primarily from intradomain interactions. This may also be true of the transferase and deaminase domains of FTCD. If interdomain interaction is minimal within the subunit, the physical characteristics of each domain should not change upon separation. To determine the extent of interaction between the transferase and deaminase domains in octameric FTCD, we examined the properties of the isolated and linked transferase and deaminase domains and their stabilities to urea-induced dissociation and denaturation.

Spectral Characterization of FTH_6, CDH_6, and $FTCDH_6$. As a first step, a spectral characterization was undertaken to examine the domains inside and outside of the FTCD octamer. The intrinsic tryptophan fluorescence emission spectra of equimolar concentrations of FTH_6, CDH_6 and $FTCDH_6$ are shown in Figure 4.1. $FTCDH_6$ has a total of 4 tryptophans, 3 in the transferase domain (W158, W171 and W224) and one in the deaminase domain (W462). In keeping with this distribution, the transferase domain exhibits significantly higher fluorescence emission intensity per mole of enzyme than does the deaminase domain. Native FTH_6 and CDH_6 have emission spectra with maximum emission wavelengths of approximately 342 nm (Fig. 4.1A) and 326-328 nm (Fig. 4.1B), respectively. The maximum emission wavelength for $FTCDH_6$, at 338 nm (Fig. 4.1C), is intermediate between those of the two domains. Similarly shaped spectra were observed using excitation wavelengths ranging from 270 to 295 nm, indicating that tryptophan provides most of the fluorescence intensity, and no evidence of conformational change was observed over a range of 10 nM to 2 μM monomer concentration for all three protein species (data not shown). While the emission spectrum of an equimolar mixture of FTH_6 and CDH_6 and the additive spectrum of the individual domains are identical (Murley & MacKenzie, 1995), the mixture of domains has slightly more (1.07 times) maximal fluorescence intensity than octameric $FTCDH_6$. Upon denaturation in 8 M urea, all three species undergo a decrease in fluorescence intensity and their λ_{max} values are redshifted

FIGURE 4.1. Intrinsic fluorescence spectra of native and denatured proteins. A. FTH_6, **B.** CDH_6, and **C.** $FTCDH_6$. The emission spectra of 1 μM enzyme in Buffer A +/- urea. (_), native enzyme; (__ __ __), enzyme renatured for 24 hours; (_._._.), enzyme in 5 M urea; (···), enzyme in 8 M urea; (—), equimolar mixture of native FTH_6 and CDH_6 (panel C only).


A
B
C
Relative Fluorescence
Wavelength (nm)
70
60
50
40
30
20
10
0
15
5
320
330
340
350
360
370

to between 354 and 357 nm, indicating that all tryptophan residues have become exposed to the solvent. Denaturation is essentially a reversible process since all three proteins regained most of their native fluorescence intensity and activity after 24 hours of renaturation. As observed for other multidomain proteins (Garel, 1992) the isolated domains renature more quickly than the full length enzyme (data not shown). This may be related to aggregation of the larger protein or slow pairing of domains and subunits as has been observed during renaturation of other multidomain or oligomeric enzymes (Jaenicke, 1987).

Both polar and nonpolar molecules can quench the tryptophan fluorescence of FTH_6 and $FTCDH_6$, suggesting that at least one of the transferase domain tryptophans is solvent exposed (data not shown). In contrast, the single tryptophan in CDH_6 is not accessible to the polar quenchers Cs^+ or I^-, but can be effectively quenched with acrylamide. This observation, combined with the λ_{max} value of CDH_6, suggests that W462 is in a hydrophobic environment (Lakowicz, 1983), either buried within the deaminase domain or at a subunit interface.

CD spectra were obtained for the full length enzyme and both independent domains (Figure 4.2). The resulting spectra were analyzed to estimate the secondary structure content of each protein (Table 4.1), using the K2D method (Andrade et al., 1993). Table 4.1 also contains secondary structure predictions made from amino acid sequence data. The transferase domain and the full length protein are estimated to contain significant amounts of both α helix and β sheet, while the deaminase domain appears to be predominantly α helical in nature. Thus, the helix signal at 222 nm was chosen to monitor urea-induced unfolding.

Urea-induced denaturation of FTH_6, CDH_6 and $FTCDH_6$; loss of transferase and deaminase activities. As the transferase and deaminase activities exist in separate domains, catalytic activity can serve as a domain-specific conformational probe. The isolated transferase domain is inactivated (Fig. 4.3A) in a cooperative transition occurring between 1.25 and 2.5 M urea. The isolated deaminase domain is less stable than the transferase domain (ΔG_{H2OF}, Table 4.2). Inactivation of this domain occurs between 0.75

FIGURE 4.2. Circular dichroic spectra of native $FTCDH_6$ and the isolated domains. The molar elipticity of: (- - -), FTH_6; (.....), CDH_6; (—), $FTCDH_6$; in Buffer A at 25°C.

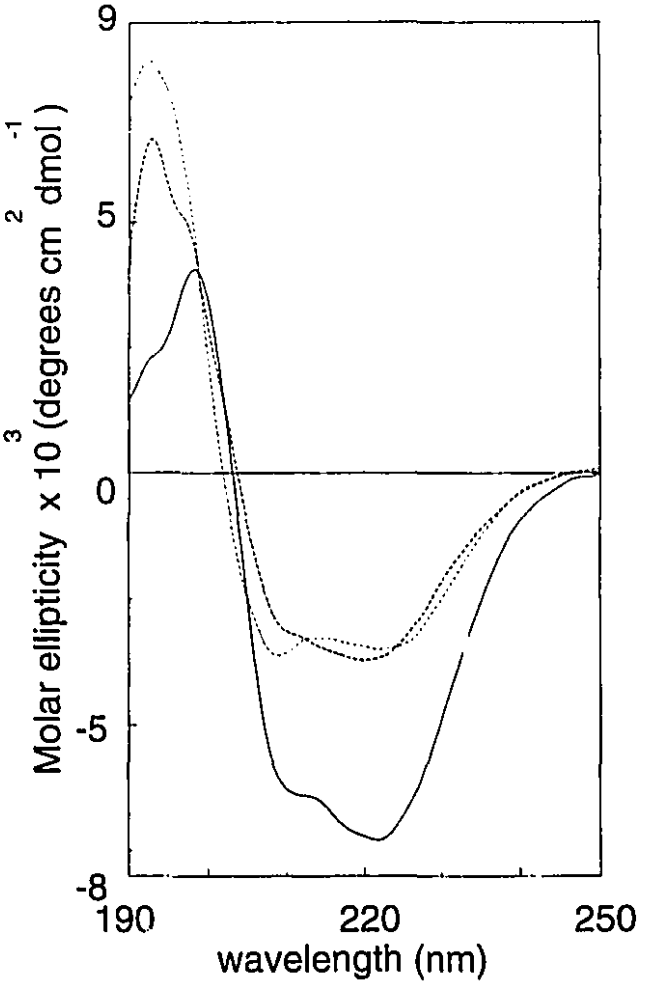

9
5
0
-5
-8
Molar ellipticity x 10 (degrees cm dmol)
3
2
-1
190
220
250
wavelength (nm)

TABLE 4.1. Secondary structure estimation

	Method	% helix	% sheet	% coil	% turn
from CD data					
FTH_6	K2D[a]	38	5	56	n/a
CDH_6	K2D	60	7	33	n/a
$FTCDH_6$	K2D	43	16	41	n/a
from sequence data					
FTH_6	Garnier[b]	53	16	18	14
	Deleage[c]	49	9	39	3
CDH_6	Garnier	72	12	9	8
	Deleage	95	0	5	1
$FTCDH_6$	Garnier	58	15	16	12
	Deleage	55	6	36	3

[a] Andrade et al., 1993.

[b] Garnier & Robson, 1989.

[c] Deleage & Roux, 1987; A double prediction method, where structural class is first determined from amino acid composition. FTH_6 and $FTCDH_6$ are predicted to be α/β-α class and CDH_6 is predicted to be all α class.

FIGURE 4.3. Percent original signal remaining during urea-induced denaturation of proteins. **A.** FTH_0, **B.** CDH_0, and **C.** $FTCDH_0$. Final concentrations were 1 μM (fluorescence and activity) or 10 μM (circular dichroism) enzyme in Buffer A plus the appropriate amount of urea. Symbols: (○), percent original transferase activity; (●), percent original deaminase activity; (□■), percent relative fluorescence = $(F-F_5)/(F_0-F_5)$, where F is the maximal peak height, F_5 is the fluorescence in 5 M urea and F_0 is the fluorescence in 0 M urea; (Δ▲), percent original CD signal at 222 nm; ◆, position of λ_{max} in nm.

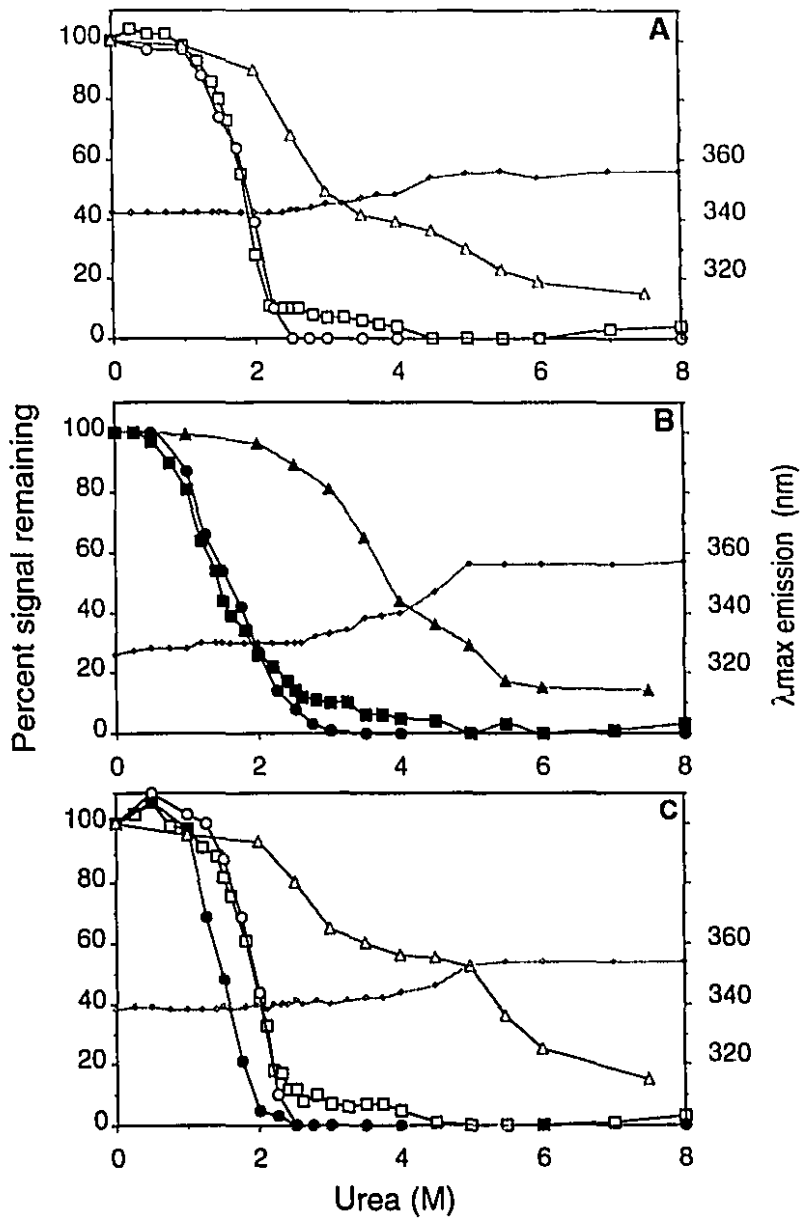

A
B
C
100
80
60
40
20
0
0
2
4
6
8
360
340
320
Percent signal remaining
λmax emission (nm)
Urea (M)

TABLE 4.2. Analysis of urea-induced denaturation

	Probe	$\Delta G_{H2OF\bullet}$[a] (kcal mol^{-1})	$m_{F\bullet}$[a] (kcal $mol^{-1}M^{-1}$)	C_m[b] (M urea)
FTH_6	transferase	4.54±0.40	2.46±0.22	1.85
	fluorescence	4.52±0.29	2.48±0.16	1.82
CDH_6	deaminase	2.55±0.14	1.60±0.08	1.59
	fluorescence	1.91±0.14	1.26±0.09	1.52
$FTCDH_6$	transferase	6.29±0.41	3.29±0.21	1.91
	deaminase	4.31±0.28	2.93±0.19	1.47
	fluorescence	4.25±0.23	2.24±0.11	1.90

[a] $\Delta G_{H2OF\bullet}$ is the free energy of the conformational change in tertiary structure in the absence of denaturant and $m_{F\bullet}$ is the urea dependency of this free energy. $\Delta G_{H2OF\bullet}$ and $m_{F\bullet}$ values were obtained from the intercept and slope of the linear extrapolation provided by Equation 2. The data in the transition region of the activity and fluorescence denaturation curves for FTH_6, CDH_6 and $FTCDH_6$ shown in Figure 3 were fit to equation 2 as described in Materials and Methods, and errors represent the standard deviation of that fit.

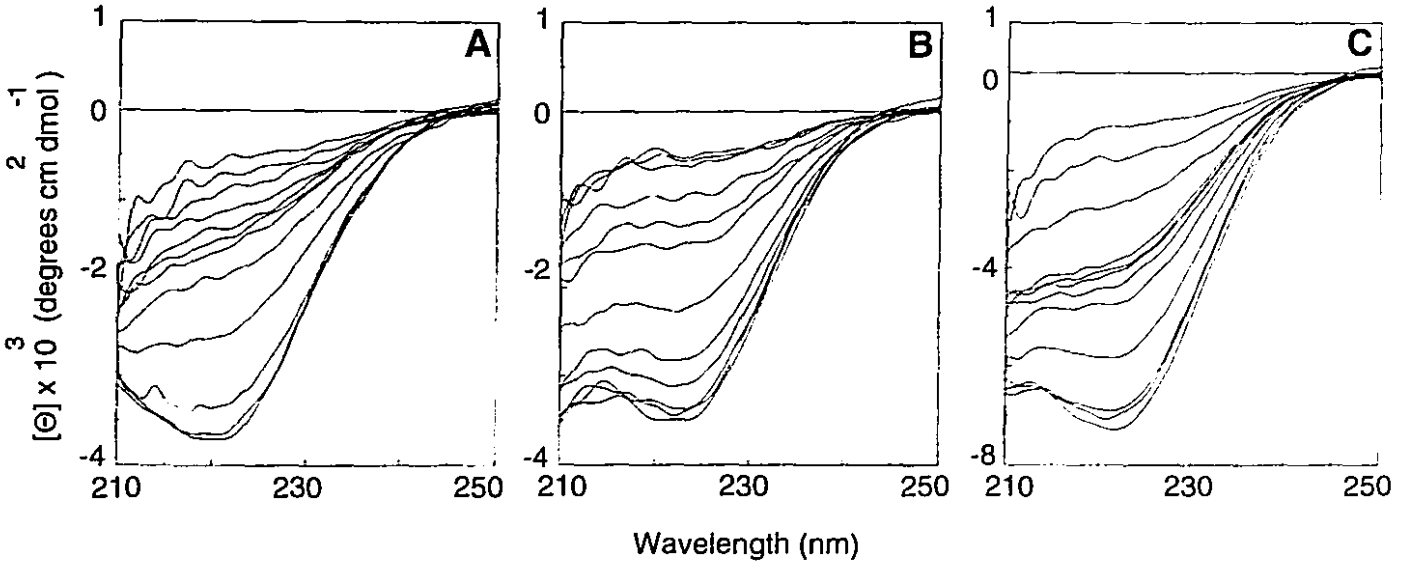
[b] C_m represents the urea concentration at which half of the protein has undergone the conformational change, $C_m = \Delta G_{H2OF\bullet}/m_{\bullet}$

and 2.5 M urea, and has a midpoint of approximately 1.6 M urea (Fig. 4.3B, C_m, Table 4.2). While the isolated and covalently linked domains are inactivated under similar concentrations of denaturant, these transitions occur more cooperatively in the octamer (m_{eq}, Table 4.2). As the ΔG_{H2O} values (Table 4.2) are slightly higher when the transferase and deaminase are covalently linked, both domains undergo modest increases in stability as part of the octamer.

Changes in intrinsic fluorescence upon denaturation in urea. As shown in Figure 4.3, the fluorescence intensities of $FTCDH_6$ and the isolated domains are quenched in response to increasing urea. The loss of FTH_6 fluorescence intensity occurs simultaneously with inactivation (Fig. 4.3A). The wavelength of maximum emission is redshifted between 2.5 and 4.5 M urea, suggesting increased solvent exposure of the tryptophan residues. Likewise, CDH_6 undergoes fluorescence intensity loss which correlates closely with inactivation (Fig. 4.3B), and a large redshift in λ_{max} occurs between 2.8 and 5 M urea, as the single tryptophan becomes more solvent accessible. When $FTCDH_6$ is subjected to urea denaturation, the fluorescence intensity loss correlates closely with loss of the transferase activity (Fig 4.3C). There is a gradual increase in λ_{max} between 2.5 and 4 M urea which then rises sharply between 4 and 5 M urea. Thus, each of these proteins undergoes conformational change at low concentrations of urea, as monitored by loss of enzyme activity and quenching of intrinsic tryptophan fluorescence.

Changes in secondary structure. To determine if the conformational change indicated by loss of fluorescence intensity and catalytic function involves protein unfolding, far UV circular dichroism was used to monitor for related changes in secondary structure. The transferase domain undergoes a biphasic decrease in the α helix CD signal at 222 nm. In 2 M urea, where FTH_6 has undergone both inactivation and a major decrease in fluorescence intensity, more than 90% of the native CD signal at 220 nm remains. Between 2 and 3 M urea significant unfolding occurs as indicated by a 40% loss in the helix signal (Fig. 4.3A, 4.4A). A small plateau is observed around 3.5 and 4 M urea, suggesting that a partly unfolded transferase may be stable at this concentration range.

FIGURE 4.4. Change in far UV CD spectra upon urea-induced denaturation of proteins. **A.** FTH_6, **B.** CDH_6, and **C.** $FTCDH_6$. Molar elipticity of enzyme in Buffer A +/- urea. Urea concentrations of 0, 1, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 7.5 M produced spectra with decreasing absolute molar elipticity at 222 nm.


A
B
C
1
0
-2
-4
-8
210
230
250
[Θ] x 10 (degrees cm dmol)
3
2
-1
Wavelength (nm)

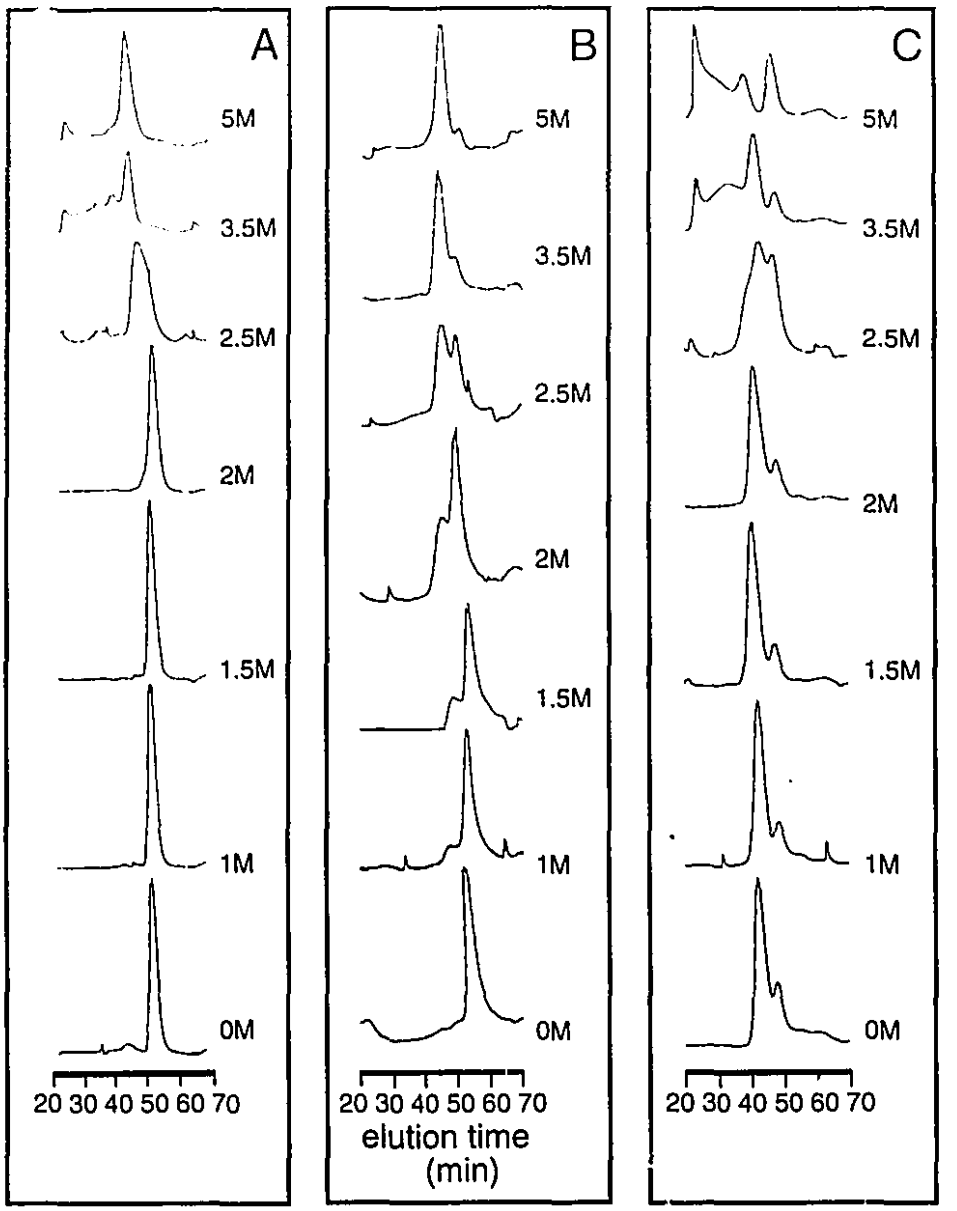
Further unfolding occurs between 4.5 and 6 M urea. The shape of the CD spectrum changes between 2.5 and 3 M urea, where the local minimum at 222 nm is lost (Fig. 4.4A).

The deaminase domain exhibits a 75% decrease in α helix signal between 2.5 and 5.5 M urea (Fig. 4.3B, 4.4B), after most of the fluorescence and activity loss has occurred. The midpoint of this transition, between 3.5 and 4 M urea, correlates with the loss of the local minimum at 222 nm (Fig. 4.4B). By 5.5 M urea, very little helix signal remains.

$FTCDH_6$ also retains most of the helix signal at 222 nm up to 2 M urea (Fig. 4.3C, 4.4C); 35% of the signal is lost between 2 and 3.5 M urea, as is the local minimum at 222 nm. A second decrease of almost 40% is observed between 5 and 7.5 M urea. Thus fluorescence and activity changes of the domains and intact enzyme occur at urea concentrations where most of the secondary structure remains intact. This suggests that the initial conformational changes are limited to changes in tertiary structure.

Urea-induced changes observed by size exclusion chromatography. Findlay and MacKenzie (1987) were able to correlate urea-induced loss of fluorescence intensity and catalytic activity with changes in quaternary structure of wildtype FTCD, which suggested that inactivation might result from subunit dissociation. Therefore, we wanted to ascertain the quaternary structure of the isolated domains and the full length enzyme at different concentrations of urea. Proteins were subjected to size exclusion chromatography to detect volume changes which might signify subunit dissociation (an increase in time of elution) or protein unfolding (a decrease in time of elution). Figure 4.5A shows the chromatographic profiles of FTH_6 in 0 to 5 M urea. At 0 M urea, the transferase domain elutes at 58.2 min, consistent with a molecular mass of 79 kDa, indicating that it behaves as a dimer. At 2 M urea there is very little change in its elution position indicating that the inactivation and loss in fluorescence intensity observed between 1 and 2 M urea result from a conformational change within the dimer, and not from subunit dissociation. At 2.5 M urea, the protein peak has broadened and moved to 52.6 min. This change in elution time coincides with unfolding of the domain as detected by far UV CD and increased

FIGURE 4.5. Size exclusion chromatography profiles of proteins in increasing urea. A. FTH_6. **B.** CDH_6, and **C.** $FTCDH_6$. Following denaturation for 2.5 hours at room temperature, samples were chromatographed at 4°C on a superose 6 column equilibrated in Buffer A containing different concentrations of urea.


A
B
C
5M
3.5M
2.5M
2M
1.5M
1M
0M
20 30 40 50 60 70
elution time
(min)

solvent exposure of the tryptophan residues. It is difficult to differentiate whether dissociation leads to unfolding of an unstable monomeric transferase or whether unfolding of the subunit induces dissociation of dimeric FTH_6. In either case, a monomeric FTH_6 is most likely unstable and prone to further unfolding. At 3.5 M urea, the elution time has further decreased to 45.3 min and the equilibrium has shifted towards a partly unfolded monomeric intermediate with an expanded volume. Partially folded intermediates have been identified in several folding and unfolding pathways (Das et al., 1995; Narhi et al., 1993; Palleros et al., 1993; Uversky & Ptitsyn, 1994; reviewed in Fink, 1995) and have different origins. At 37 kDa, the transferase domain is relatively large and might contain subdomains which are differentially folded at 3.5-4 M, analagous to the differentially folded domains within the molten globule of α-lactalbumin (Wu, et al., 1995). Alternatively, the compact intermediate may include a hydrophobic core consisting of the remaining secondary structure and solvent protected trp residue(s), surrounded by extended unfolded polypeptide chain, as has been proposed in the equilibrium unfolding of DnaK (Palleros, et al., 1993). The domain continues to unfold and expand as the transferase peak has moved forward to 44.1 min in 5 M urea. Even at 5 M urea, the domain may not be fully unfolded since 30% of θ_{222} is still retained. Protein species at 3.5 and 5 M urea which elute in the void volume or prior to the main peak may result from aggregation of the partly unfolded domain. The transferase domain is more hydrophobic than the average protein (Findlay et al., 1989) and dissociation and partial denaturation might reveal hydrophobic faces which would promote nonspecific aggregation.

The denaturation of the deaminase domain was also monitored by size exclusion chromatography (Fig. 4.5B). In the absence of urea, CDH_6 exists as a dimer, eluting with a molecular mass of 53 kDa, approximately twice its subunit size. At 1.5 M urea, the peak becomes more asymmetric, suggesting that some monomer is present. The small shoulder preceding the peak could be either an unfolded monomer or an unfolded dimer. At 2 M urea, the dimer remains the major species although a faster eluting peak has increased. As seen with the transferase domain, a conformational change producing an inactive dimer with decreased intrinsic tryptophan fluorescence appears to precede

dissociation of CDH_n. Folded monomer (57.2 min), inactive dimer (53.4 minutes) and unfolded monomer (49.6 minutes) coexist at 2.5 M urea although the major species appears to be unfolded CDH_n. This suggests that the CDH_n unfolding pathway includes a dissociation step which produces a folded monomer from inactive dimer. This CDH_n monomer is not stable at 2.5 M urea and undergoes partial unfolding. If denaturation occurred by unfolding a dimeric CDH_n, a folded monomer would not be observed. Unfolding of this domain is not very cooperative since a small concentration of dimer may persist with the unfolded species, even at 5 M urea.

Findlay and MacKenzie (1987) proposed a model for the dissociation of octameric FTCD which involved an initial dissociation to dimers followed by, at higher urea concentrations, further dissociation to monomers. They suggested that activity and fluorescence changes correlated with dissociation of FTCD to a dimeric species. In contrast to this, denaturation of the independently expressed domains indicates that a conformational change, as indicated by loss of activity and fluorescence intensity, precedes dissociation in both domains. This conformational change also precedes dissociation in the intact enzyme. Native $FTCDH_6$ elutes predominantly as an octamer with a smaller fraction existing as a smaller oligomer, perhaps a dimer or tetramer (estimated Mr = 200,000) (Fig. 4.5C). There is a small decrease in the elution time of the octamer as the denaturant increases to 2 M urea, and this may be due to loosening of its tertiary structure (Palleros et al, 1993). Between 2 and 2.5 M urea, dissociation occurs at one type of interface and the population of dimeric intermediates eluting at 46.8 min increases. This is followed by partial unfolding between 2 and 3.5 M urea, as shown by a 40% decrease in the α helix signal. At 3.5 M urea, a larger, unfolded dimer, eluting at 41.9 min is the major species. The far UV CD profile suggests that this dimer may be an unfolding intermediate analogous to the partly unfolded transferase monomer (Fig. 4.3C, 4.4C). Like the transferase domain, this unfolded dimer is susceptible to aggregation. As the urea concentration is increased to 5 M, a second dissociation event occurs producing a partly-folded monomer which elutes at 47.2 min. The $FTCDH_6$ monomer undergoes further unfolding as the urea concentration is increased to 7.5 M urea.

Chemical cross-linking with bis-succinimidyl suberate was attempted in order to

confirm urea-induced changes in quaternary structure (data not shown). The ability to cross-link FTH_6 is lost between 2 and 2.5 M urea. Cross-linking results are consistent with dissociation of native $FTCDH_6$ to dimers between 1.5 and 2.5 M urea, and further dissociation to monomers between 3.5 and 5 M urea. However conditions for efficient cross-linking of the deaminase dimer could not be obtained.

To summarize, whether expressed as an isolated domain or constrained within the octameric structure, each domain undergoes a conformational change at the same concentration of urea (as revealed by C_m). These apparent changes in tertiary structure result in a loss of activity without dissociation of the subunit interfaces. The octameric structure enhances the cooperativity of the changes but has only a small effect on domain stability. At higher concentrations of urea, both isolated domains dissociate and unfold within a narrow range of denaturant concentration. Within the FTCD octamer one type of subunit interface is disrupted under the same urea concentrations which promote dissociation of the FTH_6 and CDH_6 dimers. The second type of interface is more stable within the octamer and is maintained at higher concentrations of urea. Previously, Findlay & MacKenzie (1987) suggested that the interface between the transferase domains is lost in the first dissociation event, partly because a proteolytically derived transferase-active fragment underwent dissociation between 2 and 3 M urea as well. Now that we have determined that the isolated deaminase domain also dissociates in this concentration range, it is difficult to differentiate whether the transferase or the deaminase interface is less stable to urea. Despite their apparent inability to interact, the presence of both domains within the intact, bifunctional enzyme enhances the stability of one of the subunit interfaces within the octamer.

ACKNOWLEDGEMENTS

We wish to thank Dr. J. Turnbull, Department of Chemistry, Concordia University, for access to the spectropolarimeter and secondary structure prediction programs and Dr. J. Turnbull, as well as, Drs. J. Cromlish Department of Biochemistry, McGill University, for reading and discussion of this manuscript.

CHAPTER 5

GENERAL DISCUSSION

Octameric formiminotransferase-cyclodeaminase is arranged as a planar ring-shaped tetramer of dimers. The unusual quaternary structure of this bifunctional enzyme has prompted questions concerning the arrangement of subunits, the organization of catalytic activities and the mechanism of substrate channelling between the two types of catalytic sites. Previous investigators have shown that the transferase and deaminase activities are kinetically independent and possibly located on different domains. Expression of each activity was proposed to coincide with retention of a different subunit interface. Isolation of the cDNA encoding FTCD and determination of the amino acid sequence and domain structure of FTCD would allow testing of this hypothesis, and would provide the necessary tools and information for further analysis of the structural organization of this enzyme.

5.1 Sequence analysis of porcine liver FTCD

The porcine liver cDNA described in Chapter 2 of this thesis contains an open reading frame that encodes a 541 amino acid protein, in agreement with the apparent Mr of 60 kDa previously observed for the FTCD subunit. The deduced primary structure of FTCD includes sequences which correspond to two amino acid sequences identified by N-terminal sequencing of cyanogen bromide cleavage fragments of FTCD, confirming the identity of this clone. Porcine FTCD is N-terminally blocked to Edman degradation. Thus it is likely that the initiator methionine is removed and the next residue, a serine, has undergone a post translational modification such as acetylation. The FTCD cDNA was cloned into a prokaryotic expression vector and recombinant FTCD was expressed. The recombinant enzyme can catalyze both the transferase and deaminase activities.

Initial searches of nucleotide and amino acid data banks indicated that this sequence was novel as it displayed no significant homology existed with other known sequences. More recently, several human liver Expressed Sequence Tags, which share 64-90% identity with the porcine FTCD nucleotide sequence, have been described. Also, two groups have isolated cDNA's encoding homologues of FTCD from rat (G. Bloom, personal communication) and chicken (D. Hennig, personal communication). Both groups

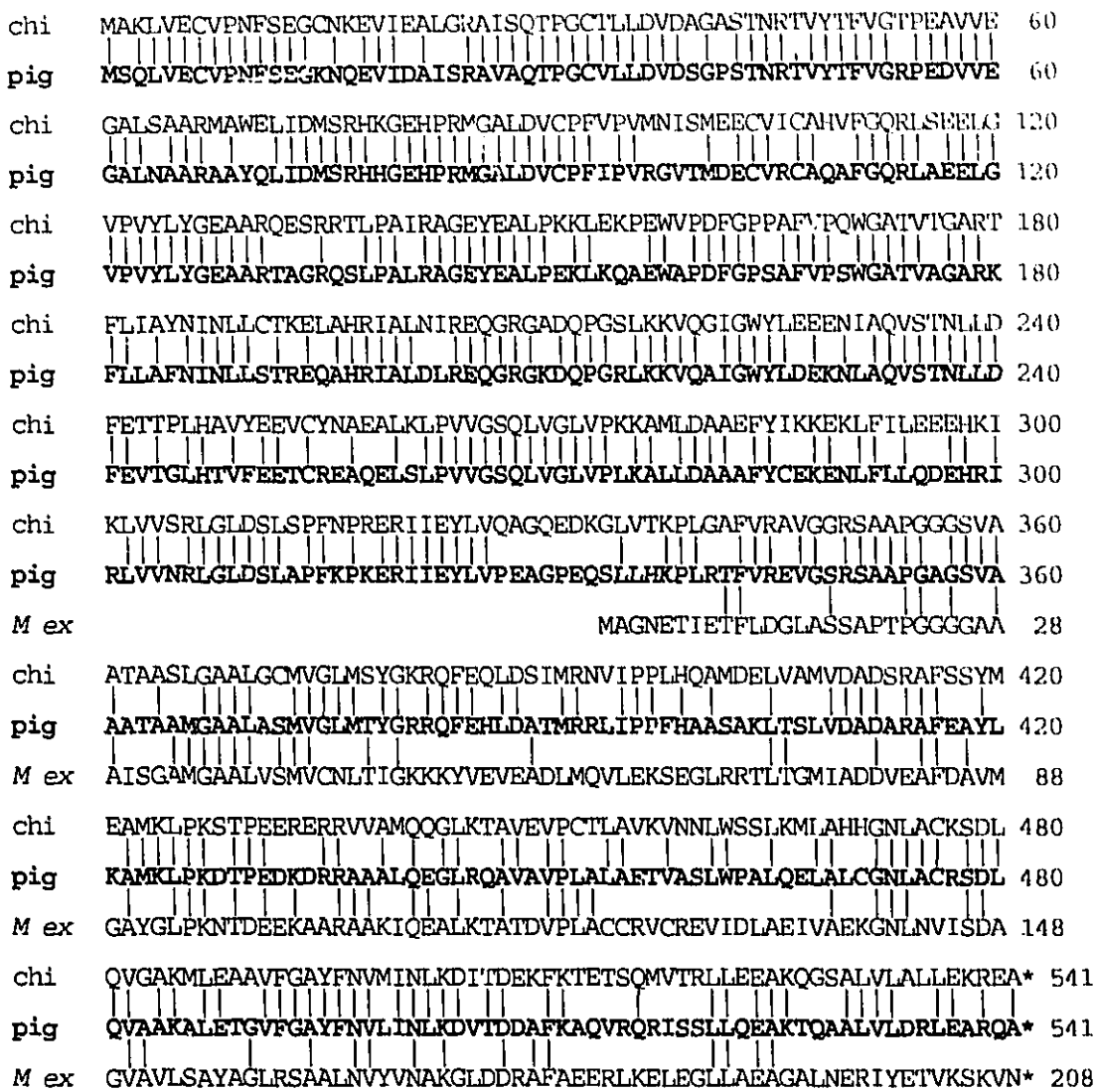
were attempting to clone a 58-60 kDa microtubule associated protein (Bloom & Brashear, 1989). It is quite likely that FTCD binds to the polyglutamates which covalently modify tubulin (Redeker et al., 1991). Like the porcine enzyme, the recombinant rat and chicken proteins are bifunctional. The deduced amino acid sequences for the chicken and porcine enzymes are compared in Figure 5.1. Both the chicken and porcine sequences encode a 541 amino acid protein. They share 68% identity and 82% homology when conservative residues are considered. Thus, a high degree of sequence conservation is observed between chicken, pig and human FTCD.

An open reading frame (*orf4*) in the beginning of a serine cycle gene cluster in *Methylobacterium extorquens* AM1 chromosome (Christserdova & Lidstrom, 1994) can encode a previously unknown 211 amino acid sequence which shares 34% identity (53% homology) with the C-terminal half of FTCD (Figure 5.1). *Methylobacterium extorquens* is a gram-negative facultative methylotroph. This serine cycle gene cluster contains genes encoding proteins involved in the assimilation of C1 units, including the folate dependent 5,10-methyleneH_4PteGlu dehydrogenase (*mtdA*). Mutants containing insertion mutations in *orf4* are unable to grow on C1 or C2 compounds but will grow in the presence of glyoxylate. Christserdova & Lidstrom (1994) have postulated that this gene product may be involved in the conversion of acetyl coenzyme A to glyoxylate. While its similarity to the deaminase domain of FTCD suggests that the *orf4* gene product may be folate-dependent, our present inability to determine which residues are responsible for binding folate must temper such an interpretation.

Bifunctional formiminotransferase-cyclodeaminase has only been identified in some eukaryotes. *E. coli* and yeast do not appear to express either activity. A formiminoH_4PteGlu-dependent cyclodeaminase activity has been described in *Clostridia* (Rabinowitz and Pricer, 1956a;b). The sequence for the gene encoding this deaminase is not yet available, and its relationship to the eukaryotic cyclodeaminase domain is not known.

All FTCD cDNA's isolated to date have been obtained from liver cDNA libraries. Northern analysis (Appendix 1) of total mRNA isolated from mouse tissues supports the argument that FTCD is liver specific. However, low amounts of FTCD activity have also

FIGURE 5.1. Alignment of the predicted amino acid sequence of porcine FTCD with its chicken homolog and *orf4* from *Methylobacterium extorquens*. Identical residues are connected by a bar. Abbreviations: pig, porcine FTCD; chi, chicken FTCD; *M ex*, *orf4* from *Methylobacterium extorquens*


chi MAKLVECVPNFSEGCNKEVIEALGRAISQTPGCTLLDVDAGASTNRTVYTFVGTPEAVVE 60
pig MSQLVECVPNFSEGKNQEVIDAISRAVAQTPGCVLLDVDSGPSTNRTVYTFVGRPEDVVE 60
chi GALSAARMAWELIDMSRHKGEHPRMGALDVCPFVPVMNISMEECVICAHVFGQRLSEELG 120
pig GALNAARAAYQLIDMSRHHGEHPRMGALDVCPFIPVRGVTMDECVRCAQAFGQRLAEELG 120
chi VPVYLYGEAARQESRRTLPAIRAGEYEALPKKLEKPEWVPDFGPPAFVPQWGATVTGART 180
pig VPVYLYGEAARTAGRQSLPALRAGEYEALPEKLKQAEWAPDFGPSAFVPSWGATVAGARK 180
chi FLIAYNINLLCTKELAHRIALNIREQGRGADQPGSLKKVQGIGWYLEEENIAQVSTNLLD 240
pig FLLAFNINLLSTREQAHRIALDLREQGRGKDQPGRLKKVQAIGWYLDEKNLAQVSTNLLD 240
chi FETTPLHAVYEEVCYNAEALKLPVVGSQLVGLVPKKAMLDAAEFYIKKEKLFILEEEHKI 300
pig FEVTGLHTVFEETCREAQELSLPVVGSQLVGLVPLKALLDAAAFYCEKENLFLLQDEHRI 300
chi KLVVSRLGLDSLSPFNPRERIIEYLVQAGQEDKGLVTKPLGAFVRAVGGRSAAPGGGSVA 360
pig RLVVNRLGLDSLAPFKPKERIIEYLVPEAGPEQSLLHKPLRTFVREVGSRSAAPGAGSVA 360
M ex MAGNETIETFLDGLASSAPTPGGGGAA 28
chi ATAASLGAALGCMVGLMSYGKRQFEQLDSIMRNVIPPLHQAMDELVAMVDADSRAFSSYM 420
pig AATAAMGAALASMVGLMTYGRRQFEHLDATMRRLIPPFHAASAKLTSLVDADARAFEAYL 420
M ex AISGAMGAALVSMVCNLTIGKKKYVEVEADLMQVLEKSEGLRRTLTGMIADDVEAFDAVM 88
chi EAMKLPKSTPEERERRVVAMQQGLKTAVEVPCTLAVKVNNLWSSLKMLAHHGNLACKSDL 480
pig KAMKLPKDTPEDKDRRAAALQEGLRQAVAVPLALAETVASLWPALQELALCGNLACRSDL 480
M ex GAYGLPKNTDEEKAARAAKIQEALKTATDVPLACCRVCREVIDLAEIVAEKGNLNVISDA 148
chi QVGAKMLEAAVFGAYFNVMINLKDITDEKFKTETSQMVTRLLEEAKQGSALVLALLEKREA* 541
pig QVAAKALETGVFGAYFNVLINLKDVTDDAFKAQVRQRISSLLQEAKTQAALVLDRLEARQA* 541
M ex GVAVLSAYAGLRSAALNVYVNAKGLDDRAFAEERLKELEGLLAEAGALNERIYETVKSKVN* 208

been described in other tissues, most notably kidney and intestine (Tabor & Wyngarden, 1959, McLain et al., 1975), and low levels of the transcript may be present in these tissues. The FTCD gene and promoter region have not yet been studied so little is known about possible mechanisms of transcriptional regulation. Secondary structure analysis of the porcine FTCD cDNA indicates that the 5' end of this cDNA can form a stem-loop structure with the initiator AUG codon situated within the loop. The transcriptional start site of the FTCD mRNA has not been mapped and the longest 5' noncoding region of the porcine FTCD cDNA includes only 27 nucleotides, so the true extent of secondary structure within this region is not known. However, this represents a possible mechanism for translational control of FTCD expression.

FASTA and Blast searches of the SWISSPROT and PIR databank searches have not identified any significant regions of amino acid homology between FTCD and other folate-binding enzymes. However, visual inspection of different folate-dependent sequences has suggested some similarities. AAAAAVSGA is the N-terminal sequence of rabbit liver 5,10-methenylH_4PteGlu synthetase (Maras et al., 1994). It has been suggested that consensus sequences could be read in reverse order (Sternlicht et al., 1987) and, interestingly, residues 356-364 of FTCD are AGSVAAATA which maps to the N-terminal region of the cyclodeaminase domain. Both the deaminase and the synthetase activities produce 5,10-methenylH_4PteGlu, and perhaps this sequence plays a similar role in both proteins.

5.2 The folate binding site

While many folate binding proteins have been cloned or sequenced, a general tetrahydrofolate binding consensus sequence has not been identified. This may reflect evolutionary pressure to develop binding sites which are specific for the C1 unit bound to the pteridine substrate, or for a particular tail length. Cook et al. (1991) have identified a putative 10-formylH_4PteGlu binding consensus sequence, $XPS(X)_2P(X)_{2-3}G$, which has been observed in 10-formylH_4PteGlu synthetases, GAR transformylases and 10-formylH_4PteGlu dehydrogenases from several different sources (Cook et al., 1991; Nour and Rabinowitz, 1992; Rankin et al., 1993). A similar sequence, GPSAFVPSWG, is found

in porcine FTCD at residues 163-172 and may play a role in H_4PteGlu recognition (Figure 5.2). The corresponding sequence in chicken FTCD retains the proline and glycine residues present in the consensus sequence. As both the transferase and the deaminase activities appear to independently bind 5-formiminnoH_4PteGlu, a second folate binding site is likely present within the primary structure of FTCD. A possible candidate for this second sequence is found between residues 338 and 348, as shown in Figure 5.2. This sequence is located at the amino-terminus of the C-terminal deaminase domain and shows similarities to the 10-formyl binding site and the putative FTCD folate binding site. Unfortunately little is known about either FTCD folate binding site.

Structural similarities between folate binding sites in different types of enzymes are not particularly evident at the tertiary structure level either. Chen has compared the crystal structures of GAR transformylase, thymidylate synthase, dihydropteridine reductase and dihydrofolate reductase (Chen et al., 1992). Similarity is mostly limited to within the nucleotide binding regions of the three nucleotide binding enzymes. Both GAR transformylase and DHFR appear to bind substrate in a cleft between two domains. Perhaps a common binding mechanism will become obvious once more crystal structures of folate-binding proteins are solved.

5.3 The domain structure of formiminotransferase-cyclodeaminase

To further analyze the structure of FTCD we wanted to determine if the transferase and deaminase activities localize to separate domains (Chapter 3, this thesis). Like the full length enzyme, a 39 kDa transferase-active proteolytic fragment was N-terminally blocked (Findlay et al., 1989), suggesting that the transferase activity resides within the N-terminal region of FTCD. Comparison of the amino acid composition of this proteolytic fragment (Findlay et al., 1989) and that of the N-terminal 350 residues of FTCD supported this assignment. Deletion mutagenesis was performed on the cDNA encoding porcine FTCD to produce constructs which would express N-terminal proteins ending at different positions between amino acids 318 and 350, and C-terminal fragments which initiated at different residues within the same region. To identify domains, the resulting fragments were characterized in terms of their catalytic activities and their solubility. These

FIGURE 5.2. Comparison of putative folate binding sequences in the transferase and deaminase domains of FTCD. FTCD residues that are identical to the 10-formylH_4PteGlu binding consensus sequence are in **bold**. Residues which are common to the scquences in both the transferase and the deaminase domain are in *italics*.

	*V**P***SX_2PX_{2-3}***G***	10-formylH_4PteGlu binding sequence
pig	163 G***P***-SA***FV***PSW***G*** 172	transferase domain
chick	163 G***P***-PA***FV***PQW***G*** 172	
pig	338 K***P***LRT***FV***REV***G*** 348	deaminase domain
chick	338 K***P***LGA***FV***RAV***G*** 348	

experiments indicated that FTCD is composed of a separable N-terminal transferase domain and a C-terminal deaminase domain connected by a minimal linker sequence comprising residues 326-333. The C-terminal boundary of the transferase domain was limited by the insolubility of shorter N-terminal fragments, perhaps resulting from improper folding. The N-terminal boundary of the deaminase domain was limited by a decrease in intrinsic deaminase activity in shorter species. Interestingly, decreasing deaminase activity was observed in fragments starting at residues 337, 339 and 341, which are immediately prior to or within the first half of a putative folate recognition sequence described in section 5.2. Inactive fragments beginning at residue 344 and 350 are missing over half or all of this sequence.

The linker region separating the transferase and deaminase domains is minimally 8 and maximally 12 residues in length. Comparison of the amino acid sequences for pig and chicken FTCD suggests that residues 327-334, as the least conserved sequence within FTCD, may act as a connecting region. This corresponds well with the linker boundaries determined by deletion analysis. The linker region may promote proper folding of FTCD simply by separating the transferase and deaminase domains as they fold, since deletion of six proposed linker residues (327 to 332, inclusive) produces a bifunctional protein which is less soluble than recombinant FTCD. Extending this deletion to include residues 327 to 338, produces a protein which is even less soluble and no longer catalyzes either reaction. While the sequence of the linker region is not conserved between pig and chicken, its length apparently is. The linker region may play a role in keeping the transferase and deaminase domains properly oriented to optimize substrate channelling between the two types of active sites. If channelling involves a swinging anchor mechanism, as described in Chapter 1, changing the length or the composition of the linker sequence may affect both the efficiency of channelling and the optimal polyglutamate tail length. It may prove informative to determine if the deletion mutant (missing six linker amino acids) preferentially channels pentaglutamylated intermediates or shorter derivatives, and if replacing the pig linker sequence with the corresponding residues from chicken affects channelling behaviour.

Both the transferase and deaminase domains can be expressed separately in *E.coli*,

demonstrating that they fold independently. For the most part, there is little change in the secondary and tertiary structure of the transferase and deaminase domains upon separation. The fluorescence spectra of the full length enzyme and the mixed transferase and deaminase domains are comparable, as are the CD spectrum of the full length enzyme and the additive spectrum of the isolated domains, between 200-240 nm (Chapter 4). However both isolated domains may be susceptible to minor conformational changes, perhaps resulting from an increased freedom of movement as the domains are no longer fettered by the spatial constraints imposed by the octamer. Kinetic characterization of the isolated transferase and deaminase domains indicated that while the domains and the full-length enzyme bind formiminoglutamate and monoglutamylated folates with similar affinities, both types of isolated domains have turnover numbers which are only one half to two-thirds the values reported for the full-length enzyme (Chapter 3). The chymotryptic transferase-active fragment also retained only 58-67% of the original activity (MacKenzie et al., 1980; Findlay & MacKenzie, 1987) following proteolysis.

The catalytic activities of many other multifunctional proteins are arranged on separable domains which can be expressed independently (Chapter 1). This may reflect an evolutionary history which involves ligation of the genes which encode the component activities. It is possible that FTCD evolved through fusion of genes which encoded monofunctional transferase and deaminase activities. Consequent evolution events may have allowed the enzyme to develop or optimize its ability to channel polyglutamylated substrate. However, since there is little available evidence to support the existence of monofunctional transferases and deaminases, it is premature to assign such a mechanism to FTCD. Only two possible candidates for a primitive deaminase activity have been described; *orf4* from *Methylobacterium extorquens*, and a 5-formiminoH$_4$PteGlu dependent cyclodeaminase that is associated with a 5,10-methenylH$_4$PteGlu cyclohydrolase activity and is involved in degradation of purines in *Clostridia* (Uyeda & Rabinowitz, 1967). *Clostridia* contain a separate formiminotransferase which transfers the formimino group from formiminoglycine to tetrahydrofolate (Uyeda & Rabinowitz, 1965). This protein is a potential precursor to the transferase domain. Further studies to determine the distribution of monofunctional counterparts of these activities in nature, combined with

the isolation and characterization of the enzymes responsible and cloning of the genes encoding these activities would be necessary to further develop the evolutionary history of this enzyme. The fusion of the transferase and deaminase activities should provide FTCD with certain advantages, many of which have been observed with other multifunctional enzymes (discussed in Chapter 1). The transferase activity produces formiminoH$_4$PteGlu, which can only be metabolized by the cyclodeaminase activity. This product is not stable at physiological pH (Paquin et al., 1985) and has the potential to interfere with other folate-dependent enzymes. Coexpression of the transferase and deaminase on a single polypeptide chain ensures the presence of both activities, preventing a build up of the formimino-intermediate in the cell. As well, the proximity of the two types of active sites and the enzyme's unique ability to channel polyglutamylated intermediate allow for the efficient transfer of the formimino species between active sites. As well be discussed later in this chapter, formation of the bifunctional enzyme results in increased stability of one type of subunit interaction within the FTCD octamer.

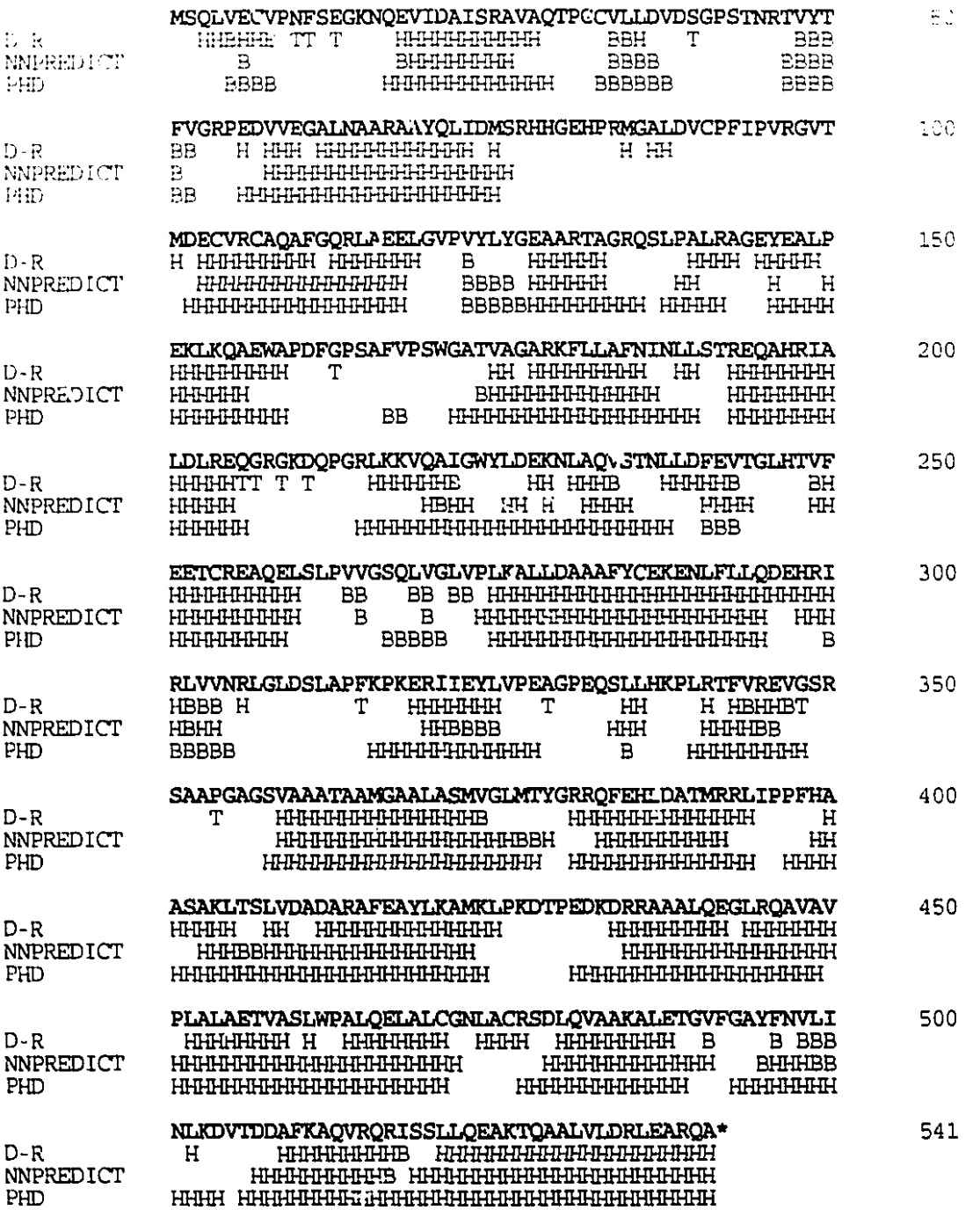
5.4 Secondary structure predictions

Secondary structure analysis of the CD spectra of the full length enzyme and both isolated domains suggested that FTCD, and the deaminase domain in particular, have a significant α-helical content (Chapter 4, Table 4.1). Secondary structure predictions based on amino acid sequence data support a predominantly α-helical deaminase domain (Chapter 4, Table 4.1; Figure 5.3a). Helical projections (Figure 5.3b) indicate that several predicted helices have hydrophobic faces with repeating alanines. Gernet et al. (1995) have described an antiparallel coiled-coil, called the alacoil, where alanines at positions a and e in the helix form interhelical contacts. The alanine rich faces in the deaminase domain may play a similar role in assembly of tertiary structure within FTCD.

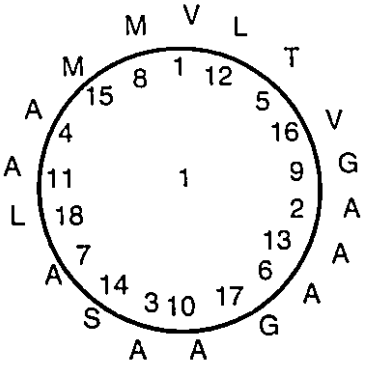
5.5 The role of the polyglutamate tail in substrate binding and channelling.

FTCD preferentially binds and channels polyglutamylated substrates. Polyglutamylation improves the binding of both the folate substrates and FIGLU (Paquin

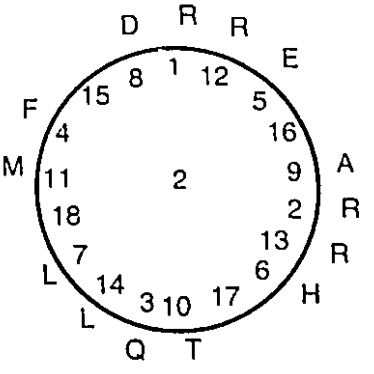
FIGURE 5.3. Secondary structure prediction for FTCD based on amino acid sequence. A H represents helix; B, sheet; T, turn; blank, coil. D-R is the method of Deleage and Roux (1989); NNPREDICT is the method of Kneller, et al. (1990); PHD is the method of Rost & Sanders (1993). **B**. Helical wheel projections of the predicted helices within the deaminase domain.



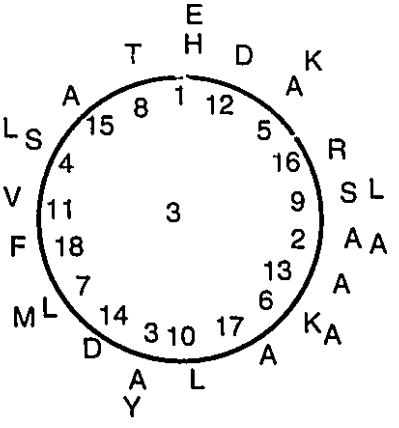
1. RESIDUES 359-376



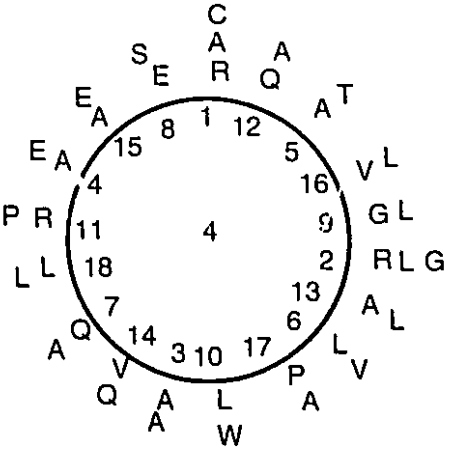
2. RESIDUES 381-394



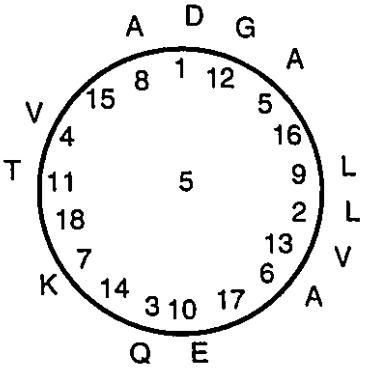
3. RESIDUES 399-425



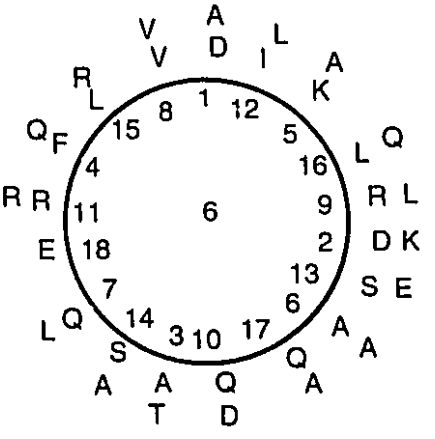
4. RESIDUES 435-472



5. RESIDUES 479-491



6. RESIDUES 507-541



et al., 1985; Findlay et al., 1989), and binding polyglutamylated substrates induces a conformational change which prevents limited proteolysis of FTCD by chymotrypsin (MacKenzie et al., 1980), and dissociation of the octamer in 3 M urea (Findlay & MacKenzie, 1988). The deaminase domain but not the transferase domain shows a preference for pentaglutamylated substrate, indicating that the polyglutamate binding site resides within this domain (Chapter 3, this thesis). This is further supported by Findlay and MacKenzie's (1988) observation that, upon urea-induced dissociation of FTCD into monofunctional dimers, only the deaminase-active dimer retained specificity for the pentaglutamylated intermediate. Previously, Paquin et al. (1985) demonstrated the presence of 4 high affinity polyglutamate binding sites within the octamer, or one per dimer. The existence of only one polyglutamate binding site per pair of deaminase domains, suggests that this site (and possibly the deaminase active site, as well) may be located at the subunit interface formed between deaminase domains.

As polyglutamylated folates are negatively charged species, it is quite likely that electrostatic interactions play an important role in binding these substrates to their target enzymes. Several polyglutamate binding enzymes also interact with other negatively charged polymers (the cyclodeaminase domain of FTCD binds to heparin; the trifunctional D\C\S binds to single stranded DNA, Wahls et al., 1993), presumably through their polyglutamate binding sites. Basic amino acids have been implicated in the binding of polyglutamylated substrates to several folate-dependent enzymes. Stroud's group has determined the crystal structures of thymidylate synthases from different sources (Kamb et al.; 1992; Finer-Moore et al., 1994, Stroud, 1994) and suggest that the polyglutamate binding site consists of a shallow cleft on the protein surface, lined with hydrophilic sidechains. While the amino acids that form this cleft are not highly conserved, several basic residues which make important contacts with the polyglutamate tail have been identified. Rabinowitz and colleagues have identified a 53 residue sequence which may form part of the synthetase polyglutamate binding site (Whitehead & Rabinowitz, 1988; Nour & Rabinowitz, 1992). This sequence is specific to 10-formylH_4PteGlu synthetases which bind longer polyglutamates and contains several doublets of basic amino acids. The primary structure of the deaminase domain of FTCD includes many basic residues

distributed throughout the sequence, as reflected by the estimated pI of 9.6. However, several doublets of basic residues are present (H336, K337; R381, R382; R392, R393; R435, R436) some of which may influence the binding of polyglutamate to the FTCD octamer.

FTCD preferentially channels pentaglutamylated but will also channel other polyglutamylated intermediates with tails of 4 or more glutamates in length, albeit with less efficiency. While channelling efficiency favours the pentaglutamylated substrate, longer polyglutamates are more tightly bound (Paquin et al., 1985), suggesting that the mechanism of substrate channelling within FTCD involves a steric component. MacKenzie and colleagues have developed a "swinging arm" model whereby the polyglutamate binds to a site on the octamer while the more mobile pterin moiety can interact with either type of catalytic site (MacKenzie & Baugh, 1980; Paquin et al., 1985). The pentaglutamate tail would provide an optimal length such that the pterin moiety could reach both the transferase and a deaminase active site without first being released from the polyglutamate binding site. This model only allows for preferential transfer of intermediate within a single molecule of the enzyme. As expected, the separated transferase and deaminase domains cannot channel the pentaglutamylated intermediate. Other enzymes channel intermediates via a swinging arm mechanism (pyruvate dehydrogenase and fatty acid synthase), however the intermediate is covalently bound to the enzyme and a specific domain within the protein functions as the swinging arm.

A different channelling mechanism was described by Knighton et al. (1994) to explain the channelling of polyglutamylated dihydrofolate within thymidylate synthase-dihydrofolate reductase from *Leishmania major*. Upon examination of the crystal structure of this bifunctional enzyme, the authors observed an unusual distribution of charged residues which extended across the surface of the protein. This surface may electrostatically guide the intermediate from the synthase to the reductase active site. Interestingly this surface is conserved in monofunctional thymidylate synthases, where it most likely functions as a polyglutamate binding site (Stroud, 1994). The relative location of the transferase and deaminase catalytic sites and their orientation within the FTCD octamer must first be defined before determining if either of these channelling

mechanisms applies to FTCD

5.6 The quaternary structure of FTCD and confirmation that the octamer is the functional unit

Cross-linking and proteolysis studies have shown that FTCD is arranged as a tetramer of dimers. Findlay and MacKenzie (1987; 1988) performed a series of dissociation and renaturation studies to determine whether a dimer could serve as the functional unit of FTCD. Instead they found that the isolated dimers were invariably monofunctional, expressing only one or the other catalytic activity, and therefore proposed that the octamer is the functional unit of FTCD. Analysis of the quaternary structure of the independent transferase and deaminase domains provides an explanation for this observation. Both the transferase and deaminase domains exist as dimers in solution, indicating that each domain contains sequences governing subunit interaction. Therefore each subunit within the octamer consists of an N-terminal transferase domain which forms a subunit interface with the transferase domain of the adjacent subunit, and a C-terminal deaminase domain which forms a second subunit interface with the deaminase domain of the other adjacent subunit. This explains why FTCD is arranged as a tetramer of dimers with alternating subunit interfaces, and confirms that the octamer is the functional unit of this enzyme.

As both the transferase and the deaminase domains exist as dimers, it is quite likely that FTCD contains two types of isologous interactions. Dimers generally do not exhibit heterologous interactions because this would leave two binding surfaces available for further aggregation, producing a less stable protein. While an argument could be made that a heterologous dimer could undergo a conformational change upon association that would exclude further association, the literature suggests that such an arrangement is very uncommon. However, an asymmetric subunit arrangement has been observed for hexokinase (Anderson et. al. 1974; Anderson & Steitz, 1975) and proposed for creatine kinase (reviewed in Degani & Degani, 1980).

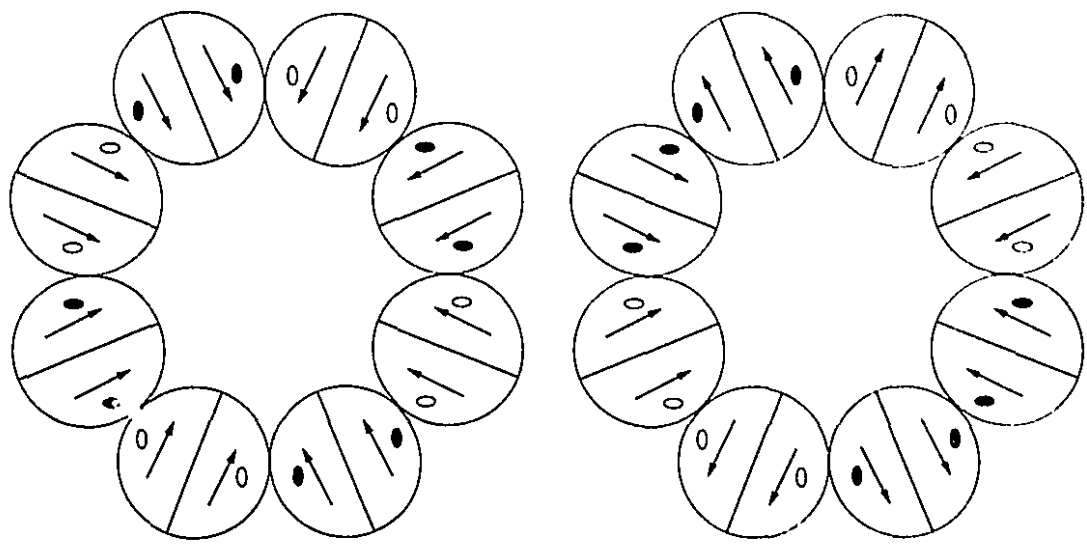
In some oligomeric enzymes, the active sites are located at the subunit interface, and require residues from both monomers. An isologous dimer could contain two

complete active sites at a single subunit interface (glutathione reductase, thymidylate synthase), or one catalytic site within each monomer (glutamate dehydrogenase). In both cases, one would expect the presence of two catalytic sites per dimer. FTCD appears to contain only one high affinity polyglutamate binding site per deaminase dimer suggesting that binding of the substrate may induce some sort of negative cooperativity, perhaps by inducing a conformational change which prevents binding of a second molecule. Alternatively if the two active sites were close, the size of the polyglutamylated substrate may prevent binding of a second molecule of the substrate, for steric reasons. Findlay (1988) suggested that curvature of the octamer might constrain one of the two catalytic sites such that only one site remained functional per interface. If this were true, we might expect the isolated dimers to have turnover rates which were twice those observed for the octameric enzyme, and in fact, the opposite was observed. A slightly different interpretation is that the curvature present in the octamer limits the flexibility of the deaminase domain such that one active site acquires a more efficient conformation, where it can bind substrates much more tightly and/or catalyze the deamination reaction more rapidly. If this latter model were accurate, the isolated deaminase domain might be able to bind two polyglutamates per dimer, but function with lower catalytic efficiency than the full length enzyme. As shown in Chapter 3, the deaminase domain has lower affinity for formiminoH$_4$PteGlu$_5$ than the full length enzyme, and has a turnover number that is only two-thirds the value described for the full length enzyme. It will be interesting to determine the shape and symmetry of the subunit interfaces within this octamer, and the relative orientations of the different sets of dimers. Figure 5.4. shows two types of orientations that are possible for a planar tetramer of dimers.

5.7 Interactions within and between domains

The transferase and deaminase domains can fold independently and show no indication of interacting once separated (Chapter 3). This suggested that the domains also function independently within the subunit. However, three pieces of evidence (both independent domains have decreased turnover numbers relative to the full length enzyme, FTCD has the ability to channel substrate, and polyglutamylated substrate protects both

FIGURE 5.4. Two arrangements for a planar, circular tetramer of dimers. Both arrangements involve isologous interactions, between the same type of domain. The stippled half of each subunit represents one type of domain and the white half represents the other type of domain. The black and white ovals represent subunits which are orientated in and out of the page, respectively.



activities against urea-induced inactivation) argue that some sort of interdomain interaction is necessary. An equilibrium unfolding study demonstrated that the tertiary structure of both domains was only slightly less stable outside of the octamer, suggesting the presence of minor domain interactions only (Chapter 4). A differential scanning calorimetry study of the full length enzyme and both isolated domains would provide a more complete analysis of interdomain interactions within FTCD.

More interestingly, these studies indicated that one type of subunit interface within the octamer becomes stabilized. The presence of the remainder of the polypeptide appears to strengthen the subunit interaction. We cannot presently distinguish whether the transferase or the deaminase interface becomes stabilized, or even why this stabilization occurs. Findlay and MacKenzie were able to trap a transferase-active dimer in 1.5 M urea during the refolding of FTCD. However one should not assume that this kinetic folding intermediate contains the more stable interface. Determining the crystal structure of this enzyme will provide some insight into interactions between the transferase and deaminase domains which could result in this type of stabilization.

Both isolated domains undergo a change in tertiary structure prior to dissociation. Dissociation and unfolding of secondary structure are coordinated, occurring under similar concentrations of denaturant. Interactions between subunits likely play a substantial role in stability of both domains and subunit assembly may involve major conformational changes. Both domains contain a high proportion of amino acids such as bulky hydrophobic residues and arginines, that are commonly found at subunit interfaces. The deaminase domain contains 53% hydrophobic residues and 7% arginines, and the transferase domain contains 44% hydrophobic and 7% arginines, as compared to approximately 40% hydrophobic residues and 5% arginines for 23 oligomeric proteins and 36% hydrophobic residues and 3% arginines for 37 monomeric proteins (average percentages taken from Table 3 of Janin, Miller & Chothia, 1988).

Unfolding of the transferase domain as monitored by circular dichroism suggest that a monomeric partially folded intermediate accumulates between 3.5 and 4 M urea. This intermediate is more compact than species at higher concentrations of denaturant and retains approximately 40% of its secondary structure (Chapter 4). Structural analysis of

this intermediate may provide information regarding which interactions form the hydrophobic core of this domain. Comparison of the structure of this intermediate and the fully folded species may elucidate which structural properties are required to make the domain competent for dimerization. Currently, it appears that association requires the presence of most of the α-helical secondary structure. Some alterations in tertiary structure can be tolerated, however, the extent of these changes is presently unknown. A close examination of the refolding and association pathways of each isolated domain would also be useful.

5.8 Perspectives for the future

While FTCD is known to be widely expressed in mammals and perhaps other eukaryotic species, only porcine FTCD has undergone extensive investigation. In humans, FTCD deficiency has been associated with a disease state, however it is controversial whether inactivation of this enzyme is the causative factor (reviewed by Rosenblatt, 1989). Two classes of phenotypes have been described. One type involves physical and mental retardation while a second, less severe type shows no mental retardation and massive excretion of FIGLU. Investigators have suggested that the more severe phenotype may result from inactivation of the deaminase activity and the mild type from loss of the transferase activity. The major difficulty with diagnosing this disorder stems from the lack of expression of FTCD outside of the liver. Expressed Sequence Tags which correspond to human FTCD could be used to isolate the human FTCD gene, and develop an alternative screening method for FTCD deficiency.

As a multifunctional enzyme, FTCD exemplifies one form of enzyme organization. The literature suggests that several enzymes involved in folate metabolism (Caperelli et al., 1980) and synthesis of deoxynucleotide precursors (Mathews, 1991; Reddy & Fager, 1993) may form higher order complexes, or metabolons, within the cytoplasm. FTCD might associate with other enzymes involved in degradation of histidine, or with the trifunctional D\C\S. It would be interesting to determine if immunoprecipitation of FTCD from liver extracts would also precipitate enzymes from the histidine degradation or folate pathways.

Throughout this chapter, I have pointed out instances where knowledge of the crystal structure of FTCD would help resolve questions that are currently unanswerable, or at least point us in the right direction. The most important of these include determining the symmetry of the subunit interactions, the number and location of active sites within the octamer, and whether this arrangement could support a swinging arm mechanism of channelling. A different approach would be to biochemically determine active site residues. Previous chemical modification studies indicate that modification of a histidine results in inactivation of the transferase (MacKenzie & Baugh, 1980) and modification of a cysteine results in loss of the deaminase (Drury & MacKenzie, 1977). Five of the six histidines in the transferase domain are conserved between chicken and pig while only one of the two cysteines (C476) in the deaminase domain are conserved. Replacement of these histidines and cysteines with conservative amino acids by site-directed mutagenesis should indicate whether these residues are required for activity.

Paquin et al. (1985) have shown that there are only four high affinity polyglutamate binding sites per octamer, or one per pair of deaminase domains. This forces the following question: what is the ratio of polyglutamate binding sites to transferase and deaminase active sites? The number of monoglutamylated folate binding sites within the octamer could not be determined by equilibrium binding studies, presumably because the dissociation constants were too high. Perhaps isothermal titration studies would provide the number of folate binding sites that are required to develop a model for the arrangement of active sites within the subunit.

5.9 Conclusion

The studies described in this thesis have provided new insight into the structural arrangement of formiminotransferase-cyclodeaminase at several different levels of structural organization. A novel cDNA encoding FTCD provided the primary structure of this enzyme. Deletion analysis of this cDNA demonstrated that FTCD is composed of an N-terminal transferase domain and a C-terminal deaminase domain. Both domains contain sequences governing dimerization, confirming that FTCD is arranged as a tetramer of dimers, and that the functional unit of FTCD is the octamer. The isolated transferase and

deaminase domains can no longer channel pentaglutamylated substrate, as predicted by the swinging arm mechanism of channelling. While the independent transferase and deaminase domains can fold stably and show no indication of interacting, several lines of evidence suggest that the transferase and deaminase domains do interact within the subunit, the most compelling being the increased stability of one type of subunit interface within the octamer.

CONTRIBUTIONS TO ORIGINAL KNOWLEDGE

1. A porcine liver cDNA encoding formiminotransferase-cyclodeaminase has been isolated. Sequence analysis indicates that this cDNA encodes a novel protein of 541 amino acids. Northern analysis suggests that FTCD is encoded by a single, liver-specific transcript of approximately 1.9 kb.

2. Bifunctional, octameric FTCD was expressed in *E. coli*. The purification of the recombinant enzyme includes a novel step, pseudo-affinity chromatography on affigel-15-polyglutamate resin.

3. Each FTCD subunit consists of an N-terminal transferase-active domain and a C-terminal deaminase-active domain, separated by a short linker region.

4. Both domains can be expressed in E. coli, indicating that each can fold independently and achieve a functional conformation. Both domains, as well as the full length enzyme, have been expressed as C-terminally histidine-tagged proteins to simplify their purification. Kinetic characterization suggests that the histidine tag has little effect on the catalytic function of these proteins.

5. Both types of domains exist as dimers. Therefore, within the octameric (tetramer of dimers) structure of FTCD, one type of subunit interface can be formed by interaction of the transferase domains, the second by interaction of the deaminase domain.

6. While the recombinant FTCD retains the ability to directly transfer the pentaglutamylated intermediate between the transferase and deaminase active sites, a mixture of isolated domains no longer exhibits this channelling behaviour. This confirms that the transferase and deaminase domains must be physically associated for channelling to occur. (The isolated domains show no indication of interacting with each other.) The isolated deaminase domain retains specificity for the pentaglutamylated substrate,

confirming that the polyglutamate binding site resides within this domain

7 The domains and the full-length enzyme exhibit multistate denaturation profiles. Both isolated domains undergo a change in tertiary structure at low concentrations of urea (below 2 M urea) to form inactive dimers. This is followed by dissociation and unfolding at slightly higher urea concentrations. While both isolated domains undergo dissociation to monomer between 2 and 2.5 M urea, only one type of subunit interface in the octamer is disrupted at this urea concentration. Dissociation of the second interface occurs between 3.5 and 5 M urea, indicating that one domain achieves increased stability within the full length enzyme.

REFERENCES

Aceto, A., Caccuri, A.M., Sacchetta, P., Bucciarelli, T., Dragani, B., Rosato, N., Federici, G., and Di Ilio, C. (1992) *Biochem. J.* **285**, 241-245.

Aimi, J.H., Qui, H., Williams, J., and Zalkin, H. (1990) *Nucleic Acids Res.* **18**, 6665-66672.

Allen, J.R., Lasser, G.W., Goldman, D.A., Booth, J.W., and Matthews, C.K. (1983) *J. Biol. Chem.* **258**, 5746-5753.

Anderson, W.F., Fletterick, R.J., and Steitz, T.A. (1974) *J. Mol. Biol.* **86**, 261-269.

Anderson, W.F., and Steitz, T.A. (1975) *J. Mol. Biol.* **92**, 279-287.

Andrade, M.A., Chacón, P., Merelo, J.J., and Morán, F. (1993) *Protein Eng.* **6**, 383-390.

Anfinsen, C.B. (1966) *Harvey Lect.* **61**, 95-1116.

Anfinsen, C.B. (1973) *Science* **181**, 223-230.

Antony, A.C. (1992) *Blood* **79**, 2807-2820.

Appling, D.R. (1991) *FASEB J.* **5**, 2645-2651.

Argos, P. (1990) *J. Mol. Biol.* **211**, 943-958.

Atkinson, D.E. (1969) *Curr. Top. Cell. Reg.* **1**, 29-43.

Auffray, C., and Rougeon, F. (1980) *Eur. J. Biochem.* **107**, 303-314.

Aviv, H., and Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408-1412.

Baggott, J.E., and Krumdieck, C.L. (1979) *Biochemistry* **18**, 1036-1041.

Bai, Y., Sosnick, T.R., Mayne, L., and Englander, S.W. (1995) *Science* **269**, 192-197.

Baker, D., and Agard, D.A. (1994) *Biochemistry* **33**, 7505-7509.

Bannerjee, R., Mande, S.C., Ganesh, V., Das, K., Dhanaraj, V., Mahenta, S.K., Suguna, K., Surolia, A., and Vijayan, M. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 227-231.

Barlowe, C.K., and Appling, D. (1988) *Biofactors* **1**, 171-176.

Batke, J. (1991) *J. Theor. Biol.* **152**, 41-46.

Beasty, A.M., Hurle, M.R., Manz, J.T., Stackhouse, T., Onuffer, J.J., and Matthews, C.R. (1986) *Biochemistry* **25**, 2965-2974.

Beaudet, R., and MacKenzie, R.E. (1975) *Biochim. Biophys. Acta* **410**, 252-261.

Beaudet, R., and MacKenzie, R.E. (1976) *Biochim. Biophys. Acta* **453**, 151-161.

Bélanger, C., and MacKenzie, R.E. (1989) *J. Biol. Chem.* **264**, 4837-4843.

Bennet, M.J., Choe, S., and Eisenberg, D. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3127-3131.

Bennet, M.J., Schlunegger, M.P., and Eisenberg, D. (1995) *Protein Sci.* **4**, 2455-2468.

Bensadoun, A., and Weinstein, D. (1976) *Anal. Biochem.* **70**, 241- 250.

Bevitt, D.J., Cortes, J., Haydock, S.F., and Leadlay, P.F. (1992) *Eur. J. Biochem.* **204**, 39-49.

Birnstiel, M.L., Busslinger, M., and Strub, K. (1985) *Cell* **41**, 349-359.

Blond, S., and Goldberg, M.E (1985) *J. Mol. Biol.* **182**, 597-606.

Blond, S., and Goldberg, M.E. (1986) *Proteins Struct. Funct. Genet.* **1**, 247-255.

Blond, S., and Goldberg, M.E. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1147-1151.

Bloom, G.S., and Brashear, T.A. (1989) *J. Biol. Chem.* **264**, 16083-16092.

Bordo, D., and Argos, P. (1990) *J. Mol. Biol.* **211**, 975-988.

Bowie, J.U., and Sauer, R.T. (1989) *Biochemistry* **28**, 7139-7143.

Bradford, M.M. (1976) *Anal. Biochem.* **72**, 248-254.

Brandts, J.F., Hu, C.Q., Lin, N.L., and Mas, M.T. (1989) *Biochemistry* **28**, 8588-8596.

Brewer, S., Tolley, M., Trayer, I.M., Barr, G.C., Dorman, C.J., Hannavy, K., Higgins, C.F., Evans, J.S., Levine, B.A., and Wormald, M.T. (1990) *J. Mol. Biol.* **216**, 883-895.

Brown, H.R., and Schachat, F.H. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2359-2363.

Buchwalder, A., Szadkowski, H., and Kirschner, K. (1992) *Biochemistry* **31**, 1621-1630.

Bycroft, M., Matouschek, A., Kellis, J.T., Serrano, L., and Fersht, A.R. (1990) *Nature* **346**, 837-845.

Caperelli, C.A., Benkovic, P.A., Chettur, G., and Benkovic, S.J. (1980) *J. Biol. Chem.* **255**, 1885-1890.

Carra, J.H., Griko, Y.V., and Privalov, P.L. (1995) *Protein Sci.* **4** (suppl. 2), 69.

Carrey, E.A. (1993) *Biochem. Soc. Trans.* **21**, 191-195.

Case, G.L., Kaisaki, and P.J., Steele, R.D. (1988) *J. Biol. Chem.* **263**, 10204-10207.

Chaffotte, A.F., Cadieux, C., Guillou, Y., and Goldberg, M.E. (1992) *Biochemistry* **31**, 4303-4308.

Chaffotte, A.F., Guillou, Y., Delpierre, M., Hinz, H-J., and Goldberg, M.E. (1991) *Biochemistry* **30**, 8067-8074.

Chanarin, I. (1969) in *The Megaloblastic Anaemias*, Blackwell Scientific Publications, Oxford.

Chen, P., Schilze-Gahmen, U., Stura, E.A., Inglese, J., Johnson, D.L., Marolewski, A., Benkovic, S.J., and Wilson, I.A. (1992) *J. Mol. Biol.* **227**, 284-292.

Cherfils, J., Duquerroy, S., and Janin, J. (1991) *Proteins Struct. Funct. Genet.* **11**, 271-280.

Chothia, C. (1984) *Ann. Rev. Biochem.* **53**, 537-572.

Chothia, C. and Finkelstein, A.V. (1990) *Ann. Rev. Biochem.* **59**, 1007-10039.

Christoserdova, L.V., and Lidstrom, M.E. (1994) *J. Bacteriol.* **176**, 7398-7404.

Cichowicz, D.J., and Shane, B. (1987) *Biochemistry* **26**, 513-521.

Coggins, J. R., and Hardie, D. G. (1986) in *Multidomain Proteins - Structure and Evolution* (Hardie, D. G., and Coggins, J. R., eds.) pp 1-12, Elsevier, Amsterdam.

Cohen, G.N., and Dautry-Varsat, A. (1980) in *Multifunctional Proteins* (Bisswanger, H., and Sceminke-Ott, E., eds.) pp 49-122, John Wiley and Sons, New York.

Cohen-Addad, C., Pares, S., Sieker, L., Neuberger, M., and Douce, R. (1995) *Nature Struct. Biol.* **2**, 63-68.

Coleman, P.F., Suttle, D.P., and Stark, G.R. (1977) *J. Biol. Chem.***252**, 6378-6385.

Cook, R.J., Lloyd, R.S., and Wagner, C. (1991) *J. Biol. Chem.* **266**, 4965-4973.

Cook, R.A., and Koshland, D.E. (1969) *Proc. Natl. Acad. Sci. U.S.A.* **64**, 247-254.

Cossins, E.A. (1984) in *Folates and Pterins: Chemistry and Biochemistry of Folates* (Blakley, R.L., and Benkovic, S.J., eds.) Vol. 1, pp. 1-60, John Wiley and Sons, New York.

Crawford, I., Clarke, M., van Cleemput, M., and Yanofsky, C. (1987) *J. Biol. Chem.* **262**, 239-244.

Creighton, T.E. (1988) *Biophys. Chem.* **31**, 155-162.

Creighton, T.E. (1993) *Proteins Structures and Molecular Properties*, Freeman, New York.

Cushley, W. (1986) in *Multidomain Proteins - Structure and Evolution* (Hardie, D. G., and Coggins, J. R., eds.) pp 13-53, Elsevier, Amsterdam.

Das, .B.K., Bhattacharyya, T., and Roy, S. (1995) *Biochemistry* **34**, 5242-5247.

Daubner, S.C., Lohse, D.L., and Fitzpatrick, P.F. (1993) *Protein Sci.* **2**, 1452-1460.

Daubner, S.C., Schrimsher, J.L., Schendel, F.Y., Young, M., Henikoff, S., Patterson, D., Stubbe, J., and Benkovic, S.J. (1985) *Biochemistry* **24**, 7059-7062.

Degani, Y., and Degani, C. (1980) *TIBS*. **5**, 337-341.

Deleage, G., and Roux, B. (1987) *Protein Eng.* **1**, 289-294.

Dill, K.A., and Shortle, D. (1991) *Ann. Rev. Biochem.* **60**, 795-825.

Dixon, K.H., Lanpher, B.C. Chiu, J., Kelley, K., and Cowan, K.H. (1994) *J. Biol. Chem.* **269**, 17-20.

Dobson, C.M., Evans, P.A., and Radford, S.E. (1994) *TIBS* **19**, 31-37.

Drury, E.J., Bazar, L.S., and MacKenzie, R.E. (1975) *Arch. Biochem. Biophys.* **169**, 662-668.

Drury, E.J., and Mackenzie, R.E. (1977) *Can. J. Biochem.* **55**, 919-923.

Dyson, H.J., and Wright, P.E. (1991) *Ann. Rev. Biophys. Biophys. Chem.* **20**, 519-538.

Eberhard, M., Tsai-Pflugfelder, M., Bolewska, K., Hommel, U., and Kirschner, K. (1995) *Biochemistry*, **34**, 5419-5428.

Evans, D.R., Bein, K., Guy, H.I., Liu, X., Molina, J.A., and Zimmerman, B.H. (1993) *Biochem. Soc. Trans.* **21**, 186-191.

Evans, J.S., Levine, B.A., Trayer, I.P.., Barr, G.C., Dorman, C.J., and Higgins, C.F. (1986) *FEBS Lett.* **208**, 211-216.

Feinberg, A.P., and Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266-267.

Fersht, A.R., (1993) *FEBS Lett.* **325**, 5-16.

Findlay, W.A. (1988) *Ph.D Thesis*, Department of Biochemistry, McGill University.

Findlay, W.A., and MacKenzie, R.E. (1987) *Biochemistry* **26**, 1948-1954.

Findlay, W.A., and MacKenzie, R.E. (1988) *Biochemistry* **27**, 3404-3408.

Findlay, W.A., Zarkadas, C.G., and MacKenzie, R.E. (1989) *Biochim. Biophys. Acta* **999**, 52-57.

Finer-Moore, J.S., Maley, G.F., Maley, F., Montfort, W.R., and Stroud, R.M. (1994) *Biochemistry* **33**, 15459-15468.

Fink, A.L. (1995) *Annu. Rev. Biophys. Struct.* **24**, 495-522.

Finkelstein, A.V., Gutun, A.M., and Badretdinov, A.Y. (1993) *FEBS Lett.* **325**, 23-28.

Flanagan, J.M., Kataoka, M., Fujisawa, and T. Engelman, D.M. (1993) *Biochemistry* **32**, 10359-1070.

Flanegan, J.B., and Greenberg, G.R. (1977) *J. Biol. Chem* **252**, 3019-3027

Freire, E., and Murphy, K.P. (1991) *J. Mol. Biol* **222**, 687-698

Freire, E., and Murphy, K.P., Sanchez-Ruiz, J.M., Galisteo, M.L., and Privalov, P.L. (1991) *Biochemistry* **31**, 250-256.

Frieden, C. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4413-4416.

Friedrich, P. (1985) in *Organized Multienzyme Systems* (Welch, G.R., ed.) pp. 141-176, Academic Press, Orlando.

Galaktos, N.G., and Walsh, C.T. (1987) *Biochemistry* **26**, 8475-8480.

Garcia-Martinez, L.F., and Appling, D.R. (1993) *Biochemistry* **32**, 4671-4676.

Garel, J-R. (1992) in *Protein Folding* (Creighton, T.E., ed.) pp. 405-454, Freeman, New York.

Garel, J-R., and Dautry-Varsat, A. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3379-3383.

Garnier, J., and Robson, R. (1989) in *Prediction of protein structure and the principles of protein conformation* (Fasman, G.D., ed.) pp 417-465, Plenum Press, New York

Garrow, T.A., Admon, A., and Shane, B. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9151-9155.

Gernert, K.M., Surles, M.C., Labean, T.H., Richardson, J.S., and Richardson, D.C. (1995) *Protein Sci.* **4**, 2252-2260.

Gerstein, M., Lesk, A.M., and Chothia, C. (1994) *Biochemistry* **33**, 6739-6749.

Gill, S.C., and von Hippel, P.H. (1989) *Anal. Biochem.* **182**, 319-326.

Gittelman, M.S., and Matthews, C.R. (1990) *Biochemistry* **29**, 7011-7020.

Goldberg, M.E., Semisotnov, G.V. Friguet, B., Kuwajima, K., Ptitsyn, O.B., Sugai, S. (1990) *FEBS Lett*, **263**, 51-56.

Goldenberg, D.P., and Creighton, T.E. (1983) *J. Mol. Biol.* **165**, 407-413.

Green, J.M., MacKenzie, R.E., and Matthews, R.G. (1988) *Biochemistry* **27**, 8014-8022.

Griko, Y.V., Venyaminov, S.Y., and Privalov PL. (1989) *Febs. Lett.* **244**, 276-278.

Grishen, N.V., and Phillips, M.A. (1994) *Protein Sci.* **3**, 2455-2458.

Grossman, S.H., Pyle, J., and Steiner, R.J. (1981) *Biochemistry*, **20**, 6122-6128.

Guy, H.I., and Evans, D.R. (1994) *J. Biol. Chem.* **269**, 23808-23816.

Halsted, C.H. (1989) *J. Amer. Coll. Nutr.* **8**, 650-658.

Hardie, D.G., and Coggins, J.R. (1986) in *Multidomain Proteins - Structure and Evolution* (Hardie, D.G., and Coggins, J.R., eds.) pp 333-334, Elsevier, Amsterdam.

Harrison, S.C., and Durbin, R. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4028-4030.

Hawkins, A.R., Moore, J.D., and Lamb, H.K. (1993) *Biochem. Soc. Trans.* **21**, 181-186.

Henikoff, S. (1987) *Methods Enzymol.* **155**, 156-165.

Hermann, R., Jaenicke, R., and Rudolph, R. (1981) *Biochemistry* **20**, 5195-5201.

Herold, M., and Kirschner, K. (1990) *Biochemistry* **29**, 1907-1913.

Herold, M., and Leistler, B. (1992) *FEBS Lett.* **308**, 26-29.

Herold, M., Leistler, B., Hage, A., Luger, K., and Kirschner, K. (1991) *Biochemistry* **30**, 3612-3620.

Horne, D.W. (1993) *Proc. Soc. Exp. Biol. Med.* **202**, 385-391.

Horne, D.W., Holloway, R.S., and Said, H.M. (1992) *J. Nutr.* **122**, 2204-2209.

Horne, D.W., Patterson, D., and Cook, R.J. (1989) *Arch. Biochim. Biophys.* **270**, 729-733.

Hubbard, S.J., and Argos, P. (1994) *Protein. Sci.* **3**, 2194-2206.

Hubbard, S.J., Gross, K-H., and Argos, P. (1994) *Protein Eng.* **7**, 613-626.

Hum, D.W., and MacKenzie, R.E. (1991) *Protein Eng.* **4**, 493-500.

Hum, D.W., Bell, A.W., Rozen, R., and MacKenzie, R.E. (1988) *J. Biol. Chem.* **263**, 15946-15950.

Huynh, T.V., Young, R.A., and Davis, R.W. (1985) in *DNA Cloning: A Practical Approach* (Glover, D.M., ed.) Vol. 1, pp. 49-78, IRL Press, Oxford.

Hyde, C.C., Ahmed, S.A. Padlan, E.A., Miles, E.W., and Davies, D.R. (1988) *J. Biol. Chem.* **263**, 17857-17871.

Ikeguchi, M., Kuwajima, K., Mitani, M., and Sugai, S. (1986) *Biochemistry* **25**, 6965-6972.

Inglese, J., Koch, J.K., Caron, M.G., and Lefkowitz, R.J. (1992) *Nature* **359**, 147-150.

Islam, S.A., Luo, J., and Sternberg, M.J.E. (1995) *Protein Eng.* **8**, 513-525.

Itzhaki, L.S., Evans, P.A., Dobson, C.M., and Radford, S.E. (1994) *Biochemistry* **33**, 5212-5220.

Jackson, R.C., and Niethammer, D. (1979) in *Chemistry and Biology of Pteridines* (Kisliuk, R.L., and Brown, G.M., eds.) pp. 665-670, Elsevier, New York.

Jaffe, J.J., Chrin, L.R., and Smith, R.B. (1980) *J. Parasitol.* **66**, 428-433.

Jaenicke, R. (1982) *Biophys Struct. Mech.* **8**, 231-256.

Jaenicke, R. (1987) *Prog. Biophys. Mol. Biol* **49**, 117-237.

Jaenicke, R. (1991) *Biochemistry* **30**, 3147-3161.

Janin, J., Miller, S., and Chothia, C. (1988) *J. Mol. Biol.* **204**, 155-161.

Janin, J., and Wodak, S.J. (1983) *Prog. Biophys. Molec. Biol.* **42**, 21-78.

Jecht, M., Tomschy, A., Kirschner, K., and Jaenicke, R. (1994) *Protein Sci.* **3**, 411-418.

Jencks, D.A., and Matthews, R.G. (1987) *J. Biol. Chem.* **262**, 2485-2493.

Jennings, P.A., Finn, B.E., Jones, B.E., and Matthews, C.R. (1993) *Biochemistry* **32**, 3783-3789.

Johnson, C.R., Morin, P.E., Arrowsmith, C.H., and Freire, E. (1995) *Biochemistry* **34**, 5309-5316.

Jones, B.E., Jennings, P.A., Pierre, R.A., and Matthews, C.R. (1994) *Biochemistry* **33**, 15250-15258.

Kamb, A., Finer-Moore, J., Calvert, A.H., and Stroud, R.M. (1992) *Biochemistry* **31**, 9883-9890.

Kantrowitz, E.R., and Lipscomb, W.N. (1988) *Science* **241**, 669-674.

Karplus, M., and Weaver, D.L. (1979) *Biopolymers* **18**, 1421-1437.

Khyse-Anderson, J. (1984) *J. Biochem. Biophys. Methods* **10**, 203-209.

Kiil, J., Jagerstad, M. and Elborg, L (1979) *Int. J. Vit. Nutr. Res.* **49**, 296-306

Kim, P.S., and Baldwin, R.L. (1990) *Ann Rev. Biochem.* **59**, 631-660.

Kirschner, K., and Bisswanger, H. (1976) *Ann. Rev. Biochem.* **45**, 143-166.

Knighton, D.R., Kan, C-C., Howland, E., Janson, C.A., Hostomska, Z., Welsh, K.M., and Matthews, D.A (1994) *Nature Struct. Biol.* **1**, 186-194.

Kneller, D.G., Cohen, F.E., and Langridge, R. (1990) *J. Mol. Biol.* **214**, 171-182.

Knowles, J.R. (1987) *Science* **236**, 1252-1258.

Korn, A.P., and Burnett, R.M. (1991) *Proteins Struct. Funct. Genet.* **9**, 37-55.

Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125-8148.

Krause, K.L., Volz, K.W., and Lipscomb, W.N. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1643-1647.

Krause, K.L., Volz, K.W., and Lipscomb, W.N. (1987) *J. Mol. Biol.* **193**, 527-533.

Kreimer, D.I., Shnyrov, Villar, E., Silman, I., and Weiner, L. (1995) *Protein Sci.* **4**, 2349-2357.

Kroczek, R.A., and Siebert, E. (1990) *Anal. Biochem.* **184**, 90-95.

Krumdieck, C.L., Eto, I., and Baggott, J.E. (1991) *Ann. N.Y. Acad. Sci.* **669**, 44-59.

Krupenko, S.A., Wagner, C., and Cook, R.J. (1995) *J. Biol. Chem.* **270**, 519-522.

Kruschwitz, H.L., McDonald, D., Cossins, E.A., and Schirch, V. (1994) *J. Biol. Chem.* **269**, 28757-28763.

Kurganov, B.I. (1993) *TIBS* **18**, 405-406.

Kutzbach, C., and Stokstad, E.L.R. (1971) *Biochim. Biophys. Acta* **250**, 459-477.

Kuwajima, K., Garvey, E.P., Finn, B.E., Matthews, C.R., and Sugai, S. (1991) *Biochemistry* **30**, 7693-7703.

Kuwajima, K., Hirokai, Y., Ikeguchi, M., and Sugai, S. (1985) *Biochemistry* **24**, 874-881.

Kuwajima, K., Mitani, M., and Sugai, S. (1989) *J. Mol. Biol.* **206**, 547-561.

Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., and Nagamura, T. (1987) *FEBS Lett.* **221**, 115-118.

Kunkel, T.A., Roberts, J.D., and Zakour R.A. (1987) *Methods Enzymol.* **154**, 367-382.

Laemmli, U.K. (1970) *Nature* **227**, 680-685.

Lakowicz, J.R. (1983) *Principles of Fluorescence Spectroscopy*, pp. 341-382, Plenum Press, New York.

Laurents, D.V., Subbiah, S., and Levitt, M. (1994). *Protein Sci.* **3**, 1938-1944.

Lawrence, M.C., and Colman, P.M. (1993) *J. Mol. Biol.* **234**, 946-950.

Leatherbarrow, R.J. (1987) *Enzfitter*, Biosoft, Cambridge, UK.

Lee, L., Kelly, R.E., Pastra-Landis, S.C., and Evans, D.R. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6802-6808.

Lehrer, S.S., and Qian, Y. (1990) *J. Biol.Chem.* **265**, 1134-1138.

Lesitler, B., and Perham, R.N. (1994) *Biochemistry* **33**, 2773-2781

Leslie, A.W. (1990) *J. Mol. Biol.* **213**, 167-186

Levinthal, C. (1968) *J. Chem. Phys.* **65**, 44-45

Levy, H.L. (1989) in *The Metabolic Basis of Inherited Disease, 6th edition* (Scriver, C.R., ed.) pp. 563-566. McGraw-Hill, New York.

Lewis, P.N., Go, N., Go, M. Kotelchuk, D., and Scheraga, H.A. (1970) *Proc. Natl. Acad. Sci. U.S.A.* **65**, 810-815.

Li, Y-J., Rothwarf, D.M., and Scheraga, H.A. (1995) *Nature Struct. Biol.* **2**, 489-494

Liljas, A., and Rossman, M.G. (1974) *Ann. Rev. Biochem.* **43**, 475-507

Lim, W.A., and Sauer, R.T. (1989) *Nature* **339**, 31-36.

Lin, B-F., and Shane, B. (1994) *J. Biol. Chem.* **269**, 9705-9713.

Lin, B-F., Huang, R.F.S., and Shane, B. (1993) *J. Biol. Chem.* **268**, 21674-21679

Lu, Y-Z., Aiello, P.D., and Matthews, R.G. (1984) *Biochemistry* **23**, 7166-7172.

Lu, J., and Dahlquist, F.W. (1992) *Biochemistry* **31**, 4749-4756.

Ludwig, M.L., Metzger, A.L., Pattridge, K.A., and Stallings, W.C. (1991) *J. Mol. Biol.* **219**, 335-358.

Lyle, S., Ozeran, D., Stanczak, J., Westley, J., and Schwartz, N.B. (1994b) *Biochemistry* **33**, 6822-6827.

Lyle, S., Stanczak, J., Ng, K., and Schwartz, N.B. (1994a) *Biochemistry* **33**, 5920-5925.

MacKenzie, R.E. (1979) in *Chemistry and Biology of Pteridines* (Kisliuk, R.L., and Brown, G.M, eds.) pp 443-446, Elsevier/North Holland, Amsterdam.

MacKenzie, R.E. (1980) *Methods Enzymol.* **66**, 626-630.

MacKenzie, R.E., Aldridge, M., and Paquin, J. (1980) *J. Biol. Chem.* **255**, 9474-9478.

MacKenzie, R.E., and Baugh, C.M. (1980) *Biochim. Biophys. Acta* **611**, 187-195.

Maras, B., Stover, P., Valiante, S., Barra, D., and Schirch, V. (1994) *J. Biol. Chem.* **269**, 18429-18433.

Marolewski, A., Smith, J.M., and Benkovic, S.J. (1994) *Biochemistry* **33**, 2531-2537.

Martel, A., and Garel, J-R. (1984) *J. Biol. Chem.* **259**, 4917-4921.

Matouschek, A., Serrano, L., Meiering, E.M., Bycroft, M., and Fersht, A.R. (1992) *J. Mol. Biol.* **224**, 837-845.

Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035-10038.

Mathews, C.K. (1991) *J. Theor. Biol.* **152**, 25-28.

Mattevi, A., Obmolova, G., Schulze, E., Kalk, K.H., Westphal, A.H., de Kok, A., and Hol, W.G.J. (1992) *Science* **255**, 1544-1550.

Matthews, B.W. (1993) *Ann. Rev. Biochem.* **62**, 139-160.

Matthews, C.R. (1993) *Ann. Rev. Biochem.* **62**, 653-683.

Matthews, R.G., (1984) in *Folates and Pterins: Chemistry and Biochemistry of Folates* (Blakley, R.L., and Benkovic, S.J., eds.) Vol. 1, pp. 497-554, John Wiley and Sons, New York.

Matthews, R.G., and Baugh, C.M. (1980) *Biochemistry* **19**, 2040-2045.

Matthews, R.G., Lu, Y-Z., Green, J.M., and MacKenzie, R.E. (1985) in *Proceedings of the Second Workshop on Folyl and Antifolyl Polyglutamates* (Goldman, I.D., ed.) pp. 65-75, Praeger Publishers, U.S.A.

Maurizi, M.R., and Ginsburg, A. (1982) *J. Biol. Chem.* **257**, 7246-7251.

Mayr, E-M., Jaenicke, R., and Glockshuber, R. (1994) *J. Mol. Biol.* **235**, 84-88.

McLain, L.D., Carl, G.F., and Bridgers, W.F. (1975) J. Neurochem. 24, 719-722.

Mejia, N.R., Rios-Orlandi, E.M., and MacKenzie, R.E. (1986) *J. Biol. Chem.* **261**, 9509-9513.

Miller, S. (1989) *Protein Eng.* **3**, 77-83.

Minard, P., Hall, L., Missiakas, D., and Yon, J.M. (1989) *Protein Eng.* **3**, 55-60.

Miranker, A., Radford, S.E., Karplus, M., and Dobson, C.M. (1991) *Nature* **349**, 633-636.

Missiakis, D., Betton, J-M., Minard, P., and Yon, J.M. (1990) *Biochemistry* **29**, 8683-689.

Montgomery, D., Jorden, R., McMacken, R., and Freire, E. (1993) *J. Mol. Biol.* **232**, 680-692.

Mueller, W.T., and Benkovic, S.J. (1981) *Biochemistry* **20**, 337-344.

Murase, S., Kawata, Y., and Yumoto, N. (1993) *J. Biochem.* **114**, 393-397.

Murley, L.L., and MacKenzie, R.E. (1995) *Biochemistry* **34**, 10358-10364.

Murley, L.L., Mejia, N.R., and MacKenzie, R.E. (1993) *J. Biol. Chem.* **268**, 22820-22824.

Murray-Brelier, A., and Goldberg, M (1989) *Proteins Struct. Funct. Genet.* **6**, 395-404.

Nada, M.A., Rhead, W.J., Sprecher, H., Schulz, H., and Roe, C.R. (1995) *J. Biol. Chem.* **270**, 530-535.

Nagy, P.L., Marolewski, A., Benkovic, S.J., and Zalkin, H. (1995) *J. Bacteriol.* **177**, 1292-1298.

Nagy, P.L., McCorkle, G.M., and Zalkin, H. (1993) *J. Bacteriol.* **175**, 7066-7073.

Neet, K.E., and Timm, D.E. (1994) *Protein Sci.* **3**, 2167-2174.

Neuberger, M., Jourdain, A., and Douce, R. (1991) *Biochem. J.* **278**, 765-769.

Nordhoff, A., Bücheler, U.S., Werner, D., and Schirmer, R.H. (1993) *Biochemistry* **32**, 4060-4066.

Nour, J.M., and Rabinowitz, J.C. (1992) *J. Biol. Chem.* **267**, 16292-16296.

Nour, J.M., and Rabinowitz, J.C. (1991) *J. Biol. Chem.* **266**, 18363-18369.

Nygard, P., and Smith, J.M. (1993) *J. Bacteriol.* **175**, 3591-3597.

Ohgushi, M., and Wasa, A. (1983) *FEBS Lett.* **164**, 21-24.

Olins, P.O. and Rangwala, S.H. (1989) *J. Biol. Chem.* **264**, 16973-16976.

Ovádi, J. (1991) *J. Theor. Biol.* **152**, 1-22.

Owers Narhi, L., Rosenfeld, R., Wen, J., Arakawa, T., Prestrelski, S.J., and Philo, J.S (1993) *Biochemistry* **32**, 10819-10825.

Pace, C.N. (1986) *Methods Enzymol.* **131**, 266-280.

Pakula, A.A., and Sauer, R.T. (1989) *Proteins. Struct. Funct. Genet.* **5**, 202-210.

Palleros, D.R., Shi, L., Reid, K.L., and Fink, A.L. (1993) *Biochemistry* **32**, 4314-4321.

Paquin, J., Baugh, C.M., and MacKenzie, R.E. (1985) *J. Biol. Chem.* **260**, 14925-14931.

Parr, G.R., and Hammes, G.G. (1975) *Biochemistry* **14**, 1600-1608.

Patel, M.S., and Roche, T.E. (1990) *FASEB J.* **4**, 3224-3233.

Pearson, W.R. and Lipman, D.J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444-2448.

Pecorari, F., Minard, P., Desmadril, M., and Yon, J.M. (1993) *Protein Eng.* **6**, 313-325.

Pelletier, J.N., and MacKenzie, R.E. (1995) *Biochemistry* **34**, 12673-12680.

Peng, Z. and Kim, P.S. (1994) *Biochemistry* **33**, 2131-2141.

Perham, R.N. (1991) *Biochemistry* **30**, 8501-8511.

Perham, R.N., and Packman, L.C. (1989) *Ann. N.Y. Acad. Sci.* **573**, 531-538.

Peri, K.G., and MacKenzie, R.E. (1991) *FEBS Lett.*, **294**, 113-115.

Pen, K.G., and MacKenzie, R.E. (1993) *Biochim. Biophys. Acta* **1171**, 281-287.

Pfeil, W., Bychkova, V.E., and Ptitsyn, O.B. (1986) *FEBS Lett.* **189**, 287-291.

Pfeil, W., and Privalov, P.L. (1976) *Biophys. Chem.* **4**, 33-41.

Privalov, P.L. (1979) *Adv. Prot. Chem.* **33**, 167-239.

Privalov, P.L. (1982) *Adv. Prot. Chem.* **35**, 1-104.

Privalov, P.L. (1989) *Ann. Rev. Biophys. Biophys. Chem.* **50**, 497-532.

Privalov, P.L. (1992) in *Protein Folding* (Creighton, T.E., ed.) pp. 83-126, Freeman, New York.

Privalov, P.L., and Potekhin, S.A. (1986) *Methods Enzymol.* **134**, 4-51.

Ptitsyn, O.B. (1992) in *Protein Folding* (Creighton, T.E., ed.) pp. 243-300, Freeman, New York.

Ptitsyn, O.B., Bychkova, V.N., and Uversky, V.N. (1995) *Phil. Trans. R. Soc. Lond. B.* **348**, 35-41.

Ptitsyn, O.B., Pain, R.H., Semisotnov, G.V., Zerovnik, E., and Razgulyaev, O.I. (1990) *FEBS Lett.* **262**, 20-24.

Ptitsyn, O.B., and Uversky, V.N. (1994) *FEBS Lett.* **341**, 15-18.

Rao, S.T., and Rossman, M.G. (1973) *J. Mol. Biol.* **76**, 241-256.

Rabinowitz, J.C., and Pricer, W.E. (1956b) *J. Am. Chem. Soc.* **78**, 5702-5704.

Rabinowitz, J.C., and Pricer, W.E. (1956a) *J. Am. Chem. Soc.* **78**, 4176-4178.

Rabinowitz, J.C., and Tabor, H. (1958) *J. Biol. Chem.* **233**, 252-255.

Radford, S.E., Dobson, C.M., and Evans, P.A. (1992) *Nature* **385**, 302-307.

Radford, S.E., Laue, E.D., Perham, R.N., Martin, S.R., and Apella, E. (1989) *J. Biol. Chem.* **264**, 767-775.

Rankin, C.A., Haslam, G.C., and Himes, R.H. (1993) *Protein Sci.* **2**, 197-205.

Rashin, A.A. (1981) *Nature* **291**, 85-87.

Rashin, A.A., Iofin, M., and Honig, B. (1986) *Biochemistry* **25**, 3619-3625.

Ratanasthien, K., Blair, J.A., Leeming, R.J., Cooke, W.T., and Melikian, V. (1974) *J. Clin. Path.* **27**, 875-879.

Richardson, J.S. (1981) *Adv. Prot. Chem.* **34**, 167-339.

Reddy, G.P., and Fager, R.S. (1993) *Crit. Rev. Euk. Gene Replic.* **3**, 255-277.

Reddy, G.P.V., and Pardee, A.B. (1980) *Proc. Natl. Acad. Sci.* **77**, 283-437.

Redeker, V., Le Caer, J-P., Rossier, J., and Prome, J-C. (1991) *J. Biol. Chem.* **266**, 23461-23466.

Reed, L.J. (1974) *Acc. Chem. Res.* **7**, 40-46.

Reidhaar-Olsen, J.F., and Sauer, R.T. (1988) *Science* **241**, 53-57.

Risse, B., Stemfer, G., Rudolph, R., Möllering, H., and Jaenicke, R. (1992) *Protein Science* **1**, 1699-1709.

Roder, H., Elöve, G., and Englander, S.W. (1988) *Nature* **335**, 700-704.

Rose, G.D., and Roy, S. (1980) *Proc .Natl. Acad. Sci. U.S.A.* **77**, 4643-4647.

Rosenblatt, D.S. (1989) in *The Metabolic Basis of Inherited Disease, 6th edition.* (Scriver, C.R., ed.) pp. 2049-2064, McGraw-Hill, New York.

Rossman, M.G., and Argos, P. (1981) *Annu. Rev. Biochem.* **50**, 497-532.

Rudolph, R., Nesslauer, G., Siebendritt, R., Sharma, A.K., and Jaenicke, R. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4625-4629.

Rudolph, R., and Jaenicke, R. (1976) *Eur. J. Biochem.* **63**, 409-417.

Rost, B., and Sander, C. (1993) *J. Mol. Biol.* **232**, 584-599.

Russel, R.B. (1994) *Protein Eng.* **7**, 1407-1410.

Sacchetta, P., Aceto, A., Bucciarelli, T., Dragani, B., Santarone, S., Allocati, N., and Di Ilio, C. (1993) *Eur. J. Biochem.* **215**, 741-745.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.

Sancho, J. and Fersht A.R. (1992) *J. Mol. Biol.* **224**, 741-747.

Sanger, F., Nicklen, S., and Coulsen, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467.

Schepartz, B. (1973) in *Regulation of Amino Acid Metabolism in Mammals.* Saunders, Philadelphia.

Schevitz, R.W., Otwinowski, Z., Joachmiak, A., Lawson, C.L., and Sigler, P.B. (1985) *Nature* **317**, 782-786.

Schirch, D., Villar, E., Maras, B., Barra, D., and Schirch, V. (1994) *J. Biol. Chem.* **269**, 24728-24735.

Schirch, V., and Strong, W.B. (1989) *Arch. Biochem. Biophys.* **269**, 371-380.

Schirmer, T., and Evans, P.R. (1990) *Nature* **343**, 140-145.

Schlindler, T., Herrier, M., Marahiel, M.A. Schmid, F.X. (1995) *Nature Struct. Biol.* **2**, 663-673.

Schumann, J., and Jaenicke, R. (1993) *Eur. J. Biochem.* **213**, 1225-1233.

Scrutton, N.S., Deonarain, M.P., Berry, A., Perham, R.N. (1992) *Science* **258**, 1140-1143.

Semistotnov, G.V., Rodinova, N.A., Razglyaev, O.I., Uversky, V.N., Gripas, A.F., and Gilmanshin, R.I. (1991) *Biopolymers* **31**, 119-128.

Shane, B. (1989) *Vitam. Horm.* **45**, 263-335.

Shane, B., and Stokstad, E.L.R. (1984) in *Folates and Pterins: Chemistry and Biochemistry of Folates* (Blakley, R.L., and Benkovic, S.J., eds.) Vol. 1, pp. 433-455, John Wiley and Sons, New York.

Shane, B., and Stokstad, E.L.R. (1985) *Ann. Rev. Nutr.* **5**, 115-141.

Shannon, K.W., & Rabinowitz, J.C. (1988) *J. Biol. Chem.* **263**, 7717-7725.

Sharma, A.K., Minke-Gogl, V., Gohl, P., Siebendritt, R., Jaenicke, R., and Rudolph, R. (1990) *Eur J. Biochem.* **194**, 603-609.

Shastry, M.C., and Udgaonkar, J.B. (1995) *J. Mol. Biol.* **247**, 1013-1027.

Shen, H., Schmuck, M., Pilz, I., Gilkes, N.R., Kilburn, D.G., Miller, R.C., and Warren, R.A.J. (1991) *J. Biol. Chem.* **266**, 11335-11340.

Shortle, D., and Abergunawardana, C. (1993) *Curr. Biol.* **1**, 121-134.

Shortle, D., and Meeker A.K. (1986) *Proteins. Struct. Funct. Genet.* **1**, 81-89.

Shortle, D., and Meeker, A.K. (1989) *Biochemistry* **28**, 936-944.

Shortle, D., Stites,W.E., and Meeker, A.K. (1990) *Biochemistry* **29**, 8033-8041.

Simmer, J.P., Kelly, R.E., Scully, J.L., Grayson, D.R., Rinker, A.G., Bergh, S.T., and Evans, D.R. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4382-4386.

Srere, P.A. (1985) in *Organized Multienzyme Systems* (Welch, G.R., ed.) pp. 1-61, Academic Press, Orlando.

Srere, P.A. (1987) *Ann. Rev. Biochem.* **56**, 89-124.

Sternlicht, H., Yaffee, M.B., and Farr, G. (1987) *FEBS Lett.* **214**, 226-235.

Stifel, F.B., Taunton, O.D., Greene, H.L., Lufkin, E.G., Hagler, L., and Herman, R.H.

(1974) *Biochim. Biophys. Acta* **354**, 194-205.

Stoops, J.K., Kolodzieu, S.J., Schroeter, J.P., Bretaudiere, J-P., and Wakil, S.J. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6585-6598.

Stover, P., and Schirch, V. (1990) *J. Biol. Chem.* **265**, 14227-14233.

Stover, P., and Schirch, V. (1993) *TIBS* **18**, 102-106.

Stroud, R. M. (1994) *Nature Struct. Biol.* **1**, 131-134.

Sugawara, T., Kujiwama, K., and Sugai, S. (1991) *Biochemistry* **30**, 2698-2706.

Sun, A-Q., Ümit Yüksel, K., and Gracy, R.W. (1992) *J. Biol. Chem.* **267**, 20168-20174.

Tabor, H., and Rabinowitz, J.C. (1956) *J. Am. Chem. Soc.* **78**, 5705-5706.

Tabor, H., and Wyngarden, L. (1959) *J. Biol. Chem.* **234**, 1830-1846.

Tabor, S. and Richardson, C.C. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1074-1078.

Tanford, C. (1968) *Adv. Prot. Chem.* **23**, 121-282.

Tasayco, M.L., and Carey, J. (1992) *Science* **255**, 594-597.

Thigpen, A.E., West, M.G., and Appling, D.R. (1990) *J. Biol. Chem.* **265**, 7907-7913.

Thornton, J.M., Jones, D.T., MacArthur, M.W., Orengo, C.M., Swindells, M.B. (1995) *Phil. Trans. R. Soc. Lond. B.* **348**, 71-79.

Thornton, J.M., and Sibanda, B.L. (1983) *J. Mol. Biol.* **167**, 443-460.

Tendian, S.W., Myszka, D.G., Sweet, R.M., Chaiken, I.M., and Brouillette, C.G. (1995) *Biochemistry* **34**, 6464-6474.

Teschner, W., Rudolph, R., and Garel, J-R. (1987) *Biochemistry* **26**, 2791-2796.

Toyama, H., Esaki, N., Yoshimura, K., Tanizawa, K., and Soda, K. (1991) *J. Biochem.* **110**, 279-283.

Traut, T.W. (1986) *Mol. and Cell Biochem.* **70**, 3-10.

Tremblay, G. (1995) *Ph.D Thesis*, Department of Biochemistry, McGill University.

Turner, S.L., Russel, G.C., Williamson, M.P., and Guest, J.R. (1993) *Protein Eng.* **6**, 101-108.

Udgaonkar, J.B., and Baldwin, R.L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8197-8201.

Ullman, A., Jacob, F., and Monod, J. (1967) *J. Mol. Biol.* **24**, 339-343.

Usha, R., Savithri, H.S., and Rao, N.A. (1992) *J. Biol. Chem.* **267**, 9289-9293.

Uversky, V.N., and Ptitsyn, O.B. (1994) *Biochemistry* **33**, 2782-2791.

Uyeda, K., and Rabinowitz, J.C. (1965) *J. Biol. Chem.* **240**, 1701.

Uyeda, K., and Rabinowitz, J.C. (1967) *J. Biol. Chem.* **242**, 24-31.

Waddle, J.J., Johnston, T.C., and Baldwin, T.O. (1987) *Biochemistry* **26**, 4917-4921.

Wahls, W.P., Song, J.M., and Smith, G.R. (1993) *J. Biol. Chem.* **268**, 23792-23798.

Wakil, S.J. (1989) *Biochemistry* **28**, 4523-4530.

Wakil, S.J., Joshi, V.C., and Stoops, J.K. (1983) *Ann. Rev. Biochem.* **52**, 537-539.

Welch, G.R., and Gaertner, F.H. (1980) *Curr. Top. Cell. Regul.* **16**, 113-162.

Welch, G.R., and Marmillot, P.R. (1991) *J. Theor. Biol.* **152**, 29-33.

Wetlaufer, D.B. (1973) *Proc. Natl. Acad, Sci., U.S.A.* **70**, 697-701.

Wetlaufer, D.G. (1981) *Adv Prot. Chem.* **34**, 61-92.

Whitehead, T.R., and Rabinowitz, J.C. (1988) *J. Bacteriol.* **170**, 3255-3261.

Williams, F.M.R., Murray, R.C., Underhill, T.M., and Flinthoff, W.F. (1994) *J. Biol. Chem.* **269**, 5810-5816.

Wilmanns, M., Priestle, J.P., Niermann, T., and Jansonius, J.N. (1992) *J. Mol. Biol.* **223**, 477-507.

Witkowski, A., Rangan, V.S., Randhawa, Z.I., Amy, C.M., and Smith, S. (1991) *Eur. J. Biochem.* **198**, 571-579.

Wodak, S.J., and Janin, J. (1981) *Biochemistry* **20**, 6544-6552.

Wolynes, P.G., Onuchic, J.N., and Thirumalai, D. (1995) *Science* **267**, 1619-1620.

Wong, S.C., Proefke, S.A., Bhushan, A., and Matherly, L.H. (1995) *J. Biol. Chem.* **270**, 17468-17475.

Wu, L.C., Peng, Z., and Kim, P.S. (1995) *Nature Struct. Biol.* **2**, 281-286.

Xanthoudakis, S., Miao, G.G., and Curran, T. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 23-27.

Yang, X-M., and MacKenzie, R.E. (1993) *Biochemistry* **32**, 11118-11123.

Yang, Y.R., and Schachman, H.K. (1993a) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11980-11984.

Yang, Y.R., and Schachman, H.K. (1993b) *Protein Sci.* **2**, 1013-1023.

Yao, Q-Z., Tian, M. and Tsou, C-L. (1984) *Biochemistry* **23**, 2740-2744.

Young, L., Jernigan, R.L., and Covell, D.G. (1994) *Protein Sci.* **3**, 717-729.

Zeigler, M.M., Goldberg, M.E., Chaffotte, A.F., and Baldwin, T.O. (1993) *J. Biol. Chem.* **268**, 10760-10765.

Zhang, T., Bertelsen, E., Benvegnu, D., and Alber, T. (1993) *Biochemistry* **32**, 12311-12318.

APPENDIX A

THE PORCINE FTCD mRNA IS LIVER SPECIFIC AND PRESENT AS A SINGLE SPECIES

The purpose of this appendix is to document the Northern blot analysis described in Chapter 2

INTRODUCTION

Formiminotransferase-cyclodeaminase activity has been identified in the liver and kidney of all mammalian species tested (Tabor & Wyngarden, 1959). The transferase activity has not been detected in insects, bacteria and yeast, however both activities have been identified in filaria (Jaffe et al., 1980) and an avian homolog has been identified in chicken liver (Hennig, personal communication, this thesis, Chapter 5).

Northern analysis was performed prior to isolation of the full length cDNA in order to determine if the putative partial cDNA hybridized to a liver specific message, and to estimate the size of the full length clone.

MATERIALS AND METHODS

Materials. DNA modification enzymes were obtained from Bethesda Research Laboratories, Boehringer Mannheim, Pharmacia or New England Biolabs. Nylon membranes used for Northern blot analysis were from Amersham Corporation as were radioisotopically labelled nucleotides. Diagnostic Xomat AR film was purchased from Kodak. Oligo dT cellulose was from Pharmacia. All other chemicals used were of reagent grade.

All cDNA probes were labelled by the random primer method using [α-^{32}P]dCTP as the labelled nucleotide (Feinberg & Vogelstein, 1984). Unless otherwise noted, protocols used were as outlined in Sambrook et al. (1989).

RNA isolation. Porcine liver was removed during surgical procedures and immediately frozen in liquid nitrogen. Liver total RNA was isolated using the urea/LiCl method (Auffray & Rougeon, 1980). Poly A+ mRNA was isolated by two passages over oligo dT cellulose, essentially as described (Aviv & Leder,1971).

Northern Analysis. Total and messenger RNA were electrophoresed on 1.25% formaldehyde agarose gels in MOPS Buffer and vacuum blotted onto Hybond N membranes (Kroczek & Siebert, 1990). Hybridization was carried out overnight at 42° in hybridization buffer containing 50% formamide, 5 X SSPE (1 x SSPE = 0.15 M NaCl, 10 mM NaH_2PO_4, 1.3M EDTA, pH 7.4), 5% Denhardt's solution, 1% dextran sulfate, 50 mM sodium phosphate, 0.5% SDS, 100 ug/ml denatured salmon sperm DNA and 2.5x 10^6cpm/ml of cDNA probe derived from the EcoR1 insert of FT2e. Membranes were washed once in 1 X SSC, 0.5% SDS, 42°C for 20 minutes; once in 0.1 X SSC, 0.5% SDS, 42°C for 20 minutes; once in 0.1 X SSC, 0.5% SDS, 55°C for 20 minutes. The mouse tissues mRNA blot shown in Figure A.2 had previously been used by K. Peri (Peri and MacKenzie, 1991), and was stripped of any remaining signal, prior to hybridization with the FT2e probe.

RESULTS AND DISCUSSION

The largest partial cDNA (FT2e) isolated from the λgt11 porcine liver cDNA library, was only 1.1 kb and was missing sequences from both the 5' and 3' ends. However, this fragment could be positively identified as encoding FTCD because the deduced amino acid sequence included sequences identical to those in the porcine enzyme, as determined by N-terminal sequencing of cyanogen bromide fragments of FTCD (Chapter 2). This cDNA hybridized to a 1.9 kb message on Northern Blots of porcine liver RNA (Figure A.1). This is the same size as the longest cDNAs later isolated from the λgt10 porcine liver cDNA library, suggesting that these latter cDNAs are approximately full length. Only a single FTCD transcript is observed in pig liver.

Northern analysis of mRNA from different mouse tissues indicates that this transcript is liver specific (Figure A.2), as expected. FTCD message was not detected in other tissues, even kidney suggesting that it is present at very low levels, if at all in these tissues. Interestingly, the FTCD cDNA also hybridizes to a second liver specific transcript of approximately 3 kb, suggesting that a related transcript may be present in mouse liver.

FIGURE A.1. Northern blot analysis of porcine liver RNA. **A.** Porcine liver polyA+ mRNA (2 μg) and **B.** Porcine liver total RNA (20 μg) was electrophoresed on a 1.25% formaldehyde-agarose gel in MOPS buffer and vacumn blotted onto Hybond N. Hybridization with the ^{32}P-labelled partial cDNA, FT2e, was performed as described under *Materials and Methods*, and radioactivity was detected by fluorography.

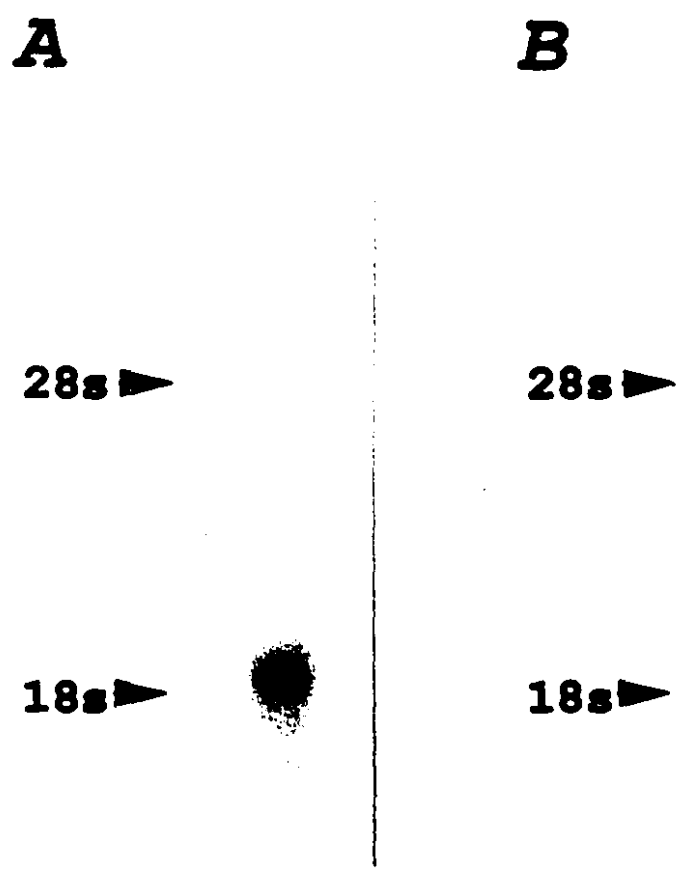
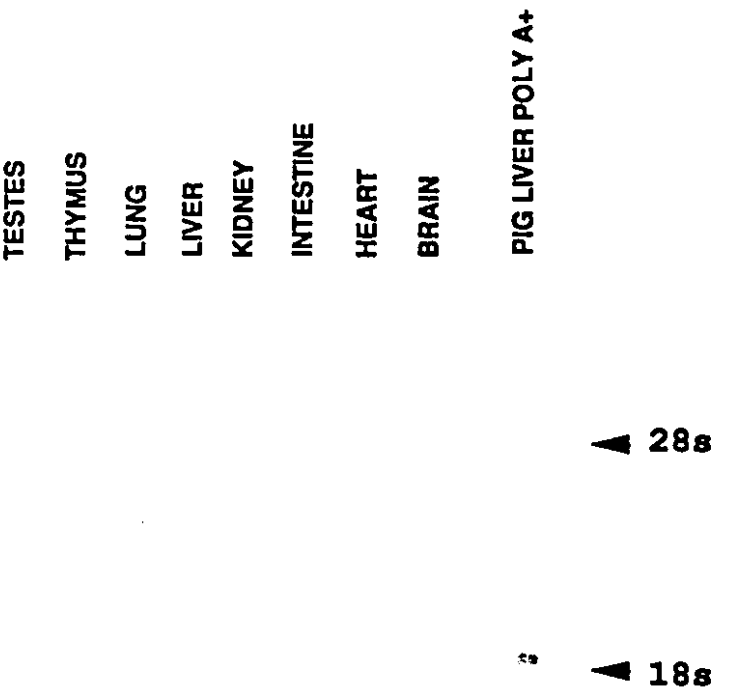

A
B
28s
28s
18s
18s

FIGURE A.2. Tissue distribution of FTCD mRNAs. Each lane contains 2 μg of polyA$^+$ mRNA isolated from different mouse tissues as indicated or 2 μg of polyA$^+$ mRNA from pig liver. Hybridization with the ^{32}P-labelled partial cDNA, FT2e, was performed as described under *Materials and Methods*, and radioactivity was detected by fluorography.

MOUSE TISSUES TOTAL MRNA



ACKNOWLEDGEMENTS

The mouse tissues mRNA blot was a generous gift from K. Peri. Porcine liver was kindly provided by Dr. Gamal Selmy of the Royal Victoria Hospital, Department of Urology.