INDEPENDENT DOMAINS OF A TRIFUNCTIONAL ENZYME

by

Dean W. Hum

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

Department of Biochemistry
McGill University
Montréal, Québec (Canada)

September, 1990

DEDICATION

To my father and in memory of my mother,

for their love and encouragement.

ABSTRACT

The human folate-dependent trifunctional enzyme methyleneH_4folate deydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase catalyzes three sequential reactions in the interconversion of one-carbon folate derivatives. In its native conformation the enzyme is composed of an amino-terminal dehydrogenase/cyclohydrolase domain and a larger carboxyl-terminal synthetase domain. A cDNA clone encoding the human trifunctional enzyme was isolated and characterized. Sequence analysis yielded the complete deduced primary structure of 935 amino acid residues, and revealed the general location of a putative linker region joining the two major functional domains of the enzyme. The human cDNA clone, when expressed in *E. coli* can produce the active full length trifunctional protein. As well, the two major domains of the human protein can be expressed as independent functional enzymes. A rapid protocol was derived to purify the full length enzyme and the dehydrogenase/cyclohydrolase domain, when over-expressed in *E. coli*. In the native enzyme, the two functional domains are separated by an interdomain region of 19 residues which are not essential for catalytic activity. The human trifunctional enzyme and the two major domains can separately rescue a *Saccharomyces cerevisiae ade3* mutant.

RESUME

L'enzyme humain trifonctionnel méthylènetétrahydrofolate déshydrogénase-méthènyltétrahydrofolate cyclohydrolase-formyltétrahydrofolate synthétase catalyse trois réactions consécutives dans l'interconversion des dérivés folates porteurs d'une unité de carbone. Dans son état natif, l'enzyme est composé de deux domaines majeurs: un premier domaine comprend les activités déshydrogénase et cyclohydrolase et se trouve dans la partie amino-terminale alors qu'un second domaine, présentant une masse moléculaire plus grande, contient l'activité synthétase et est situé dans la partie carboxy-terminale. Un clone d'ADNc codant pour cet enzyme humain trifonctionnel a été isolé et charactérisé. L'analyse de sa séquence a permis de déduire que la structure primaire de cet enzyme comprend 935 acides aminés. De plus, il a été possible de localiser un "linker" potentiel joignant les deux domaines majeurs de l'enzyme. L'expression de l'ADNc d'origine humaine chez *E. coli* produit une protéine trifonctionnelle qui est active. Les deux domaines majeurs de la protéine humaine peuvent aussi être exprimés individuellement chez *E. coli*. Un protocole expérimental a été établi pour purifier l'enzyme trifonctionnel ainsi que le domaine déshydrogénase-cyclohydrolase lorsqu'ils sont surexprimés chez *E. coli*. Une région de 19 acides aminés comprise entre les deux domaines majeurs s'est révélée non essentielle à l'activité catalytique de l'enzyme. L'enzyme trifonctionnel ainsi que les deux domaines majeurs peuvent complémenter un mutant Saccharomyces cerevisiae *ade3*.

TABLE OF CONTENTS

FOREWORD

Parts of this thesis include the text of original papers submitted for publication. In compliance with the Faculty of Graduate Studies and Research "Guidelines concerning Thesis Preparation", the text of section 7 entitled "Manuscripts and authorship" is cited below:

"The candidate has the option, subject to the approval of the Department, of including as part of the thesis the text, or duplicated published text (see below), of an original paper, or papers. In this case the thesis must still conform to all other requirements explained in Guidelires Concerning Thesis Preparation. Additional material (procedural and design data as well as descriptions of equipment) must be provided in sufficient detail (e.g. in appendices) to allow a clear and precise judgement to be made of the importance and originality of the research reported. The thesis should be more than a mere collection of manuscripts published or to be published. It must include a general abstract, a full introduction and literature review and a final overall conclusion Connection texts which provide logical bridges between different manuscripts are usually desirable in the interests of cohesion.

It is acceptable for theses to include as chapters authentic copies of papers already published, provided these are duplicated clearly on regulation thesis stationery and bound as an integral part of the thesis. Photographs or other materials which do not duplicate well must be included in their original form. In such instances, connecting texts are mandatory and supplementary explanatory material is almost always necessary.

The inclusion of manuscripts co-authored by the candidate and others

is acceptable but the candidate is required to make an explicit statement on who contributed to such work and to what extent, and supervisors must attest to the accuracy of the claims, e.g. before the Oral Committee. Since the task of the Examiners is made more difficult in these cases, it is in the candidate's interest to make the responsibilities of authors perfectly clear. Candidates following this option must inform the Department before it submits the thesis for review.

This format for thesis preparation has been approved by the Department of Biochemistry, McGill University. Each chapter has its own numeration of figures and tables. All references are compiled at the end of the thesis. A single list of abbreviations has been prepared and is included following the list of Tables.

ACKNOWLEDGEMENTS

I would like to express my most profound gratitude to Dr. Robert E. MacKenzie who has patiently guided me throughout the course of my work in his laboratory. I greatly appreciate his always taking the time to discuss problems and questions, be it scientific or otherwise As well, I appreciate his good sense of humour which, amongst other things, stimulated lively and productive discussion, and provided a most enjoyable working environment in the lab. His profound knowledge of scientific concepts, and expert criticisms of experimental results have been most helpful and have helped me to mature as a scientist. As well, I thank him for his critical and diligent proof-reading of my thesis. Bob MacKenzie has been for me both a friend and a mentor who has shown that to be a first-rate scientist need not prevent one from also being a kind and generous person.

I thank the past and present members of Dr. MacKenzie's laboratory for providing a most enjoyable and collaborative working atmosphere, and for their camaraderie. I am grateful to Carole Bélanger, Xiao-Ming Yang and Drs. Ethel Rios-Orlandi, Wendy Findlay and Krishna Peri for the many hours of helpful, informative and intellectually stimulating discussions. I am indebted to Narciso Mejia for preparing an exceptional polyclonal antiserum which was used to initially isolate a cDNA clone As well, I thank him for always promptly providing us with needed materials and for maintaining the lab in working order. I thank the newer members, Gilles Tremblay, Joelle Pelletier and Laura Lea Murley for proof-reading sections of my thesis, and for adding a most enjoyable new and refreshing dimension to the lab. I thank Carole Bélanger for the french translation of the

thesis abstract and for typing chapters 1 and 5.

I thank the students and faculty members of the Biochemistry Department, many of whom have helped me in numerous aspects. I very much appreciated the collaborative atmosphere in the department and the willingness of most people to discuss scientific problems. I especially thank Dr. Nicole Beauchemin for providing the lambda gt11 library; Marc Laroche and Dr. Howard Bussey (Biology Dept.) for technical assistance with the yeast work; and Drs. William Scheffield, Isaac Edery , Jerry Pelletier and Michael Boylan for their constructive and generous technical advice. As well, I thank Kathy Teng for photography and Gisele de Souza for typing of manuscripts.

Finally, I would like to express my gratitude to my fiancée, friends and family for their continued love, support and encouragement. I am especially indebted to my sister Sharon, my brother Frank and my friends Isabel Rambaldi and Robert Esguerra.

PUBLICATION OF THE WORK PRESENTED IN THIS THESIS

Primary Structure of a Human Trifunctional Enzyme: Isolation of a cDNA Encoding Methylenetetrahydrofolate Dehydrogenase-Methenyl-tetrahydrofolate Cyclohydrolase-Formyltetrahydrofolate Synthetase. Hum, D.W., Bell, A.W., Rozen, R., and MacKenzie, R.E. (1988) *The Journal of Biological Chemistry* 263, 15946-15950.

Expression of Active Domains of a Human Folate-Dependent Trifunctional Enzyme in *Escherichia coli*. Hum, D.W., and MacKenzie, R.E. (1990) submitted to *Protein Engineering*.

Complementation of a Yeast *ade3* Mutant by the Functional Domains of a Human Trifunctional Enzyme. Hum, D.W., and MacKenzie, R.E. (in preparation).

CONTRIBUTIONS TO ORIGINAL KNOWLEDGE

This thesis deals with the human folate-dependent trifunctional enzyme methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase, studied using a molecular biology approach.

1. A polyclonal antiserum specific for the porcine trifunctional enzyme cross-reacts with the rat and human enzymes, and was used to isolate a cDNA encoding the human trifunctional protein. This was the first mammalian dehydrogenase-cyclohydrolase-synthetase enzyme to have been cloned. Northern analysis demonstrated that the human enzyme is encoded by a single transcript of approximately 3.1 kb.
2. Sequence analysis revealed the complete deduced primary structure of the protein which contains 935 amino acid residues, and allowed for the first time, sequence comparison with the enzyme from *Saccharomyces cerevisiae*. The primary structure of the human enzyme is homologous with that of the yeast enzyme. The two enzymes share an overall sequence identity of 58% although, the amino-terminal dehydrogenase/cyclohydrolase domain (48% identical) is less homologous than thecarboxyl-terminal synthetase domain (63% identical). Sequence analysis identified the general location of a putative linker region, and the synthetase domain contains two regions that are homologous to consensus sequences for an ATP-binding site.
3. Expression of the active human trifunctional enzyme, and separately the dehydrogenase/cyclohydrolase and synthetase domains in *E. coli*, demonstrated that the functional domains are

independent units which can fold separately to attain active conformations. Expression of independent domains confirmed the location of the interdomain region and defined the functional boundaries of the domains. When expressed independently, the dehydrogenase/cyclohydrolase domain exhibit similar kinetic properties as in the native enzyme.

4. A simple and rapid protocol was established to purify the human trifunctional enzyme and the dehydrogenase/cyclohydrolase domain when over-expressed in *E. coli*.

5. When expressed in *S. cerevisiae* the human trifunctional enzyme can rescue an *ade3* mutant thus demonstrating the similarity between the enzymes. The dehydrogenase/cyclohydrolase domain can also complement the *ade3* mutant, demonstrating that the domain is sufficient to interconvert 10-formylH_4folate for *de novo* purine synthesis and that the synthetase domain is not essential for survival of *S. cerevisiae*. Expression of the synthetase domain alone can also complement the *ade3* mutant by the incorporation of supplemented formate.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AFP	alpha-fetoprotein
AICAR	5-aminoimidazole-4-carboxamide ribonucleotide
AIR	5-aminoimidazole ribonucleotide
ATC	aspartate transcarbamylase
bp	base pair(s)
BPTI	bovine pancreatic trypsin inhibitor
CAD	glutamine-dependent carbamylphosphate synthetase-aspartate transcarbamylase-dihydroorotase
CPS	carbamylphosphate synthetase
D/C	dehydrogenase/cyclohydrolase
DEPC	diethyl pyrocarbonate
DHO	dihydroorotase
GAR	glycinamide ribonucleotide
H_4folate	tetrahydrofolate
IMP	inosine monophosphate
kb	kilo base pair(s)
kDa	kilodaltons
MOPS	4-morpholinepropanesulfonic acid
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NEM	N-ethylmaleimide
PAGE	polyacrylamide-gel electrophoresis
PMSF	phenylmethyl-sulfonyl fluoride
RBS	ribosome binding site

SACAIR	5-aminoimidazole-4-(N-succinylo-carboxamide) ribonucleotide
SDS	sodium dodecyl sulfate
Syn	synthetase
UMP	uridine monophosphate

CHAPTER I

GENERAL INTRODUCTION

STRUCTURE AND ORGANIZATION OF ENZYMES

Multifunctional enzymes, where two or more activities are properties of a single type of polypeptide, are found in all classes of organisms and represent one form of enzyme organization. While the mechanism for the evolution of multifunctional enzymes is debated, it is clear that regions of the multifunctional polypeptide are usually related and homologous to corresponding monofunctional proteins found in other organisms. It is becoming apparent that the different activities of a multifunctional enzyme are organized as functional domains in the native structure of the protein. In some cases the multifunctional protein can be cleaved to yield independent domains which retain stable active conformations. Presumably, a polypeptide fragment, representative of a functional domain, can fold independently to attain an active conformation. This thesis deals with methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase which catalyses three sequential reactions involved in the interconversion of one-carbon folate derivatives which have several roles in folate-mediated metabolism. The activities of the eukaryotic trifunctional enzyme are catalyzed by monofunctional proteins in lower organisms and it has been demonstrated that the trifunctional enzyme can be cleaved to yield both a bifunctional and a monofunctional domain. The folate-dependent trifunctional protein provides an excellent system for studying the evolution, independent domain folding and relationship between structure and function of a multifunctional enzyme.

Following a general overview on several aspects of the hierarchy of protein structure and the current theories for the initiation of protein folding, I will briefly review the domain structure of globular proteins including multifunctional enzymes. This will be followed by a review and

description of several multifunctional enzymes involved in one-carbon folate-mediated metabolism, with specific emphasis on the properties of methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase.

Hierarchic Organization of Protein Structure

The process of formation of functional protein following synthesis is dependent on the ability of a given polypeptide to attain the proper three dimensional conformation in its native milieu of the cell. Protein structure can be viewed as a hierarchic organization with primary structure at the first level and the complete tertiary or quaternary structure of the protein as the highest level.

Primary Structure - The primary structure of a protein refers to the amino acid sequence of its polypeptide chain(s) which is specified by the sequence of codons in its messenger RNA (mRNA). The role of the mRNA is to provide a template for the process of translation which assembles the polypeptide chain. During or following translation of a given transcript the polypeptide backbone of the resulting amino acid sequence is spatially arranged to attain secondary, supersecondary, tertiary and in some cases quaternary structure.

Secondary Structure - The formation of secondary structure is dependent largely on hydrogen bonding within the polypeptide and does not necessarily take into account the conformation of side groups.

The α-helix is one of the most prominent polypeptide regular structures which was first described by Linus Pauling (Pauling *et al.*, 1951) based on geometrical parameters. The side-chains of the residues project out of the helix and generally do not interfere with helix formation except for the imino acid proline which is a known helix-breaker. While different residues have weak preferences for or against

existing in an α-helix, some residues such as Ala, Glu, Leu and Met are considered good helix formers while Pro, Gly, Tyr and Ser are poor helix formers (Levitt, 1978). Although the α-helix is the most prominent helix in protein, other structures such as short segments of the 3_{10}-helix (Venkatachalan, 1968) have been located at the ends of a few alpha-helices.

Along with the α-helix, Pauling (Pauling and Corey, 1951) also proposed a different secondary structure known as the β-pleated sheet. β-Sheets are composed of β-strands where the backbone of the polypeptide chain is nearly fully extended. β-Strands can interact in either a parallel or antiparallel orientation and the formed sheets are "pleated" with each successive α-carbon being a little above or below the plane of the sheet. β-Strands can combine to produce purely parallel or antiparallel sheets or, less frequently, a mixed-sheet composed of smaller parallel and antiparallel sheets (Richardson, 1977). In both the parallel and antiparallel orientations the neighbouring side groups on separate strands extend to the same side of the sheet (Richardson, 1981). Although the regular secondary structures of helices and β-sheets are major elements of a protein, other less understood non repetitive secondary structures such as reverse turns (also known as β-bends, β-turns, tight turns, etc.), coil conformation and disordered structure comprise a significant part of an average globular protein. Peptide units which comprise a reverse turn can be as short as 3 to 4 residues and have distinctly recognizable conformations (Crawford *et al.*, 1973). Reverse turns provide a means for a β-strand to reverse itself to form the next anti-parallel strand in a β-sheet. In practice the term "coiled

conformation" is used to describe structures which do not fit into the categories of helix or β-structure. However, coiled structures are highly organized and have defined non-repetitive backbone conformations which are not random or disorganized (Richardson, 1981). In contrast to the well-ordered coil structures, there exists in protein regions which are genuinely disordered. These disordered structures are either absent on electron density maps or appear with much lower density than the rest of the protein. The disordered regions in a protein can be caused by multiple alternative conformations being constantly assumed by that region of polypeptide chain (for review see Richardson, 1981). Some of the non-repetitive structures can be viewed as links along a polypeptide chain which can assume conformations between secondary structural elements and allow for folding of the polypeptide to form super-secondary structures.

Supersecondary structure - In keeping with the hierachic organization of protein structure, Rao and Rossman (1973) coined the term "supersecondary structure" to describe frequently recurring folds of polypeptide chain which are consistently found in different but structurally related proteins. These folds consist of a sequence of secondary structural elements which are combinations of α and/or β-structures. One example is the β-α-β supersecondary structure which has been observed in several proteins including alcohol dehydrogenase, triose-phosphate isomerase and lactate dehydrogenase (Fothergill-Gilmore, 1986). Although there are many possible combinations of α-helices and β-strands, the most prominent supersecondary structures of two or three adjacent segments of secondary structure are: $\alpha\alpha$, $\beta\beta$, $\beta\beta\beta$, and $\beta\alpha\beta$ (Levitt and Chothia, 1976). It has

been suggested that the occurrence of these recurring folds (supersecondary structures) in different proteins is the result of pressure to find energetic stability through packing of the secondary structures involved (Rossman and Argos, 1981). In defining supersecondary structures it was emphasized that these structures are a higher level secondary structure but do not constitute either the complete tertiary structure or entire structural domains.

Tertiary Structure - The tertiary structure of a protein refers to the three dimensional arrangement of the entire polypeptide chain and takes into account the spatial disposition of the side groups. The three dimensional arrangement of secondary structural elements into compact conformations occurs during formation of the protein tertiary structure. The principles which determine the conformations and packing of polypeptide chains into proteins have been reviewed by Chothia (1984).

The packing of secondary structures, which follow different packing models, is governed by the size, shape, and conformation of side chains. One model for packing of α-helices is the "ridges into grooves" model (Chothia *et al.*, 1977, 1981). In this model, the side chains of residues on the surfaces of helices form ridges which are separated by grooves. The helices pack together by the ridges of one helix fitting into the grooves of the other and vice-versa.

When α-helices are found packed onto β-sheets there is a strong tendency for the axes of the helices to be nearly parallel to the sheet strands. One model where this alignment of axes is observed is the "complementary twist" model for helix-sheet packing (Chothia *et al.*, 1977,

Janin and Chothia, 1980). When the helix axis is parallel to the strand direction, helix residues form a twisted surface that is complementary to that of the twisted beta-sheet.

While these secondary structural elements of a polypeptide chain fold to yield the tertiary structure of protein, the folding of polypeptide chains sometimes involves the formation of compact globular clusters along the chain which were coined "domains" by Edelman and coworkers (Edelman, 1973).

Protein Domains - A domain is composed of secondary structural elements which are folded into a unit of stable tertiary structure. Domains of a single polypeptide chain are structurally independent, usually have a specific function and can sometimes be isolated as stable fragments (Rossmann and Argos, 1981). As an example, each subunit of glyceraldehyde-3-phosphate dehydrogenase is folded into two well defined domains (Biesecker *et al.*,1977). The amino-terminal domain contains the residues involved in binding NAD+ while the carboxyl-terminal domain is involved in catalysis and binding glyceraldehyde-3-phosphate. The tertiary structure of the amino-terminal domain of glyceraldehyde-3-phosphate is very similar to the structures of domains in many other proteins which bind dinucleotide coenzymes. These domains are generally known as the "nucleotide binding domain" or the "Rossman fold" and are formed by a super-secondary structure of alternating α-helices and β-strands (Rossman *et al.*, 1975).

Quaternary Structure - Many globular proteins such as glyceraldehyde-3-phosphate (α_4) and hemoglobin ($\alpha_2\beta_2$) are oligomeric and are composed of aggregates of two or more polypeptide chains. Polypeptide chains, sometimes referred to as subunits, associate through non-covalent interactions, and in some cases, through disulfide bonds. The subunits associate in a geometrically specific manner and the quaternary structure refers to the spatial arrangement of these subunits in the native conformation of an oligomeric protein, without regard to the tertiary structure of each subunit. Each subunit is folded into an apparently independent globular conformation which then interacts with other subunits. Forces which interact at the subunit interfaces resemble those of the interior of a protein. The interfaces between subunits are closely packed and involve hydrophobic and ionic interactions as well as hydrogen bonds (Miller, 1989). Based on a review of over 300 oligomeric proteins, Klotz *et al.* (1977) observed that most proteins possessing quaternary structure are either dimers or tetramers of identical subunits. Proteins which are composed of non-identical subunits are usually arranged as dimers or tetramers.

Protein Folding

The folding of a polypeptide chain into its native three dimensional conformation in the aqueous environment of the cell involves non-covalent interactions which include hydrophobic interaction, electrostatic forces, hydrogen bonds and Van der Waals forces (Creighton, 1984b).

For globular proteins, the folding of the polypeptide chain results in a compact unit with an overall shape which is roughly spherical. In

general, the interior of a protein is close-packed, hydrophobic and devoid of large empty spaces (Chothia, 1984).

Through denaturation and renaturation studies of polypeptide chains it has been shown that, in optimal aqueous environments, the primary structure of a protein contains all the necessary requirements to direct folding and yield the native tertiary structure (Anfinsen, 1966). In their classical experiments, Haber and Anfinsen (1961) and White (1961), demonstrated the complete denaturation and renaturation of bovine pancreatic ribonuclease. Following denaturation with 8M urea and 2-mercaptoethanol, the polypeptide refolded under optimal conditions of pH and protein concentration to exhibit the native physical and enzymatic properties (Anfinsen and Haber, 1961).

The "thermodynamic hypothesis" put forth by Anfinsen (1973) for the basis of protein folding and for processes of self assembly, states that the tertiary structure of the native protein in its normal physiological environment is the conformation with the lowest Gibbs free energy. In other words, a polypeptide chain in its native state of folded conformation is more stable than in the unfolded state. Interestingly, the differences in free energy between the folded and unfolded states of a polypeptide chain is of the order of only 25-60 kJ mol^{-1} (Pfeil, 1981).

It is apparent that the folding of polypeptides is not a random process. As pointed out by Karplus and Weaver (1976), the time taken for a random polypeptide of 100 residues to sample all possible conformations would be $\sim 10^{50}$ years. To understand the process of a protein folding to its active conformation it is necessary to decipher the pathway between the unfolded and folded states. The pathway of folding would describe the

intermediate states and the transition states which exist between the unfolded and folded polypeptide. Recently, Murry-Brelier and Goldberg (1989) proposed two possible folding pathways for the β_2-subunit of *E. coli* tryptophan synthase.

One of the few proteins for which an elaborate folding pathway has been elucidated is bovine pancreatic trypsin inhibitor (BPTI) (for review, see Creighton, 1978). BPTI has a polpeptide chain of fifty-eight residues and in its native folded conformation is stabilized by three disulfide bonds.

Kim and Baldwin (1982) reviewed the evidence for specific intermediates in the refolding of a number of small proteins and summarized the various models proposed for domain folding. From the large number of models suggested, two generalized working models were proposed (Kim and Baldwin, 1982). In the "framework" model, elements of secondary structure are formed early in folding and are rearranged to generate the final tertiary structure, whereas in the "modular assembly" model, different regions of the polypeptide fold independently at different rates and then interact to form the overall structure.

Initiation of Folding - Anfinsen (1973) had proposed that the initiation of polypeptide folding occurs at regions of a protein chain that can serve as "nucleation sites". He presented the notion that portions of a protein chain that can serve as nucleation sites for folding will be those that can "flicker" in and out of the conformation that they occupy in the final protein, and that they will form a relatively rigid structure, stabilized by a set of cooperative interactions. While it is generally accepted that

for most proteins the polypeptide chain contains all the necessary requirements for folding, there are at least three models for the initiation of protein folding (Baldwin, 1989). The three models, which will be briefly discussed in this section, propose that protein folding initiates from **A)** formation of specific interactions; **B)** formation of secondary structures or **C)** a hydrophobic collapse.

A - Based on thermodynamic arguments Creighton (1988) proposed a correlation between the ordered formation of disulfide bonds in BPTI and the pathway of folding, which supports the hypothesis of specific interactions involved in initiating and directing folding. As part of the folding pathway, the polypeptide chain of BPTI in its unfolded state initially forms a single but random disulfide bond which can interchange between different cysteine residues. It was suggested that formation of the correct 30-51 disulfide bond in BPTI should stabilize a folded conformation favoring that disulfide bond (Creighton, 1988).

Oas and Kim (1988) provided experimental evidence which supports the proposal that a specific interaction along a polypeptide chain can lead to a folded conformation which favors the initial interaction. It was shown that synthetic peptide fragments of BPTI, which correspond to residues 20 to 33 and 43 to 58, formed a disulfide bond between residues 30 and 51, similarly to the 30-51 intermediate formed during folding of the full length polypeptide. In aqueous solution the linked peptides displayed secondary and tertiary structures similar to those found in the native protein. As well, the structure can be unfolded with increasing temperature and no substantial folded structure was detected when the

disulfide bond was reduced. Thus, it was demonstrated that a specific interaction which joins two segments of a polypeptide chain can stabilize folding. However, it was pointed out that these results do not rule out the possibility of specific secondary structure interactions and non-specific hydrophobic interactions having a prominent role in protein folding (Oas and Kim, 1988).

B) - An alternate model for initiation of polypeptide folding (Lewis *et al.*, 1970), sometimes referred to as the framework model or the kernel theory, postulates that the first step in protein folding is the formation of local secondary structures. These structures would be stabilized by packing and would interact to form more complex structures. Experimental evidence in support of this model was provided by the observations of rapid and somewhat stable secondary structure formation in unfolded polypeptides and peptide fragments (for review, see Wright *et al.*, 1988).

From the studies of several globular proteins, including lysozyme and ribonuclease, Lewis *et al.* (1971) proposed that one of the initial steps during the folding process is the stabilization of secondary structures. As proposed based on mathematical probability, the formation of a β-bend can direct two distant α-helical regions along the polypeptide chain into close proximity to form a stabilized pair of α-helices (Lewis *et al.*, 1971).

Finkelstein and Ptitsyn (1976) performed a study based on the calculated probability of the polypeptide chain of several globular proteins to form helical or other irregular secondary structures. After comparing this with the native secondary structures based on X-ray data,

it was concluded that "the α-helices do pre-exist in an unfolded chain, and therefore the tertiary structure can be assembled from their fluctuating embryos which serve as blocks".

Based on his analysis of energy, kinetics and diffusion factors of the primary structure of globular proteins, Lim (1978) hypothesized that:

> "(i) Short α-helices bearing hydrophobic clusters on their surfaces must be formed first. (ii) These must then be united into a highly helical intermediate globule. (iii) A native structure must be formed by the subsequent transition of α-helices of the highly helical intermediate globule into different types of secondary structure (this transition is not accompanied by any considerable shifts of the material of α-helices relative to each other)".

In support of the framework model of folding Kuwajima *et al.* (1987) studied the refolding of cytochrome C and β-lactoglobulin by stopped-flow circular dichroism. By monitoring the refolding kinetics of the proteins, accomplished by dilution of the denaturant (guanidine hydrochloride), the instantaneous formation of secondary structure was observed.

C) - The model for initiation of protein folding which currently seems to be the most widely accepted, is one which postulates the polypeptide chain to first collapse under the influence of hydrophobic interactions with the exclusion of H_2O. A hydrophobic core is formed from the collapse and would further induce the formation of regular and irregular secondary structures. This theory was first put forth by Rose and Roy (1980) and they stated "our evidence suggests that linear chain regions rich in

hydrophobic residues serve as small clusters that fold against each other, with concomitant or even later fixation of secondary structure". In their studies, the "packing profiles" of several proteins, which is a measure of the protein/protein contact density about each residue (α-carbon) in the sequence, were compared to their hydrophobicity profile. The comparisons revealed that regions of high hydrophobicity along the polypeptide chain corresponded to regions of high density while regions of low hydrophobicity corresponded to regions of low density. Rose (1978) had also shown that the minima in the hydrophobicity profile correspond to β-turns and solvent-exposed parts of helices of the polypeptide chain. These observations suggested that local sites of maximal hydrophobicity fold together to establish the packed interior core of the protein, thereby bringing intervening sites of lower hydrophobicity to the solvent accessible surface.

Lim and Sauer (1989) suggested that the most essential feature of the core sequence of lambda repressor, for attaining an active conformation, is the hydrophobicity of its residues. It was suggested that the simple pattern of hydrophobic residues along the polypeptide chain may dictate its folding.

Molten Globule State - A parallel has been drawn between the polypeptide structure of a hydrophobic collapse and an intermediate of polypeptide folding, the "molten globule state" (Baldwin, 1989; for review see Kuwajima, 1989; Ohgushi and Wada, 1983). The term "molten globule state" was coined by Dolgikh *et al.* (1981) to describe the conformational state of α-lactalbumin when partially denatured in acid. The protein was

described as being in a compact state with fluctuating tertiary structure.

In a study of several globular proteins, Ptitsyn *et al.* (1990) demonstrated in each case a transient intermediate which forms a compact structure during refolding. It was proposed that the molten globular state occurs early on the pathway of folding of all globular proteins.

As discussed by Baldwin (1989), the models for initiation of protein folding have certain aspects which make them complementary. It is possible that the generation of a compact structure from a hydrophobic collapse favors formation of secondary structures. On the other hand, the formation of secondary structures can produce a situation where a hydrophobic side of an α-helix or β-strand is buried towards the interior in the native protein. Thus, the formation of secondary structures would generate hydrophobic surfaces which can interact in the collapsed state. From characterization of partly folded guinea pig α-lactalbumin by ^{1}H-NMR, Baum *et al.* (1989) concluded that native-like secondary structures are also found in the molten globule state.

The role of structural domains as folding units in the process of folding was emphasized by Yon *et al.* (1988) with their studies on the folding pathway of the two-domain protein phosphoglycerate kinase. It was suggested that the carboxyl-terminal domain, being more stable, folds first and may influence the folding of the amino-terminal domain.

From NMR studies of a peptide model, Staley and Kim (1990) concluded that the proper folding of BPTI is dependent on the early formation of native-like subdomains and suggested that understanding the hierarchy of

subdomains in native proteins may be an important part of understanding protein folding.

Domain Structure of Globular Proteins

While there is disagreement on the correct model for the general pathway of protein folding, there is little doubt that attaining the native structures of globular proteins involves the folding of polypeptides into structural domains. The X-ray structure of lysozyme (Phillips, 1966), which was the first for an enzyme, and that of papain (Drenth *et al.*, 1968) were the first to emphasize the domain structure of native proteins. The structure of both of these single chain proteins revealed two distinct globular units which were separated by a cleft.

As previously mentioned, the work of Porter (1973) and Edelman (1973) on the immunoglobulins also support the concept of domains. From his studies based on peptide fragments of *E. coli* β-galactosidase, Goldberg (1969) suggested that the native structure of the enzyme is composed of domains or "globules" folded around independent nucleation centers.

With evi ence of domains found in several globular proteins Wetlaufer (1973) proposed a hypothesis that domains are the result of independent assembly processes occurring along the polypeptide chain:

> "Distinct structural regions have been found in several globular proteins composed of single polypeptide chains. The existence of such regions and the continuity of peptide chain within them, coupled with kinetic arguments, suggests that the

early stages of the three-dimensional structure formation (nucleation) occur independently in separate parts of these molecules. A nucleus can grow rapidly by adding peptide chain segments that are close to the nucleus in amino acid sequence. Such a process would generate three-dimensional (native) protein structures that contain separate regions of continuous peptide chain.... One of the most searching experimental tests for an independent continuous region would be to demonstrate self-assembly of just that region, with a high degree of fidelity".

In support of his theory, Wetlaufer (1981) published a review on the folding of protein fragments where the independent folding of protein domains to native-like structures was demonstrated with fragments of several globular proteins.

Characteristics of Protein Domains - Drawing from several studies of domains in many globular proteins (Rossmann and Argos, 1981; Richardson, 1981) Coggins and Hardie (1986) compiled a list of features that are characteristic of domains:

"1. Similar structures may be recognised elsewhere, either within the same protein (as with the immunoglobulins) or in another protein (as with the dehydrogenases). These similar folds or domains are clearly identifiable at the level of tertiary structure, even when primary structure homology is

not obvious.

2. Domains are generally made up from single continuous stretches of polypeptide chain, typically about 100 amino acids long, although some much larger domains are known.

3. Domains are often spatially well separated from each other and always form a recognisable, compact cluster of amino acid residues.

4. Domains usually have a specific catalytic or binding function.

5. The active sites of multidomain proteins frequently occur at the interface between domains; this permits the simple functions carried by each domain to be brought together to form a more complex structure.

6. Limited proteolytic cleavage can sometimes be used to identify and isolate the individual domains of a protein."

In light of the observation that similar structural patterns or structural domains are found among many globular proteins, Richardson (1981) classified proteins according to their pattern(s) of tertiary structure. In this system, proteins were classified on the basis of their domain structure. Approximately 100 different structural domains were grouped into four broad categories, each of which had several subgroupings. The four categories were 1) antiparallel α-domains 2) parallel α/β-domains 3) antiparallel β-domains 4) small disulfide-rich or metal ligand-rich domains.

As previously mentioned, the nucleotide-binding domain,

alternatively known as the "Rossmann fold" is an example of a recurring domain found in several enzymes which exhibit similar function and tertiary structure (Rossmann *et al.*, 1975). Based on sequence comparisons between proteins which bind pyridine dinucleotides, the sequence Gly-X-Gly X-X-Gly has been proposed as the "fingerprint" for the nucleotide-binding β-α-β unit (Wierenga *et al.*, 1986; Scrutton *et al.*, 1990). The sequence is usually located in the region between the first β-strand and the amino-terminus of the α-helix.

Although it is widely accepted that proteins with similar sequences would fold into closely similar structures, it has been proposed that during the course of evolution the tertiary structure of protein domains may be more highly conserved than their primary structure (Rossmann *et al.*, 1975; McLachlan, 1987).

From the high-resolution crystallographic structures of glycolytic enzymes it has been deduced that active sites in multidomain enzymes are usually located in clefts or at the interface formed between domains (Fothergill-Gilmore, 1986). In some cases where the multidomain enzyme is oligomeric, such as glyceraldehyde-3-phosphate dehydrogenase, the active site can involve the interface formed between the subunits.

The polypeptide chain of the monomeric yeast hexokinase folds into two globular domains. From the X-ray structure of hexokinase B it has been suggested that glucose binds in a cleft which is formed in between the two domains (Anderson *et al.*, 1978). It was subsequently demonstrated that the binding of glucose at the cleft induces a conformational change to a more compact structure (Bennett and Steitz, 1978). This demonstrates the ability of multidomain proteins to change conformation by rotation of

domains which are linked through flexible regions of polypeptide commonly referred to as "hinges".

Another example of a hinge region was demonstrated through the controlled proteolysis of the tetrameric yeast flavocytochrome b2 (Pompon and Lederer, 1976). Treatment of the enzyme with chymotrypsin and subtilisin cleaved each subunit polypeptide to yield two major domains. However, the enzyme maintained some catalytic activity without destruction of the oligomeric state of the protein, thus suggesting some degree of domain interaction.

Evidence for the independent folding of structural domains and interdomain interaction was demonstrated by proteolysis of the two domain protein, pig pancreatic elastase (Ghelis *et al.*, 1978). Following proteolysis and denaturation, a fragment of polypeptide corresponding to one of the domains was able to refold independently back to the structure of the domain in the native protein. As well, *in vitro* complementation of the two domains led to a molecule with a similar conformation as the native protein and exhibited weak enzymatic activity.

Yon *et al.* (1988) suggested that the two domains of horse muscle phosphoglycerate kinase each fold at different rates. The C-terminal domain which is more stable was shown to fold before the N-terminal domain. Subsequently, Minard *et al.* (1989) demonstrated that yeast phosphoglycerate kinase is composed of two independent domains. By recombinant DNA techniques the amino-terminal and carboxyl-terminal domains were expressed separately in yeast, and the carboxyl-domain retained the ability to bind nucleotides.

Size of Protein Domains - Hardie and Coggins (1986) suggested that for a monomeric polypeptide to form a stable tertiary structure and bind some ligand with specificity, the polypeptide should have at least 100 residues. This lower limit for the size of a stable domain is in contrast to the proposal of "modules" which was used to designate units of structure smaller than domains.

Following the suggestion of Gilbert (1978) that "genes-in-pieces" can provide additional mechanisms for protein evolution, Blake (1978, 1979) proposed the hypothesis that exons encode for units of protein domains or subdomains. This hypothesis was supported by studies which demonstrated that the separate exons of the immunoglobulin heavy chain gene encode the separate protein domains and the hinge region (Sakano *et al.*, 1979). The size of an average exon is approximately 140 base pairs (Blake, 1983) and encodes polypeptides of approximately 45 amino acid residues. A polypeptide of 50 residues is below the suggested minimum of 100 residues to form a stable domain, and as pointed out by Hardie and Coggins (1986), domains appear to be encoded by two or more exons (Blake, 1983), which brings us back to the concept of modules.

The term "module" was first used to describe the compact structural units found in mouse hemoglobin (Go, 1981) and chicken lysozyme (Go, 1983). In these studies, the modules of each protein were suggested to correspond well with the exons of the relevant gene.

In his arguments to suggest that proteins are composed of modules, Traut (1986) defines a module as "a minimal polypeptide chain that can assume a stable folded structure and bind some ligand with high specificity." Ligands in this case include substrates, effectors and

enzyme subunits. He states that modules are 1) smaller than protein subunits and domains and 2) have an average size of 5 kDa. From the analysis of several well characterized enzymes Traut related the size of a protein subunit to the number of unique ligand binding functions described for the particular enzyme. Based on his hypothesis it is predicted that 1) each module has a unique function such as binding a ligand and 2) the size of an enzyme subunit is dependent on the number of modules required to accomplish the enzyme's biological role (Traut, 1986).

The hypothesis of exons encoding units of protein structure (Gilbert, 1978; Blake, 1978, 1979, 1981, 1983) is supported by the hypothesis of modules since an average module of 5 kDa can be encoded by an exon which has an average size of 140 bp and encodes a peptide segment of approximately 5.4 kDa. It is also consistent with the suggestion of Wetlaufer (1981) that 20-40 amino acids is the smallest unit of protein with folded structure and some type of function (Traut, 1988a). In providing support for this hypothesis Traut (1988a) cited several studies where exons have been suggested to correspond to units of protein structure as well as units of function. He also argued (Traut, 1988a) that during the course of evolution the deletion of introns and changes in the position of intron/exon junctions can explain why existing patterns of exons might not consistently relate to units of structure or function in proteins.

Traut (1988b) conducted a review on enzymes involved in nucleotide metabolism in which he undertook to relate "the significance of subunit size and polymer size for biological function and regulatory properties." As the data set, enzymes which function in the synthesis, interconversion

or catabolism of nucleotides were included. In correlating subunit size to the hypothesis of modules it was revealed that, for the enzymes involved in nucleotide metabolism, the average molecular weight per ligand binding function (module) is approximately 50% larger than previously determined. The proposed explanations were that 1) enzymes involved in nucleotide binding are not sufficiently characterized and 2) in larger proteins, which may result from the fusion of two or more domains, some binding functions can become lost or masked (Traut, 1988b). Included in his study were several examples of multifunctional enzymes involved in nucleotide metabolism. One observation which is not disputed is that, for multifunctional proteins, the increase in subunit molecular weight corresponds with additional catalytic functions of the protein.

Multifunctional Proteins

Multifunctional proteins are generally characterized as consisting of a single type of polypeptide chain having multiple catalytic or binding functions. They are very similar to multienzyme complexes in that the active sites for several catalytic activities are associated in close proximity. However, while multienzyme complexes are composed of different types of polypeptides associated by non-covalent interactions, the multifunctional enzymes consist of active sites covalently linked within a single polypeptide chain.

Multifunctional enzymes have been found in many if not all classes of prokaryotic and eukaryotic organisms and examples of these enzymes have been well documented (Kirschner and Bisswanger, 1976; Schmincke-Ott and Bisswanger, 1980; Coggins and Hardie, 1986). Multifunctional enzymes

usually catalyze two or more consecutive reactions in a metabolic process. However, for those which do not catalyze consecutive reactions the enzyme activities are still usually from a single metabolic pathway. It is apparent that multifunctional polypeptides are more common in eukaryotes and are involved in metabolic pathways which are located in the cytoplasm (Coggins and Hardie, 1986). The pathways for *de novo* synthesis of pyrimidines and purines are two good examples of multifunctional enzymes catalyzing several reactions in metabolic pathways.

In mammalian cells, five of the six enzymes involved in *de novo* pyrimidine biosynthesis, from glutamine to uridine monophosphate (UMP), are components of multifunctional proteins. The first three steps in the pathway are catalyzed by the trifunctional enzyme (CAD) with the activities carbamoyl phosphate synthetase II (CPS), aspartate transcarbamoylase (ATC) and dihydroorotase (DHO) (Coleman *et al.*, 1977). The last two activities, orotate phosphoribosyltransferase and orotidine monophosphate decarboxylase are properties of a mammalian bifunctional enzyme (McClard *et al.*, 1980) while the enzyme catalyzing the fourth step, dihydroorotate dehydrogenase, is monofunctional. In *S. cerevisiae* (Lue and Kaplan, 1969) and *N. crassa* (Williams *et al.*, 1970) the CPS and ATC activities exist as a bifunctional enzyme while in *E. coli* the three activities of the mammalian CAD enzyme are catalyzed by three non-associated monofunctional proteins (Evans, 1986).

In the biosynthetic pathway of *de novo* purine synthesis from 5-phosphoribosyl 1-pyrophosphate to inosine monophosphate (IMP), seven of the ten activities are properties of three multifunctional enzymes in avian liver while in *E. coli* the activities seem to be properties of

different and separate polypeptides. In chicken liver, the bifunctional enzyme [5-aminoimidazole ribonucleotide (AIR) carboxylase]-[5-aminoimidazole-4-(N-succinylocarboxamide) ribonucleotide (SACAIR) synthetase] catalyzes reactions 6 and 7 (Patey and Shaw, 1973), while another bifunctional enzyme [5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) transformylase]-[IMP cyclohydrolase] catalyzes steps 9 and 10 (Benkovic, 1984). The activities glycinamide ribonucleotide (GAR) synthetase, GAR transformylase and AIR synthetase form a trifunctional enzyme as isolated from chicken liver (Daubner *et al.*, 1985) however, in this case only the first two activities catalyzing reactions two and three are sequential while the third activity catalyzes reaction five in the purine biosynthetic pathway.

Functional and Structural Domains in Multifunctional Enzymes - Kirschner and Bisswanger (1976) proposed that multifunctional enzymes are composed of functional units of activity which correspond to discrete structural "domains" within the folded polypeptide chain. As well, it has been observed that the areas in between functional regions of multifunctional enzymes are susceptible to proteolysis.

Earlier experiments to support the hypothesis of domains relied on limited proteolysis of the full length polypeptide chain to yield separate peptide fragments which retained individual activities of the full length protein. The "Klenow fragment" of *E. coli* DNA polymerase I was obtained by mild proteolysis with subtilisin (Jacobsen *et al.*, 1974). Proteolysis of *E. coli* DNA polymerase I produced a small amino-terminal fragment which exhibited only the 5'-3' exonuclease activity and a large fragment which

retained the polymerase as well as the 3'-5' exonuclease activity.

Evidence that the active sites of *E. coli* aspartokinase I-homoserine dehydrogenase I were on separate domains was obtained when an ochre mutation in its gene led to expression of a truncated amino-terminal peptide fragment with only the aspartokinase activity (Janin *et al.*, 1967). Subsequently, limited proteolysis of this bifunctional enzyme with chymotrypsin produced an active homoserine dehydrogenase fragment while the kinase activity was destroyed (Truffa-Bachi, 1974).

It has been demonstrated that proteolysis of the hamster CAD enzyme with trypsin can give rise to three separate functional domains (Davidson *et al.*, 1981; Grayson and Evans, 1983; Rumsby *et al.*, 1984) with each domain expressing one activity of the CAD enzyme. The sum of the molecular weights of the three domains is similar to that of the undigested monomer (Davidson *et al.*, 1981).

Further evidence to support the existence of functional domains in multifunctional enzymes is provided by sequence homology between monofunctional polypeptides of lower organisms and the corresponding domain regions in the multifunctional proteins of higher organisms. Tryptophan synthase in yeast is composed of two functional domains A and B, while in prokaryotes it is an $\alpha_2\beta_2$ tetramer. From sequence homology it was determined that domain A of the yeast enzyme corresponds to the α-subunit and that domain B corresponds to the β-subunit (Crawford *et al.*, 1987).

Simmer *et al.* (1989) demonstrated that the amino acid sequence of the ATC domain in the hamster CAD enzyme is homologous to the sequence of the monofunctional ATC of *E. coli* and *B. subtilis*.

From the isolation of genes and cDNAs encoding multifunctional enzymes, it has been possible to manipulate the coding region and specifically express fragments of polypeptide corresponding to putative functional domains. Maley and Davidson (1988) isolated several cDNA fragments of the hamster CAD cDNA which complemented an *E. coli* mutant (*pyr B*) lacking the ATCase activity. The cDNA fragments which contained the region homologous to the *E. coli pyr B* gene were able to encode stable proteins which folded independently and expressed ATCase activity. As well, their studies showed that the cDNA fragment which was analogous to the entire *pyr B* gene, plus 12 codons of CAD sequence upstream of this region, was the most active. Larger fragments which include more of the upstream CAD sequence were less active and less stable while a smaller fragment which was missing five codons at the 5'-end of the analogous region with *pyr B* was also less active. It was suggested that the extra residues, which included seven prolines, might have hindered the domain folding, interfered with the formation of functional multimers and/or destabilized aggregates. The shorter fragment may also cause these effects or may be missing critical residues for catalytic reaction (Maley and Davidson, 1988).

Chen *et al.* (1990) recently isolated the cDNA clone encoding the avian bifunctional AIR carboxylase-SAICAR synthetase, by complementation of *E. coli pur* mutants. Deletion of the 3'-region of the cDNA clone destroyed the carboxylase activity, and a protein domain expressing only the synthetase activity was obtained.

A fragment of cDNA encoding the thioesterase and acyl carrier protein domains of chicken fatty acid synthase have been expressed in *E.*

coli (Pazirandeh *et al.*, 1989). The recombinant thioesterase was found to be enzymatically active and had the same substrate specificity and kinetic properties as in the native multifunctional synthase. Proteolysis with chymotrypsin cleaved the recombinant acyl carrier protein from the thioesterase domain at the same location as in the native enzyme. This demonstrated the folding of functionally active domains which was apparently independent of the other domains of the multifunctional enzyme.

Linker Regions in Multifunctional Enzymes - The ability to obtain stable functional domains from multifunctional enzymes demonstrates that there exist regions of the polypeptide chain, commonly called "linker regions" or "connecting regions" which are highly susceptible to proteolysis and serve to link the domains of the protein together. It has been suggested that the conformational flexibility of linker regions allow for the correct folding of independent functional domains as well as the entire polpeptide chain of multifunctional enzymes (Hardie and Coggins, 1986; Zalkin *et al.*, 1984).

The domain in the hamster CAD enzyme which corresponds to the aspartate transcarbamylase activity was isolated and characterized (Grayson and Evans, 1983). A proteolytic fragment of 40 kDa containing aspartate transcarbamylase activity was obtained following controlled proteolysis of the 240 kDa CAD polypeptide. The quaternary structure and kinetic properties of the domain were similar to those of the CAD enzyme and suggested that the native tertiary structure of the transcarbamylase region was preserved in the isolated domain. However, in the absence of stabilizing agents, the activity of the domain had a much shorter half-

life than of that in the full length protein and thus suggested that interactions with other regions of the molecule may stabilize the transcarbamylase domain in the intact enzyme.

By analysis of the cDNA clone encoding the hamster CAD protein and through complementation of *E. coli pyr B* mutants, Shigesada *et al.* (1985) succeeded in localizing the aspartate transcarbamylase domain to the 3'-end of the cDNA. Subsequently, Simmer *et al.* (1989) sequenced the region of cDNA spanning the 3'-region of the dihydroorotase domain to the 5'-region of the transcarbamylase domain. Through sequence analysis it was suggested that the dihydroorotase and transcarbamylase domains are joined together through a linker region of 133 amino acid residues (14 kDa). The linker region was shown to contain 28 prolines (21%) and 31 charged residues (23%), it is very hydrophilic and secondary-structure calculations predicted a low propensity to form helical or sheet structures. These results taken together suggested that the linker was an exposed segment of polypeptide on the surface of the protein molecule and correlates with the vulnerability of linker regions to proteolytic digestion.

It has been suggested that the function of a linker region is not necessarily dependent on a specific sequence (Zalkin *et al.*, 1984). It was demonstrated that the linker region of 11 amino acid residues joining the two domains of the bifunctional enzyme anthranilate synthase :indole-3-glycerol phosphate synthase share little homology between the enzymes from yeast and *Neurospora crassa*. The proposal received further support from studies on the linker region of the yeast bifunctional enzyme tryptophan synthase (Crawford *et al.*, 1987). Amino acid sequence

comparison between the yeast tryptophan synthase and the corresponding α- and β-subunits of the prokaryotic enzyme revealed a 66-residue interdomain segment of the bifunctional enzyme with little or no similarity to the prokaryotic polypeptides. A deletion of 18 residues from this linker region inactivated the enzyme; the activity was partially restored by the insertion of 14 residues, whose type and sequence had no relation to the 18 residues deleted. As well, a spontaneous duplication of 48 bases which resulted in a spontaneous duplication of 16 amino acid residues in the linker region of the inactivated enzyme restored activity to a greater extent. In relation to the hypothesis that the linker region is critical in directing or permitting proper folding of the polypeptide chain, the changes in the linker region leading to greater or lesser loss of enzyme activity can be due to interference with proper folding of domains. While it is clear that the residues of the linker region play no direct role in catalysis, these results suggest that the length of the linker region is more critical than its sequence in maintaining function.

Advantages of Multifunctional Enzymes - Hardie and Coggins (1986) suggested several advantages of having sequential enzymes in a pathway associated together either as a multienzyme complex or as a multifunctional enzyme (for review see Welch, 1977). One advantage is the increased stability attributed to aggregated polypeptides chains of the multienzyme complex and to the stable close packing of multidomain structures of the multifunctional enzyme. As the size of a globular protein increases, its volume increases more rapidly than its surface area which allows for more internal hydrophobic stabilizing interactions and

therefore larger proteins would be expected to be more stable than smaller ones.

The association of active sites on different polypeptides of an enzyme complex, or on separate domains of a multifunctional enzyme may lead to conformational changes of the protein such that the catalytic efficiencies of the associated active sites are increased relative to the unassociated active sites. Intuitively, the close proximity of active sites decreases the time required for intermediates to diffuse from one active site to the next (transit time), which consequently decreases the time for the overall sequence of reactions.

The intermediate between two reactions can be "channelled" from one active site to the next, where channelling is defined as the preferential transfer of the product from one active site to the active site of the next activity without equilibration with the medium. Without the necessity for substrate to equilibrate with the medium, potentially reactive or unstable intermediates could be protected. As well, substrate channelling eliminates potential interference or competition between pathways where the substrate for one activity of a pathway can inhibit or be utilized by an activity on another pathway (for review see Gaertner, 1978).

Matchett (1974) demonstrated that the bifunctional enzyme tryptophan synthase of *N. crassa* channels the intermediate indole from one active site to the other to synthesize tryptophan in two catalytic steps.

In the case of the multifunctional enzyme fatty acid synthase, channelling involves a covalently bound intermediate. For both the mammalian (α_2) and yeast ($\alpha_6\beta_6$) enzymes the lengthening fatty acyl chain

is transfered between active sites for successive reactions while attached to the sulfhydryl group of the phosphopantotheine prosthetic group of the acyl carrier protein domain (Hardie and McCarthy, 1986).

Ljungcrantz *et al.* (1989) reported substrate channelling in an artificial bifunctional enzyme. The genes encoding β-galactosidase and galactose dehydrogenase were fused together to express a bifunctional protein in *E. coli*. The hybrid protein was found in multimeric forms and was suggested to channel galactose from the β-galactosidase active site to that of galactose dehydrogenase.

The association of active sites also allows for the possibility of coordinate regulation of enzyme activities. Such a control would require an effector molecule to bind a single site (effector site) and mediate its effect through conformational changes of the multifunctional protein or protein complex. An example of such an effect is seen with *E. coli* aspartokinase I-homoserine dehydrogenase I where it was suggested that an allosteric effector (threonine) regulated both activities by binding at a single allosteric site (Cohen and Dantry-Varsat, 1980).

In the case of multifunctional enzymes, where the different activities are encoded by a single transcript, there is a coordinate expression of the individual enzyme activities. This allows for coordinate regulation of expression at the genetic level and ensures stoichiometric amounts of each activity which together catalyze a related set of reactions.

Evolution of Multidomain Proteins

The observation that similar structural and functional units are found in different proteins (Rossmann *et al.*, 1975) correlates with the suggestion that domains are the fundamental units of protein evolution (Hardie and Coggins, 1986; Janin and Wodak, 1983) and is consistent with the proposal that new protein molecules can be constructed from parts of pre-existing ones (Gilbert, 1978).

Stemming from the hypothesis that exons code for units of protein structure and function, the exon recombination hypothesis suggests that novel proteins can be formed and evolve from the shuffling and recombination of exons (Gilbert, 1978; Blake, 1978, 1979, 1981). While an exon encoding a domain with a specific function can recombine randomly in a given gene, only exons which recombine in the correct arrangement for expression of a useful product would be favourably selected. An advantage of such a mechanism is that new domains with the same function would not need to be developed for each different protein (Traut, 1988b). In support of the hypothesis, Blake (1978) argued that in a new enzyme, formed by the combination of two or more domains, the only location for a novel interaction of amino acid residues to confer a new function is in the region of the domain interface.

Another mechanism for the creation of novel enzymes/proteins is the duplication of an existing functional gene so that one of the two copies could diverge via mutations, until it acquired a different activity and would be stabilized by selection (Traut, 1988b). In this process a stable tertiary structure would be already established for a functional protein and mutations could lead to substitution of one type of substrate binding

activity for another. Support for this mechanism exists in the examples of isozymes which are expressed from separate genes (Markert, 1975).

Scrutton *et al.* (1990) demonstrated that the specificity of an enzyme can be modified by a few amino acid changes. Modification of several residues in the β-α-β fold of the dinucleotide binding domain in glutathione reductase changed the coenzyme specificity of the enzyme from NADP+ to NAD+.

The modular assembly model of protein evolution proposes that larger multidomain proteins evolved by fusion of smaller pre-existing functional units (Janin and Wodak, 1983). The mechanism of gene duplication followed by gene fusion leads to the creation of larger proteins containing two or more domains having a common structure and sometimes a common function. Examples of such proteins include calmodulin which has 3 domains (Simmen *et al.*, 1985), ovomucoid with 3 domains (Stein *et al.*, 1980) and ubiquitin with 7 domains (Finley and Varchavsky, 1985).

It has been proposed that the mechanism of duplication followed by mutation and fusion can give rise to enzymes with expanded regulatory functions (Traut, 1988b). It has been suggested that the mammalian phosphofructokinase has evolved by duplication of a bacterial ancestor gene, where the duplicated catalytic center has subsequently been modified into an allosteric regulatory site (Poorman *et al.*, 1984).

Albumin and α-fetoprotein (AFP), which are members of a multigene family, have evolved by gene duplication (Young and Tilghman, 1986). Through sequence analysis and prediction of secondary structures, it was proposed that both proteins were composed of three homologous domains each divided further into two subdomains. Evidence that albumin and AFP arose

from the triplication of a primordial domain was provided when the single copy gene for each protein was isolated. Analysis of the gene structure revealed a three fold repeat of an exon pattern. As well, there are homologies between the equivalent exons in each repeat and the location of exon boundaries are conserved. The exons have been classified according to their location in each repeat and can be correlated to recurring units in the three repeating domains of the protein. Despite sequence and structural homologies between the three domains, each domain of serum albumin has diverged to bind different ligands (Blake, 1985).

Isolation and characterization of the mouse *mdr1* cDNA and gene, suggested that the mouse gene evolved from the duplication and subsequent fusion of an ancestral gene (Raymond and Gros, 1989). The deduced amino acid sequence revealed that the protein is composed of two homologous halves. It was also suggested that 10 of the 12 putative transmembrane domains of the protein are encoded by individual exons and the two ATP-binding sites, one in each half of the protein, are each encoded by three sequential exons. From a comparison of exons and introns between several enzymes known to contain the same ATP-binding consensus sequence as in the *mdr1* P-glycoprotein, it was suggested that the exons involved in ATP-binding share a common ancestor and have been used by unrelated genes to encode a similar ATP-binding domain in different proteins (Raymond and Gros, 1989).

Evolution of Multifunctional Enzymes - It has been suggested that multifunctional enzymes, which are composed of functional domains, evolved by gene fusion events which could have involved the previously mentioned

mechanisms (Schmincke-Ott and Bisswanger, 1980). Support for the suggestion that multifunctional enzymes have evolved from fusion of ancestor monofunctional genes is provided when functional domains in the multifunctional enzymes of higher organisms correspond in function, size and sequence to monofunctional polypeptides found in lower organisms.

In eukaryotes, the AROM multifunctional enzyme catalyzes five sequential reactions while in prokaryotes the five reactions are catalyzed by separate proteins encoded by genes scattered throughout the genome (for review, see Coggins and Boocock, 1986). The gene fusion hypothesis for the *Neurospora crassa* AROM polypeptide is attractive considering that the subunit molecular weight of the corresponding polypeptides in *E. coli* add up to 167 kDa compared to 165 kDa for the AROM subunit of the *N. crassa* protein (Coggins and Boocock, 1986). Isolation of the *ARO1* gene encoding the AROM enzyme of *Saccharomyces cerevisiae* revealed that the yeast protein is a "mosaic" of five functional domains, each of which is homologous to a monofunctional *E. coli* polypeptide (Duncan *et al.*, 1987).

Fatty acid synthase in vertebrates is a multifunctional enzyme which catalyzes 6 reactions and has an acyl-carrier function. In fungi (including yeast), the 6 activities and the acyl-carrier function are distributed between two non-identical polypeptide chains while in bacteria the activities are properties of separate monofunctional enzymes (Schweizer, 1986; Hardie and McCarthy, 1986). Based on the isolation of active fragments from the multifunctional enzymes, sequence homology between active domains and monofunctional enzymes, and genetic evidence to support independent active domains in the yeast enzyme, it has been argued that the vertebrate enzyme as well as the yeast enzyme evolved through

gene fusion (Hardie *et al.*, 1986; Schweizer, 1986; Hardie and McCarthy, 1986). Although both the yeast and vertebrate enzymes have been postulated to have envolved through a similar mechanism, the observation that the order of active domains along the polypeptide chain is different among the proteins indicates that they arose by independent gene fusion events (McCarthy *et al.*, 1983; Hardie and McCarthy, 1986).

Genes which encode enzymes for *de novo* purine synthesis have been mapped on the *E. coli* chromosome (Bachmann and Low, 1980) and except for *pur H(J)D*, *pur MN* and *pur EK* operons, the genes are dispersed on the chromosome. This is in sharp contrast to the situation in *Bacillus subtilis* where the enzymes catalyzing ten enzymatic steps of *de novo* purine synthesis are encoded by a cluster of twelve genes organized as an operon (Ebbole and Zalkin, 1987). Comparison of the amino acid sequence between the bacterial monofunctional enzymes and the corresponding domains in multifunctional enzymes revealed significant homology. GAR synthetase which is monofunctional in *B. subtilis* is part of a bifunctional enzyme (GAR synthetase-AICAR synthetase) in *S. cerevisiae* and part of a trifunctional enzyme (GAR synthetase-AIR synthetase-GAR transformylase) in *Drosophila*. Alignments of the GAR synthetase from the three organisms demonstrated homology in monofunctional, bifunctional, and trifunctional enzymes from a prokaryote, lower eukaryote and invertebrate (Ebbole and Zalkin, 1987). Henikoff *et al.* (1986) reported that AIR synthetase, in the *Drosophila* trifunctional enzyme, contains two duplicate copies of an AIR synthetase sequence. The amino acid sequence of the monofunctional AIR synthetase in *E. coli* and *B. subtilis* is highly conserved and is similar to the single-copy sequence in the yeast bifunctional enzyme and

to both copies of the duplicated AIR synthetase sequence in *Drosophila*. In *B. subtilis* the genes *pur MN*, which encode AIR synthetase and GAR transformylase, are overlapping and suggest a possible step in the evolution of the gene encoding the trifunctional enzyme. As proposed by Ebbole and Zalkin (1987), an insertion of one nucleotide or deletion of 2 nucleotides at or immediately upstream of an ancestral *pur M* translation termination codon (TGA) would fuse the *pur M* and *pur N* coding sequences to yield an inframe AIR synthetase-GAR transformylase multifunctional protein.

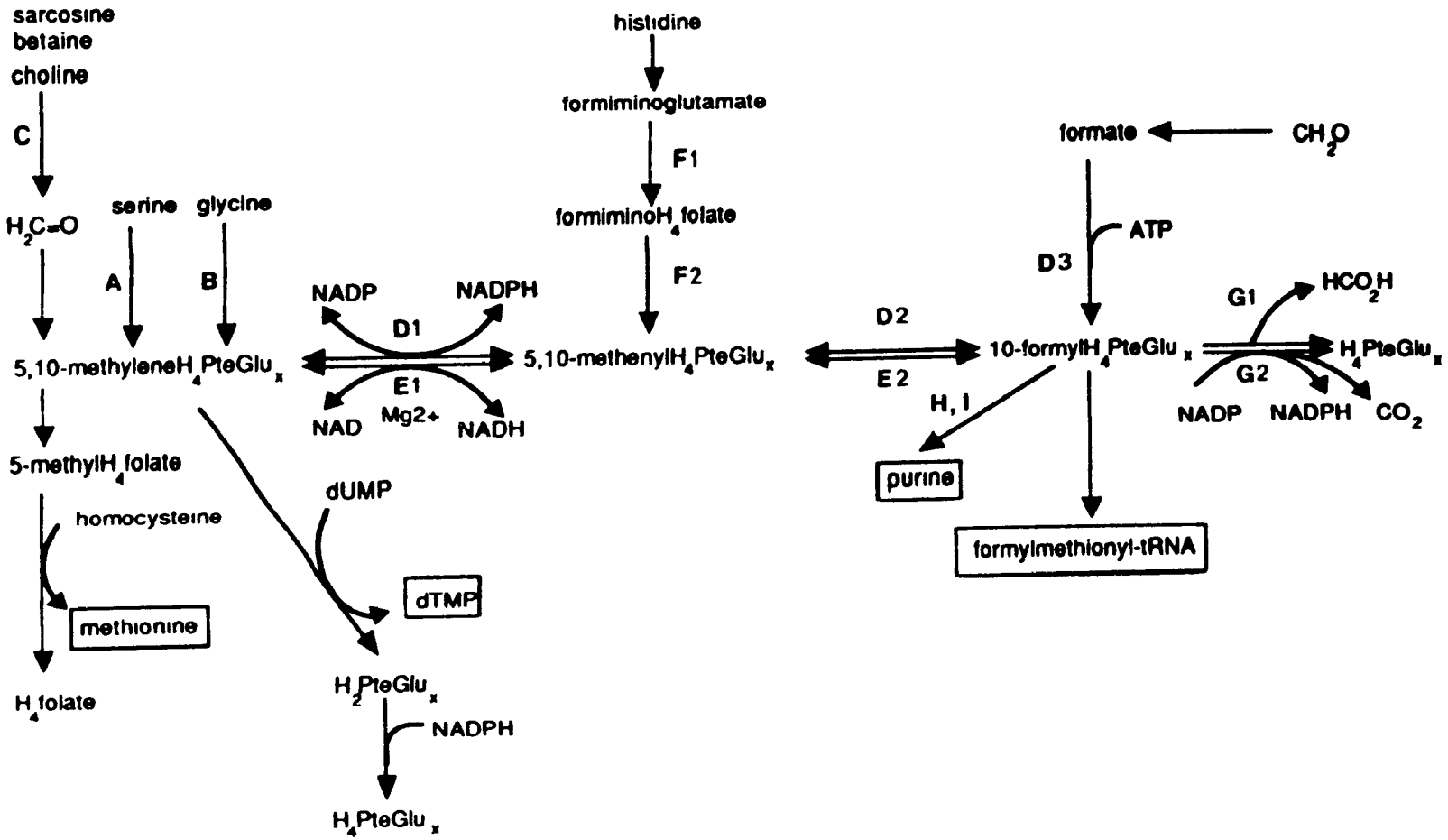
Several of the enzymes involved in *de novo* purine biosynthesis are found in a multienzyme complex in avian liver along with a trifunctional enzyme which interconverts one-carbon tetrahydrofolate derivatives required for purine synthesis (Caperelli *et al.*, 1980; Smith *et al.*, 1980). From recent findings that several of these enzymes are multifunctional, at least 9 activities are found in the multienzyme complex. Henikoff *et al.* (1986) suggested that, in the purine biosynthetic pathway, the existence of multifunctional enzymes and their organization into an enzyme complex must offer some selective advantage to the organism. Pathways organized in this way allow for more efficient genetic regulation and also for more efficient biochemical regulation of metabolism. Kirschner and Bisswanger (1976) discussed the possibility that the fusion of contiguous structural genes may occur before the corresponding enzymes have developed an affinity for one another. Conversely, a multienzyme complex conferring a selective advantage to the organism may have evolved before fusion of the corresponding structural genes.

Multifunctional Enzymes Involved in Folate-mediated Metabolism

The subject of this thesis is the domain structure of a folate-dependent trifunctional enzyme which I have studied by sequence comparison, *in vivo* expression and functional complementation. This is only one example of a multifunctional enzyme in the broad area of folate-mediated metabolism, so I will briefly describe some examples and characteristics of other multifunctional enzymes in the folate pathways before reviewing in more detail the known properties of methylenetetrahydrofolate dehydrogenase methenyltetrahydrofolate cyclohydrolase-formyltetrahydrofolate synthetase.

Overview - Folate and its derivatives are involved in several biosynthetic as well as degradative pathways (for a review, see MacKenzie, 1984). An outline of some pathways which utilize folate derivatives is shown in Fig. 1 where the activities of a given multifunctional enzyme are indicated by the same letter. The overall scheme involves degradative reactions which produce one-carbon derivatives of tetrahydrofolates at different oxidation states of the one-carbon unit that are then used in biosynthetic pathways. Quantitatively, serine and glycine are the most important contributors of one-carbon units for the synthesis of substituted folates. Serine is incorporated into the active one-carbon pool in the cytosol by cleavage with serine hydroxymethyltransferase to yield glycine and 5,10-methyleneH_4folate. It has been proposed that in mitochondria, the glycine cleavage system and catabolism of choline and sarcosine yield 5,10-methyleneH_4folate which may be used by serine hydroxymethyltransferase to convert glycine to serine for transport of one-carbon units to the

Figure 1. **Pathways of folate-mediated one-carbon metabolism.** Multifunctional enzymes are indicated by the same letter. **A**, serine hydroxymethyltransferase; **B**, glycine cleavage system; **C**, dimethylglycine and sarcosine dehydrogenase; **D1**, NADP-dependent 5,10-methyleneH_4folate dehyrogenase; **D2**, 5,10-methenylH_4folate cyclohydrolase; **D3**, 10-formyl-H_4folate synthetase; **E1**, NAD-dependent 5,10-methyleneH_4folate dehydrogenase; **E2**, 5,10-methenylH_4folate cyclohydrolase; **F1**, 5-formimino - H_4folate:glutamate formiminotransferase; **F2**, 5-formiminoH_4folate cyclodeaminase; **G1**, 10-formylH_4folate hydrolase; **G2**, 10-formylH_4folate dehydroganase. **H, I**; GAR transformylase and AICAR transformylase, each is part of a multifunctional enzyme involved in *de novo* purine biosynthesis and uses 10-formylH_4folate as cofactor.


sarcosine
betaine
choline
C
$H_2C{=}O$
serine
glycine
A
B
5,10-methyleneH_4PteGlu$_x$
5-methylH_4folate
homocysteine
methionine
H_4folate
NADP
NADPH
D1
E1
NAD
Mg2+
NADH
dUMP
dTMP
H_2PteGlu$_x$
NADPH
H_4PteGlu$_x$
histidine
formiminoglutamate
F1
formiminoH_4folate
F2
5,10-methenylH_4PteGlu$_x$
formate
CH_2O
D3
ATP
D2
E2
10-formylH_4PteGlu$_x$
H, I
purine
formylmethionyl-tRNA
G1
HCO_2H
G2
H_4PteGlu$_x$
NADP
NADPH
CO_2

cytosol. 5,10-MethyleneH_4folate is either used for synthesis of thymidylate, is reduced to 5-methylH_4folate for methionine synthesis or is oxidized by two activities of methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase to 10-formylH_4folate, which is a substrate in *de novo* purine synthesis.

It is thought that the conversion of 5,10-methyleneH_4folate to 10-formylH_4folate occurs in the cytosol although, this is presently an area of vigourous debate. Barlowe and Appling (1988a) suggested that, like most cytosolic NADP-linked dehydrogenases, NADP-dependent methyleneH_4folate dehydrogenase should catalyze the conversion of methenylH_4folate to methyleneH_4folate. In a coupled enzyme assay system with rabbit methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase and serine hydroxymethyltransferase, Strong and Schirch (1989) demonstrated the conversion of formate to serine. It was recently reported that folate-dependent reactions in the mitochondria of mammalian transformed cells (Mejia and MacKenzie, 1988), yeast cells (Shannon and Rabinowitz, 1986) and rat liver cells (Barlowe and Appling, 1988) can interconvert 5,10-methyleneH_4folate and 10-formylH_4folate. It has been suggested that a mitochondrial enzyme, such as 10-formylH_4folate synthetase or hydrolase, may convert 10-formylH_4folate to H_4folate and formate which can be transported to the cytosol; in the presence of H_4folate and ATP the formate would be incorporated into the active one-carbon pool by cytosolic 10-formylH_4folate synthetase to provide 10-formylH_4folate for *de novo* purine synthesis (Barlowe and Appling, 1988; Mejia and MacKenzie, 1988).

Strong *et al.* (1990) suggested that the inhibition of cytosolic

serine hydroxymethyltransferase would not block purine and thymidylate synthesis and thus argued that the mitochondrial enzymes may convert glycine and serine to formate which would be incorporated into the active one-carbon pool in the cytosol. Scanlon *et al.* (1981) demonstrated that inhibition of serine hydroxymethyltransferase with tetrahydrohomofolate did not significantly inhibit either thymidylate or purine biosynthesis. Mutants of CHO cells which lack mitochondrial serine hydroxymethyltransferase were reported to be auxotrophic for glycine (Chasin *et al.*, 1974) thus suggesting that this enzyme is important for synthesizing glycine for use in the cytosol. From the studies published thus far it is apparent that the pathways involved in the interconversion of one-carbon folate derivatives are more complex than originally thought, and deciphering the overall scheme will prove to be a challenging and exciting task.

Multifunctional Enzymes - The folate-dependent bifunctional enzyme formiminoglutamate formiminotransferase - 5-formiminoH_4folate cyclodeaminase catalyzes two sequential reactions in the pathway of histidine degradation which incorporates a carbon unit into the activated one-carbon pool of folate derivatives (for review, see Shane and Stokstad, 1984). Formiminotransferase catalyzes the transfer of the formimino group from formiminoglutamate to the N^5 position of H_4folate with a release of free glutamate. The cyclodeaminase activity then cyclizes the 5-formiminoH_4folate with a release of ammonia to yield 5,10-methenylH_4folate. This bifunctional enzyme is found in the liver and kidneys of all mammalian species tested.

Most of the work on formiminotransferase-cyclodeaminase has been performed using enzyme purified from pig liver (Tabor and Wyngarden, 1959; Drury *et al.*, 1975; Findlay *et al.*,1989). The native enzyme from pig liver is an octomer composed of identical subunits of 62 kDa which are arranged in a planar ring (Drury *et al.*, 1975; Beaudet and MacKenzie; 1975).

It has been demonstrated that the two activities of the enzyme occur at separate catalytic sites and that the intermediate can be channeled between the active sites. (Drury and MacKenzie, 1977; MacKenzie *et al.*, 1980; MacKenzie and Baugh, 1980). Channelling experiments with the substrate tetrahydropteroyl(glutamate)n having 1, 3, 5 or 7 gamma-glutamyl residues showed that when the pentaglutamate substrate was used, no lag and no accumulation of intermediate was observed, corresponding to 100% channelling. As well, by mixing enzyme which lacked the cyclodeaminase activity (Drury and MacKenzie, 1977) with enzyme which lacked the transferase activity it was shown that channelling required that the active sites be present on the same molecule (MacKenzie and Baugh, 1980).

It was proposed that the long polyglutamate chain would serve as a type of anchor to attach the intermediate to the enzyme molecule and that the pentaglutamate is the optimal chain length to allow the transfer of the tetrahydropteroyl moiety between the two active sites without its dissociation from the enzyme (MacKenzie and Baugh, 1980). Subsequently it has been shown that there exists only one polyglutamate binding site for every pair of catalytic sites (Paquin *et al.*, 1985).

Through the partial denaturation of formiminotransferase-cyclodeaminase, Findlay and MacKenzie (1987) demonstrated that isolated

active deaminase-dimers and transferase-dimers have different subunit-subunit interfaces, one of which is required for each activity. Renaturation of denatured enzyme in the presence of 1.5 M urea led to the formation of transferase-dimers only, while the deaminase activity was recovered by the removal of urea and correlated with the formation of octamer (Findlay and MacKenzie, 1988). This suggested that the native octameric structure is required for expression of both activities and therefore for channelling of the intermediate. The polyglutamate binding site was shown to be associated with the deaminase-dimer and it was also demonstrated that the transferase active site is part of a structural/functional domain capable of refolding independently to form an active dimer (Findlay and MacKenzie, 1988). These experiments demonstrate that the quaternary structure of the enzyme is required for bifunctionality.

10-Formyltetrahydrofolate dehydrogenase-hydrolase is a folate-dependent bifunctional enzyme which catalyzes two non-sequential reactions. The NADP-dependent dehydrogenase reaction releases carbon dioxide while the hydrolase reaction releases formate, and both reactions catalyze the conversion of 10-formylH_4folate to H_4folate (for review, see MacKenzie, 1984). It has been suggested that the metabolic role of the enzyme is to regenerate H_4folate in the absence of active reactions that utilize 10-formylH_4folate (Osborn *et al.*, 1957).

Zamierowski and Wagner (1977) identified, and Cook and Wagner, (1982) purified and partially characterized a rat liver folate-binding protein (Cytosol I) which was determined to be 10-formylH_4folate dehydrogenase-hydrolase (Min *et al.*, 1988). The dehydrogenase-hydrolase

activity is found mainly in liver and kidney, and has not been detected in other tissues. The enzyme was purified to homogeneity from rat liver and was apparently a tetramer with identical subunits of 108 kDa (Scrutton and Beis, 1979). Rios-Orlandi *et al.* (1986) purified the enzyme from pig liver and determined the subunit size to be 92.5 kDa. They demonstrated that both activities of the bifunctional enzyme occur simultaneously and apparently at different sites. However, Case *et al.* (1988) reported that 10-formylH_4folate dehydrogenase and hydrolase are separate enzymes in rat liver where the dehydrogenase activity was localized in the cytosol and the hydrolase activity was exclusively localized in the mitochondria. Johlin *et al.* (1989) demonstrated immunologically that the rat and human liver 10-formylH_4folate dehydrogenase possess similar properties.

NAD-dependent methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase is a bifunctional enzyme which catalyzes two sequential reactions involved in the interconversion of one-carbon units from 5,10-methyleneH_4folate to 10-formylH_4folate. Based on enzyme activity assays (Mejia and MacKenzie, 1985) and Northern blot analyses (Bélanger and MacKenzie, 1989) it has been reported that the NAD-dependent dehydrogenase activity was detectable in many transformed and established cell lines as well as in developing tissues but not in normal differentiated cells. The enzyme was purified 6000-fold from Ehrlich ascites tumour cells and was determined to be a 68 kDa dimer with identical subunits of 34 kDa (Mejia *et al.*, 1986). The NAD-dependent dehydrogenase required Mg^{2+} for activity and was stimulated by inorganic phosphate.

By isolation of the cDNA encoding the bifunctional enzyme (Bélanger and MacKenzie, 1989) and by assays of subcellular fractions of murine cell

lines (Mejia and MacKenzie, 1988) it was revealed that NAD-dependent methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase is a mitochondrial enzyme. The deduced amino acid sequence revealed a signal peptide of 34 amino acids which had characteristics of other known mitochondrial presequences. As well, enzyme activity assays and Western analyses localized the enzyme to mitochondria.

The folate derivative 10-formylH_4folate is a substrate for two activities which are both properties of different multifunctional proteins in the *de novo* purine pathway. As previously mentioned, the folate-dependent GAR transformylase activity is part of the trifunctional enzyme GAR synthetase-AIR synthetase-GAR transformylase (Daubner *et al.*, 1985) and the other folate-dependent activity, AICAR transformylase forms a bifunctional enzyme with IMP cyclohydrolase (Mueller and Benkovic, 1981). It has been suggested that in avian liver these two multifunctional enzymes and the trifunctional enzyme methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase associate to form a multienzyme complex involved in *de novo* purine synthesis (Smith *et al.*, 1980; Caperelli *et al.*, 1980; Benkovic, 1984).

The activities methyleneH_4folate dehydrogenase, methenylH_4folate cyclohydrolase and formylH_4folate synthetase catalyze three sequential reactions involved in the interconversion of activated one-carbon derivatives of H_4folate (Fig. 2) (for review, see MacKenzie, 1984). The dehydrogenase reaction, which is dependent on a pyridine dinucleotide as cofactor, catalyzes the oxidation of 5,10-methyleneH_4folate to 5,10 methenylH_4folate (Fig. 2). The dehydrogenase activity of the *S. cerevisiae* (Tatum *et al.*, 1977) and the porcine (Green *et al.*, 1986) enzymes have

been shown to stereospecifically transfer a hydride to the *pro*-R-position of NADPH. The cyclohydrolase activity converts 5,10-methenylH_4folate to 10-formylH_4folate while the ATP-dependent synthetase reaction synthesizes 10-formylH_4folate from formate and H_4folate.

In prokaryotes, the three activities are properties of separate proteins or, as shown for *Clostridium thermoaceticum*, the dehydrogenase and cyclohydrolase activities can form a bifunctional enzyme while the synthetase activity is separate and monofunctional. However, in some prokaryotes which include *E. coli.* (Dev and Harvey. 1978, 1982), *Pseudomonas aeruginosa*, *Klebsiella pneumonia* and *Salmonella typhimurium*, the 10-formylH_4folate synthetase activity was not detectable (Shannon *et al.*, 1986; Whitehead *et al.*, 1988).

The co-purification of the three activities from porcine liver led to the suggestion that the three enzymes occurred as a complex (Mackenzie, 1973) and was later determined to be a trifunctional enzyme. The eukaryotic trifunctional enzyme has been isolated and characterized from ovine liver (Paukert *et al.*, 1976), porcine liver (Tan *et al.*, 1977), *Saccharomyces cerevisiae* (Paukert *et al.*, 1977; Staben *et al.*, 1987), rabbit liver (Schirch, 1978), avian liver (Caperelli *et al.*, 1980), mouse liver (Gardam *et al.*, 1987), human liver (Mackenzie *et al.*, 1988) and rat liver (Cheek and Appling, 1989). In all cases the enzyme was apparently a dimer with identical subunits of approximately 100 kDa. The similarity between several of the eukaryotic trifunctional enzymes and the monofunctional 10-formylH_4folate synthetase of *Clostridium acidiurici* have been demonstrated by immunological cross-reactivity (Staben and Rabinowitz, 1983). A polyclonal antiserum raised against the porcine

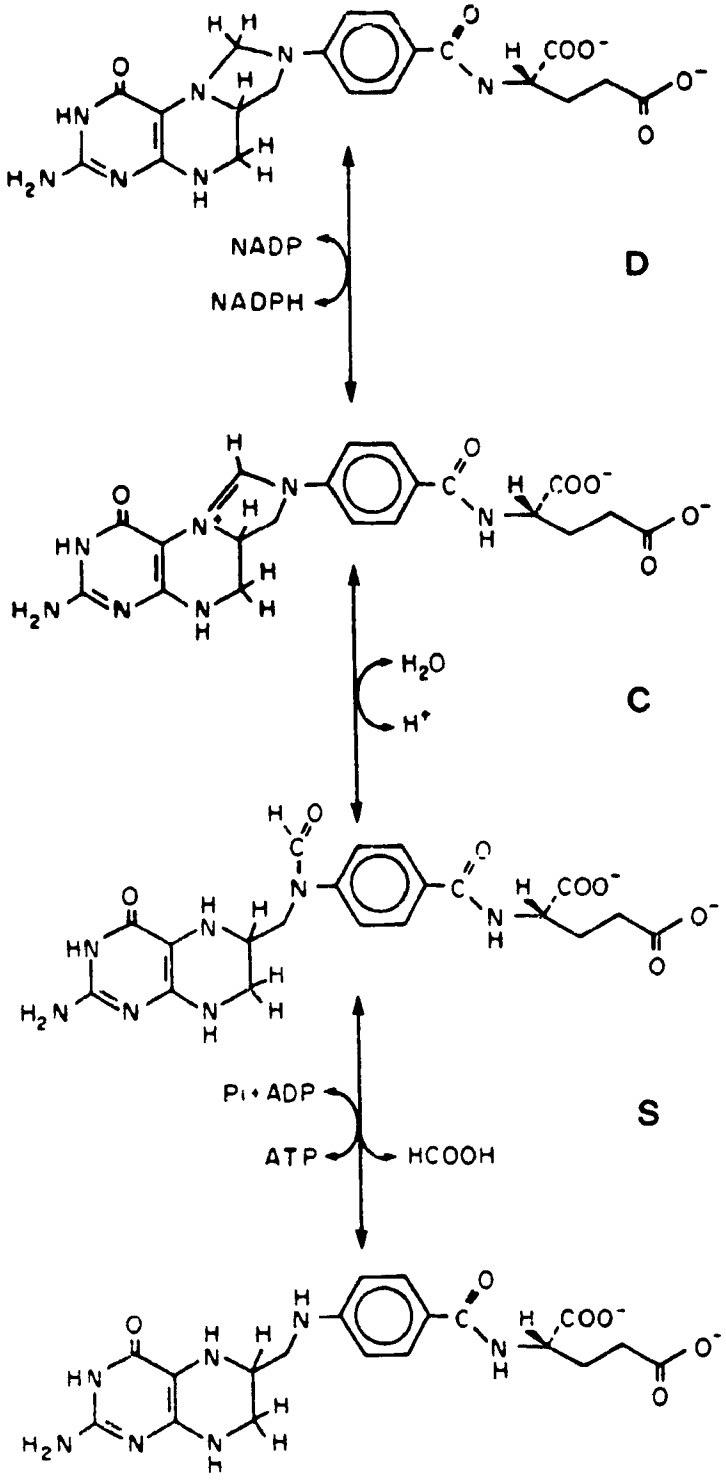


Figure 2. The three sequential reactions catalyzed by 5,10-methyleneH_4folate dehydrogenase (D), 5,10-methenylH_4folate cyclohydrolase (C) and 10-formylH_4folate synthetase (S).

enzyme also cross-reacts with the rat and human enzymes (Rozen *et al.*, 1985). The enzyme is found primarily in the cytosol although, Shannon and Rabinowitz (1986) also isolated a mitochondrial isozyme from *S. cerevisiae*. The mitochondrial enzyme was shown to have a signal sequence of 34 amino acid residues and is encoded in the yeast nuclear genome.

Paukert *et al.* (1977) demonstrated that the synthetase activity of the yeast trifunctional enzyme can be isolated as an independent domain. Limited proteolysis of the full length enzyme with trypsin coordinately destroyed the dehydrogenase and cyclohydrolase activities and yielded a homodimer with subunits of 76 kDa which contained the synthetase activity. Subsequently, Tan and MacKenzie (1977, 1979) showed that the porcine enzyme is composed of two major functional domains. Limited proteolysis of the protein with trypsin yielded a 33 kDa fragment which retained dehydrogenase/cyclohydrolase activities while proteolysis with chymotrypsin coordinately inactivated dehydrogenase and cyclohydrolase activities but yielded a synthetase-active fragment of 67 kDa. Immunodiffusion analysis showed nonidentity between the two peptide fragments and amino-terminal sequence analysis of the two peptides and the uncleaved enzyme indicated that the dehydrogenase-cyclohydrolase domain was located at the amino-terminal region while the synthetase domain was at the carboxyl-terminal region of the polypeptide (Tan and MacKenzie, 1979). Although it was not possible to isolate an active dehydrogenase/cyclohydrolase fragment of the yeast enzyme by proteolysis (Paukert *et al.*, 1977), a mutant of *S. cerevisiae* was isolated which did not express synthetase activity but retained the dehydrogenase/cyclohydrolase activities (deMata and Rabinowitz, 1980). Limited treatment of the rabbit

trifunctional enzyme with trypsin allowed for the simultaneous isolation of a 36 kDa dehydrogenase-cyclohydrolase-active fragment and a 66 kDa synthetase-active fragment (Villar *et al.*, 1985). It was suggested that each subunit of the native enzyme is composed of a small bifunctional domain and a large monofunctional domain which are structurally and functionally independent.

Cohen and MacKenzie (1978) demonstrated a functional interaction between the dehydrogenase and cyclohydrolase active site(s) of the porcine enzyme. By measuring the time course of appearance of 5,10-methenylH_4folate and of 10-formylH_4folate from the dehydrogenase and cyclohydrolase activities in enzyme assays, it was determined that, in the two sequential reactions, 5,10-methenylH_4folate did not equilibrate with the solution. 10-FormylH_4folate was produced without a lag period and it was determined that 60% of the intermediate was preferentially channelled from the dehydrogenase to the cyclohydrolase active site to yield 10-formylH_4folate. As well, the dehydrogenase activity displayed an ordered bi-bi kinetic mechanism. Subsequently, Wasserman *et al.* (1983) demonstrated a similar channelling of intermediate between the dehydrogenase and cyclohydrolase active sites of the avian enzyme.

It was shown that the efficiency of substrate channelling between the dehydrogenase and cyclohydrolase active site(s) of the porcine trifunctional enzyme was not affected by the number of glutamates on the tetrahydropteroyl(glutamate)n substrates (MacKenzie and Baugh, 1980; Ross *et al.*, 1984). However, the folylpolyglutamates did bind with a slightly higher affinity. Wasserman *et al.* (1983) suggested that channelling within the avian enzyme was more efficient with a triglutamate substrate.

It has been suggested that the synthetase domain of the rabbit trifunctional enzyme undergoes a conformational change in the presence of MgATP and H_4pteroylpolyglutamates (Schirch *et al.*, 1986; Strong *et al.*, 1987). Studies revealed that, in the presence of MgATP, the binding of the tri-, tetra-, or pentaglutamate of H_4pteroylglutamates caused a conformational change of the synthetase domain to a more thermally stable state and abolished the cold lability of the domain. As well, the rate of domain formation by chymotrypsin proteolysis of the enzyme was increased although the H_4pteroylpolyglutamates stabilized the synthetase domain from further digestion by chymotrypsin. When H_4pteroyltriglutamate was used as substrate in place of the H_4pteroylmonoglutamate, the Km values for MgATP and formate were lowered 3.6- and 520-fold, respectively.

The close association between the dehydrogenase and cyclohydrolase active site(s) of the porcine enzyme was further demonstrated when Smith and MacKenzie (1983) showed that the cyclohydrolase activity of the trifunctional enzyme was inhibited by NADP+, which is a substrate of the dehydrogenase activity. This showed quite clearly that these two activities cannot function simultaneously and are not kinetically independent. From the observations of both activities in a small 33 kDa domain, the channelling of an intermediate, the non-independent nature of the two activities and the identical inactivation and ligand protection of both activities, it was suggested that the dehydrogenase and cyclohydrolase acitvities overlap to some extent or actually share a common active site (Smith and MacKenzie, 1983). Chemical modification of the enzyme with diethylpyrocarbonate (DEPC) inactivated the dehydrogenase and cyclohydrolase in a coordinate manner but did not inactivate the

synthetase activity. Folate was able to protect both activities against inactivation by DEPC, and NADP+ potentiated the protection against inactivation for both the dehydrogenase and cyclohydrolase. Chemical modification with phenylglyoxal also inactivated both activities at the same rate and to the same extent while the synthetase activity was affected to a much lesser extent. As well, the combination of NADP+ and folate protected the dehydrogenase and cyclohydrolase activities in a coordinate manner. A similar study of chemical modification with the yeast trifunctional enzyme also suggested an overlapping site for the dehydrogenase and cyclohydrolase reactions (Appling and Rabinowitz, 1985) although there were differences in the rate of inactivation of the dehydroganase and cyclohydrolase activities both by DEPC and by N-ethylmaleimide (NEM). While NADP+ alone did not protect the dehydrogenase and cylcohydrolase activities of the porcine enzyme against inactivation by DEPC or phenylglyoxal (Smith and MacKenzie, 1983), Appling and Rabinowitz (1985a) demonstrated that both activities of the yeast enzyme were protected against inactivation by DEPC and NEM by NADP+ although, the dehydrogenase activity was protected to a slightly greater extent. Based on the lack of any significant accumulation of 5,10-methenylH_4folate during the dehydrogenase and cyclohydrolase reaction, the identical heat stability and protection by NADP+ for these two activities, and the observation of similar affinity constants of the two activities for different H_4folate compounds, Schirch (1978) also suggested a single site for the dehydrogenase and cyclohydrolase activities of the rabbit enzyme.

Smith and MacKenzie (1985) provided direct evidence to suggest that a single folate site is implicated in both the dehydrogenase and

cyclohydrolase activities of the porcine enzyme. By using carbodiimide activated [^{3}H]-folate, which inactivated in parallel both the dehydrogenase and cyclohydrolase activities in the presence of NADP+. it was demonstrated that one mole of reagent bound and led to the inactivation of one mole of the bifunctional dehydrogenase/ cyclohydrolase domain.

The three activities of the trifunctional enzyme supply activated one-carbon units required in biosynthetic pathways which are indispensable for survival of the organism. Expression of the enzyme is probably ubiquitous and constitutive in all cells. By manipulating the diet of chickens (Wasserman *et al.*, 1984) and incubating yeast under various growth conditions (Appling and Rabinowitz, 1985b) it has been suggested that the amount of the trifunctional enzyme can be influenced by nutritional requirements where the regulation of expression occurs at a pretranslational level. Barlowe and Appling (1988b) reported that exposure of rat liver tissue to nitrous oxide decreases the expression of the trifunctional enzyme and suggested that regulation of expression occurs at the translational or pretranslational level. Based on enzyme activity assays, Western analyses (Cheek and Appling, 1989) and Northern analyses (Thigpen *et al.*, 1990) it has been reported that the level of the trifunctional enzyme varies among the different tissues in rat.

The cDNA encoding methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase has been isolated from several species and will be discussed later on in the thesis. It is apparent that isolation and characterization of the cDNA and gene will provide an invaluable tool to further study the structure and function of the protein.

CHAPTER 2

PRIMARY STRUCTURE OF A HUMAN TRIFUNCTIONAL ENZYME. ISOLATION OF A cDNA ENCODING METHYLENETETRAHYDROFOLATE DEHYDROGENASE-METHENYLTETRAHYDROFOLATE CYCLOHYDROLASE-FORMYLTETRAHYDROFOLATE SYNTHETASE

SUMMARY

A DNA clone complementary to the messenger RNA encoding the human trifunctional enzyme 5,10-methylenetetrahydrofolate dehydrogenase-5,10-methenyltetrahydrofolate cyclohydrolase- 10-formyltetrahydrofolate synthetase has been isolated from a lambda gt10 library. *In vitro* transcription-translation of the 3.1 kb cDNA clone yields a protein of 101 kDa which is of identical size and exhibits the same immunoreactivity as the enzyme purified from human liver. A coding region of 2805 bp in the cDNA encodes a protein of 935 amino acids. The initiator methionine is absent from the purified enzyme and the amino terminal 30 amino acids derived by automated sequence analysis are identical (Arg at position 18 was not identified) with that deduced from the nucleotide sequence. The amino acid sequence of the human enzyme shows extensive homology with that of the yeast enzyme, although the amino-terminal bifunctional dehydrogenase-cyclohydrolase domain is less homologous than the carboxyl-terminal synthetase domain. A region was identified which probably serves as a link between these two major domains of the human enzyme. The synthetase domain contains two regions that are homologous to consensus sequences for an ATP-binding site.

INTRODUCTION

The enzymes 5,10-methyleneH_4folate dehydrogenase (E.C.1.5.1.5), 5,10-methenylH_4folate cyclohydrolase (E.C.3.5.4.9) and 10-formylH_4folate synthetase (E.C.6.3.4.3) catalyze three sequential reactions in the interconversion of one-carbon derivatives of tetrahydrofolate which are substrates for methionine, thymidylate and *de novo* purine syntheses. Because they exist in nature both as single enzymes and as different enzyme conjugates, these three activities provide a good system in which to examine the evolution and advantages of enzyme organization. In procaryotes the activities exist either as three separate proteins or the dehydrogenase-cyclohydrolase activities can form a bifunctional enzyme (MacKenzie, 1984). In eukaryotes the three activities are properties of a single protein, a homodimer of polypeptides of approximately 100 kDa. The eukaryotic trifunctional enzyme consists of two major domains, an amino terminal portion containing the dehydrogenase-cyclohydrolase activities and a larger synthetase domain. Proteolysis of the mammalian enzymes has yielded both domains as active species (Tan and MacKenzie, 1977; Tan and MacKenzie, 1979; Villar *et al.*, 1985), while an active synthetase domain has been obtained from the yeast protein (Paukert *et al.*, 1977). The dehydrogenase-cyclohydrolase activities of the enzyme from pig liver are not kinetically independent (Smith and MacKenzie, 1983), and appear to share a common folate-binding site (Smith and MacKenzie, 1985), but it is not known to what extent these properties are shared by the protein from yeast. These two trifunctional proteins show quantitative differences in the relative rates of inactivation of the dehydrogenase and cyclohydrolase by histidine specific chemical

modification, with the mammalian enzyme showing almost coincident loss of both activities (Smith and MacKenzie, 1983) while about a two-fold difference in rate of inactivation is observed with the yeast enzyme (Appling and Rabinowitz, 1985). This might suggest less interdependency between the dehydrogenase and cyclohydrolase activities in the yeast protein, as has been described for the mammalian NAD-dependent dehydrogenase-cyclohydrolase enzyme (Rios-Orlandi and MacKenzie, 1988). The similarities or differences between the yeast and the mammalian enzymes can also be compared at the level of the primary structure of the respective proteins. Although X-ray crystallography studies will be required to deduce the complete structures of the proteins, the primary structure can be used to further define the two functional domains and determine the region which links them.

MATERIALS AND METHODS

Restriction enzymes and DNA modification enzymes were obtained from Bethesda Research Laboratories or Boehringer Mannheim. All reagents and enzymes for DNA sequencing were from United States Biochemical Corporation. Nitrocellulose and nylon membranes were from Amersham and radioisotopically labelled nucleotides were from Amersham or New England Nuclear. Diagnostic X-Omat AR film was from Kodak. Protein A-Sepharose CL-4B was from Pharmacia. All common chemicals were of the highest grade commercially available.

Unless specified, all recombinant DNA techniques followed the procedures outlined by Maniatis *et al.* (1982). All cDNA probes were labelled by the random primer method with [^{32}P]dCTP to an average specific activity of 10^9 cpm/μg of DNA (Feinberg and Vogelstein, 1984).

Isolation of cDNA Clones - A polyclonal antibody was raised by subcutaneous injection of purified porcine trifunctional protein (Smith and Mackenzie, 1983; Tan *et al.*, 1977) in Freund's adjuvant into a New Zealand white rabbit. Injections were administered at 2-week intervals; two weeks after the third injection, the rabbit was bled and serum was collected and stored at -70^{0}C. A lambda gt11 human liver cDNA library that was kindly provided by Dr. D.W. Stafford, University of North Carolina, Chapel Hill, U.S.A., was screened with the polyclonal antibody following the method described by Huynh *et al.* (1985). The antiserum was diluted 1/200 and was not preabsorbed with *E. coli* extract prior to screening. After incubating with antiserum the membranes were treated with ^{125}I-protein A (80-90 μCi/μg) using 10^7 cpm/ml. Positives were selected,

plaques were purified and the cDNA inserts were sized by digestion of purified lambda gt11 DNA with *Eco*RI prior to electrophoresis in 1% agarose gels.

Two positives designated H-1 and H-2 were isolated; the insert from H-1 (230 bp) was subcloned into the *Eco*RI site of the pUC 13 vector (Vieira and Messing, 1982) and used as a probe to screen a lambda gt10 cDNA library made from the human colonic adenocarcinoma cell line LS-180. (The library was a generous gift from Drs. N. Beauchemin and C.P. Stanners (Beauchemin *et al.*, 1987). The membranes were incubated for 12 h at 42°C in 50% formamide, 5 x SSPE, 2 x Denhardt's solution, 100 μg/ml denatured salmon sperm DNA and the cDNA probe at 5 x 10^5 cpm/ml (1 x SSPE = 0.15 M NaCl, 10 mM NaH_2PO_4, 1.3 mM EDTA, pH 7.4). They were washed for 15 min with 1 x SSC, 0.1% SDS at room temperature, and then with 0.1 x SSC, 0.1% SDS for 15 min at room temperature and twice for 45 min at 65°C (1 x SSC = 0.15 M NaCl, 15 mM sodium citrate). Thirteen positive recombinants were isolated from the LS-180 library of which two (HT-2, HT-12) were full length.

Protein Sequencing - Automated Edman degradations of purified human trifunctional enzyme were executed on a Model 470A Gas Phase Sequencer equipped with an on-line Model 120A Phenylthiohydantoin Analyzer (Applied Biosystems Inc., ABI) employing the general protocol of Hewick *et al.* (1981). The protein sample was applied to a filter coated with 1.5 mg of polybrene plus 0.1 mg NaCl (Biobrene Plus) and standard programs (03RPRE and 03RPTH, ABI) were employed for precycling and sequencing.

cDNA Sequencing - HT-2 was subcloned into the *Eco*RI site of the Bluescript-SK$^+$ and Bluescript-KS$^+$ vectors (Stratagene Cloning Systems) and progressive unidirectional deletions using exonuclease III/mung bean nuclease were carried out as outlined by the supplier. After the deletions the DNA was subjected to electrophoresis in a 0.8% agarose gel, appropriately sized bands were excised, the DNA was purified and the ends repaired with the Klenow fragment of *E. coli* DNA polymerase I prior to ligation with T4-DNA ligase. The coding strand of the cDNA was sequenced from single-stranded DNA templates while the sequence of the non-coding strand was determined using double-stranded DNA. All sequences were derived from dideoxy sequencing (Sanger *et al.*, 1977) using [^{35}S]dATP sequencing reactions with Sequenase™ (United States Biochemical Corp.).

Northern Analysis - Total RNA from LS-180 cells grown in culture was extracted by the method of Chirgwin *et al.* (1979). Poly(A)$^+$ mRNA was isolated by hybridization to Hybond-mAP paper (Amersham Corp.) as described by the manufacturer. Total and messenger RNA were electrophoresed on 1.25% formaldehyde-agarose gels in MOPS and blotted onto nitrocellulose. Hybridization was carried out overnight at 42^{0}C with probes (5 x 10^5 cpm/ml) derived from the 230 bp H-1 fragment and the 1.26 kb *Hae*II fragment of HT-2 in 35% formamide, 10% dextran sulfate, 5 x SSC, 4 x Denhardt's solution, 10 mM Tris•Cl (pH 7.5), 0.1% SDS, and 100 μg/ml denatured salmon sperm DNA. After hybridization the filters were washed with 2 x SSC, 0.1% SDS for 30 min at 42^{0}C, and with 0.1 x SSC, 0.1% SDS for 10 min at 50^{0}C and rinsed with 0.1 x SSC at room temperature.

In vitro Transcription - RNA transcripts of the HT-2 cDNA were synthesized using T3- and T7-RNA polymerase following standard protocols as outlined by Stratagene Cloning Systems. [^{3}H]CTP (≥20 μCi/mmol) was incorporated into the transcripts and 5 μg of RNA was electrophoresed in a 1% formaldehyde-agarose gel. The gel was impregnated with EN^{3}HANCE (NEN) and dried prior to fluorography.

In vitro Translation - The T3 and T7 transcripts (20 ng/μl of translation mix) were translated in a cell-free, message dependent, rabbit reticulocyte lysate system (Pelham and Jackson, 1976) in the presence of 1 μCi/ml of [^{35}S]methionine (≥800 μCi/mmol). For immunoprecipitation the translated samples were diluted with 9 volumes of 1% Triton X-100, 20 mM methionine, 0.02% sodium azide, 10 mM EDTA (pH 7.4), all dissolved in phosphate buffered saline (pH 7.4). The samples were centrifuged for 45 min at 120,000 x g and the supernatants were incubated with 5 μl of polyclonal antiserum and 0.4 volumes of 4 M NaCl for 8 h at 4^{0}C. The immunocomplexes were then precipitated by incubating with Protein A-Sepharose CL-4B at room temperature for 60 min, and washed 3 times with the above Triton X-100 solution and twice with 0.9% NaCl. The washed pellets were resuspended in 0.1 M Tris•HCl (pH 6.8), 4% SDS, 1.4 M 2-mercaptoethanol, 20% glycerol, 0.1% bromophenol blue, boiled for 3 min, cooled, centrifuged in an Eppendorf microfuge and supernatants applied to an 8% SDS-polyacrylamide gel (Laemmli, 1970). In experiments where the antibodies were preabsorbed with purified human enzyme (Gardam *et al.*, 1988), 5 μl of antiserum was incubated with the protein for 8 h at 4^{0}C using the buffer conditions as described above prior to incubation with

translation products.

RESULTS AND DISCUSSION

A human liver lambda gt11 cDNA library was screened with a rabbit polyclonal antibody raised against the trifunctional enzyme isolated from porcine liver. An initial screening of 2.7×10^6 recombinants yielded two positive clones of 230 bp which cross-hybridized. One of the positive clones (H-1) was used to screen approximately 3.5×10^5 recombinants of a human lambda gt10 cDNA library and thirteen positive clones varying in sizes from 800 bp to 4 kb were isolated. All 13 clones were digested with *Eco*RI and were found to contain a common 3'-end by Southern analysis using the H-1 fragment as a probe. The longest insert of 4 kb was rejected because it was found to contain two unrelated sequences of DNA, presumably a result of blunt end ligation during construction of the cDNA library. A 154 bp probe from the 5'-end of the 4 kb insert, when used to rescreen the library, selected 16 additional clones, but none of them hybridized with the original 230 bp H-1 DNA. A clone of about 3.1 kb (HT-2) was chosen for sequence analysis; the deduced sequence of the first 30 amino acids after the first AUG codon was identical to that obtained from the NH_2-terminal sequence of the purified protein with the exception that in the protein-sequence analysis the predicted Arg at cycle 18 was not identified.

Northern blot analyses of mRNA isolated from the human cell line LS-180 were carried out with the 230 bp H-1 insert and the 1.26 kb *Hae*II fragment of HT-2 as probes. Locations of the probes in the cDNA are shown in Fig. 1. Both hybridized to a common mRNA, a single band of 3.1 kb (Fig. 2A, lane 1) which is of sufficient length to code for a 100 kDa protein. These results are indicative of the protein being translated

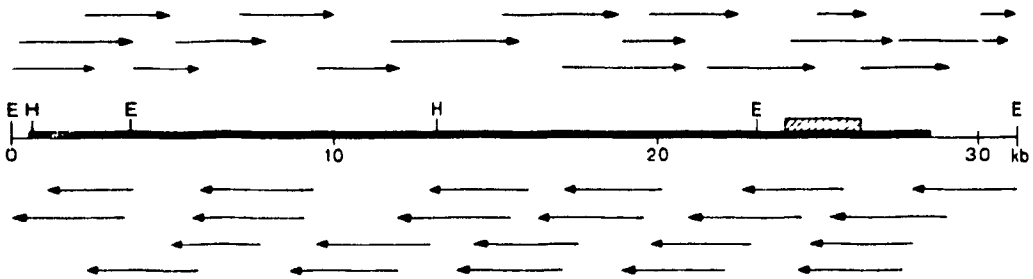
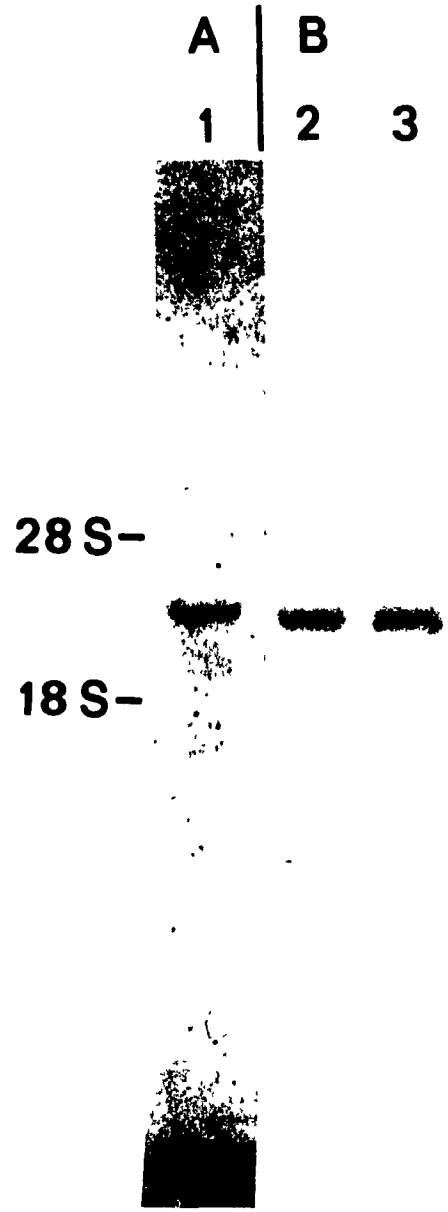


Figure 1. **Physical map and sequencing strategy for HT-2 cDNA.** The diagram shows (from left to right) the 5'-untranslated region (thin line), the amino acid coding region (heavy line), the 230 bp H-1 cDNA (hatched box) and the 3'-untranslated region (thin line). The arrows indicate the direction and extent of the DNA sequence determined. Abbreviations for restriction endonuclease sites: E, *Eco*RI; H, *Hae*II.

Figure 2. **Northern analysis of human mRNA (A) and in vitro transcription of the cDNA clone (B).** (A) 3 μg of LS180 poly(A)$^+$ mRNA was electrophoresed on a 1.25% formaldehyde-agarose gel in MOPS buffer and transferred onto nitrocellulose. Hybridization with the ^{32}P-labelled 1.26 kb *Hae*II fragment of HT-2 was as described in Materials and Methods and radioactivity was detected by fluorography. (B) HT-2 was subcloned into the Bluescript-SK+ and Bluescript-KS+ vectors and transcribed with T7 (lane 1) and T3 (lane 2) RNA polymerase. Transcripts (5 μg) containing incorporated [^{3}H]CTP were electrophoresed on a 0.9% agarose gel and were detected by fluorography.

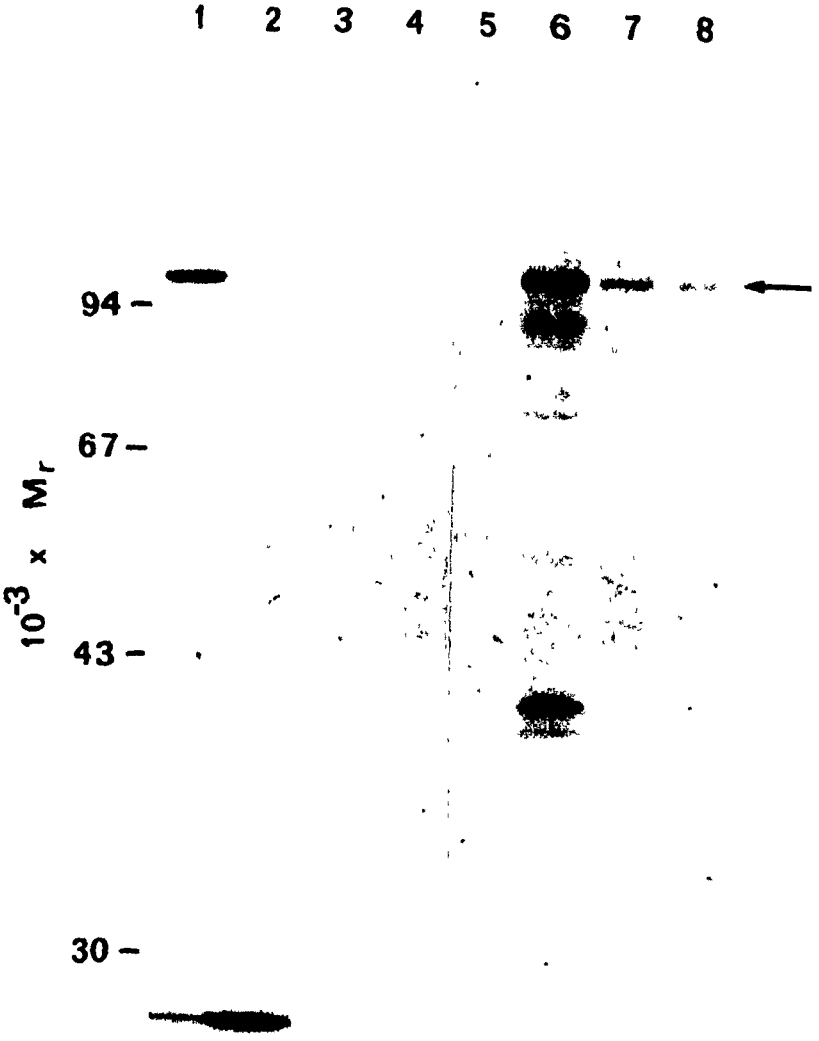

A
B
1
2
3
28 S–
18 S–

from a single population of transcripts which are essentially identical in size with the isolated cDNA clone.

Southern blot analyses of genomic DNA isolated from human leukocytes were carried out using the 1.93 kb *Eco*RI fragment from HT-2. The probe hybridizes to DNA fragments which add up to approximately 40 kb (results not shown) indicating the presence of large introns in the gene; however subsequent studies have localized genes on two separate chromosomes (Rozen *et al.*, unpublished).

Transcripts of the clone were obtained by subcloning the HT-2 cDNA into the *Eco*RI site of Bluescript-SK+ as well as Bluescript-KS+ and carrying out transcription *in vitro* with either T3 or T7 polymerase. The full length transcripts of 3.1 kb are shown in Fig. 2B showing that a single transcript was obtained using either promoter. These transcripts were then translated in a cell-free, rabbit reticulocyte lysate *in vitro* translation system to produce a major protein product of approximately 100 kDa which co-migrates with the purified human trifunctional enzyme on SDS-PAGE; however some minor lower molecular weight products were also produced (Fig. 3). The major protein (100 kDa) as well as the minor polypeptides were shown to be immunoprecipitated by a polyclonal antibody raised against the purified enzyme. Furthermore, the immunoprecipitation of all the synthesized proteins could be inhibited by preabsorbing the antiserum with purified human trifunctional enzyme prior to incubation with proteins synthesized *in vitro* (Fig. 3). The minor polypeptides of lower M_r are most probably the result of premature termination of translation by the *in vitro* translation system or the erroneous initiation of translation at an AUG codon located downstream of the actual start

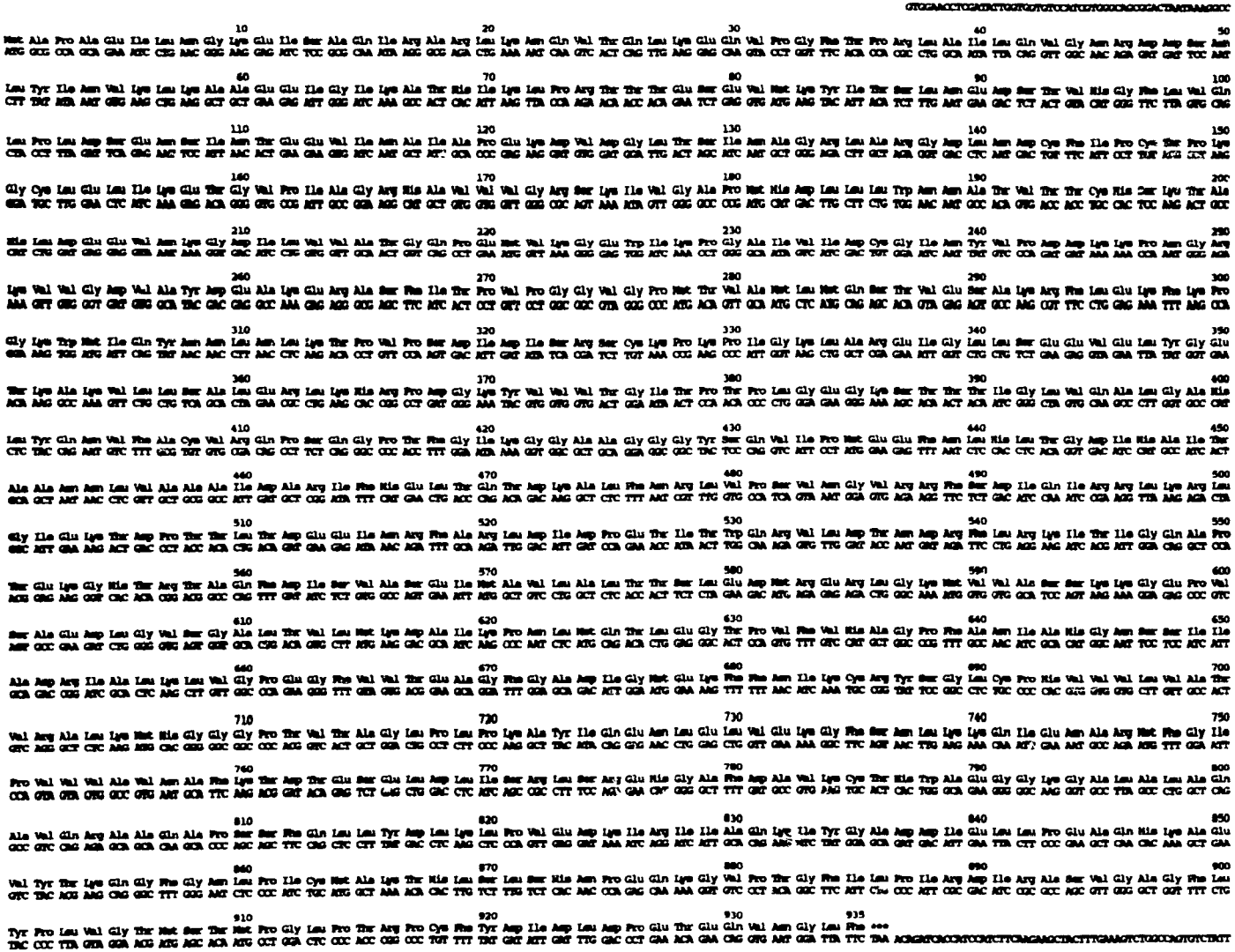
Figure 3. **In vitro translation of HT-2 transcripts and the immunoprecipitation of translation products.** Total translation products (lane 1) and immunoprecipitated proteins (lanes 4-8) were electrophoresed on a SDS-polyacrylamide gel and detected by fluorography. Equal amounts of translation products were immunoprecipitated with non immune serum (lane 4); no serum (lane 5); untreated antiserum (lane 6); or with antiserum preabsorbed with 5 μg (lane 7); or 10 μg (lane 8) of purified trifunctional enzyme. Lanes 2 and 3 are controls for translation and immunoprecipitation of products with no added transcript. The arrow indicates the position of the purified human trifunctional enzyme.


1 2 3 4 5 6 7 8
94 –
67 –
43 –
30 –
10^{-3} x M_r

codon, although some proteolysis of the translation product could also explain these bands. *In vitro* translations programmed with the non-coding transcript prepared in the same fashion did not produce any labelled protein products (data not shown).

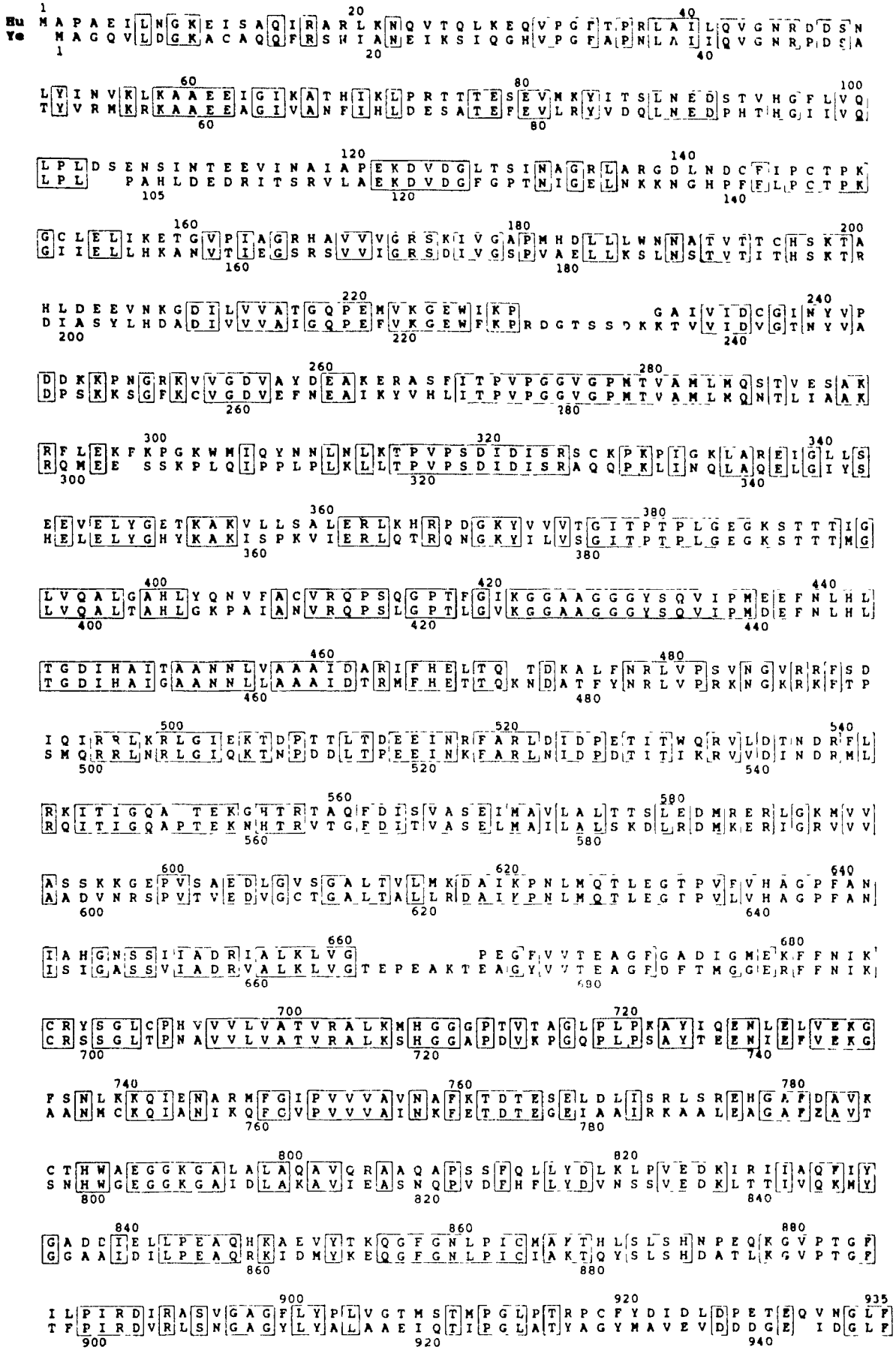
The complete sequence of the cDNA was determined on each strand of the DNA according to the strategy outlined in Fig. 1. The nucleotide sequence and the deduced amino acid sequence are shown in Fig. 4. The open reading frame of 2805 bp codes for 935 amino acids which is of sufficient length to encode the trifunctional polypeptide. The other two reading frames contain 57 and 65 termination codons distributed throughout the sequence. The deduced amino acid sequence has a calculated M_r of 101 423 excluding the initiator methionine. The calculated M_r agrees well with the estimate made from SDS-polyacrylamide gel electrophoresis (Tan *et al.*, 1977). The coding region is flanked by a 5'-untranslated region of 53 bp and a 3'-untranslated region of 259 bp which is followed by poly-A. We could deduce the location of the initiator methionine, which is not present in the purified enzyme, from the protein sequence data which begins with the amino acids Ala-Pro-Ala-Glu. Twenty-nine of the first 30 amino acids of the purified protein correspond to the sequence deduced from the cDNA. The nucleotide sequence flanking the first AUG codon in the clone is consistent with the consensus sequence proposed for initiation of translation in higher eukaryotes (Kozak, 1987). The putative translation termination codon TAA is found at position number 2806 and the 3'-untranslated region contains the polyadenylation signal AATAAA which is found 16 nucleotides upstream of the poly-A (Fig. 4).

Figure 4. **DNA sequence of HT-2 cDNA and the deduced amino acid sequence.** The amino acid sequence is numbered sequentially from the amino terminus of the predicted protein starting from the first methionine. The stop codon is indicated by (***) and the putative polyadenylation signal is underlined.



We have compared the deduced amino acid sequence of the human trifunctional enzyme to that of the yeast enzyme (Staben and Rabinowitz, 1986) and found extensive homology. When gaps are introduced into both sequences to attain maximum alignment they are 58% identical (Fig. 5), although the human form of the protein contains eleven fewer amino acids. Multifunctional enzymes have been postulated to have arisen from smaller monofunctional proteins and each enzyme activity exists as a separate active domain in the multifunctional protein. Good examples of this are the pentafunctional AROM enzyme of *Saccharomyces cerevisiae* (Duncan *et al.*, 1987) and the fatty acid synthase of most eukaryotes (Hardie and McCarthy, 1986). For both of these proteins it has been shown that the polypeptide is a mosaic of functional domains joined by linking regions and it has been proposed that the genes encoding these enzymes arose from the fusion of the smaller genes encoding the monofunctional proteins. Previous work on the mammalian trifunctional enzyme isolated from pig liver has shown that an active domain of 33 kDa from the NH_2-terminal, expressing only dehydrogenase and cyclohydrolase activities, could be isolated following limited proteolysis of the native protein with trypsin (Tan and MacKenzie, 1977). Proteolysis of the enzyme with chymotrypsin yielded a COOH-terminal domain of 67 kDa which retained only the synthetase activity (Tan and MacKenzie, 1979). Extending this work to the sequence at hand we have located a region which potentially separates the two functional domains. Secondary structure analysis using the model of Robson (Garnier *et al.*, 1978) and the LKB-PROSIS computer program has revealed that the sequence between amino acids 299 and 333 exhibits a region of disordered conformation that is flanked by α-helices and could

Figure 5. **Comparison of the deduced amino acid sequence of the human trifunctional enzyme with that of yeast.** Sequence of the human protein (Hu) is the top line and the yeast sequence (Staben and Rabinowitz, 1986) on the bottom line is labelled Ye. Both sequences are numbered sequentially from the amino terminus starting at the first methionine; the last digit of the number indicates the residue. Homologous residues are boxed.


Hu
Ye

be susceptible to proteolysis. This region corresponds to that which has been suggested to separate the two domains of the yeast enzyme (Staben and Rabinowitz, 1986) and the sequences in part of this region (Hu 299-313) are not highly homologous. This agrees well with a previous proposal that the connector regions are important for maintaining structural integrity of multifunctional proteins, but that their sequence is less important (Zalkin *et al.*, 1984). As well the proposed connector regions in both proteins correspond to a relatively hydrophilic region on the polypeptide. This suggests that the region is exposed on the surface of the protein and may be available for proteolysis. In aligning sequences to attain homology between the two enzymes we have confirmed that the two functional domains are arranged in the same order in both the mammalian and yeast enzymes, as previously determined (Tan and MacKenzie, 1979; Jones, 1977; de Mata and Rabinowitz, 1980). The amino acid sequence of the NH_2-terminal domain (Hu 2-305) is 48% homologous to the sequence of the enzyme from yeast. However, the sequences of the synthetase domains (Hu 306-935) are 63% identical and contain several extensive stretches of identical residues. The lower homology between NH_2-terminal domains of the two enzymes may explain the quantitative differences observed when studying the rates of inactivation of the dehydrogenase and cyclohydrolase with diethyl pyrocarbonate (Smith and MacKenzie, 1983; Appling and Rabinowitz, 1985). The extensive sequence homology in the synthetase domain suggests a more stringent conservation of amino acid sequence which concurs with the observation by Staoen and Rabinowitz (1983) of immunological cross-reactivity between the trifunctional eukaryotic enzymes and the monofunctional prokaryotic 10-formyltetrahydrofolate synthetases.

The formyltetrahydrofolate synthetase activity which is found in the carboxyl-terminal domain of the trifunctional enzyme utilizes MgATP as cofactor. We have identified two regions in the synthetase domain which are homologous to sequences (A and B) believed to be involved in nucleotide binding in enzymes that bind ADP or ATP (Walker *et al.*, 1982; Husain *et al.*, 1986). The sequence T-P-L-G-E-G-K-S-T-T-T-I-G-L-V at residues 380-394 of the human enzyme is similar to the A consensus sequence G-X_4-G-K-T/S-X_6-I/V. The B consensus sequence R-X_3-G-X_3-L-(hydrophobic)$_4$-D is located at residues 892-904 of the human protein and is R-A-S-V-G-A-G-F-L-Y-P-L-V-G. Both sequences correspond to homologous regions in the yeast enzyme (Fig. 5). Fry *et al.* (1986) have proposed that the glycine-rich A sequence is important for conformational change in addition to nucleotide binding. This proposal is interesting in light of the observation by Strong *et al.* (1987) that the synthetase domain of the rabbit trifunctional enzyme undergoes a detectable conformational change in the presence of the substrates tetrahydropteroylpolyglutamate and MgATP, as well as ammonium ion.

The enzyme from human sources is a total of eleven amino acids shorter than that from yeast, with significant deletions occurring in the dehydrogenase-cyclohydrolase domain after amino acid 229 and in the synthetase domain after amino acid 660. These two deletions in the human protein are also readily identified by two horizontal displacements in the hydropathy profiles of the proteins (not shown). When these areas are taken into account the hydropathy profiles of the human and yeast proteins are very similar, reflecting the sequence homology as well as a large number of conservative changes in non-homologous positions. It will be

interesting to determine if these deletions alter the properties of either domain, especially with respect to the kinetic properties of the dehydrogenase-cyclohydrolase.

CHAPTER 3

EXPRESSION OF ACTIVE DOMAINS OF A HUMAN FOLATE-DEPENDENT TRIFUNCTIONAL ENZYME IN *ESCHERICHIA COLI*

SUMMARY

The cDNA encoding the human trifunctional enzyme methylene-tetrahydrofolate dehydrogenase-methenyltetrahydrofolate cyclohydrolase-formyltetranydrofolate synthetase was engineered to contain a procaryotic ribosome binding site and was expressed under the bacteriophage T7 RNA polymerase promoter in *Escherichia coli*. Site-directed mutagenesis was used to prepare constructs that encode separately the dehydrogenase/cyclohydrolase (D/C) domain as amino acid residues 1-301, and the synthetase (Syn) domain as residues 304-935. Both domains formed active enzymes thereby demonstrating their ability to fold independently. The full length enzyme, D/C and Syn domains were expressed at levels 4, 55, and 3 fold higher than the specific activities found in liver. Additional mutagenesis and independent expression of domains further defined the interdomain region to include amino acids 292 to 310. The D/C domain was purified to homogeneity by a single affinity chromatographic step, and the full length protein in a two step procedure. The kinetic properties of the D/C domain appear unaltered from those of the trifunctional enzyme.

INTRODUCTION

The enzymes 5,10-methyleneH_4folate dehydrogenase (EC 1.5.1.5), 5,10-methenylH_4folate cyclohydrolase (EC 3.5.4.9) and 10-formylH_4folate synthetase (EC 6.3.4.3) catalyze three sequential reactions in the interconversion of 10-formylH_4folate required for purine biosynthesis, and methyleneH_4folate which is used in the synthesis of thymidylate or, after reduction, to provide the methyl group of methionine. In certain prokaryotes the three activities are separate proteins or, as is the case in *Clostridium thermoaceticum*, the dehydrogenase and cyclohydrolase can exist as a bifunctional enzyme and the synthetase as a separate protein (MacKenzie, 1984). In eukaryotes the three activities are properties of a single polypeptide of approximately 100 kDa which forms a homodimeric trifunctional enzyme as isolated from yeast (Paukert *et al.*, 1977) and the livers of rabbit (Schirch, 1978), sheep (Paukert *et al.*, 1976), pig (Tan *et al.*, 1977) and chicken (Caparelli *et al.*, 1978).

Proteolysis experiments with the mammalian trifunctional enzymes (Tan and MacKenzie, 1977, Tan and MacKenzie, 1979, Villar *et al.*, 1985) demonstrated that the protein can be separated into two functional domains: a carboxyl-terminal domain which contains only the ATP-dependent synthetase activity and corresponds to about two-thirds of the original polypeptide, and an amino-terminal domain of approximately one-third the native polypeptide which contains the cyclohydrolase and NADP-dependent dehydrogenase activities. Chemical modification of the porcine (Smith and MacKenzie, 1983) and yeast (Appling and Rabinowitz, 1985) proteins with diethyl pyrocarbonate suggested that the dehydrogenase and cyclohydrolase active sites overlap at least partially. Smith and Mackenzie (1983)

demonstrated that these two activities in the enzyme isolated from pig liver are not kinetically independent and appear to share a common folate binding site (Smith and MacKenzie, 1985).

Comparison of the primary structures of the yeast (Staben and Rabinowitz, 1986) and human (Hum *et al.*, 1988) trifunctional enzymes, deduced from the nucleotide sequences, reveals extensive homology between the two proteins. A region of low homology was proposed as a putative linker region joining the two major domains of the human enzyme (Hum *et al.*, 1988) and this area overlaps that which had been suggested to separate the two domains of the yeast enzyme (Staben and Rabinowitz, 1986). Barlowe *et al.* (1989) converted several cysteine residues to serine in the dehydrogenase/cyclohydrolase domain of the yeast enzyme by site-specific mutagenesis of the gene. The parallel effects of these substitutions on the dehydrogenase and cyclohydrolase activities support the hypothesis of overlapping active sites.

In this paper we describe the expression in *E. coli* of the full length human enzyme, and separately, its two major domains. By site-directed mutagenesis we have localized the interdomain region and demonstrated that the domains can fold independently into enzymatically active species.

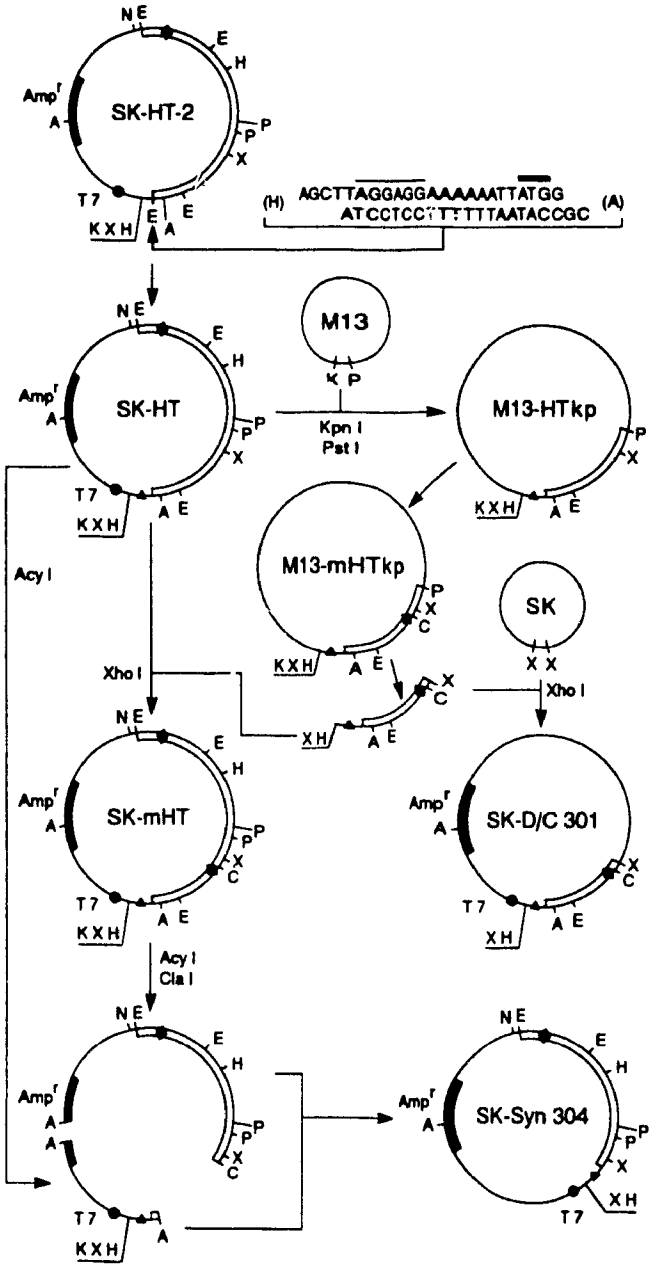
MATERIALS AND METHODS

Restriction endonucleases and DNA modification enzymes were obtained from Bethesda Research Laboratories and Boehringer Mannheim. Reagents and enzymes for DNA sequencing were from United States Biochemical Corp. All matrices for column chromatography were purchased from Pharmacia and nitrocellulose membranes were from Schleicher and Schuell. Oligonucleotides used in mutagenesis were synthesized and purified by the Regional DNA Synthesis Laboratory, University of Calgary. All common chemicals used were of the highest grade commercially available. Unless specified otherwise standard recombinant DNA techniques were carried out as described by Maniatis *et al.* (1982).

Bacterial Strains and Plasmids - The plasmid pBluescript-SK+ was used as a cloning vector in *E. coli* XL1-Blue, both obtained from Stratagene. pBluescript-SK+ was also the basis for constructs used to express exogenous proteins in *E. coli*. The K38 strain of *E. coli*, which harbours the pGP1-2 plasmid, and employed to express the proteins was a generous gift from Dr. Charles C. Richardson (Tabor and Richardson, 1985). cDNA constructs in M13mp18 were introduced into *E. coli* RZ1032 (*dut*$^-$ *ung*$^-$) (Kunkel *et al.*, 1987) to produce single stranded uracil-containing DNA templates for oligonucleotide-directed mutagenesis.

Construction of Expression Plasmids - The full length cDNA clone HT-2 described earlier (Hum *et al.*, 1988) was used for construction of expression plasmids as outlined in Fig. 1.

Figure 1. **Construction of the expression plasmids SK-HT, SK-D/C 301 and SK-Syn 304.** The oligonucleotides used to insert the RBS are shown; the RBS is *overlined* and the initiator ATG is *overlined in bold*. The bacteriophage T7 RNA polymerase promotor is indicated by ● , the inserted RBS by ▲ and the stop codon by ◆ . Abbreviations for restriction endonuclease sites: H, *Hind*III; K, *Kpn*I; N, *Not*I; P, *Pst*I; X, *Xho*I; A, *Acy*I; C, *Cla*I; E, *Eco*RI.


SK-HT-2
Amp^r
T7
KXH
(H) AGCTTAGGAGGAAAAAATTATGG
ATCCTCCTTTTTTAATACCGC (A)
M13
Kpn I
Pst I
M13-HTkp
SK-HT
M13-mHTkp
SK
Acy I
Xho I
SK-mHT
SK-D/C 301
Acy I
Cla I
SK-Syn 304

The complementary oligonucleotides *5'*AGCTTAGGAGGAAAAAATTATGG*3'* and *5'*CGCCATAATTTTTTCCTCCTA*3'* were mixed and hybridized at 37°C for 2 hrs in 50mM Tris-Cl pH 7.5, 10mM $MgCl_2$ to form a synthetic *Hind*III-*Acv*I fragment This fragment was designed to insert an efficient ribosome binding site (RBS) into the 5'-untranslated region of the HT-2 cDNA. The *Hind*III-*Not*I fragment of pBluescript-SK+ and the *Acy*I-*Not*I fragment of HT-2 were mixed at approximately equal molar ratio with the synthetic fragment and ligated to construct the vector (SK-HT) used for expression of the full length trifunctional protein (HT). The correct sequence of the RBS was confirmed by sequence analysis. This construct contained the coding sequence, stop codon and the entire 3'-untranslated region of the cDNA.

Oligonucleotide-directed mutagenesis and subsequent subcloning were used to construct vectors that would express separately the dehydrogenase/ cyclohydrolase (D/C) and the synthetase (Syn) domains. The terminology used incorporates the amino acid residue numbers of the human enzyme. The amino-terminal domain represented by D/C 301 is a protein that terminates following amino acid residue 301 at an introduced stop codon Syn 304, on the other hand, represents a domain where, through mutagenesis, initiation begins just prior to residue 304 and terminates at the normal stop codon (Fig. 2). These proteins are expressed by the corresponding SK-D/C 301 and SK-Syn 304 plasmids.

SK-D/C 301 was constructed in the following way. The 1.3 kb *Kpn*I-*Pst*I fragment from SK-HT was subcloned into the multiple cloning site of M13mp18. Single stranded template was prepared and the mutagenic oligonucleotide *5'*GGTTGTTATACTGAATCATCGATTATCCTGGCTTAAATTTCTCC*3'* was used to introduce both a stop codon (TAA) and a *Cla*I restriction site in M13-

mHTkp. The 1.1 kb *Xho*I-*Xho*I fragment from M13-mHTkp was subcloned into the *Xho*I site of pBluescript-SK to form SK-D/C 301. The same fragment was ligated with the 5.0 kb *Xho*I-*Xho*I fragment of SK-HT to form SK-mHT. SK-D/C 301 was used for expression of the dehydrogenase-cyclohydrolase domain D/C 301.

Construction of SK-Syn 304 for expresssion of the synthetase domain Syn 304 was achieved by ligation of the 4.0 kb *Acy*I-*Cla*I fragment of SK-mHT with the 1.1 kb *Acy*I-*Acy*I fragment of SK-HT. Using this strategy only the transformants which contain the insert in the correct orientation are ampicillin resistant. This manipulation introduced an additional codon for alanine after the initiator methionine, in frame with the coding sequence for Syn 304 (Fig. 2).

Oligonucleotide-directed mutagenesis was carried out essentially as described by Zoller and Smith (1987) using uracil-rich single-stranded DNA from recombinant M13-HTkp following the method of Kunkel *et al.* (1987). Mutants were selected by plaque-lift hybridization with the labelled mutagenic oligonucleotide, followed by sequential washes of increasing stringency. The 1.1 kb *Xho*I-*Xho*I insert in SK-D/C 301 was sequenced to confirm the desired base changes and to ensure that there were no secondary mutations. All sequences were derived from dideoxy sequencing (Sanger *et al* , 1977) of double stranded DNA.

Plasmids for expression of D/C 312 and Syn 315, D/C 296 and Syn 300, as well as D/C 291, were constructed as described above for D/C 301 and Syn 304 and used synthetic oligonucleotides complementary to sequences shown in Table 1.

Figure 2. **The strategy employed for expression of the full length protein, the amino-terminal dehydrogenase/cyclohydrolase domain (D/C 301) and the carboxyl-terminal synthetase domain (Syn 304).** The amino acid residues (299-313) of the proposed interdomain region are shown on the top line and the initiator methionine is designated number 1. The second line represents the corresponding nucleotide coding sequence of the cDNA and the adenine of the initiator ATG is designated number 54 as previously reported for the full length cDNA (Hum *et al.*, 1988). The sequence of the RBS is indicated by the *broken line* and the *broken line in bold* shows the *Acy*I restriction site. Nucleotide changes in the linker region are indicated by * . *Underlined* is the introduced stop codon and *underlined in bold* is the restriction site recognized by *Cla*I. Translation of the amino-terminal domain is terminated after residue 301 as indicated by ● . The expression plasmid SK-Syn 304 was constructed by religation of the compatible *Acy*I and *Cla*I sites.

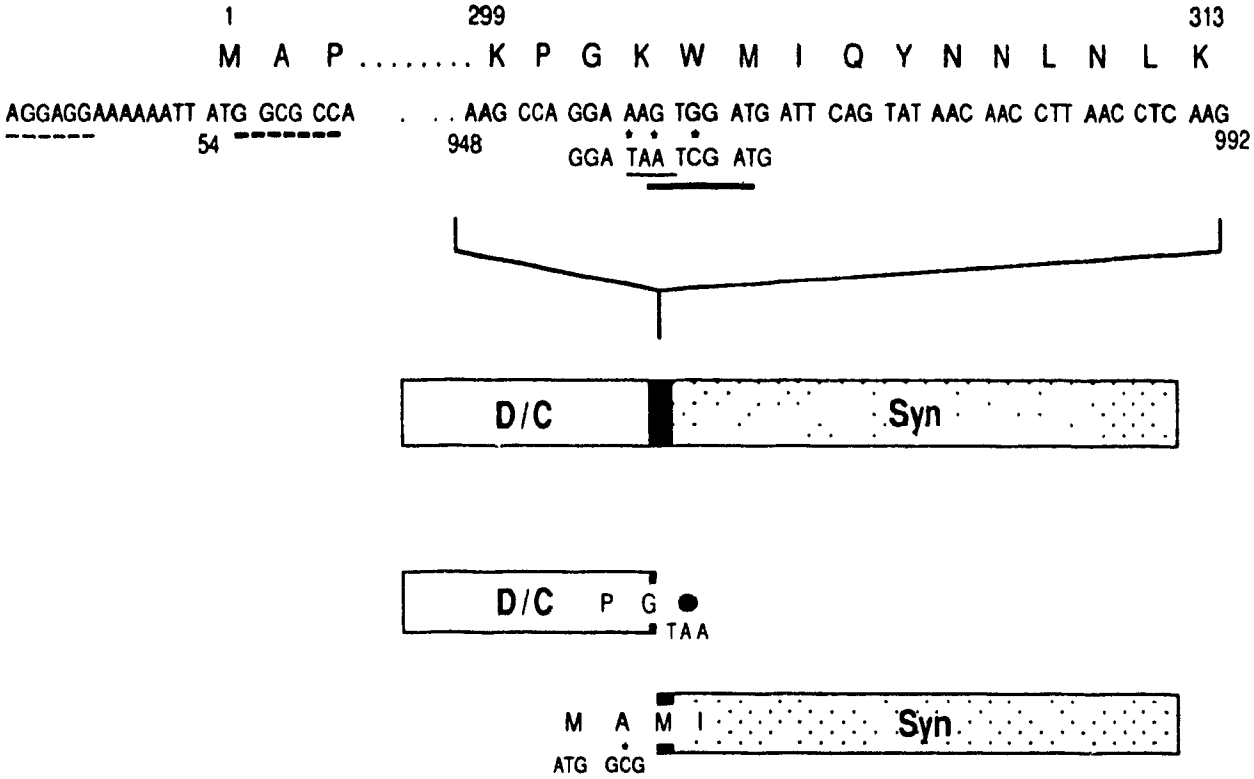

1
299
313
M A P ..... K P G K W M I Q Y N N L N L K
AGGAGGAAAAAATT ATG GCG CCA . . . AAG CCA GGA AAG TGG ATG ATT CAG TAT AAC AAC CTT AAC CTC AAG
54
948
992
GGA TAA TCG ATG
D/C
Syn
D/C P G
TAA
M A M I
Syn
ATG GCG

Table 1 Synthetic oligonucleotides used for mutagenesis

Plasmid	Coding Sequence of Mutants	Restriction Site
SK-D/C 312 SK-Syn 315	980 5' C CTT AAC CTC T*AG G*C*G CC*T GTT CCA AGT G	Acy I
SK-D/C 301 SK-Syn 304	938 5' G GAG AAA TTT AAG CCA GGA T*A*A T*CG AT*G ATT CAG TAT AAC AAC C	Cla I
SK-D/C 296 SK-Syn 300	929 5' G CGT TTC CTG GAG T*AA TT*G A*CG CCA GGA AAG TGG	Acy I
SK-D/C 291	915 5' GTA GAG AGT GCC T*A*A G*C*T TTC CTG GAG	Hind III
SK-Syn 309	961 5' GG ATG ATT CAG TAT G*T*T AAC CTT AAC CTC	Hpa I / Hinc II
SK-Syn 310	965 5' G ATT CAG TAT AAC C*A*G C*T*G AAC CTC AAG ACA	Pvu II
SK-Syn 311	(deletion mutant)	
SK-Syn 316	(deletion mutant)	
	915 5' GTA GAG AGT GCC AAG CGT TTC CTG GAG AAA TTT AAG CCA GGA AAG TGG ATG ATT CAG TAT AAC AAC CTT AAC CTC AAG ACA CCT GTT CCA AGT G 3' 1008	

Complementary oligonucleotides made to the sequences shown, were used to construct the various plasmids. Base changes are indicated by * , and *underlined* sequences are restriction endonuclease sites. Plasmids SK-Syn 311 and SK-Syn 316 were constructed by nucleotide deletions using exonuclease III. The unmutated nucleotide coding sequence is shown at the bottom.

To construct the plasmids for expression of Syn 309 and Syn 310, oligonucleotide-directed mutagenesis was used to introduce a unique restriction site (Table 1) in the proposed interdomain region following the procedure described above for M13-mHTkp. The mutated DNA was digested at the introduced restriction site and with *Nco*I to remove the region of the cDNA encoding the dehydrogenase/cyclohydrolase domain and the ends of the DNA were repaired with the Klenow fragment of *E. coli* DNA polymerase I. Appropriately sized DNA fragments were purified and ligated with T4 DNA ligase. Following transformation of *E. coli*, DNA was isolated and sequenced to confirm the proper sequence for the RBS and for initiation of translation at the desired position.

To construct the plasmids for expression of Syn 311 and Syn 316, SK-mHT was restricted with *Cla*I and nested-deletions were introduced using exonuclease III and mung bean nuclease as outlined by the supplier. Appropriately sized DNA fragments were purified, digested with *Nco*I and the ends were repaired with Klenow fragment prior to ligation with T4 DNA ligase. DNA isolated from 48 transformants was sequenced to identify constructs which initiated translation at the desired position.

Expression of the Constructs - Expression of the full length trifunctional enzyme as well as the D/C and Syn domains was carried out essentially as described by Tabor and Richardson (1985) for expression of the bacteriophage T7 gene 5 protein, using a coupled T7 RNA polymerase/promotor system. The pBluescript-SK+ vector carrying the DNA of interest downstream of the T7 RNA polymerase promotor was introduced into K38 cells which harbour the pGP1-2 vector. This vector contains the

gene for the heat-sensitive lambda repressor, CI857, and the gene for T7 RNA polymerase under the control of the inducible lambda P_L promotor. Transformed K38 cells were grown to an $O.D._{600}$ of 0.8-1.0 at 30°C in LB-broth containing 100 μg/ml ampicillin and 50 μg/ml kanamycin. The temperature was increased to 42°C for 30 min to induce T7 RNA polymerase expression. Cells were then incubated at 30°C for an additional 40 min to allow continued transcription by T7 RNA polymerase. After cooling on ice for 10 min the K38 cells were harvested by centrifugation at 10 000 x g for 10 min and cell pellets were stored at -70°C.

Enzyme Assays and Channelling Experiments - The dehydrogenase, cyclohydrolase and synthetase assays were performed as previously described (Tan *et al.*, 1977) as were channelling experiments (Cohen and MacKenzie, 1978; Rios-Orlandi and MacKenzie, 1988). The efficiency of channelling was calculated as a ratio of the rate of 10-formylH_4folate production to the rate of 5,10-methyleneH_4folate utilization and is expressed as percentage. Protein samples were precipitated by the method of Bensadoun and Weinstein (1976) and protein determination was by the method of Lowrey *et al.* (1951).

Enzyme Purification - Approximately 4 ml of packed *E. coli* cells were sonicated in 8 ml of lysis buffer (0.1 M potassium phosphate, pH 7.3, 35 mM 2-mercaptoethanol, 1 mM PMSF, 1 mM benzamidine, 0.5% nonidet P-40) for 2 min (8 X 15 sec separaced by 30 sec on ice). The cell suspension was diluted to 0.05 M potassium phosphate by the addition of 8 ml of dilution buffer (35 mM 2-mercaptoethanol, 1 mM PMSF, 1 mM benzamidine), sonicated

for an additional minute and glycerol was added to a final concentration of 20 %. The total extract was clarified by centrifugation at 20 000 x g for 30 min at 4°C. After centrifugation the resulting pellet was resuspended in lysis buffer and was sonicated as described above. The supernatants from both extractions were pooled, protamine sulfate (10 mg/ml, pH 7.0) was added to a final concentration of 1 mg/ml, stirred for 30 min on ice and centrifuged at 20 000 x g for 30 min at 4°C to remove precipitated nucleic acids. The supernatant was applied to a 2.6 x 3 cm column of 2',5'-ADP Sepharose (168 ml/hr) that was equilibrated with buffer A (50 mM potassium phosphate, pH 7.3, 35 mM 2-mercaptoethanol, 1 mM PMSF, 1 mM benzamidine, 20% glycerol). The column was washed (168 ml/hr) sequentially with 56 ml (3.5 column volumes) of buffer A, 32 ml of a linear gradient of 0.05 M to 2 M potassium phosphate, pH 7.3 in buffer A, 32 ml of 2 M potassium phosphate, pH 7.3 in buffer A and 48 ml of buffer A. The enzyme was eluted (156 ml/hr) with buffer A containing 7 mM NADP. The eluate was applied directly onto a 1.6 x 2 cm column of DEAE-sepharose (connected in line) and the flow-through was collected in 4 ml fractions. This procedure produced homogeneous dehydrogenase/cyclohydrolase protein.

Purification of the full length trifunctional enzyme required the following additional steps. Flow-through fractions from the DEAE-Sepharose column which contained active enzyme were pooled (24 ml) and applied to a G-25 desalting column (2.6 x 28 cm) equilibrated and eluted at 4°C with buffer C (25 mM KPi, pH7.3, 35 mM BME, 1 mM PMSF, 20% glycerol). The active enzyme was collected (approx. 45 ml) and applied (60 ml/hr) to a Mono Q column equilibrated with the same buffer. The column was washed with 25 ml of buffer C and a 20 ml linear gradient of

0 M to 0.1 M KCl in buffer C was used to effect elution.

Western Analysis - Supernatant from *E. coli* extract as well as purified proteins were precipitated by the method of Bensadoun and Weinstein (1976) and were separated on SDS-polyacrylamide slab gels as described by Laemmli (1970). Proteins were transfered electrophoretically onto nitrocellulose paper following the general method of Towbin *et al.* (1979). A polyclonal antibody raised against the trifunctional protein (Hum *et al.*, 1988) was diluted 1:500 and was used as the primary antibody. Protein bands were visualized using an alkaline phosphatase conjugated anti-rabbit second antibody, employing the Protoblot immunoscreening system from Promega Biotec. Alternatively, visualization of protein bands with ^{125}I-protein A followed the procedure used for screening a lambda gt11 library as was described previously (Hum *et al.*, 1988).

RESULTS

Expression of the Trifunctional Enzyme and its Two Major Domains - Construction of the plasmids is described in detail in Materials and Methods. Expression of all the cDNA constructs used in these studies was accomplished by transcription from the original T7 RNA polymerase promotor of the pBluescript-SK+ vector. A ribosome binding site (RBS) was inserted between the T7 promotor and the initiator ATG to construct SK-HT, which was used for expression of the complete coding region of HT-2 (Fig. 1). All expression plasmids contained the same RBS and the identical nucleotide sequence through the initiator ATG except for SK-Syn 309, 310, 311 and 316 in which two cytosines replaced two thymidines immediately 5' of the initiator ATG.

The construction of SK-D/C 301 for initial expression of the amino-terminal dehydrogenase/cyclohydrolase domain (D/C 301) utilized oligonucleotide-directed site-specific mutagenesis to introduce a stop codon (TAA) and a unique restriction site (*Cla*I) into the region of the cDNA which was proposed to code for the interdomain region of the trifunctional enzyme (Fig. 2). The plasmid SK-Syn 304 for initial expression of the carboxyl-terminal synthetase domain (Syn 304) was constructed by deleting the region of cDNA encoding the D/C domain (Fig. 1, Fig. 2).

SK-HT, SK-D/C 301 and SK-Syn 304 were transfected separately into K38 cells. A significant amount of the human trifunctional enzyme was expressed and was only partially soluble in *E. coli* extracts (Fig. 3A). All three constructs expressed proteins that could be detected in the soluble fractions of *E. coli* extracts by Western analysis as bands of 100

kDa (HT), 34 kDa (D/C 301) and 67 kDa (Syn 304) in good agreement with predicted values of Mr (Fig 3B). The 67 kDa protein expressed by SK-Syn 304 is the least stable of the exogenous proteins and was partially degraded in *E. coli* extract yielding a break-down product of 38 kDa. The proteins detected by the polyclonal antibody were not found in control extracts of *E. coli* cells transformed with pBluescript-SK+.

Cells transformed with SK-HT and shown to express a 100 kDa protein contain all three activities of the trifunctional enzyme at levels significantly above control K38 cells (Table 2). Crude extracts containing the exogenous 34 kDa polypeptide have markedly increased levels of dehydrogenase and cyclohydrolase activities but no detectable synthetase activity. On the other hand, cells transformed with SK-Syn 304 and which expressed a 67 kDa polypeptide contained significant synthetase activity.

Figure 3. **Immunoblot analysis of the supernatant fraction of the lysates of E. coli transformed with SK-HT, SK-D/C 301 and SK-Syn 304.** In A, 100 μg of protein from whole cell lysates (lanes 2 and 4) and from supernatant obtained after centrifugation at 20 000 x g for 30 min (lanes 1 and 3) were analyzed by Western blot. K38 cells transformed with SK-HT were either induced (lanes 1 and 2) or not induced (lanes 3 and 4) for expression of exogenous proteins. In B, each lane contains 100 μg of protein from the supernatant fractions of the lysates of *E. coli* transformed with pBluescript-SK+ (lane 1), SK-HT (lane 2), SK-D/C 301 (lane 3) and SK-Syn 304 (lane 4). The indicated protein molecular weight markers are indicated in kilodaltons.

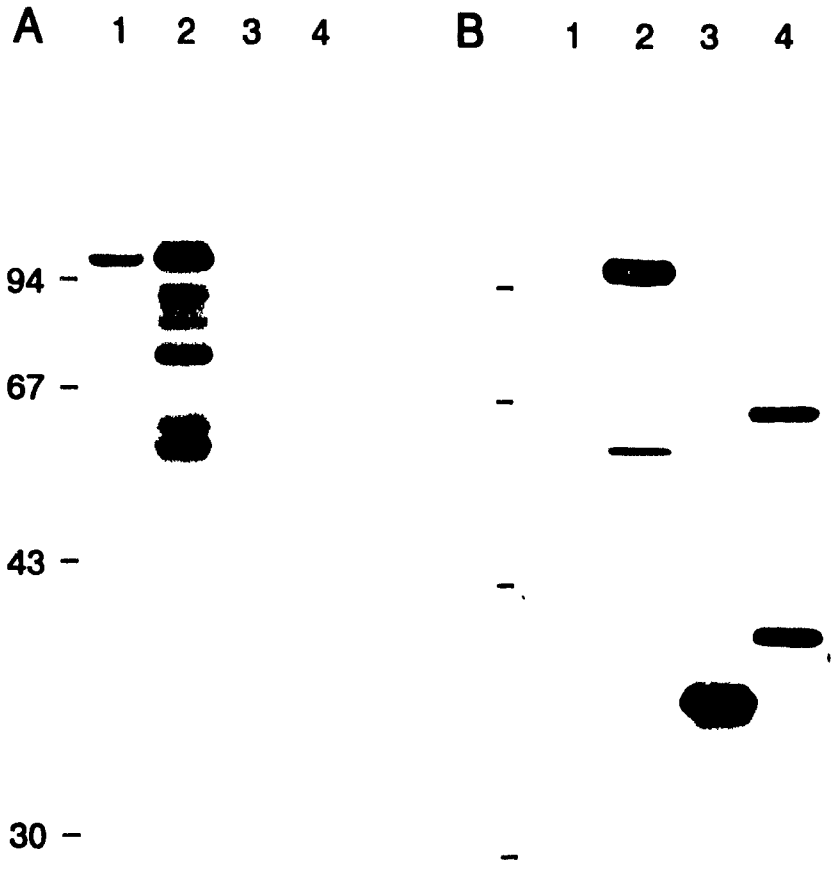

A
1
2
3
4
B
1
2
3
4
94
67
43
30

Table 2	Specific Activities of Enzymes		
Enzyme	Dehydrogenase	Cyclohydrolase	Synthetase
HT (E) [a]	0.038	0.18	0.033
(P) [b]	10.0	32.0	13.8
D/C 301 (E)	0.52	3.1	ND [c]
(P)	32.7	120	ND
SYN 304 (E)	---	---	0.064
E. coli (E)	0.002	0.003	ND

The specific activities are expressed as umol $min^{-1} mg^{-1}$

[a] Total extract from the supernatant fraction of lysates.

[b] Purified protein

[c] ND, activity not detected

The values for the purified proteins are averages of two separate experiments in duplicate. The variations between measurements were within 10%.

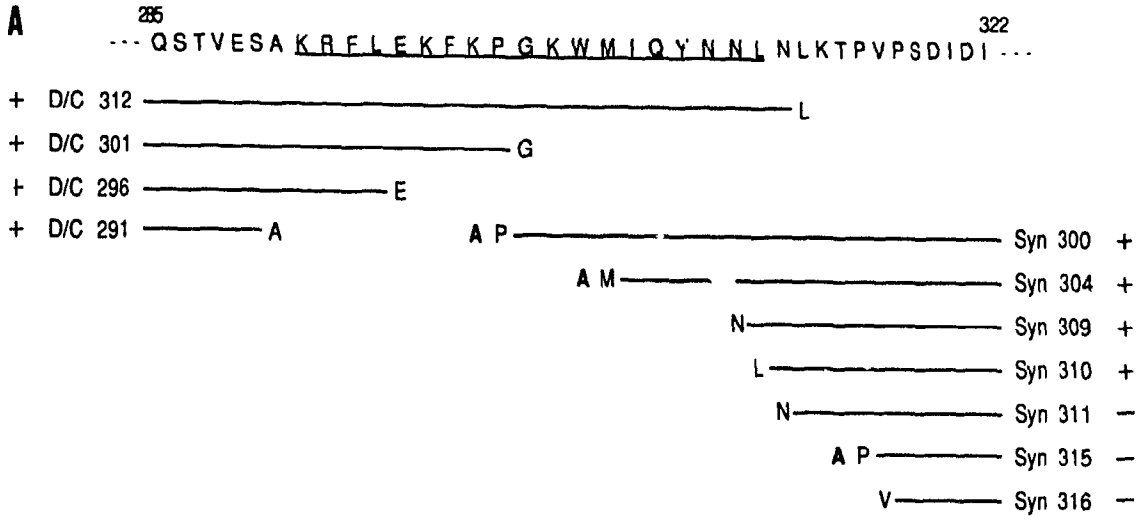
Location of Domain Boundaries - Because the initial domain constructs D/C 301 and Syn 304 expressed functional enzymes, it was possible to explore further the extent of the interdomain region. The nucleotide sequence in the proposed interdomain region was mutated to allow for expression of various lengths of the amino-terminal and carboxyl-terminal polypeptides (Fig. 4A). For expression of the dehydrogenase/cyclohydrolase domain, nonsense codons were introduced at different positions in this region to express polypeptides which differed in length at their carboxyl-end. Polypeptides which terminated after residue 291, where the stop codon replaced a highly conserved lysine residue (Bélanger and MacKenzie, 1989), retained both the dehydrogenase and cyclohydrolase activities. Construct SK-D/C 291 expressed a polypeptide that was extensively degraded and was highly insoluble in *E. coli* extracts. However, when the low amount of enzyme that remained soluble was concentrated following affinity chromatography on 2',5'-ADP Sepharose, it was found to be active This is in sharp contrast to D/C 296, D/C 301 and D/C 312 which are active, relatively stable and mostly soluble proteins. When the extracts of cells were chromatographed on a 2',5'-ADP Sepharose column, and protein corresponding to 0.006 units of dehydrogenase activity were electrophoresed on SDS-PAGE, the amounts of the dehydrogenase/ cyclohydrolase proteins were approximately equal as visualized by Western analysis (Fig. 4B).

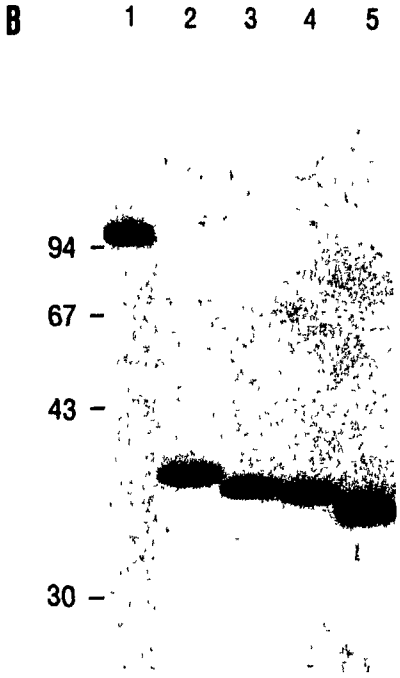
To determine the boundary of the synthetase domain, constructs were made to express carboxyl-terminal polypeptides which varied in length at their amino-ends where translation was initiated at different positions in the proposed interdomain region and ended at the original stop codon.

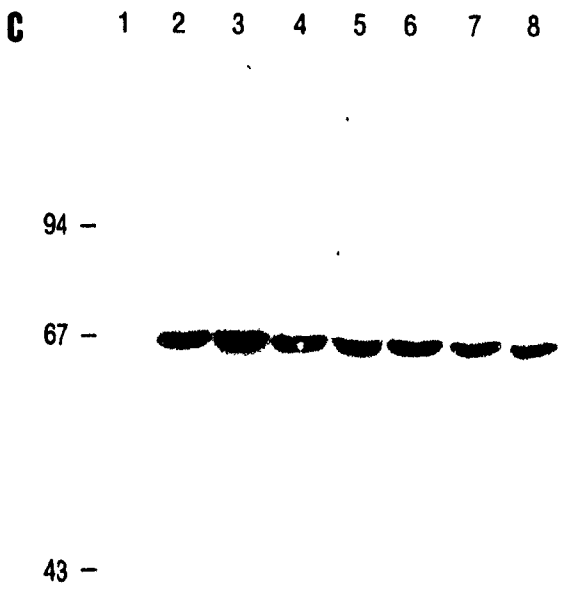
Figure 4. **Analysis of the interdomain region.** A: The sites of termination of translation for expression of the amino-terminal D/C domain and the sites of initiation of translation for expression of the carboxyl-terminal Syn domain are indicated (initiator methionine not shown). *E. coli* K38 cells transformed with the appropriate expression plasmids were induced for expression and the total extracts were assayed for enzyme activity. The presence and absence of detectable enzyme activity are indicated by (+) and (-) respectively. The deduced interdomain region of 19 amino acid residues is *underlined*.

B: Extracts from cells induced to express D/C 312 (lane 2), D/C 301 (lane 3), D/C 296 (lane 4) and D/C 291 (lane 5) were purified on 2',5'-ADP Sepharose columns. Aliquots representing 0.006 units of dehydrogenase activity were analyzed by Western blot. Proteins were visualized by fluorography. Lane 1 contains 0.005 units of dehydrogenase activity of the purified human trifunctional enzyme.

C: Aliquots of the supernatant fraction of whole cell lysates of *E. coli* cells induced to express Syn 300 (lane 2), Syn 304 (lane 3), Syn 309 (lane 4), Syn 310 (lane 5), Syn 311 (lane 6), Syn 315 (lane 7) and Syn 316 (lane 8) were analyzed by Western blot. Proteins in lane 1 were from cells transformed with pBluescript-SK+. Lanes 2 and 3 contained 30 μg and lanes 4-8 75 μg of protein.


A
285
--- Q S T V E S A K R F L E K F K P G K W M I Q Y N N L N L K T P V P S D I D I ---
322
+ D/C 312
+ D/C 301
+ D/C 296
+ D/C 291
L
G
E
A
A P
Syn 300 +
A M
Syn 304 +
N
Syn 309 +
L
Syn 310 +
N
Syn 311 –
A P
Syn 315 –
V
Syn 316 –

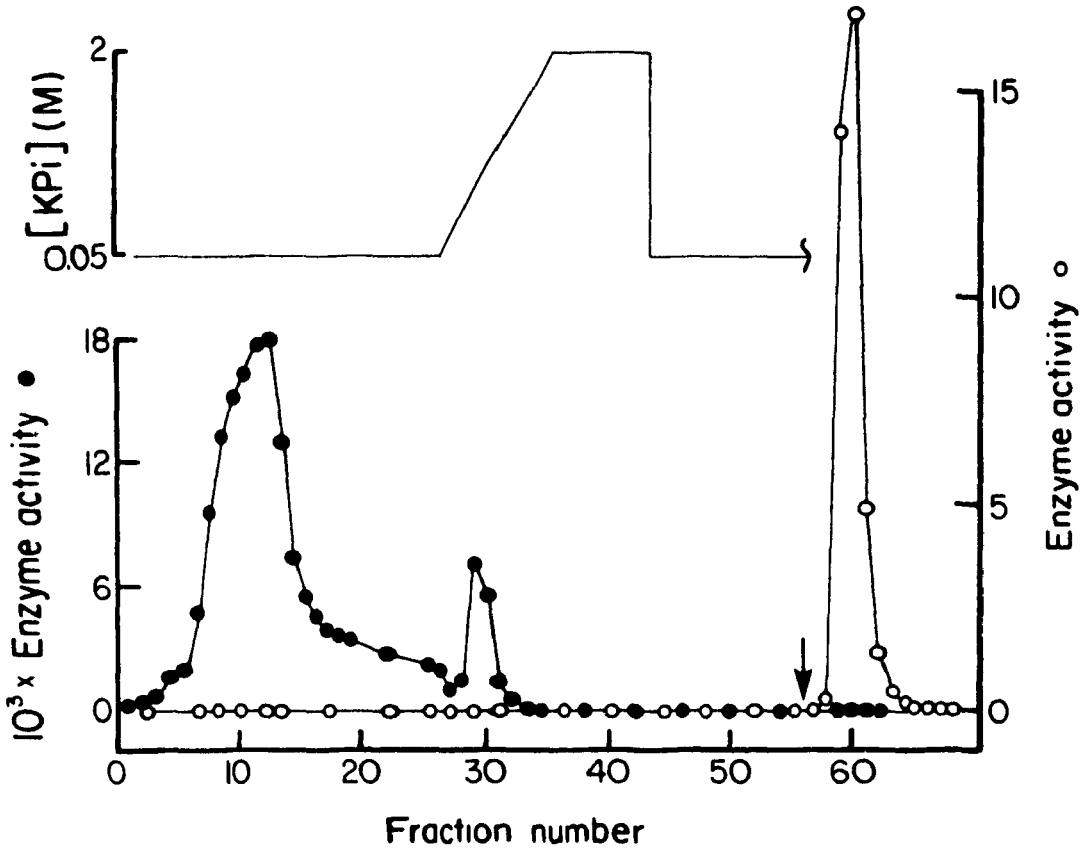

B
1 2 3 4 5
94 –
67 –
43 –
30 –


C
1 2 3 4 5 6 7 8
94 –
67 –
43 –

All constructs expressed soluble synthetase protein as shown by the Western analysis of *E. coli* extracts in Fig. 4C. Constructs where initiation began at or before residue 310 expressed proteins which retained detectable synthetase actvity. However, no synthetase activity could be detected in *E. coli* extracts containing Syn 311, Syn 315 or Syn 316. Under the conditions used in these studies a specific activity of 0.2 nmol min^{-1} mg^{-1} of protein was taken as the minimum to unambiguously establish expression of activity. The control extracts of cells transformed with pBluescript-SK+ exhibit an apparent specific activity of not more than 0.05 nmol min^{-1} mg^{-1}. The specific activities of the active synthetase domains (Syn 300, 304, 309 and 310) in *E. coli* extracts were at least 1 nmol min^{-1} mg^{-1} and the activities of Syn 311, 315 and 316 were indistinguishable from that of the pBluescript-SK+ control.

Purification of Enzymes - Determination of the functional boundaries of the two domains showed that D/C 301 terminated near the middle of the sequence linking functional domains, perhaps in the true "linker" sequence, and was selected as appropriate for purification. The D/C domain, as well as the full length trifunctional protein, both over-expressed in *E. coli*, were purified by FPLC. Separation of the exogenous human enzymes from the endogenous *E. coli* enzyme relied on their differences in affinity for a 2',5'-ADP Sepharose column (Fig. 5). The endogenous methyleneH$_4$folate dehydrogenase of *E. coli* binds only weakly to 2',5'-ADP Sepharose, and under the conditions used, most of the enzyme either did not bind to the column or was eluted by low ionic strength at the beginning of the phosphate gradient.

Figure 5. **Separation of the exogenous dehydrogenase/cyclohydrolase domain (D/C 301) from the endogenous dehydrogenase enzyme in E. coli cell extracts.** The supernatant fraction of whole lysate of cells transformed with SK-D/C 301 (350 mg of protein) was applied to a 2.6 x 3 cm 2',5'-ADP Sepharose column. The column was washed and eluted as described in Materials and Methods and the eluted fractions (4 ml each) were assayed for dehydrogenase activity (○). The addition of 7 mM NADP to the column to effect elution of D/C 301 is indicated by (↓). An extract from cells transformed with pBluescript-SK+ (380 mg of protein) was processed under identical conditions and the dehydrogenase activity (●) in the eluted fractions was determined. The gradient of phosphate buffer (KPi) from 0.05 M to 2 M is indicated on the top portion of the graph. Enzyme activities are expressed as $\mu mol/min^{-1}/ml^{-1}$.


2
[KPi] (M)
0.05
18
12
6
0
10^3 x Enzyme activity •
15
10
5
0
Enzyme activity ○
0
10
20
30
40
50
60
Fraction number

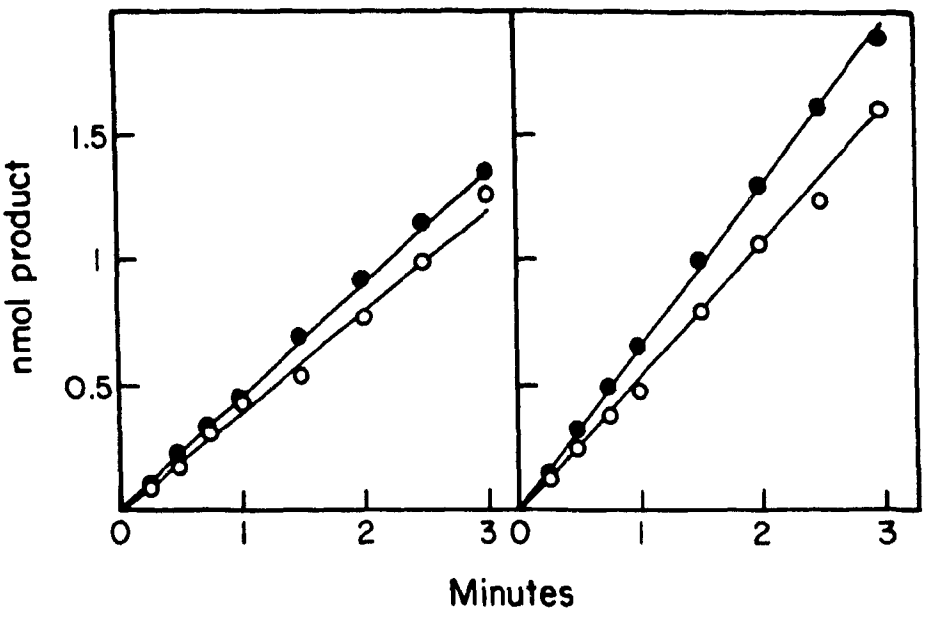
The exogenous 34 kDa polypeptide was purified in one simple chromatographic step by FPLC. From approximately 4 g of packed cells we were able to purify 5 mg of the active dehydrogenase/cyclohydrolase domain (D/C 301) with 80% recovery. Purification of the full length enzyme followed the same protocol as for the D/C domain, but two additional polypeptides of 66 kDa and 42 kDa were present in the flow through fractions of the DEAE-sepharose column (results not shown). The 100 kDa protein which was approximately 85% pure at this stage was further purified on a Mono Q column. The 100 kDa trifunctional enzyme and the 34 kDa dehydrogenase/cyclohydrolase domain were both purified to homogeneity as shown by SDS-PAGE (Fig. 6).

Experiments with the trifunctional enzyme and the D/C domain demonstrated that in each case approximately 45% of the methenylH_4folate intermediate produced by the dehydrogenase reaction was preferentially converted to 10-formylH_4folate by the cyclohydrolase reaction (channelling) (Fig. 7). As in previous studies with the trifunctional enzyme from porcine liver (Cohen and MacKenzie, 1978), the absence of a lag period in production of 10-formylH_4folate is indicative that the methenylH_4folate produced does not equilibrate with the medium before being acted upon by the cyclohydrolase. Also, as previously described for the porcine enzyme (Smith and MacKenzie, 1983), NADP, which is a cofactor for the dehydrogenase reaction, inhibits the cyclohydrolase activity of the human enzymes purified from *E. coli*. (data not shown).

Figure 6. **SDS-PAGE of fractions from the purification procedure.** Proteins were electrophoresed on a 9% acrylamide gel and were stained with Coomassie Brillant Blue. The gel contains: 100 μg of protein from the supernatant fraction of lysates of cells transformed with pBluescript-SK+ (lane 1), and SK-HT, (lane 2); 10 μg of purified full length trifunctional protein, (lane 3); 100 μg of protein from cells transformed with SK-D/C 301 which contained 0.05 units of dehydrogenase activity (lane 4). A sample of the lysate shown in lane 4 was applied to a 2',5'-ADP sepharose column and 100 μg of protein from the flow-through fraction is shown in lane 5. Lanes 6 and 7 contain 0.05 units of dehydrogenase activity and 10 μg respectively of the purified dehydrogenase/cyclohydrolase domain (D/C 301). Lane 8 contains 100 μg of protein from cells transformed with SK-Syn 304 and represents 0.006 units of synthetase activity.


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

Figure 7. **Time course of appearance of products used to determine efficiency of channelling.** The products represented are: ●, 5,10-methenylH_4folate and ○, 10-formylH_4folate. (A) shows the results for the trifunctional human enzyme (HT) and (B) the dehydrogenase/cyclohydrolase domain (D/C 301) purified from *E. coli* extracts.


nmol product
1.5
1
0.5
0
1
2
3
0
1
2
3
Minutes

DISCUSSION

We have been particularly interested in the nature of the kinetic interdependency and the possible physical overlap of the dehydrogenase/cyclohydrolase activities of this trifunctional enzyme (Smith and MacKenzie, 1983, 1985). Expression of the human enzyme in *E. coli* provides a more readily available source than human liver, with a 3-4 fold increase in level. Although some of the protein was insoluble in extracts of *E. coli*, it was possible to express approximately 0.5 mg of soluble enzyme per gram of cells. It has been reported that the level of soluble protein expressed in *E. coli* increases at lower temperatures (Schein, 1989), however, expression of the trifunctional enzyme at 30^{0}C, where T7 RNA polymerase was provided by bacteriophage T7 (CE6) infection (Studier and Moffatt, 1986), did not increase the amounts of soluble protein (results not shown). Expression of the active synthetase domain in *E. coli* was also achieved but attempts to purify the domain were unsuccessful as the 67 kDa polypeptide was largely insoluble and more labile than the full length enzyme. However, the expression of the dehydrogenase/cyclohydrolase domain at much higher levels (50 fold), coupled with a very simple purification procedure allows us to obtain good quantities of this bifunctional domain, free of possible contamination by the *E. coli* activities.

It has been postulated that multifunctional enzymes arise from smaller monofunctional proteins where each enzyme activity exists as a separate functional domain in the multifunctional protein, as exemplified by the mammalian trifunctional CAD proteins (Simmer *et al.*, 1989) and the eukaryotic fatty acid synthases (Hardie and McCarthy, 1986). The folate-

dependent eukaryotic trifunctional enzyme (dehydrogenase-cyclohydrolase-synthetase) is similar to many other multifunctional proteins where a series of functional domains are connected by protease-sensitive "linker regions" (Schmincke-Ott and Bisswanger, 1980). Previous work on the mammalian trifunctional enzyme has shown that the protein can be separated into two active domains following limited proteolysis (Tan and MacKenzie, 1977, 1979; Villar *et al.*, 1985). Combining the results of the proteolysis studies with analysis of the primary and predicted secondary structure of the protein, we previously proposed that the sequence between amino acids 299 and 313 of the human trifunctional enzyme could function as the "linker region" between the two major domains (Hum *et al.*, 1988). This sequence is predicted to exhibit a region of disordered conformation and corresponds to a relatively hydrophilic region of the polypeptide which suggests that it is exposed on the surface of the protein and thus would be susceptible to proteolytic cleavage. Staben and Rabinowitz (1986) proposed that a stretch of amino acids from position 306 to 321 of the yeast enzyme might serve as the interdomain region. More recently, Thigpen *et al.*, (1990) suggested that the interdomain region of the rat enzyme includes amino acids 304 to 317. To determine the region between functional domains in the human enzyme, the cDNA was manipulated to allow for expression of 1) amino-terminal protein fragments where translation terminated at introduced nonsense codons and 2) carboxyl-terminal peptides where translation initiated at introduced artificial translation initiation sites. The approach to define functional domains was to reduce the length of the carboxyl-terminal region of the dehydrogenase/cyclohydrolase domain and the amino-terminal region of the synthetase

domain until we observed significant effects on the physical or catalytic properties of the separate domains. The boundary of the synthetase domain was determined as a complete loss of catalytic activity of Syn 311, 315 and 316 while Syn 310 was active. Polypeptides expressed from SK-Syn 300 and SK-Syn 304 contain an additional alanine residue at the amino terminus which does not seem to affect the levels of synthetase activity. Variations in the level of expression of the synthetase domains in *E. coli* as observed in Fig. 4C can be due to differences in nucleotide sequence proximal to the initiator methionine which may affect the efficiency of initiation of translation. Because of the lability of this domain, different levels of expression of the various constructs, and the inability to purify the proteins, we would not detect more subtle changes that might occur in the shorter active proteins. On the other hand, our studies demonstrated that although D/C 291 was active, the properties of this protein were altered as evidenced by reduced solubility, increased degradation and instability in *E. coli* extracts, compared with all other D/C proteins. Therefore we assign residues 292-310 as the limits of the region linking the two domains which is located more toward the amino-terminus than our original prediction of 299-313 for the human enzyme, or the predicted interdomain regions of 304-317 and 306-321 for the rat (Thigpen *et al.*, 1990) and yeast (Staben and Rabinowitz, 1986) enzymes respectively.

It is interesting that amino acid sequence comparisons between the human enzyme and the monofunctional 10-formylH_4folate synthetase from *Clostridium acidiurici* (Shannon and Rabinowitz, 1988) and *Clostridium thermoaceticum* (Lovell *et al.*, 1990) reveal that the beginning of these

prokaryotic proteins correspond to residues 316 and 314 of the human enzyme and would incorrectly indicate the beginning of the active mammalian synthetase domain. In making this comparison one must remember that the proteins from *Clostridium* are smaller than the synthetase domain of the human enzyme and exist as a tetramer and not as a dimer (MacKenzie, 1984).

While the residues in the region between amino acids 292 and 310 of the human trifunctional enzyme are not required for catalytic activity of the functional domains, the actual "linker region" which joins the domains is most probably a shorter sequence found in the region bound by residues 292 to 310. Argos (1990) recently reported a study on the characteristics of linker regions, based on the tertiary structure of multidomain proteins. It was observed that preferred linker amino acids are usually not hydrophobic, with the exception of Ala and Pro, with the average length of the linker being 6.5 residues. Within the region between residues 292 to 310 of the human enzyme is found a sequence of 6 amino acids, Lys-Phe-Lys-Pro-Gly-Lys, where 5 of the residues are considered favourable linker constituents and may form the actual linker region. While we have succeeded in defining the general area of the linker region joining the major domains, finding its definitive location will require the tertiary structure of the protein.

Although previous studies showed that the proteolytically derived domains are active, the expression of an active 34 kDa dehydrogenase/ cyclohydrolase domain (D/C 301) from the amino-terminus and an active 67 kDa synthetase domain (Syn 304) from the carboxyl-terminus demonstrates that the two domains are able to fold independently. That the linear

polypeptide chain of each domain is able to fold into a functional three dimensional unit without the influence of the other portions of the procein is consistent with the theory proposed for modular folding of multidomain proteins, beginning at multiple nucleation centers along the polypeptide chain (Janin and Wodak, 1983). The expression of an active 34 kDa D/C domain of the human enzyme is in contrast with the yeast enzyme where Shannon and Rabinowitz (1988) suggested that an intact synthetase domain is necessary to stabilize the bifunctional D/C domain. The proteins D/C 301 and Syn 304 used to demonstrate the independence of the two domains are derived from constructs located near the center of the interdomain region and are the most suitable for further studies on the domains.

Previously it was found that isolation of the D/C domain following proteolysis of the pig enzyme yielded a fragment with a significantly lower than expected specific activity (Tan and MacKenzie, 1977) and it was not possible to determine if this was due to partial inactivation or to a fundamental change in the properties of the isolated domain. In this study, the specific activities of the purified D/C domain are in the order of three-fold higher than those of the full length enzyme as expected for a polypeptide of approximately one third the size. The similarities between the full length protein and the bifunctional domain with respect to their specific activities and extent of substrate channelling illustrate that the kinetic properties of the independently formed dehydrogenase/cyclohydrolase domain are not grossly altered.

Independent expression of the two active domains in *E. coli* and the ability to manipulate the protein primary structure provide new tools to

investigate the catalytic sites, the domain structure and the possible functional advantages of this complex human enzyme.

CHAPTER 4

COMPLEMENTATION OF A YEAST ade3 MUTANT BY THE FUNCTIONAL DOMAINS OF A HUMAN TRIFUNCTIONAL ENZYME

SUMMARY

The cDNA encoding the human trifunctional enzyme methyleneH_4folate dehydrogenase-methenylH_4folate cylcohydrolase-formylH_4folate synthetase was expressed in an *ade3* strain of *Saccharomyces cerevisiae* which exhibited decreased endogenous levels of the three enzyme activities. The human enzyme was catalytically active when expressed in yeast and was able to complement the *ade3* mutation thus alleviating the adenine auxotrophy. Independent expression of the dehydrogenase/cyclohydrolase domain also complemented the mutation in a similar manner as the full length enzyme. As well, expression of the synthetase domain alone rescued the *ade3* mutant but was dependent on supplemented formate.

INTRODUCTION

The eukaryotic trifunctional enzyme methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase catalyzes the interconversion of one-carbon folate derivatives. The NADP-dependent dehydrogenase reaction can oxidize 5,10-methyleneH_4folate to 5,10-methenylH_4folate which is then converted by the cyclohydrolase reaction to 10-formylH_4folate for *de novo* purine synthesis. In most organisms the ATP-dependent synthetase reaction is presumed to incorporate formate into the active one-carbon pool to form 10-formylH_4folate (for review see MacKenzie, 1984).

The trifunctional enzyme has been isolated from several sources (Paukert *et al.*, 1976; Tan *et al.*, 1977; Paukert *et al.*, 1977; Schirch *et al.*, 1978; Caperelli *et al.*, 1980) as a homodimer with subunits of approximately 100kDa. Limited proteolysis of the porcine (Tan and MacKenzie, 1977; Tan and MacKenzie, 1979) and rabbit (Villar *et al.*, 1985) enzymes have yielded an amino terminal dehydrogenase/cylcohydrolase domain and a carboxyl-terminal synthetase domain. cDNA clones encoding the yeast (Staben and Rabinowitz, 1986), human (Hum *et al.*, 1988) and rat (Thigpen *et al.*, 1990) proteins have been isolated and the location of a putative interdomain region joining the two major domains has been predicted in each case.

Here we report expression of the active human full length trifunctional enzyme, the dehydrogenase/cyclohydrolase domain and the synthetase domain in an *ade3* mutant strain of *Saccharomyces cerevisiae* which expresses decreased levels of enzyme activity. Rescue of the mutant yeast strain with the full length enzyme demonstrates the ability of the

human protein to attain an active conformation and complement the yeast enzyme. Complementation with the dehydrogenase/cyclohydrolase domain reveals the non-essential nature of the synthetase domain. However, expression of the synthetase domain incorporates formate into the active one-carbon pool and eliminates the adenine auxotrophy of the *ade3* mutant.

MATERIALS AND METHODS

Restriction endonucleases and DNA modification enzymes were purchased from Bethesda Research Laboratories and Boehringer Mannheim. Components used for growth media were from Difco and auxotrophic supplements were from Sigma. All common chemicals used were of the highest grade commercially available.

Strains, Plasmid, Media - *Saccharomyces cerevisiae* strains STX24-1B (*α ade3 ura1 gal2*) and X2180-1A (*a SUC2 mal mel gal2 CUP1*) were obtained from the yeast genetic stock center, University of California, Berkeley, and T9-3C (*MAT a ura3 his3 leu2*) was obtained from Dr. Howard Bussey (McGill University, Biology Dept.). The haploid strains T9-3C and STX24-1B were crossed, the resulting diploids were sporulated and asci were dissected. The spores were grown and by appropriate selection YA3-1 (*a leu2 ade3 ura3*) was isolated.

Yeast expression vectors pVT100-U and pVT100-L (a generous gift from Dr. T. Vernet) (Vernet *et al.*, 1987) were purified from *Escherichia coli* MC1061. Separate cDNA fragments which encode and express the active human trifunctional protein (HT), the dehydrogenase/cyclohydrolase domain (D/C 301) and the synthetase domain (Syn 304) were inserted separately into the expression vectors as follows: the cDNA fragment originally in pBluescript-SK+ (Stratagene) was restricted with *Nco*I, the ends of the DNA were repaired with Klenow fragment, the desired fragment was excised by restriction with *Sac*I and following agarose gel electrophoresis was purified using Gene Clean™ (Bio 101 corp.) following manufacturer's protocol. The purified cDNA fragment was ligated into the *Pvu*II-*Sac*I site

of the expression vectors using T4 DNA ligase (Fig. 1). The resulting constructs pVT100-U-HT, pVT100-L-HT, pVT100-U-D/C 301, pVT100-L-D/C 301, pVT100-U-Syn 304 and pVT100-L-Syn 304 were propagated in *E. coli* XL1-Blue (Stratagene) and inserts were analyzed by restriction analyses. The expression plasmids, isolated from XL1-Blue, were used to transform YA3-1 following a modified LiCl method (Vernet *et al.*, 1987).

The media used for the growth of yeast at 30°C were YPD complete medium (1% Bacto-yeast extract, 2% Bacto-peptone, 2% dextrose) and minimal medium (0.67% Bacto-yeast nitrogen base w/o amino acids, 2% dextrose) supplemented with the following as required, adenine (20 mg/L), histidine (20 mg/L), leucine (60 mg/L), uracil (20mg/L) and sodium formate (12.8 mM).

Assays and Western Analysis - To obtain whole cell lysates for enzyme assays and Western analyses, 50 ml cultures of *Saccharomyces cerevisiae* were grown aerobically at 30°C to mid-log phase, cells were harvested by centrifugation at 12 000 x g for 5 min and resulting cell pellets were immediately frozen at -80°C. To effect cell lysis, 240 ul of ice cold buffer A (0.1 M potassium phosphate, pH7.3, 35 mM 2-mercaptoethanol, 1 mM PMSF, 1 mM benzamidine, 0.5% nonidet-P40) and 0.25 g of glass beads (0.45-0.50 mm diameter, Fisher Scientific) were added to the frozen cell pellet and was vortexed at the maximum setting for 5 x 20 sec spaced by 20 sec on ice. The mixture was centrifuged at 12 000 x g for 2 min at 4°C and the supernatant was recovered. The spectrophotometric assays of dehydrogenase, cyclohydrolase and synthetase activities were performed as

previously described (Tan *et al.*, 1977). Protein samples were precipitated by the method of Bensadoun and Weinstein (1976) and protein determinations were by the method of Lowrey *et al.* (1951). For Western analysis the proteins were separated by SDS-PAGE as described by Laemmli (1970) and were transferred electrophoretically onto nitrocellulose paper following the general method of Towbin *et al.* (1979). A polyclonal antiserum raised against the porcine trifunctional enzyme (Hum *et al.*, 1988) was used as the primary antibody (1:500 dilution). Protein bands were visualized using an alkaline phosphatase conjugated anti-rabbit second antibody, employing the Protoblot immunoscreening system from Promega Biotec.

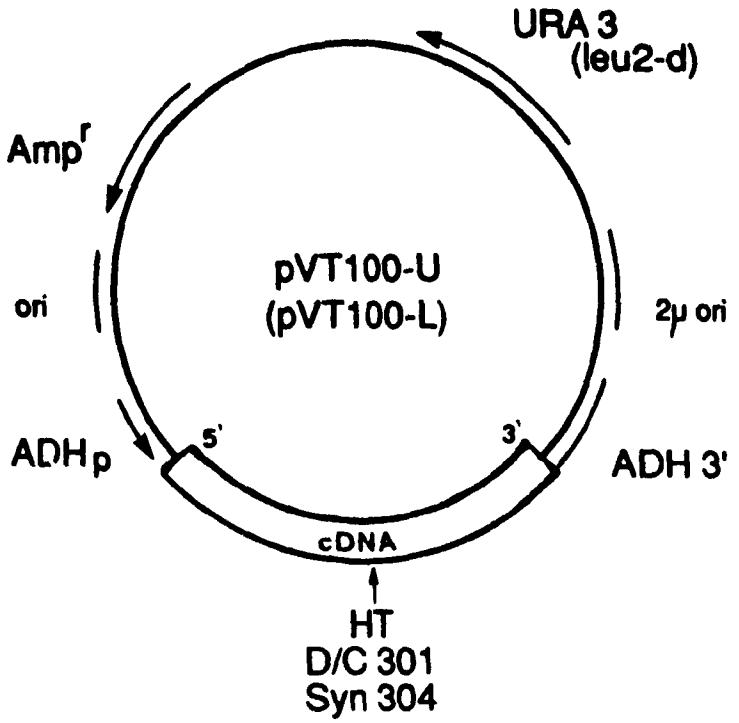


Figure 1. **Schematic representation of the expression plasmids pVT100-U-HT, pVT100-L-HT, pVT100-U-D/C 301, pVT100-L-D/C 301, pVT100-U-Syn 304 and pVT100-L-Syn 304.** The cDNA fragments were inserted downstream of the alcohol dehydrogenase promotor (ADHp) and upstream of the 3'-region of the *ADH1* gene (ADH3'). The pVT100-U vector harbors the *URA3* gene while the pVT100-L vector harbors the defective *leu2*-d gene as selective markers.

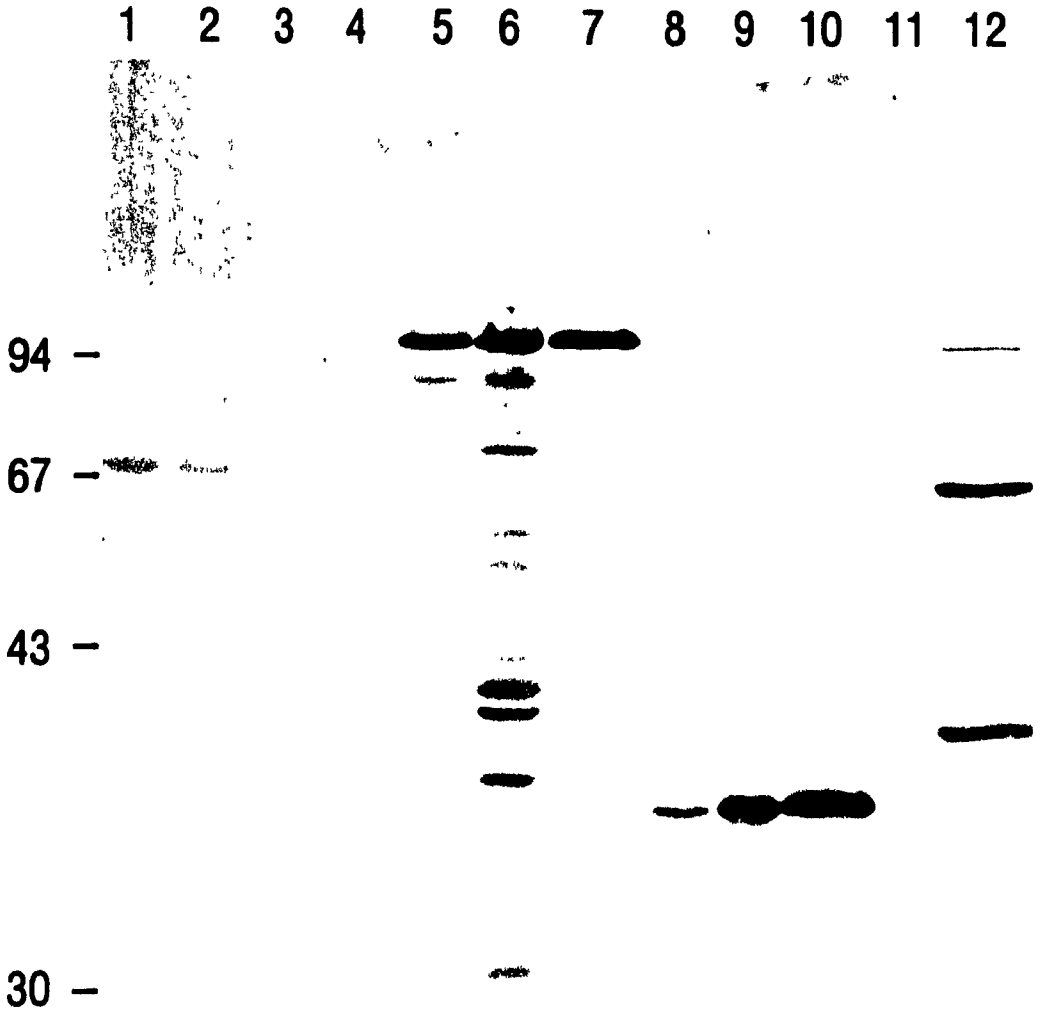
RESULTS

The cDNA (HT-2) encoding the folate-dependent human trifunctional enzyme, methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase, was subcloned into the yeast expression shuttle vectors pVT100-U and pVT100-L (Fig. 1). As well, separate cDNA fragments which encode the amino-terminal dehydrogenase/cyclohydrolase domain (D/C 301) and the carboxyl-terminal synthetase domain (Syn 304) were subcloned into the expression vectors. In all cases the cDNA was inserted within the multiple cloning site of the vectors downstream of the yeast alcohol dehydrogenase promoter.

cDNA fragments encoding HT, D/C 301 and Syn 304 subcloned in the pVT100-U and pVT100-L vectors were transfected separately into *Sacchoromyces cerevisiae* YA3-1 and expression of the human enzymes was detected in the soluble fractions of yeast extracts by Western analysis (Fig. 2). The exogenous proteins of 100, 34 and 67 kDa have molecular weights in good agreement with the predicted values for HT, D/C 301 and Syn 304 respectively and are identical in size with the proteins when expressed in *E. coli*. Increased expression of the full length protein, by transformation with pVT100-L-HT led to higher levels of degradation. Transformation of YA3-1 with pVT100-L-Syn 304 or pVT100-U-Syn 304 yielded a low number of transformants which grew as small colonies and expressed the least amount of soluble protein.

To determine the activity of the human enzymes when expressed in YA3-1, the supernatant fractions of whole yeast extracts were assayed for enzyme activity. Extract from untransformed YA3-1 expressed small amounts of dehydrogenase and cyclohydrolase activity at levels below of that

Figure 2. **Western analysis of whole cell extracts of Saccharomyces cerevisiae expressing exogenous proteins.** 100 μg of protein from whole cell extracts of *S. cerevisiae* were separated by SDS-PAGE and the Western blot was probed with polyclonal antiserum. Lanes 1 and 2 contain extract from wildtype yeast (X2180-1A) and the *ade3* mutant (YA3-1) respectively. The other lanes contain extracts of YA3-1 transformed with pVT100-U (lane 3), pVT100-L (lane 4), pVT100-U-HT (lane 5), pVT100-L-HT (lane 6), pVT100-U-D/C 301 (lane 8), pVT100-L-D/C (lane 9) and pVT100-U-Syn 304 (lane 11). Lanes 7, 10 and 12 contain samples of HT, D/C 301 and Syn 304 respectively when expressed in *E. coli*.


1 2 3 4 5 6 7 8 9 10 11 12
94 –
67 –
43 –
30 –

detected in the extract from wild-type cells (X2180-1A) (Table 1). Cells transformed with pVT100-L-HT expressed increased amounts of dehydrogenase, cyclohydrolase and synthetase activities at levels at least 25 fold above those found in wild-type *Saccharomyces cerevisiae*. Cells transformed with pVT100-L-D/C 301 also expressed increased levels of dehydrogenase and cyclohydrolase activity although, no detectable level of synthetase activity was measured. On the other hand, transformation with pVT100-L-Syn 304 yielded a slight increase in the level of synthetase activity.

To determine the ability of the human enzymes to complement the yeast trifunctional enzyme, transformed YA3-1 cells were grown on minimal medium plates in the presence or absence of required amino acids and nucleotides. Similarly to other *ade3* mutants which are auxotrophic for adenine and histidine, YA3-1 was unable to grow when supplemented only with uracil, leucine and histidine, but growth was apparent when adenine was added. Expression of the human trifunctional enzyme in this strain, by transformation with pVT100-U-HT, complemented the *ade3* mutation and cells were able to survive when supplemented with only leucine (Fig. 3A). Transformation with pVT100-L-HT also complemented the mutation when supplemented with only uracil. Expression of D/C 301 in YA3-1 was also sufficient to complement the *ade3* mutation and yield results identical to those of the full length enzyme (Fig. 3B). Transformation of YA3-1 with pVT100-U-Syn 304 yielded cells which grew very poorly when supplemented with only leucine and grew slighty better when supplemented with leucine and histidine (Fig. 3C). However, under both of these conditions, the addition of formate to the medium increased the rate of cell growth and

allowed for complementation of the *ade3* mutation by the synthetase domain.

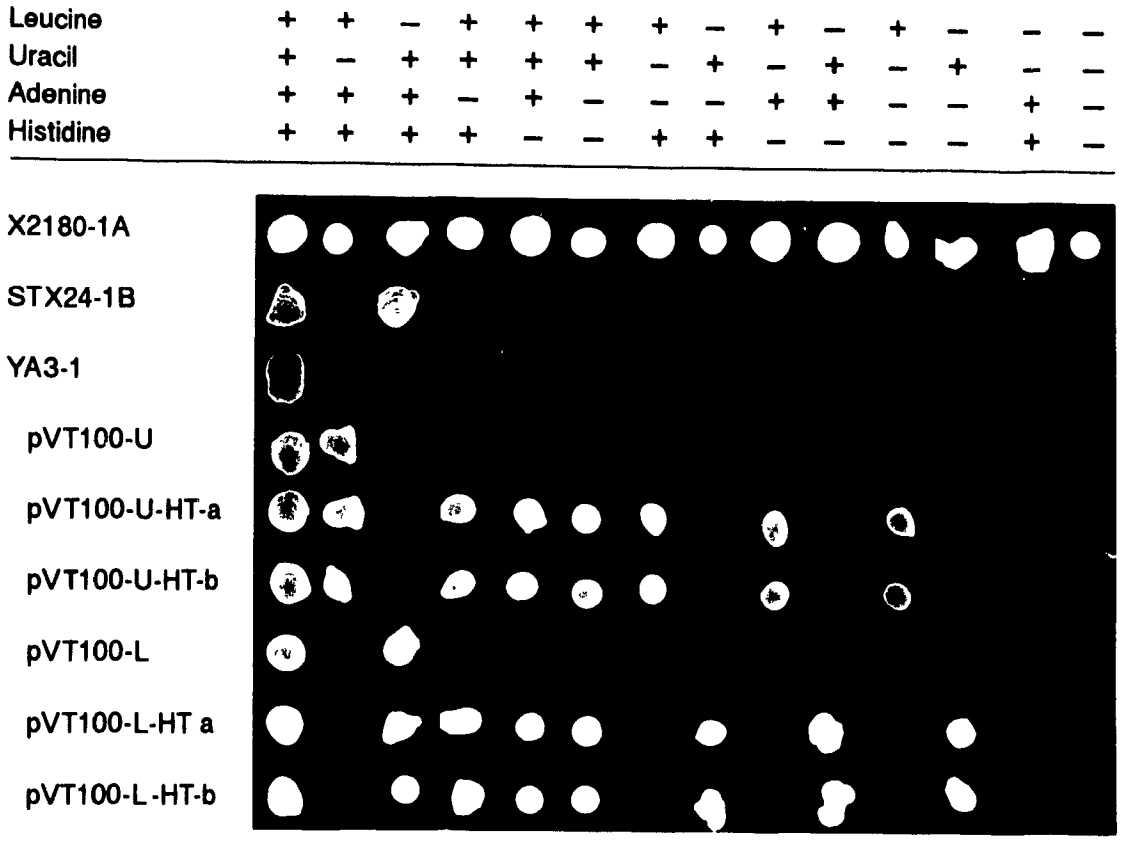
Table 1 Specific Activities of Expressed Enzymes		
Enzyme	Dehydrogenase	Synthetase
X2180-1A (Wildtype)	0.006	0.003
YA3-1	0.001	ND[a]
HT	0.151	0.231
D/C 301	0.138	ND
Syn 304	0.001	0.007

The specific activities are expressed as umol min^{-1} mg^{-1}

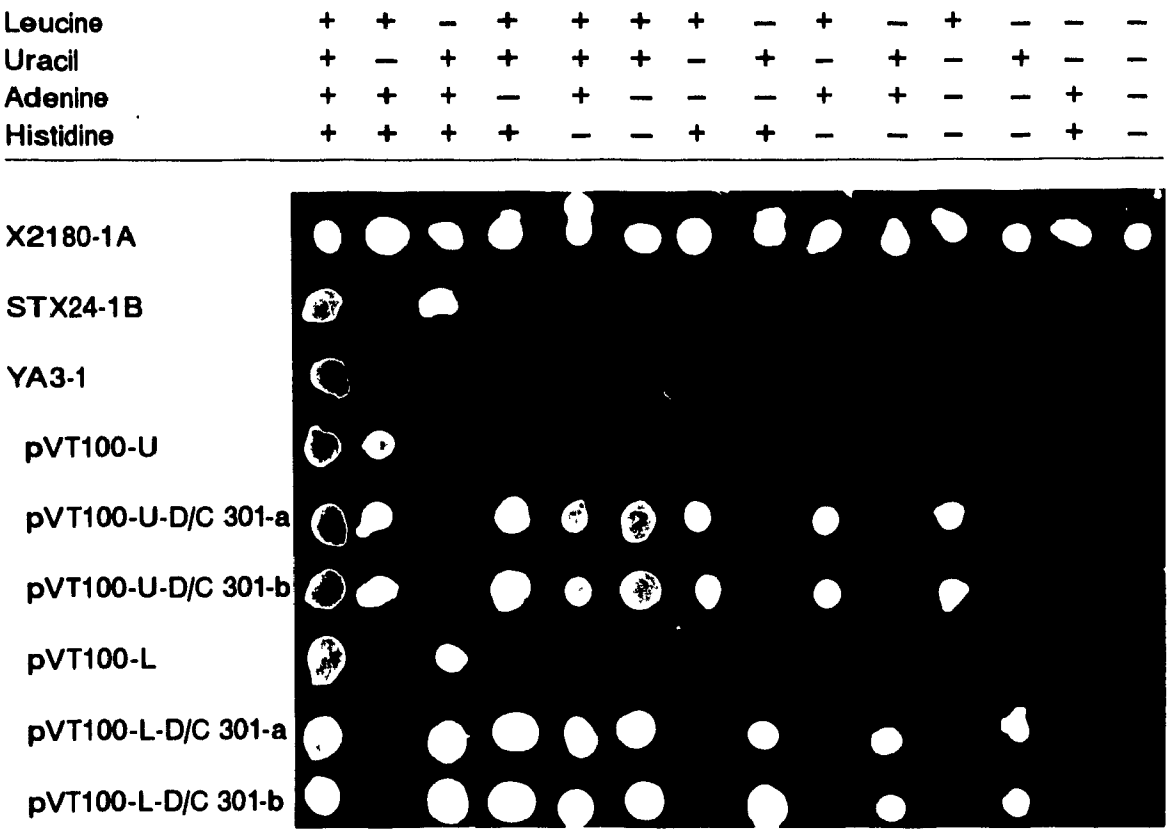
[a] ND, activity not detected

Figure 3. **Complementation of the ade3 mutant (YA3-1).** Approximately 10^5 cells of wildtype (X2180-1A) and *ade3* (STX24-1B and YA3-1) *Saccharomyces cerevisiae* as well as transformed YA3-1 were grown on minimal medium plates supplemented with the following nutrients (final concentration in mg/L): adenine (20), histidine (20), leucine (60) and sodium formate (12.8 mM). Cells were incubated at 30^0C for approximately 18 hours. At the top of each figure, (+) and (-) indicate the presence and absence of the respective nutrients in each column. The strain of cells in each row and the constructs used to transform YA3-1 are listed on the left-hand side of each panel. Panel A: complementation with the human trifunctional enzyme (HT). Panel B: complementation with the dehydrogenase/cyclohydrolase domain (D/C 301). Panel C: complementation with the synthetase domain (Syn 304).

A

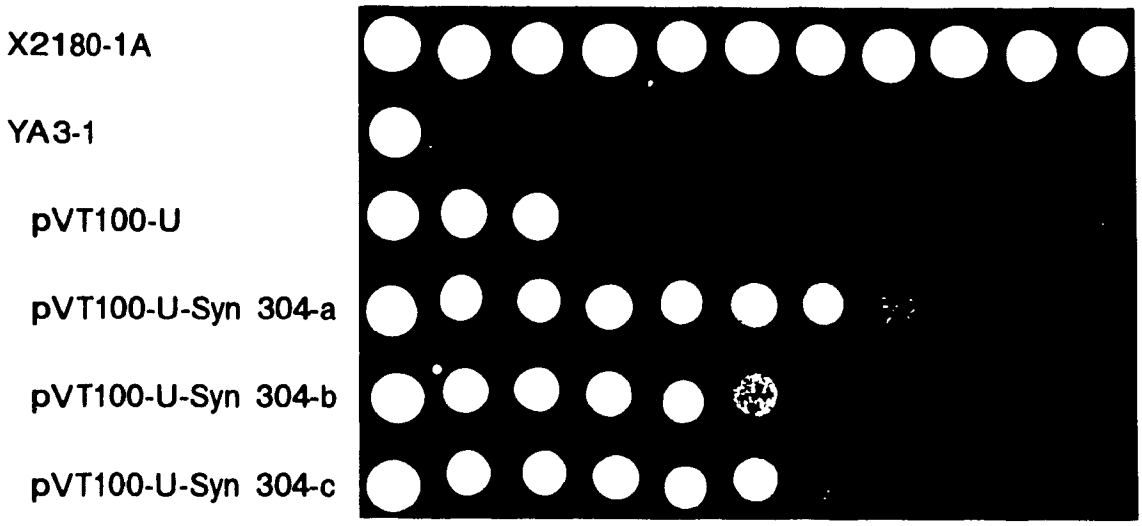

Leucine + + – + + + + – + – + – – –
Uracil + – + + + + – + – + – + – –
Adenine + + + – + – – – + + – – + –
Histidine + + + + – – + + – – – – + –
X2180-1A
STX24-1B
YA3-1
pVT100-U
pVT100-U-HT-a
pVT100-U-HT-b
pVT100-L
pVT100-L-HT a
pVT100-L-HT-b

B


Leucine + + – + + + + – + – + – – –
Uracil + – + + + + – + – + – + – –
Adenine + + + – + – – – + + – – + –
Histidine + + + + – – + + – – – – + –
X2180-1A
STX24-1B
YA3-1
pVT100-U
pVT100-U-D/C 301-a
pVT100-U-D/C 301-b
pVT100-L
pVT100-L-D/C 301-a
pVT100-L-D/C 301-b

C

Uracil	+	−	−	−	−	−	−	−	−	−	−
Formate	+	+	−	+	−	+	−	+	−	+	−
Leucine	+	+	+	+	+	+	+	+	+	−	−
Adenine	+	+	+	+	+	−	−	−	−	+	+
Histidine	+	+	+	−	−	+	+	−	−	+	+



DISCUSSION

The cDNA (HT-2) encoding human methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase, and separate cDNA fragments which encode the dehydrogenase/cyclohydrolase domain and synthetase domain have been used to express separately the proteins in *Saccharomyces cerevisiae*. Western analyses demonstrated the ability of a polyclonal antiserum, raised against enzyme purified from pig liver, to recognize the human proteins of expected molecular weight when expressed in YA3-1. Although the human and yeast proteins share several stretches of extensive amino acid sequence homology (Hum *et al.*, 1988), this antiserum did not cross-react with the yeast trifunctional protein.

It has been well documented that the *ADE3* locus of chromosome 7 encodes the trifunctional enzyme in *Saccharomyces cerevisiae* (McKenzie and Jones, 1977; de Mata and Rabinowitz, 1980). Several strains of yeast have been isolated which harbour different mutations in the *ade3* locus and were found to lack usually all three of the activities (dehydrogenase/cyclohydrolase/synthetase) (Jones and Fink, 1982). The low levels of dehydrogenase and cyclohydrolase activities detected in YA3-1 can be from the mitochondrial form of the enzyme which is encoded by the *MIS1* locus in the yeast genome (Shannon and Rabinowitz, 1988). The product of *MIS1* is non-essential for cell survival, although it is unclear if this enzyme is capable of generating one-carbon units from the mitochondria to be utilized in the cytoplasmic pathway (Shannon and Rabinowitz, 1988; Barlowe and Appling, 1988).

Our results demonstrated the ability of the human trifunctional enzyme to complement the yeast enzyme and thus rescue strain YA3-1.

Complementation of the *ade3* mutation with only the dehydrogenase/ cyclohydrolase domain reveals that the synthetase activity is non-essential for survival of *Saccharomyces cerevisiae* under the experimental conditions, and demonstrate that the bifunctional domain alone can utilize serine to generate one-carbon units for purine synthesis. This parallels the situation in *E. coli* which expresses the dehydrogenase and cyclohydrolase activities and survives in the absence of the synthetase activity (Dev and Harvey, 1978, 1982).

Expression of the carboxyl-terminal domain in YA3-1 demonstrated that the synthetase activity can complement the *ade3* mutant by incorporation of supplemented formate into the one-carbon pool for purine synthesis. In our studies, the requirement for histidine and formate in addition to the expression of the synthetase domain to complement YA3-1 may be related to the suggestion of E. W. Jones (Jones and Fink, 1982) that in the absence of the dehydrogenase and cyclohydrolase activities in an *ade3* mutant, the accumulation of 5,10-methyleneH_4folate can inhibit the fourth enzyme in histidine biosynthesis. In contrast, the rescue of YA3-1 with HT and D/C 301, where 5,10-methyleneH_4folate is a substrate of the dehydrogenase reaction, did not have a requirement for histidine.

Isolation of a cDNA clone encoding the human trifunctional enzyme provides us with a tool with which we can change specific amino acids of the full length protein and the two domains to determine the effect(s) of the manipulation(s) on structure and/or function. The ability to complement an *ade3* mutant with the human enzyme and its domains provides a sensitive method to screen for active or inactivated enzyme.

CHAPTER 5

GENERAL DISCUSSION

The eukaryotic trifunctional methyleneH_4folate dehydrogenase-methenylH_4folate cyclohydrolase-formylH_4folate synthetase has been isolated from several sources (see chapter 1). The kinetic properties of the enzymes are well characterized. However, without knowledge of the primary structure, attempts to study the nature of active sites and the domain structure of the proteins by chemical modification and proteolysis have been limited. The isolation and characterization of the cDNA encoding the human trifunctional enzyme have provided a tool to study the enzyme from a molecular biology approach. Determination of the primary structure and expression of protein fragments have furthered
our understanding of the domain structure and will support future studies of chemical modification, site-specific mutagenesis and ultimately in determining the tertiary structure of the protein.

Sequence Analysis - Isolation of the human cDNA described in this thesis represents the first mammalian dehydrogenase-cyclohydrolase-synthetase enzyme to be cloned (chapter 2). Clones containing the entire coding region of the cytosolic and mitochondrial trifunctional enzymes of *Saccharomyces cerevisiae* have been characterized (Staben and Rabinowitz, 1986; Shannon and Rabinowitz, 1988), and recently Thigpen *et al.* (1990) reported the isolation of a cDNA encoding the rat trifunctional enzyme. Sequencing of the cDNAs or genes from the various organisms have allowed for comparison of the deduced amino acid sequences of the enzymes. The open reading frames encoded by the isolated clones range from 946 residues for the yeast cytosolic enzyme, 943 residues for the mature yeast mitochondrial isozyme and 935 residues for both the human and rat enzymes.

This is in agreement with the apparent Mr of approximately 100 kDa for the isolated trifunctional proteins. The deduced amino acid sequences of the trifunctional proteins are highly homologous and share an overall identity of 60% (Thigpen *et al.*, 1990). The degree of conservation in primary structure of the trifunctional protein across species is similar to that observed for other proteins (Creighton, 1984a). As discussed in chapter 2, it is interesting that the carboxyl-region encoding the synthetase domain is more highly conserved than the amino-region encoding the dehydrogenase/cyclohydrolase domain. Following the initial report that the yeast cytosolic and human enzymes are 63% identical in the synthetase domain and 48% identical in the dehydrogenase/cyclohydrolase domain (chapter 2), Thigpen *et al.* (1990) reported a similar pattern of homology when all the deduced sequences of the folate-dependent trifunctional protein were compared. Clones encoding the murine (Bélanger and MacKenzie, 1989) and human (Peri *et al.*, 1990) mitochondrial NAD-dependent bifunctional dehydrogenase/cyclohydrolase have been isolated and their deduced amino acid sequences are homologous to the amino-terminal bifunctional dehydrogenase/cyclohydrolase domain of the trifunctional enzyme (Bélanger and MacKenzie, 1989; MacKenzie *et al.*, 1990).

The gene encoding monofunctional 10-formylH_4folate synthetase has been isolated from *Clostridium acidiurici* (Whitehead and Rabinowitz 1986, 1988) and *Clostridium thermoaceticum* (Lovell *et al.*, 1990). The deduced amino acid sequences from the two clones are 66% identical and are homologous with the carboxyl-domain of the eukaryotic trifunctional proteins (Lovell *et al.*, 1990).

MethyleneH_4folate dehydrogenase, methenylH_4folate cyclohydrolase and

formylH_4folate synthetase exist in nature both as single enzymes and as different enzyme conjugates (MacKenzie, 1984); therefore, these three activities provide a good system in which to examine the evolution and advantages of multifunctional enzymes. It has been proposed in several cases that multifunctional proteins evolved by the fusion of genes encoding the more primative monofunctional proteins (chapter 1). Homologies between the *Clostridial* monofunctional synthetase enzyme with the corresponding carboxyl domain of the trifunctional protein is consistent with the latter enzyme having evolved from gene fusion events. Although it is premature to conclude gene fusion as the evolutionary mechanism of the trifunctional enzyme, the isolation and characterization of genes encoding the corresponding monofunctional and bifunctional dehydrogenase/cyclohydrolase enzymes of lower organisms may shed more light on the subject.

The comparison of amino acid sequences of the trifunctional protein from several species have revealed regions of extensive homology which may be essential for catalysis or folding to an active conformation. To confirm initial reports that cysteine residues were essential for dehydrogenase and cyclohydrolase activities (Appling and Rabinowitz, 1985a), Barlowe *et al.* (1989) have utilized site-directed mutagenesis to change each cysteine residue to serine within the dehydrogenase/cyclohydrolase domain of the yeast trifunctional enzyme. The intimate relationship between the dehydrogenase and cyclohydrolase activities was demonstrated, as amino acid substitutions at two of the three cysteines affected both activities. Consistent with results of chemical modification of the yeast enzyme with NEM, substitution of the

cysteines affected the two activities to different extents. Interestingly, substitution of cysteine 257 enhanced the cyclohydrolase specific activity two-fold but did not affect that of the dehydrogenase activity. Similar experiments of site-directed mutagenesis are being undertaken with the cDNA encoding the human trifunctional enzyme, and the extent of overlap between the dehydrogenase and cyclohydrolase activities will be explored. It is interesting to speculate that the dehydrogenase/cyclohydrolase functional domain may be composed of two or more structural domains. The two activities may be closely associated at an inter-domain interface such that a slight change in either domain would lead to an alteration of both activities.

Sequence analysis of the eukaryotic trifunctional enzymes and comparison of protein sequences between the yeast, human and rat enzymes have led to a proposed location of a putative interdomain region joining the bifunctional dehydrogenase/cyclohydrolase domain to the synthetase domain. Earlier studies of the eukaryotic enzymes had suggested a proteolytic susceptible region which joins the two major functional domains (Paukert *et al.*, 1977; Tan and MacKenzie, 1977; Tan and MacKenzie, 1979; Villar *et al.*, 1985). The proposed location of the putative interdomain region in each of the eukaryotic enzymes was approximately in the same area. The general location of the putative interdomain region correlates with the end of the NAD-dependent dehydrogenase/cyclohydrolase enzyme (Bélanger and MacKenzie, 1988) and the start of the *Clostridial* 10-formylH_4folate synthetase (Whitehead and Rabinowitz, 1988; Lovell *et al.*, 1990). However, this approach only suggested the approximate area of the interdomain linker region.

Independent Domains - To further our characterization of the human trifunctional enzyme, the isolated cDNA was used to express the enzyme in *E. coli* (chapter 3). Utilizing the isolated yeast gene (Staben *et al.*, 1987) and the human cDNA to express the trifunctional enzyme indisputably confirmed earlier conclusions that the dehydrogenase, cyclohydrolase and synthetase activities in eukaryotes are properties of a single type of polypeptide chain.

The strategy which we employed to demonstrate the independent nature of the two functional domains was to express them separately in *E. coli*. Interestingly, at the time of a review on multifunctional proteins by Kirschner and Bisswanger (1976) there were no known reports of active domains which were expressed from genes or cDNAs separated by the artificial introduction of translation termination or re-initiation signals. However, as mentioned in chapter one, there have recently been several reports where the separate domains of some multifunctional proteins have been expressed from genes or cDNA which were manipulated by recombinant DNA techniques. By the introduction of translation termination and initiation signals into the proposed interdomain region of the human trifunctional enzyme, we succeeded in expressing the dehydrogenase/cyclohydrolase and synthetase domains separately. It was demonstrated that each separate domain can fold independently into an active conformation in the absence of the other portion of the polypeptide chain. These experiments also confirmed the location of the interdomain region and revealed that the boundaries of the functional domains in the trifunctional enzyme do not correspond exactly to those of the bifunctional NAD-dependent dehydrogenase/cyclohydrolase and the

monofunctional 10-formylH_4folate synthetase of *Clostridium acidiurici* and *Clostridium thermoaceticum*.

Based on the functional boundaries of the domains in the trifunctional enzyme, the region joining the two functional domains is approximately 19 amino acids long (chapter 3), and 9 of the residues are considered to be desirable linker constituents according to a review by Argos (1990) on characteristics of linker regions. Although these 19 residues are not essential for catalytic activity, the actual linker region is probably a subset of the 19 residue sequence. Similar to many enzymes, it is probable that removal of a few residues from the ends of the domains would not have altered its activity. To further study the nature of the linker it would be interesting to alter the primary structure and/or length of this region to determine its effect(s), if any, on catalytic function. It has been suggested that the length of a linker region may be more important than its sequence (Zalkin *et al.*, 1984) although, a significant change in the conformation of this region may potentially disrupt any interaction between the functional domains thus altering their stability and function

Initial characterization of the functional dehydrogenase/ cyclohydrolase domain and the human trifunctional enzyme isolated from *E. coli* suggests that the kinetic properties of the isolated bifunctional domain are not grossly altered. The similar substrate turn-over rate of the two proteins and their ability to channel 5,10-methenylH_4folate with equal efficiency further demonstrate the independent nature of the isolated bifunctional domain, and suggests that the isolated domain can probably fold into a tertiary structure similar to that of the native

enzyme.

One advantage of fusing the synthetase domain as part of the trifunctional enzyme might be the stabilization of the synthetase activity. Although the expressed synthetase domain exhibited enzyme activity it was unstable and highly insoluble in *E. coli* extract. It is possible that the two major domains of the trifunctional enzyme can interact to some extent and attain a more stable conformation. In *Clostridium*, the synthetase is a tetramer of four identical subunits (MacKenzie and Rabinowitz, 1971; MacKenzie, 1984) while in the eukaryotic trifunctional enzyme the synthetase domain is part of a homodimer Interaction of the synthetase domain with the dehydrogenase/cyclohydrolase domain in the homodimeric trifunctional enzyme might mimic the interactions of a tetrameric synthetase enzyme and thus stabilize the carboxyl-domain.

Villar *et al.* (1985) reported that following proteolysis of the rabbit trifunctional enzyme, the functional domains were isolated as dimers. Like many dimeric proteins which have a two-fold symmetry at their subunit interface (Miller 1989), the dimeric interfaces of the two functional domains in the native enzyme and as separate fragments presumably have a two-fold symmetry although, this has yet to be determined.

Expression of the human trifunctional enzyme and the dehydrogenase/ cyclohydrolase domain in *E. coli* provides us with a readily abundant source of both enzymes. Using the procedure described in chapter 3, it is possible to purify large amounts of the dehydrogenase/cyclohydrolase domain which makes structural analysis a feasible possibility. In

addition to the two functional domains, the trifunctional enzyme is most probably composed of several structural domains which might only be revealed by detailed structural analysis. X-ray crystal analysis of the native enzyme and the bifunctional domain will allow for comparison of their tertiary structures, and reveal the ability of the separated functional domain to attain native tertiary conformation. As well, structural analysis may reveal the nature of the overlap between the dehydrogenase and cyclohydrolase activities, and may identify the nucleotide-binding site of NADP+ in the dehydrogenase/cyclohydrolase domain. While the sequences of the binding sites from many proteins have been shown to differ, many of them share a similar β-α-β fold (Wierenga *et al.*, 1985; Hanukoglu and Gutfinger, 1989).

Comparison of primary structures between the human and yeast cytosolic trifunctional enzymes and the NAD-dependent dehydrogenase/cyclohydrolase have revealed that the latter two enzymes contain an insert of eight amino acids in the same region of their respective sequences corresponding to the dehydrogenase/cyclohydrolase domain (Bélanger and Mackenzie, 1988). It has been shown that the dehydrogenase and cyclohydrolase activities of the porcine, murine and human enzymes are not kinetically independent where NADP, which is a substrate of the first reaction, inhibits the cyclohydrolase reaction (Smith and MacKenzie, 1983; Rios-Orlandi and MacKenzie, 1988). The two activities of the NAD-dependent enzyme are kinetically independent since NAD does not inhibit the cyclohydrolase activity (Rios-Orlandi and MacKenzie, 1988). As well, chemical modification of the yeast cytosolic and porcine trifunctional enzymes suggested that the dehydrogenase and

cyclohydrolase activities are less closely associated in the yeast enzyme than in the porcine enzyme. These observations prompted us to insert the eight extra amino acids found in the yeast sequence into the corresponding region of the human dehydrogenase/cyclohydrolase domain, to determine if the insertion can lessen the association between the two activities of the human enzyme. The eight residues were inserted into the cDNA by oligonucleotide-directed mutagenesis. Subsequent expression of the cDNA in *E. coli* produced the desired soluble protein although, the insertion inactivated both activities of the dehydrogenase/ cyclohydrolase domain.

Relationship Between the Mitochondrial and Cytosolic Enzymes? - Although the micochondrial NAD-dependent bifunctional enzyme and the dehydrogenase/cyclohydrolase domain catalyze identical metabolic reactions, they differ in cofactor specificity. The bifunctional enzyme requires NAD and Mg^{2+} for activity while the bifunctional domain of the trifunctional enzyme is NADP-dependent. In total extracts of transformed cells, the NAD-dependent dehydrogenase/cyclohydrolase can be separated from the trifunctional enzyme by column chromatography (Mejia and MacKenzie, 1985), and Western blot analysis have demonstrated that the two enzymes are immunologically distinct (Gardam *et al.*, 1988). Thus far, the NAD-dependent dehydrogenase/cyclohydrolase has only been detected in transformed cells and in tissues containing undifferentiated cells (Mejia and Mackenzie, 1985; Bélanger and Mackenzie, 1989). However, basal expression of this enzyme in normal adult cells has not been ruled out. The NAD-dependent bifunctional enzyme localized in mitochondria of mammalian cells (Mejia and MacKenzie, 1988) may be analogous to the

mitochondrial isozyme of the NADP-dependent trifunctional enzyme found in *S. cerevisiae* (Shannon and Rabinowithz, 1986). The physiological role of the two mitochondrial enzymes has not been determined although they may serve similar functions. One possible explanation for expression of the mitochondrial enzyme is that the products of the cytoplasmic and mitochondrial enzymes are required in two cellular compartments but are not transported across the mitochondrial membrane. As proposed by Shannon and Rabinowitz (1986, 1988) and Mejia and MacKenzie (1988), the activities of the mitochondrial enzymes can convert 5,10-methyleneH_4folate to 10-formylH_4folate which can be used to produce formylmethionyl-tRNA to support initiation of protein synthesis in mitochondria (Galper, 1974; Bianchetti *et al.*, 1977). Interestingly, Shannon and Rabinowitz (1988) have reported that the yeast mitochondrial trifunctional enzyme is non essential for survival of the organism. In their experiments, chromosomal deletion of the gene produced viable yeast in minimal medium and had no detectable effect on cell growth.

Barlowe and Appling (1988a) proposed a pathway where folate-dependent activities (NADP-dependent dehydrogenase, cyclohydrolase, synthetase) in mitochondria of mammalian cells can interconvert 5,10-methyleneH_4folate and 10-formylH_4folate and provide one-carbon units for folate-mediated metabolism in the cytoplasm. It was proposed that 10-formylH_4folate synthetase activity in mitochondria can generate ATP, H_4folate and formate from 10-formylH_4folate. The formate would then be transported across the mitochondria membrane and be incorporated into the active one-carbon pool by cytoplasmic 10-formylH_4folate synthetase of the trifunctional enzyme. Although low amounts of dehydrogenase (NADP-

dependent), cyclohydrolase and synthetase activities were reported to be detected in the mitochondria of rat liver cells, the activities can be from residual amounts of the cytoplasmic trifunctional enzyme contaminating the mitochondrial fraction. It is clear that the mitochondrial NAD-dependent dehydrogenase/cyclohydrolase enzyme can interconvert 10-formylH_4folate but it lacks a synthetase activity. Alternately, 10-formylH_4folate hydrolase activity would convert 10-formylH_4folate to formate and H_4folate but its activity has not been detected in the mitochondria of transformed mammalian cells (Mejia and Mackenzie, 1988). However, 10-formylH_4folate hydrolase activity has been reported to be detected in rat liver mitochondria (Barlowe and Appling, 1988a; Case *et al.*, 1988).

Based on amino acid sequence homology between the NAD-dependent bifunctional enzyme and the NADP-dependent trifunctional enzyme it is tempting to speculate on the evolution of the enzymes. The genes encoding the two multifunctional proteins could have evolved by independent fusion events, as proposed for mammalian and yeast fatty acid synthase (McCarthy *et al.*, 1983). Another possibility is that one enzyme could have evolved from the other. Assuming hypothetically that the trifunctional cytosolic enzyme is the more ancient form, gene duplication followed by exon-shuffling and/or point-mutations could allow divergence to form the NAD-dependent bifunctional enzyme. Point mutations which change amino acid residues can alter the enzyme's cofactor requirement from NADP to NAD/Mg^{2+}. Scrutton *et al.* (1990) demonstrated that by changing residues in the NADP-binding domain of glutathione reductase, the enzyme acquired a novel preference for the coenzyme NAD without disrupting the specificity for

glutathione. The introduction of nonsense codons before the coding region of the synthetase domain would delete its activity from the enzyme, and exon shuffling could lead to fusion of exon(s) encoding a mitochondrial signal sequence. The corresponding trifunctional isozymes in yeast may have also resulted from gene duplication although, in this case the mitochondrial enzyme may have diverged to a lesser extent where the cofactor specificity and the synthetase activity have been conserved; as well, the mitochondrial and cytosolic trifunctional enzyme in yeast are not immunologically distinct (Shannon and Rabinowitz, 1986). While any proposal for the evolution of these cytoplasmic and mitochondrial enzymes would be pure speculation at this time, isolation and characterization of corresponding mammalian genomic clones will be useful. The conservation of intron/exon pattern is a strong indication of gene duplication as was shown for the globin gene family (Efstratiadis *et al.*, 1980). While the intron/exon arrangement of a gene encoding the mammalian trifunctional enzyme has yet to be determined, the gene encoding the mouse NAD-dependent bifunctional enzyme has been shown to contain eight exons separated by seven introns (Bélanger and MacKenzie, 1990). As well, Rozen *et al.* (1989) localized the gene encoding the human trifunctional enzyme to chromosome 14q24.

Complementation Studies - Similarities between the yeast and human trifunctional enzymes were further demonstrated by complementation studies. The ability of the human trifunctional enzyme to rescue a yeast *ade3* mutant suggests that the enzymes are similar in structure and function (chapter 4). Recently, Thigpen *et al.* (1990) reported the rescue

of a yeast *ade3* mutant with the rat trifunctional enzyme.

As described in chapter 4, complementation of the *ade3* mutant with the human dehydrogenase/cyclohydrolase domain demonstrated that the synthetase activity was non essential for survival of *S. cerevisiae*. It is presumed that 10-formylH_4folate synthetase can incorporate formate into the active one carbon pool to form 10-formylH_4folate however, the physiological significance of the activity is unclear since formate has not been considered a major one-carbon donor for most organisms (MacKenzie, 1984). While 10-formylH_4folate synthetase is widespread in nature, it is not found in *E. coli* (Dev and Harvey, 1978, 1982) or in several other bacteria (Whitehead *et al.*, 1988).

Barlowe and Appling (1988a) have suggested that the dehydrogenase activity of the trifunctional enzyme predominately converts 5,10-methenylH_4folate to 5,10-methyleneH_4folate in the cytosol. However, the rescue experiments in yeast clearly showed that the dehydrogenase and cyclohydrolase activities can interconvert 10-formylH_4folate to provide for purine synthesis and alleviated the adenine auxotrophy. Complementation of the yeast mutant with just the human 10-formylH_4folate synthetase domain was dependent on formate and strongly suggests that formate can be a one-carbon donor to synthesize 10-formylH_4folate for *de novo* purine synthesis.

The yeast *ade3* mutant can be used as a sensitive tool to screen for the human trifunctional enzyme and its two functional domains. Following mutagenesis, the ability to express the human protein in the yeast system may allow us to define specific residues and/or regions involved in the catalytic and structural functions of the trifunctional enzyme. Assuming that the enzymes from other organisms are similar to the yeast, human and

rat trifunctional enzymes, the yeast *ade3* system can be used to isolate DNA clones from other species which would help identify invariant residues, and aid in deciphering the mechanism of evolution of the eukaryotic trifunctional enzyme.

REFERENCES

Anderson, C.M., McDonald, R.C. and Steitz, T.A. (1978) *J. Mol. Biol.* **123**, 1-13.

Anfinsen, C.B. (1966) *Harvey Lect.* **61**, 95-116.

Anfinsen, C.B., and Haber, E. (1961) *J. Biol. Chem.* **236**, 1361-1363.

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

Appling, D.R., and Rabinowitz, J.C. (1985a) *Biochemistry* **24**, 3540-3547.

Appling, D.R., and Rabinowitz, J.C. (1985b) *J. Biol. Chem.* **260**, 1248-1256.

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

Bachmann, B.J., and Low, K.B. (1980) *Microbiol. Rev.* **44**, 1-56.

Baldwin, R.L. (1989) *TIBS* **14**, 291-294.

Barlowe, C.K., and Appling, D.R. (1988a) *Biofactors* **2**, 171-176.

Barlowe, C.K., and Appling, D.R. (1988b) *Biochem. and Biophys. Res. Comm.* **157**, 245-249.

Barlowe, C.K., Williams, M.E., Rabinowitz, J.C., and Appling, D.R. (1989) *Biochemistry* **28**, 2099-2106.

Baum, J., Dobson, C.M., Evans, P.A., and Hanley, C. (1989) *Biochemistry* **28**, 7-13.

Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A., and Stanners, C.P. (1987) *Mol. Cell Biol.* **7**, 3221-3230

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

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

Benkovic, S.J. (1984) *TIBS* **9**, 320-322.

Bennett, W.S., and Steitz, T.A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4848-4852.

Bensadoun, A., and Weinstein, D. (1976) *Anal. Biochem.* **70**, 421-250.

Bianchetti, R., Lucchini, G., and Sartirana, M.L. (1971) *Biochem. Biophys. Res. Commun.* **42**, 97-102.

Biesecker, G., Harris, J.I., Thierry, J.C., Walker, J.E., and Wonacott, A.J. (1977) *Nature* **266**, 328-333.

Blake, C.C.F. (1978) *Nature* **273**, 267.

Blake, C.C.F. (1979) *Nature* **277**, 598.

Blake, C. (1983) *Nature* **306**, 535-537.

Blake, C.C.F. (1981) *Nature* **291**, 616.

Blake, C.C.F. (1985) *Int. Rev. of Cyt.* **93**, 149-185.

Caperelli, C.A., Chettur, G., Lin, L.Y., and Benkovic, S.J. (1978) *Biochem. Biophys. Res. Commun.* **82**, 403-410

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

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

Chasin, L.A., Feldman, A., Konstam, M., and Urlaub, G. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 718-722.

Cheek, W.D., and Appling, D.R. (1989) *Arch. of Biochem. and Biophys.* **270**, 504-512.

Chen, Z., Dixon, J.E., and Zalkin, H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3097-3101.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979) *Biochem.* **18**, 5294-5299

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

Chothia, C., Levitt, M., and Richardson, D. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4130-4134.

Chothia, C., Levitt, M., and Richardson, D. (1981) *J. Mol. Biol.* **145**, 215-250.

Coggins, J.R., and Boocock, M.R. (1986) in *Multidomain Proteins-Structure and Evolution* (D.G. Hardie, and J.R. Coggins eds.), Elsevier Publishing Co., New York, pp. 259-279.

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

Cohen, G.N., and Dautry-Varsat (1980) in *Multifunctional Proteins* (H. Bisswanger and E. Schmincke-Ott eds.), John Wiley & Sons, Inc., New York, pp 49-121.

Cohen, L., and MacKenzie, R.E. (1978) *Biochim. Biophys. Acta* **522**, 311-317.

Coleman, P.F., Suttle, D.P., and Stark, G.R. (1977) *J. Biol. Chem.* **252**, 6379- 6385.

Cook, R.J., and Wagner, C. (1982) *Biochemistry* **21**, 4427-4434.

Crawford, J.L., Lipscomb, W.N., and Schellman, C.G. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 538-542.

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

Creighton, T.E. (1978) *Prog. Biophys. Molec. Biol.* **33**, 231-297.

Creighton, T.E. (1984a) in *Proteins*, (T.E. Creighton eds.) pp.93-126, W.H. Freeman & Co., New York.

Creighton, T.E. (1984b) in *Proteins*, (T.E. Creighton eds.) pp. 133-152, W.H. Freeman & Co., New York.

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

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

Davidson, J.N., Rumsby, P.C., and Tamaren, J. (1981) *J. Biol. Chem.* **256**, 5220- 5225.

de Mata, Z.S., and Rabinowitz, J.C. (1980) *J. Biol. Chem.* **255**, 2569-2577.

Dev, I.K., and Harvey, R.J. (1978) *J. Biol. Chem.* **253**, 4245-4253.

Dev, I.K., and Harvey, R.J. (1982) *J. Biol. Chem.* **257**, 1980-1986.

Dolgikh, D.A., Gilmanshin, R.I., Brazhnikov, E.V., Bychkova, V.E., Semisotnov, G.V., Venyaminov, S.Y., and Ptitsyn, O.B. (1981) *FEBS letters* **136**, 311- 315.

Drenth, J., Jansonius, J.N., Koekoek, R., Swen, H.M., Wolthers, B.G. (1968) *Nature* **218**, 929-932.

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

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

Duncan, K., Edwards, R.M., and Coggins, J.R. (1987) *Biochem. J.* **246**, 375-386.

Ebbole, D.J., and Zalkin, H. (1987) *J. Biol. Chem.* **262**, 8274-8287.

Edelman, G.M. (1973) *Science* **180**, 830-840.

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M. O'Connell, C., Spritz, R.A., De Riel, J.K., Forget, B.G., Weissman, S.M.,

Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, R.E., Shoulders, C.C., and Proudfoot, N.J. (1980) *Cell* **21**, 653-668.

Evans, D.R. (1986) in *Multidomain Proteins-Structure and Evolution* (D.G. Hardie and J.R. Coggins eds.), Elsevier Publishing Co., New York, pp. 283-322.

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

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. et Biophys. Acta* **999**, 52-57.

Finkelstein, A.V., and Ptitsyn, O.B. (1976) *J. Mol. Biol.* **103**, 15-24.

Finley, D., and Varshavsky, A. (1985) *TIBS* **10**, 343-347.

Fothergill-Gilmlore, L.A. (1986) in *Multidomain Proteins-Structure and Evolution* (D.G. Hardie and J.R. Coggins eds.), Elsevier Science Publishing Co., New York, pp. 85-168.

Fry, C.D., Kuby, S.A., and Mildvan, A.S. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 907-911

Gaertner, F.H. (1978) *TIBS* **3**, 63-65.

Galper, J.B. (1974) *J. Cell Biol.* **60**, 755-763.

Gardam, M.A., Mejia, N.R., and MacKenzie, R.E. (1987) *Biochem. Cell Biol.* **66**, 66-70.

Garnier, J., Osguthorpe, D.J., and Robson, B. (1978) *J. Mol. Biol.* **120**, 97-120

Ghelis, C., Tempete-Gaillourdet, M., and Yon, J.M. (1978) *Biochem. and Biophys. Res. Comm.* **84**, 31-36.

Gilbert, W. (1978) *Nature* **271**, 501.

Go, M. (1981) *Nature* **291**, 90-92.

Go, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1964-1968.

Goldberg, M.E. (1969) *J. Mol. Biol.* **46**, 441-446.

Grayson, D.R., and Evans, D.R. (1983) *J. Biol. Chem.* **258**, 4123-4129.

Green, J., Matthews, R.G., and MacKenzie, R.E. (1986) in *Chemistry and Biology of Pteridines, Pteridines and Folic Acid Derivatives*, (B.A. Cooper and V.M. Whitehead eds.), Walter de Gruyter, New York, 901-904.

Haber, E., and Anfinsen, C.B. (1961) *J. Biol. Chem.* **236**, 422-424.

Hanukoglu, I., and Gutfinger, T. (1989) *Eur. J. Biochem.* **180**, 479-484.

Hardie, D.G., and Coggins, J.R. (1986) in *Multidomain Proteins-Structure and Evolution* (D.G. Hardie and J.R. Coggins eds.), Elsevier Publishing Co., New York, pp. 333-344.

Hardie, D.G., and McCarthy, A.D. (1986) in *Multidomain Proteins-Structure and Evolution* (D.G. Hardie and J.R. Coggins, eds.), Elsevier Publishing Co., New York, pp 229-253.

Hardie, D.G., McCarthy, A.D., and Braddock, M. (1986) *Biochem. Soc. Trans.* **14**, 568-570.

Henikoff, S., Keene, M.A., Sloan, J.S., Bleskan, J., Hards, R., and Patterson, D. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 720-724.

Hewick, R.M., Hankapiller, M.W., Hood, L.E., and Dieger, W.J. (1981) *J. Biol. Chem.* **256**, 7990-7997.

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

Husain, J., Van Houten, B., Thomas, D.C., and Sancar, A. (1986) *J. Biol. Chem.* **261**, 4895-4901

Huynh, T.V., Young, R.A., and Davis, R.W. (1985) in *DNA Cloning*, vol. 1, a practical approach (Glover, D.M., ed.), pp. 49-78, IRL press, Oxford

Jacobsen, H., Klenow, H., and Overgaard-Hansen, K. (1974) *Eur. J. Biochem.* **45**, 623-627.

Janin, J., Truffa-Bachi, P., and Cohen, G.N. (1967) *Biochem. and Biophys. Res. Comm.* **26**, 430-434.

Janin, J., and Chothia, C. (1980) *J. Mol. Biol.* **143**, 95-128.

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

Johlin, F.C., Swain, E., Smith, C., and Tephly, T.R. (1989) *Mol. Pharmacology* **35**, 745-750.

Jones, E.W. (1977) *Genetics* **85**, 209-223.

Jones, E.W., and Fink, G.R. (1982) in *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, (J.N. Strathern, E.W. Jones and J.R. Broach eds.), pp. 279-299 Cold Spring Harbor Laboratory, New York.

Karplus, M., and Weaver, D.L. (1976) *Nature* **260**, 404-406.

Kim, P.S., and Baldwin, R.L. (1982) *Ann. Rev. Biochem.* **51**, 459-489.

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

Klotz, I.M., Darnall, D.W., and Langerman, N.R. (1977) in *The Proteins*, Vol I, 3rd edition (H. Neurath and R.L. Hill eds.), pp. 293-411 Academic Press, New York.

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

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

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

Kuwajima, K. (1989) *Proteins* **6**, 87-103.

Laemmli, U.K. (1970) *Nature* (London) **277**, 680-685.

Levitt, M. (1978) *Biochemistry* **17**, 4277-4285.

Levitt, M. and Chothia, C. (1976) *Nature* **261**, 552-558.

Lewis, N.L., Momany, F.A., and Scheraga, H.A. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 2293-2297.

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

Lim, V.I. (1978) *FEBS letters* **89**, 10-14.

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

Ljungcrantz, P., Carlsson, H., Mansson, M-O., Buckel, P., Mosbach, K., and Bulow, L. (1989) *Biochemistry* **28**, 8786-8792.

Lovell, C.R., Przybyla, A., and Ljungdahl, L.G. (1990) *Biochemistry* **29**, 5687-5694.

Lowrey, O.H., Rosebrough, N.N., Farr, A L., and Randall, R.J. (1951) *J. Biol. Chem.* **193**, 265-275.

Lue, P.F., and Kaplan, J.G. (1969) *Biochem. and Biophys. Res. Comm.* **34**, 426-433.

MacKenzie, R.E., and Rabinowitz, J.C. (1971) *J. Biol. Chem.* **246**, 3731-3736.

MacKenzie, R.E. (1973) *Biochem. and Biophys. Res. Comm.* **53**, 1088-1095.

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

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

MacKenzie, R.E. (1984) in *Folates and Pterins* **Vol. 1**, *Chemistry and Biochemistry of Folates* (R.L. Blakley and S.J. Benkovic eds.), John Wiley & Sons, Inc., New York, pp. 255-306.

MacKenzie, R.E., Mejia, N.R., and Yang, X-M. (1988) *Adv. in Enz. Regul.* **27**, 31-39.

MacKenzie, R.E., Bélanger, C., Peri, K.G., Hum, D.W., and Sheffield, W.P. in *Chemistry and Biology of Pteridines, Pteridines and Folic Acid Derivatives* (1990 in press).

Maley, J.A., and Davidson, J.N. (1988) *Mol. Gen. Genet.* **213**, 278-284.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Press

Markert, C.L. (1975) in *Isozymes* (C.L. Markert, ed.), Academic Press, New York, pp. 1-9.

Matchett, W.H. (1974) *J. Biol. Chem.* **249**, 4041-4049.

McCarthy, A.D., Goldring, D., and Hardie, D.G. (1983) *FEBS lett* **162**, 300-304.

McClard, R.W., Black, M.J., Livingstone, L.R., and Jones, M.E. (1980) *Biochemistry* **19**, 4699-4706.

McKenzie, K.Q., and Jones, E.W. (1977) *Genetics* **86**, 85-102.

McLachlan, A.D. (1987) in *Cold Spring Harbor Symposia on Quantitative Biology*, **Vol. LII**, Cold Spring Harbor Laboratory, pp. 411-420.

Mejia, N.R., and MacKenzie, R.E. (1985) *J. Biol. Chem.* **260**, 14616-14620.

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

Mejia, N.R., and MacKenzie, R.E. (1988) *Biochem. Biophys. Res. Comm.* **155**, 1-6.

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

Min, H., Shane, B., and Stokstad, E.L.R. (1988) *Biochim. et Biophys. Acta* **967**, 348-353.

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

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

Murry-Brelier, A., and Goldberg, M.E. (1989) *Proteins* **6**, 395-404.

Oas, T.G., and Kim, P.S. (1988) *Nature* **336**, 42-48.

Ohgushi, M., and Wada, A. (1983) *FEBS lett.* **164**, 21-24.

Osborn, M.J., Hatefi, Y., Kay, L.D., and Huennekens, F.M. (1957) *Biochim. Biophys. Acta* **26**, 208-210.

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

Patey, C.A.H., and Shaw, G. (1973) *Biochem. J.* **135**, 543-545.

Paukert, J.L., Straus, L.D., and Rabinowitz, J.C. (1976) *J. Biol. Chem.* **251**, 5104-5111.

Paukert, J.L., Williams, G.R., and Rabinowitz, J.C. (1977) *Biochem. and Biophys. Res. Comm.* **77**, 147-154.

Pauling, L., Corey, R.B., and Branson, H.R. (1951) *Proc. Natl. Acad. Sci. USA* **37**, 205-211.

Pauling, L., and Corey, R.B. (1951) *Proc. Natl. Acad. Sci. USA* **37**, 729-740.

Pazirandeh, M., Chirala, S.S., Huang, W-Y., and Wakil, S.J. (1989) *J. Biol. Chem.* **264**, 18195-18201.

Pelham, H.R.B., and Jackson,R.J. (1976) *Eur. J. Biochem.* **67**, 247-265.

Peri, K.G., Bélanger, C., and MacKenzie, R.E. (1989) *Nucl. Acids Res.* **17**, 8853.

Pfeil, W. (1981) *Mol. and Cell. Biochem.* **40**, 3-28.

Phillips, D.C. (1966) *Sci. Amer.* **215**, 78-92.

Pompon, D. and Lederer, F. (1976) *Eur. J. Biochem.* **68**, 415-423.

Poorman, R.A., Randolph, A., Kemp, R.G., and Heinrikson, R.L. (1984) *Nature* **309**, 467-469.

Porter, R.R. (1973) *Science* **180**, 713-716.

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

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

Raymond, M. and Gros, P. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6488-6492.

Richardson, J.S. (1977) *Nature* **268**, 495-500.

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

Rios-Orlandi, E.M., Zarkadas, C.G., and MacKenzie, R.E. (1986) *Biochim. Biophys. Acta* **871**, 24-35.

Rios-Orlandi, E.M., and MacKenzie, R.E. (1988) *J. Biol. Chem.* **263**, 4662-4667.

Rosa, M.D. (1979) *Cell* **16**, 815-825.

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

Rose, G.D. (1978) *Nature* **272**, 586-590.

Ross, J., Green, J., Baugh, C.M., MacKenzie, R.E., and Matthews, R.G. (1984) *Biochemistry* 23, 1796-1801.

Rossmann, M.G., Liljas, A., Branden, C-I., and Banaszak, L.J. (1975) in *The Enzymes*, 3rd ed., Chapter XI (P.D. Boyer, ed.), Academic Press, Inc., New York, pp. 61-102.

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

Rozen, R., Hum, D.W., and MacKenzie, R.E. (1985) *Can. J. Biochem. Cell Biol.* 63, 1189-1193.

Rozen, R., Barton, D., Du, J., Hum, D.W., MacKenzie, R.E., and Francke, U. (1989) *Am. J. Hum. Genet.* 44, 781-786.

Rumsby, P.C., Campbell, P.C., Niswander, L.A., and Davidson, J.N. (1984) *Biochem. J.* 217, 435-440.

Sakano, H., Rogers, J.H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R., Tonegawa, S. (1979) *Nature* 277, 627-633.

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

Scanlon, K.J., Cashmore, A.R., Moroson, B.A., Dreyer, R.N., and Bertino, J.R. (1981) *Mol. Pharmacology* 19, 481-490.

Schein, C.H. (1989) *Biotechnology* 7, 1141-1148.

Schirch, L.V. (1978) *Arch. Biochem. and Biophys.* 189, 283-290.

Schirch, V., Strong, B., Joshi, G., Lura, R., and Dennis, T. (1986) in *Chemistry and Biology of Pteridines, Pteridines and Folic Acid Derivatives*, (B.A. Cooper and V.M. Whitehead eds.), Walter de Gruyter, New York, pp. 887-891.

Schmincke-Ott, E., and Bisswanger, H. (1980) in *Multifunctional Proteins* (H. Bisswanger and E. Schmincke-Ott eds.), John Wiley & Sons, Inc.,

New York, pp. 1-29.

Schweizer, M., Roberts, L.M., Holtke, J., Takabayashi, K., Muller, G., Hoja, U., Hollerer, E., Schuh, B., and Scheizer E. (1986) *Biochem. Soc. Trans.* 14, 572-574.

Scrutton, M.C., and Beis, I. (1979) *Biochem. J.* 177, 833-846.

Scrutton, N.S., Berry, A., and Perham, R.M. (1990) *Nature* 343, 38-43.

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

Shannon, K.W., and Rabinowitz, J.C. (1986) *J. Biol. Chem.* 261, 12266-12271.

Shannon, K.W., Whitehead, T.R., Staben, C., and Rabinowitz, J.C. (1986) in *Chemistry and Biology of Pteridines, Pteridines and Folic Acid Derivatives*, (B.A. Cooper and V.M. Whitehead eds.), Walter de Gruyter, New York, pp 865-885.

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

Shigesada, K., Stark, G.R., Maley, J.A., Niswander, L.A., and Davidson, J.N. (1985) *Mol. and Cell. Biol.* 5, 1735-1742.

Simmen, R.C.M., Tanaka, T., Ts'ui, K.F., Putkey, J.A., Scott, M.J., Lai, E.C., and Means, A.R. (1985) *J. Biol. Chem.* 260, 907-912.

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

Smith, D.D.S., and MacKenzie, R.E. (1983) *Can. J. of Biochem. and Cell Biol.* 61, 1166-1171.

Smith, D.D.S., and MacKenzie, R.E. (1985) *Biochem. and Biophys. Res. Comm.* **128**, 148-154.

Smith, G.K., Mueller, T., Wasserman, G.F., Taylor, W.D., and Benkovic, S.J. (1980) *Biochemistry* **19**, 4313-4321.

Staben, C., and Rabinowitz, J.C. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6799-6803.

Staben, C., and Rabinowitz, J.C. (1986) *J. Biol. Chem.* **261**, 4629-4637.

Staben, C., Whitehead, T.R., and Rabinowitz, J.C. (1987) *Anal. Biochem.* **162**, 257-264.

Staley, J.P., and Kim, P.S. (1990) *Nature* **344**, 685-688.

Stein, J.P., Catterall, J.F., Kristo, P., Means, A.R., and O'Malley, B.W. (1980) *Cell* **21**, 681-687.

Strong, W., Joshi, G., Lura, R., Muthukumaraswamy, N., and Schirch, V. (1987) *J. Biol. Chem.* **262**, 12519-12525.

Strong, W.B., and Schirch, V. (1989) *Biochemistry* **28**, 9430-9439.

Strong, W.B., Tendler, S.J., Seither, R.L., Goldman, I.D., and Schirch, V. (1990) *J. Biol. Chem.* **265**, 12149-12155.

Studier, F.W., and Moffatt, B.A. (1986) *J. Mol. Biol.* **189**, 113-130.

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

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

Tan, L.U.L., Drury, E.J., and MacKenzie, R.E. (1977) *J. Biol. Chem.* **252**, 1117- 1122.

Tan, L.U.L., and MacKenzie, R.E. (1977) *Biochim. Biophys. Acta* **485**, 52-59.

Tan, L.U.L., and MacKenzie, R.E. (1979) *Can. J. of Biochem. and Cell Biol.* **57**, 806-812.

Tatum, C.M.Jr., Benkovic, P.A., Benkovic, S.J., Potts, R., Schleicher, E., and Floss, H.G. (1977) *Biochemistry* **16**, 1093-1102.

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

Towbin, H., Staebelin, T., and Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4350-4354.

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

Traut, T.W. (1988a) *Proc. Natl. Acad. Sci. USA* **85**, 2944-2948.

Traut, T.W. (1988b) *CRC Crit. Rev. in Biochem.* **23**, 121-169.

Truffa-Bachi, P., Veron, M., and Cohen, G.N. (1974) *CRC Crit. Rev. in Biochem.* **2**, 379-415.

Uyeda, K., and Rabinowitz, J.C. (1967) *J. Biol. Chem.* **242**, 4378-4385.

Venkatachalam, C.M. (1968) *Biopolymers* **6**, 1425-1436.

Vernet, T., Dignard, D., and Thomas, D.Y. (1987) *Gene* **52**, 225-233.

Vieira, J., and Messing, J. (1982) *Gene* **19**, 259-268.

Villar, E., Schuster, B., Peterson, D., and Schirch, V. (1985) *J. Biol. Chem.* **260**, 2245-2252.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982) *EMBO J.* **1**, 945-951

Wasserman, G.F., Benkovic, P.A., Young, M., and Benkovic, S.J. (1983) *Biochemistry* **22**, 1013-1021.

Wasserman, G.F., Mueller, W.T., Benkovic, S.J., Liao, W.S.L., and Taylor, J. (1984) *Biochemistry* **23**, 6704-6710.

Welch, G.R. (1977) *Prog. Biophys. Molec. Biol.* **32**, 103-191.

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

Wetlaufer, D.B. (1981) *Adv. in Prot. Chem.* **34**, 61-89.

White, F.H.Jr. (1961) *J. Biol. Chem.* **236**, 1353-1360.

Whitehead, T.R., and Rabinowitz, J.C. (1986) *J. Bacteriol.* **167**, 205-209.

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

Whitehead, T.R., Park, M., and Rabinowitz, J.C. (1988) *J. Bacteriol.* **170**, 995-997.

Wierenga, R.K., De Maeyer, M.C.H., and Hol, W.G.J. (1985) *Biochemistry* **24**, 1346-1357.

Wierenga, R.K., Terpstra, P., and Hol, W.G.J. (1986) *J. Mol. Biol.* **187**, 101-107.

Williams, L.G., Bernhardt, S., and Davis, R.H. (1970) *Biochemistry* **9**, 4329- 4335.

Wright, P.E., and Dyson, H.J., and Lerner, R.A. (1988) *Biochemistry* **27**, 7167- 7175.

Yon, J.M., Betton, J-M., Desmadril, M., Mitraki, A., Minard, P., Gaillard, S., Ballery, N., Missiakas, D. (1988) *J. of Chrom.* **440**, 421-437.

Young, P.R., and Tilghman, S.M. (1986) in *Multidomain Proteins-Structure and Evolution* (D.G. Hardie and J.R. Coggins eds.), Elsevier Publishing Co., New York, pp 55-81.

Zalkin, H., Paluh, J.L., van Cleemput, M., Moye, W.S., and Yanofsky, C. (1984) *J. Biol. Chem.* **259**, 3985-3992.

Zamierowski, M.M., and Wagner, C. (1977) *J. Biol. Chem.* **252**, 933-938.

Zoller, M.J., and Smith, M. (1987) *Methods Enzymol.* **154**, 329-350.