7

The Molecular Biology of Tryptophan Synthase: A Model for Protein–Protein Interaction

SIMON SWIFT AND GORDON S.A.B. STEWART

Department of Applied Biochemistry and Food Science, University of Nottingham, Faculty of Agricultural and Food Sciences, Sutton Bonington, Loughborough, UK

Review objectives

One of the major goals of modern biochemistry is to elucidate the molecular mechanisms which operate within protein structure/folding to promote the specific interaction of disparate subunits in multisubunit enzyme formation. One perspective recognizes that many of the keys to protein–protein interaction may be elucidated through the detailed understanding of a few model systems. Implicit in this view is the belief that there are a limited number of molecular motifs available for interprotein recognition, that these may be defined from limited studies and that the enormous diversity of protein interaction reflects the numerical consequence of randomly permutating a modest number of structural motifs. With the advent of protein engineering, the elucidation of these elements is of considerable importance. The opportunity to design protein interaction by conforming to a limited set of rules, would have far-reaching consequences in the development of biotechnology.

The tryptophan synthase system has been the subject of a number of recent reviews. These have concentrated mainly on enzymology (Miles, 1986; Miles, Bauerle and Ahmed, 1987), evolutionary aspects (Crawford, 1989) and the early history of the biochemical studies with tryptophan synthase (Yanofsky, 1987), how the enzyme was characterized and its role in the determination of gene–protein co-linearity and the genetic code. We believe that a considerable resource for the pursuance of protein structure/folding studies is available in the wealth of molecular and biochemical data on tryptophan synthase

Abbreviations: CD, circular dichroism; CRM, cross-reacting material; DGGE, denaturing gradient gel electrophoresis; FMT, 5-methyltryptophan; GuHCL, guanidine hydrochloride; InG, indole glycerol; InGP, indole 3-glycerol phosphate; IPP, indole 3-propanolphosphate; NMR, nuclear magnetic resonance; PCR, polymerase chain reaction; PLP, pyridoxal phosphate; SDM, site-directed mutagenesis; UV, ultraviolet.

Biotechnology and Genetic Engineering Reviews — Vol. 9, December 1991
0264–8725/91/09/229–294 $20.00 + $0.00 © Intercept Ltd, P.O. Box 716, Andover, Hampshire SP10 1YG, UK

and, for this reason, we believe it timely to review this knowledge with the particular perspective of emphasizing its potential as a model system in protein–protein interaction.

The factors that govern the selection of a model for protein–protein interaction may be simply defined. First, it must be possible to genetically or biochemically alter the primary structure of individual protein subunits in both a random and systematic fashion so as to examine the effect of primary structure on quaternary structure. Secondly, the biological function of the multisubunit system should be convenient to assay. Thirdly, it must be possible to discriminate between changes that solely affect catalytic function and those that are related to a modulation in subunit interactions.

The tryptophan synthase of *Escherichia coli* offers a model system for protein–protein interaction which not only meets these criteria, but exhibits the following additional advantages. First, the gene sequence, primary structure, multisubunit character and biochemistry are known and well documented. Secondly, the genetic elements of each subunit may be obtained as plasmid-coded derivatives, thus facilitating both the generation and the analysis of random or site-specific mutations. Thirdly, we will describe recent studies in this laboratory which demonstrate that it is possible to phenotypically select mutations specifically affecting subunit interaction as opposed to catalytic function. Such mutants can be selected without prior knowledge of the likely mutational events affecting protein–protein interaction.

The tryptophan biosynthetic pathway

A set of seven enzyme activity domains, with the *trp* genetic elements specified by the domain letter, comprise the tryptophan pathway in all organisms studied (*see Figure 1*). Basically, the E-domain, encoded by *trpE*,

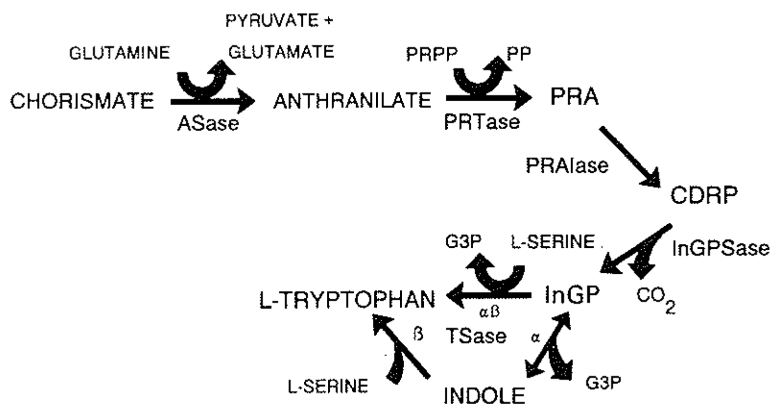


Figure 1. The tryptophan biosynthetic pathway. Chorismate is the last intermediate common to the synthesis of aromatic amino acids. Abbreviations are: ASase, anthranilate synthase; CDRP, 1-(*o*-carboxylphenylamino)-1-deoxyribulose phosphate; G3P, D-glyceraldehyde-3-phosphate; InGP, indoleglycerol phosphate; InGPSase, indoleglycerol phosphate synthase; PP, pyrophosphate; PRA, *N*-phosphoribosyl anthranilate; PRAIase, *N*-phosphoribosyl anthranilate isomerase; PRPP, 5-phosphoribosyl pyrophosphate; PRTase, phosphoribosyl anthranilate transferase; TSase, tryptophan synthase (α, β and αβ reactions indicated).

catalyses the anthranilate synthase reaction, with ammonia as the amino-group donor. The G-domain, encoded by *trpG*, interacts with the E-domain and provides the glutamine amidotransferase activity for the glutamine-dependent anthranilate synthase reaction. The D-domain, encoded by *trpD*, catalyses the phosphoribosyl transferase reaction. The F-domain, encoded by *trpF*, catalyses the phosphoribosyl anthranilate isomerase reaction. The C-domain, encoded by *trpC*, catalyses the indoleglycerol phosphate synthase reaction. The A-domain, encoded by *trpA*, catalyses the conversion of indoleglycerol phosphate (InGP) to indole. The B-domain, encoded by *trpB*, catalyses the conversion of indole to tryptophan (Hütter, Niederberger and DeMoss, 1986).

CHROMOSOMAL ARRANGEMENT OF *trp* GENES

During evolutionary divergence, the arrangement of the genetic elements into genes and the distribution of these genes on the chromosome has produced an array of different systems. For example, certain genetic elements may be fused to give single polypeptide chains having more than one enzymatic function (Crawford, 1989). Furthermore, the genes for tryptophan biosynthesis may map together in a single operon or may be found unlinked as single genes or gene clusters in smaller operons.

In those organisms studied, the chromosomal organization of *trp* genes is represented thus: genes in the same operon are placed in transcriptional order (if known) separated by a dot; presumed to be preceded by a promoter and any regulatory elements and followed by a terminator. If two genes encoding enzymatic functions are fused to give a single polypeptide, the letter for one of them is placed in parentheses. The respective functions are placed in operon order, and the placement of brackets is determined historically in most cases. Hence the *E. coli trp* operon is E.(G)D.C(F).B.A. In some cases other genes are linked to the tryptophan biosynthetic genes. Examples include the *trpI* gene from *Pseudomonas aeruginosa* which has a role in regulation (Chang and Crawford, 1990), and the as yet undefined *usg* gene from *Caulobacter crescentus* (Ross and Winkler, 1988a).

Table 1 shows the chromosomal disposition of the structural elements in bacteria studied to date. The arrangement of species follows the phylogenetic scheme of Woese (1987).

MECHANISMS OF REGULATING TRYPTOPHAN BIOSYNTHESIS

It is not our objective here to review *trp* operon regulation, this has been done by Crawford and Stauffer (1980), Yanofsky (1981) and Somerville (1988). However, a brief overview may prove helpful for subsequent discussion. In the organisms studied to date a variety of regulatory mechanisms have been identified and in many cases more than one level of control is in operation (*see Table 1* for a summary and references).

Regulation has been demonstrated at both DNA and protein levels. At the protein level, simple allosteric modification by tryptophan occurs in an

Table 1. Chromosomal arrangement and regulation of *trp* genes in bacteria

Group	Chromosomal arrangement	Regulation*
Eubacteria		
Purple Bacteria		
γ-subdivision		
Escherichis coli, etc.	E.(G)D.C(F).B.A[1–3]	Repression[4], attenuation[5]
Serratia marcescens, etc.	E.G.D.C(F).B.A[6]	Repression[4], attenuation[5]
Pseudomonas aeruginosa, etc.	E G.D.C F I.B.A[7–12]	Repression(EGDC)[12], induction(BA)[13,45]
Acinetobacter calcoaceticus	E G.D.C F.B.A[14–16]	Repression(EGDC)[14], (FBA)non-coordinate *trp* response[14]
β-subdivision		
Chromobacterium violaceum	G? F? E? D? C? B? A?	Feedback only detected[43]
Pseudomonas acidovorans	E G.D.C F.B.A[17]	
α-subdivision		
Rhizobium meliloti	E(G) D.C F.B.A[18,19]	Attenuation[44]
Caulobacter crescentus	E.D.C G? usgF.B.A[20,21]	Feedback only detected[22]
Gram-positive Eubacteria		
Low G+C subdivision		
Bacillus pumilis	E.D.C G? F? B? A?[34]	
Bacillus stearothermophilus	G? E.B.A.D.C.F[41,42]	
Bacillus subtilis, etc.	G E.D.C.F.B.A[23–25]	Attenuation[26–28]
Clostridium thermocellum	E.G.D? F? C? B? A?[40]	
Lactobacillus casei	E? G? D.C.F.B.A[29]	
High G+C subdivision		
Brevibacterium lactofermentum	E.G.D.C(F).B.A[30,31]	Repression[32], attenuation[32]
Streptomyces griseus	E? G?D? C(F).B.A[35–37]	
Spirochetes		
Spirochete subdivision		
Spirochaeta aurantia	E G? D? C? F? B? A?[6]	
Leptospira subdivision		
Leptospira biflexa	E.G D? C? F? B? A?[6]	
Radio-resistant micrococci and relatives		
Thermophile subdivision		
Thermus thermophilus	E.G.D? F? C? B.A[38,39]	Attenuation(EG)[38]
Archaebacteria		
Methanococci		
Methanococcus voltae	E? G? C? D.F.B.A[33]	

* Regulation by feedback inhibition apparently universal (Crawford, 1989).

References: (1) Crawford, Nichols and Yanofsky (1980); (2) Nichols, Blumenberg and Yanofsky (1981); (3) Yanofsky *et al.* (1981); (4) Crawford and Stauffer (1980); (5) Yanofsky (1984); (6) Crawford (1989); (7) Hedges, Jacob and Crawford (1977); (8) Essar *et al.* (1990); (9) Essar, Eberly and Crawford (1990); (10) Hadero and Crawford (1986); (11) Crawford and Eberly (1989); (12) Manch and Crawford (1981); (13) Chang and Crawford (1990); (14) Cohn and Crawford (1976); (15) Kaplan *et al.* (1984); (16) Sawula and Crawford (1972); (17) Buvinger, Stone and Heath (1981); (18) Bae, Holmgren and Crawford (1989); (19) Johnston, Bibb and Berringer (1978); (20) Winkler *et al.* (1984); (21) Ross and Winkler (1988a); (22) Ross and Winkler (1988b); (23) Band, Shimotsu and Henner (1984); (24) Henner, Band and Shimotsu (1984); (25)Gryczan, Contente and Dubnau (1980); (26) Shimotsu *et al.* (1986); (27) Kuroda *et al.* (1986); (28) Kuroda, Henning and Yanofsky (1988); (29) Natori, Kano and Imamoto (1990); (30) Matsui, Sano and Ohtsubo (1986); (31) Matsui, Miwa and Sano (1987); (32) Matsui and Sano (1987); (33) Sibold and Henriquet (1988); (34) Rivas, Jarvis and Rudner (1990); (35) Smithers and Engel (1974); (36) Meade (1985); (37) Rivero-Lezcano *et al.* (1990); (38) Sato *et al.* (1988); (39) Koyama and Furukawa (1990); (40) Sato *et al.* (1989); (41) Vallier and Welker (1990); (42) Ishiwata *et al.* (1989); (43) Wegman and Crawford (1968); (44) Bae and Crawford (1990); (45) Manch and Crawford (1982).

end-product negative-feedback system, reducing the activity of the enzyme for the first committed step (anthranilate synthase). The feedback mechanism appears to be universally employed (Crawford, 1989).

At the DNA level two basic methods of reducing transcription of the *trp* structural genes have been observed. A simple negative-acting repressor,

activated by free tryptophan in the cell, is used by a number of organisms. The gene for the repressor protein, *trpR*, is unlinked, and its product has been shown to regulate linked genes in *trp* operons (in *E. coli* for example), unlinked genes in *trp* regulons (the unlinked *trpE* and *trpGDC* regions of *P. aeruginosa*) and other amino acid biosynthetic genes. A more complex attenuation system, having many interspecies differences, involving charged $tRNA^{trp}$ levels and mRNA secondary structures, is observed in many organisms.

The co-ordinate regulation of expression in both operons and regulons allows for an efficient biosynthetic pathway. A diversity of systems exist for the regulation of tryptophan production to suit species requirements and natural environments; hence the different controls designed for either a rapid response to transient tryptophan deficiency (allosteric modification of a basal level of enzymes), or for longer term deficiency (relief of repression and attenuation to allow the energy-expensive synthesis of both biosynthetic enzymes and their products). Finally, in *Pseudomonas*, tryptophan synthase expression is induced by the presence of substrate (InGP), which is in turn controlled by repression.

In addition to the primary control of operon expression, relatively weak intra-operon promoters have been described capable of modulating the expression of individual genes (Yanofsky *et al.*, 1981).

Tryptophan synthase: the enzyme

In the following section we attempt to distil from the depth of work on tryptophan biosynthesis and tryptophan synthase those facts essential for an understanding of how to use tryptophan synthase as a model for protein–protein interaction and interpret any results gained. The simple mechanics of the bacterial tryptophan synthase system and its regulation in *E. coli* will be introduced. The extensive mutagenic work will be reappraised, with emphasis given to the effects on αβ interaction, and the intergeneric relatedness of the numerous tryptophan biosynthetic systems will be examined, paying particular interest to tryptophan synthase, its primary amino acid consensus sequence and assays of tertiary structure conservation.

Tryptophan synthase (EC 4.2.1.20), initially named tryptophan desmolase (Yanofsky, 1952), and subsequently tryptophan synthetase (Yanofsky and Bonner, 1955) catalyses the terminal reaction of tryptophan biosynthesis.

The first genetic and biochemical work with tryptophan synthase began using the *Neurospora crassa* system. An understanding of the genetics and biochemistry of tryptophan biosynthesis in commercially important yeasts, fungi, algae and higher plants is still important. Yeasts and fungi are fundamental tools in the biotechnology industry and an understanding of tryptophan biosynthesis offers much in the way of controlling expression and selecting recombinant derivatives. In higher plants, tryptophan is a precursor of many secondary metabolites, for example plant hormones such as the auxin indoleacetic acid, and so its biosynthesis and control in these plants is of considerable importance.

In the specific field of protein interaction, the fungal enzyme is of special interest since the tryptophan synthase genes are fused. The fusion appears to be unique, as evidence exists for separate α and β activities in cyanobacteria (*Anabaena variabilis*, Sakaguchi, 1970), green algae (*Chlorella ellipsoidae*, Sakaguchi, 1970) and higher plants (*Arabidopsis thaliana*, Berlyn, Last and Fink, 1989; *Hyoscyamus muticus*, Fankhauser, Pythoud and King, 1990). The implications of the fusion of the tryptophan synthase genes in the order *trpAB* (as opposed to *trpBA* in the bacterial operon) are discussed in a later section with other aspects of cloning the tryptophan synthase genes from *Saccharomyces cerevisiae, N. crassa* and *Coprinus cinereus*.

Regulation of tryptophan biosynthesis in higher plants appears to be mediated by end-product feedback inhibition of anthranilate synthase (Fankhauser, Pythoud and King, 1990). Feedback control also appears to be the major mechanism for control in fungal systems (Hütter, Niederberger and DeMoss, 1986), where enzyme levels are subject to a system for the general control of amino acid biosynthesis (Hütter and Niederberger, 1983). Nevertheless, specific control of single genes has been observed (Turner and Matchett, 1968; Zalkin and Yanofsky, 1982).

Excepting the fungal system described above, in other organisms the enzyme is an $\alpha\beta_2\alpha$ heterotetramer (Crawford, 1989). In *E. coli* the α-subunit (29 kDa, 2·7S, 268 amino acid residues; Creighton and Yanofsky, 1966; Yanofsky and Crawford, 1972) catalyses the aldolytic cleavage of indole 3-glycerolphosphate (InGP) in what is termed the α-reaction. The β-subunit (397 amino acid residues) normally exists as a β_2 dimer (90 kDa, 5·1S; Creighton and Yanofsky, 1966; Hathaway and Crawford, 1970), containing two molecules of the co-factor pyridoxal phosphate (PLP) per dimer, and catalyses the synthesis of L-tryptophan from indole and L-serine, in what is termed the β-reaction. Upon $\alpha\beta$ subunit interaction, individual α-subunits bind two identical and independent sites on the β_2-subunit with each pair contributing the same enzymatic activity in the $\alpha\beta_2\alpha$ complex (147 kDa, 6·4S; Creighton and Yanofsky, 1966). Importantly, the interacting subunits exhibit an increase in the rates of both α- and β-reactions in the range of 1–2 orders of magnitude. Furthermore, the $\alpha\beta_2\alpha$ complex catalyses the physiologically important reaction, termed the $\alpha\beta$-reaction, which is the sum of α- and β-reactions.

α-reaction (catalysed by α and $\alpha\beta_2\alpha$):
indole 3-glycerol phosphate $\longleftrightarrow$ indole + D-glyceraldehyde 3-phosphate.

β-reaction (catalysed by β_2 and $\alpha\beta_2\alpha$):
indole + L-serine $\rightarrow$ L-tryptophan.

$\alpha\beta$-reaction (catalysed by $\alpha\beta_2\alpha$):
indole 3-glycerol phosphate + L-serine $\rightarrow$ L-tryptophan + D-glyceraldehyde 3-phosphate + H_2O.

The overall structure of the tetrameric complex is an extended $\alpha\beta\beta\alpha$ subunit arrangement with an overall length of 150 Å (Hyde *et al.*, 1988). The

β_2 dimer is located at the centre and the two α-subunits are separated from each other at opposite ends of the complex. The active site of each α-subunit is near the interface with the β-subunit, whereas the β active site is buried deep within the subunit. Since each $\alpha\beta$ pair appear back to back, the formation of the tetramer does not appear necessary for catalysis. It may, however, have some conformational or stability role (Hyde and Miles, 1990).

The active sites of neighbouring α- and β-subunits are separated by a distance of about 30 Å, being connected by a tunnel having a diameter sufficient to accommodate indole. The tunnel passes through the subunit interface with a third of its length comprised of α-residues and the remainder β-residues. The tunnel provides a physical conduit for the internal diffusion of indole between the active sites, preventing its escape into the cytosol (Dunn *et al.*, 1990). Channelling may be particularly important in the case of indole since the molecule is non-polar and may escape the cell through the membrane (Manney, 1970).

The first mutational studies

Extensive mutagenesis played a central role in the initial biochemical and genetic characterization of the *trpBA*/tryptophan synthase system. Apart from providing valuable structure–function information, this work demonstrates both the usefulness of random mutagenesis in the study of protein structure–function relationships and the importance of selection regimes.

Examination of the biochemical, immunological and genetic properties of *trp* mutants during the 1950s helped to elucidate the two-component nature of tryptophan synthase. Mutation in either *trpA* or *trpB* uncouples the physiologically important reaction, with the tryptophan synthase reaction appearing as the two half reactions.

Hundreds of mutant strains with defects in α- or β-subunits were readily detected as tryptophan auxotrophs and independently isolated by penicillin selection. The majority of mutants to be included here are ultraviolet (UV) irradiated derivatives of *E. coli* K-12 and point mutations.

Initial classifications (Lerner and Yanofsky, 1957; Yanofsky and Crawford, 1959) introduced the important concept of cross-reacting material (CRM). Antibodies raised against the wild-type subunit could subsequently react with material present in mutant cell extracts in CRM$^+$ strains, but not in CRM$^-$ strains. α and β CRM formers were defective in their own intrinsic reaction and consequently the $\alpha\beta$-reaction, but retained the ability to enhance their partner subunit in its intrinsic reaction. In these mutants, therefore, catalytic function was compromised but interaction and enhancement sites were apparently intact. The CRM non-formers were unable to participate in any of the tryptophan synthase reactions, so both catalytic and interactive sites appeared to be disrupted.

In initial studies, both types of α-mutants accumulated indoleglycerol and responded to tryptophan or indole in the growth medium. The α CRM formers showed a good growth response to indole because of the stimulation of the β_2-subunit. The poor, but definite, growth response to indole of CRM$^-$

strains was explained by *trp* operon derepression allowing the overproduction of β_2 to compensate for the reduced β-reaction activity.

The first β-mutants isolated only recovered growth in response to tryptophan in the growth medium. The CRM formers could be discriminated from the β CRM non-formers on the basis of accumulation products. β CRM^- strains accumulated indoleglycerol as α alone was insufficient to process the excess InGP produced by the derepressed pathway. β CRM^+ strains produced an activated α-subunit able to perform the aldolysis of InGP at a rate sufficient to produce indole, which was accumulated. Indoleglycerol was also accumulated because of the reversible nature of the indole-forming reaction.

These first studies highlight some important points applicable to protein–protein interaction.

The catalytic and interactive sites appear separate, mutations affecting the catalytic activity of a subunit do not affect its ability to activate the other subunit. It is feasible, therefore, that mutations can be produced at the interaction site without affecting catalysis.

The loss of αβ interactive and activating properties in a CRM^- strain produce a phenotypically detectable change. The unassisted α-subunits accumulate indoleglycerol instead of indole and the unassisted β_2-subunits show a poorer growth response to indole. The effects provide an opportunity to select for non-interactive mutants.

Mutant strains with defects in α or β are readily detected as tryptophan auxotrophs and may be isolated by penicillin selection. Partially active, but nevertheless trp^+ mutants (for example of the non-interactive type), however, will not be selected by this technique (Yanofsky *et al.*, 1961). To detect these important mutants new selection protocols had to be devised, and in the remainder of this section mutants selected using more sophisticated regimes shall be described.

The α-mutants

Analysis of the properties of α-mutants has helped provide insights into the molecular interaction between α and β_2. It is possible to classify amino acid residues into three categories on the basis of interactive function, although one can imagine certain residues being in more than one group.

Separate α- and β_2-subunits, in solution, must recognize each other prior to interaction. We can therefore postulate the existence of residues on both subunits that allow recognition of the partner subunit. Post-recognition, binding will occur and we can predict the existence of residues responsible for the subunit-to-subunit contacts involved in binding. Finally, the presence of residues modulating the strength of interaction is indicated, with roles perhaps involved in the reciprocal transmission of ligand-binding effects.

In addition to the above, the study of α-mutants has demonstrated two types of inactivating mutations. The first is an alteration of a functional group where the mutation has a direct effect on the interaction site. The second type appears to reflect important structural or stability residues, the alteration of which affects conformation at the interactive site. In this case the mutation

affects function indirectly. Differentiation between these mutant types is required in order to assign residues to the binding site.

The roles of intraprotein interaction and α-folding-unit interaction will be examined below.

MUTAGENESIS OF *trpA* AND THE DEVELOPMENT OF SELECTION TECHNIQUES

Genetic and biochemical studies examining hundreds of independently isolated *trpA* mutants, surprisingly revealed only a small percentage of mis-sense mutants, of which many were repeats at the same site. Penicillin selection of *trp* auxotrophs has allowed only the identification of inactivating mutations. This technique has not detected positions having other important, but non-essential roles (for example residues involved in inter-subunit interactions). Indeed, the fact that independent mutations affecting the same site are isolated in preference to mutations at new sites, suggests that α has relatively few key sites wherein a mutation will cause catalytic inactivation. These sites were identified as Phe-22, Glu-49, Tyr-175, Leu-177, Gly-211, Gly-234 and Ser-235. Asp-60 was implicated in later studies (Milton *et al.*, 1986) and Gly-213 was implicated by second-site mutagenesis.

To identify new classes of mutants, and to further classify known essential residues, new approaches were required.

Reversion studies have led to a greater understanding of structure–function relationships for certain key residues. The various selection techniques that have been devised to study reversion are examined below and will be followed by a detailed discussion of those mutants, identified through such studies, that contribute to an understanding of αβ interaction.

Reversion is defined as a second mutational event which repairs, fully or partially, the first mutational event. Reversion may occur at the original site of mutation in primary-site reversion, with changes back to the wild-type residue or another functional residue. Second-site reversion, at other sites in the same gene, may indicate structurally close residues in the folded polypeptide chain involved in important intraprotein interactions. Similarly, second-site mutagenesis of partially active primary-site revertants, yielding inactive protein chains, has given additional information about these interactions.

Primary-site reversion of inactive mutants at α211 have identified a number of fully or partially active substitutions for the wild-type glycine. Murgola and Yanofsky (1974a) devised a selection protocol which did not discriminate between auxotrophic and prototrophic mutants, relying on α's ability to stimulate β_2 by interaction. Amber (UAG) α211 mutants are able to grow on a minimal medium supplemented with indole. This is due to the elevated expression of β_2 compensating for a low, unenhanced catalytic activity. The amber nonsense mutants form an α-fragment that is CRM$^-$ and is unable to activate β_2. The presence of a co-repressor, 5-methyltryptophan (FMT; Murgola, 1985), counteracts *trp* derepression. Consequently, low levels of β_2 are produced which can no longer support growth with limiting levels of indole. Reversion at amber 211 to mis-sense codons allows the formation of

the complete α-chain, now capable of stimulating β_2. These revertants thus regain growth potential on minimal medium containing FMT and limiting indole, whether or not the α revertant has a restored catalytic function.

The FMT/indole selection technique used in the above studies has relevance with respect to the broader study of αβ interaction in tryptophan synthase. Using the basic principle of this selection, we have been able to manipulate FMT and indole levels to produce a medium on which interacting catalytic subunits can support growth, but non-interacting catalytic subunits cannot. Mutants derived from these studies will be described in more detail on pp. 242–243.

In an attempt to identify other key sites, particularly at the active centre, the structure of tryptophan synthase has been probed using non-mutagenic techniques (for example chemical modification and X-ray crystallography). Potentially important residues identified from such studies have been further investigated using site-directed mutagenesis (Kawasaki *et al.*, 1987; Yutani *et al.*, 1987; Ahmed *et al.*, 1988; Nagata, Hyde and Miles, 1989). Four examples that elucidate the roles of Glu-49, Asp-60, Cys-81, Cys-118, Cys-154, Tyr-175 and Arg-179 will be discussed on pp. 241–247.

The reliance of selection protocols on phenotypic alteration, limits the range of α-mutants detectable to essential residues. It may, however, be predicted that pseudo-essential residues or sequences exist that, for example, determine the rate of folding or the final stability of a protein. Mis-sense mutations at these sites are not thought to affect function significantly and therefore such mutations could not have been detected in early mutagenic studies.

Milton *et al.* (1986) selected α-mutants on the basis of DNA changes using a random single base-pair saturation mutagenesis procedure described by Myers, Lerman and Maniatis (1985). Chemical mutagenesis was performed *in vitro* on subcloned *trpA* fragments, and mutagenized fragments were selected on the basis of single base-pair differences from the wild type, using denaturing gradient gel electrophoresis (DGGE). Mutations were, therefore, not selected with any functionality bias.

Examination of the first 206 base pairs (63 amino acids) yielded 20 single base alterations, 12 multiple base alterations and 17 single residue alterations. Although many alterations had apparently no effect, a variety of novel functional defects were observed and the region α28–54 appeared particularly critical for folding. The mutants produced are discussed on pp. 241–247 and shown in *Table 2*.

LOCATION OF INTERACTIVE AND CATALYTIC SITES

An important consideration in the choice of a model system for protein–protein interaction is whether discrimination between changes affecting catalytic activity, and changes affecting interaction is possible. We must be able to differentiate between changes solely affecting interaction, changes solely affecting catalytic activity, and changes affecting both. Obviously, if interaction and catalysis sites are in close proximity, or overlap, local conformational

Table 2. Mutants produced by saturation mutagenesis

Mutant	Stability	β-Activation	α-Activity	Comments
Tyr-4→Cys	0	0	0	unstable
Ser-6→Pro	+	+	+	normal
Pro-21→Thr/Ile-36→Val	+	+	+	normal
Pro-21→Ser	+	+	+	normal
Phe-22→Ser	+	+	0	class IIa
Thr-24→Met	+	+	+	normal
Pro-28→Ser	0	+	+	unstable
Pro-28→Leu	+	0	+	non-interactive
Ser-33→Leu	0	0	+	unstable, non-interactive?
Gly-44→Ser	0	0	+	unstable, non-interactive?
Ala-47→Thr/Ala-59→Thr	0	+	0	unstable, class IIa
Gly-51→Asp	0	0	+	unstable, non-interactive
Pro-53→His/Leu-58→Gln	+	0	0	double mutant
Phe-54→Ile	0	0	0	unstable
Asp-60→Asn	+	+	0	class IIa
Thr-63→Lys	+	+	+	normal

Stability was measured over a 22 hour induction period.
+, >75% wild-type activity.
0, <50% wild-type activity.

changes may prohibit the production of mutants solely affecting either property. It is helpful, therefore, to map the respective locations of both sites on the α-protein.

The α CRMs produced by many mutants can be distinguished from both each other and wild-type α on the basis of physical properties. Maling and Yanofsky (1961) identified a group of mutants, mapping in a cluster at the front of the *trpA* gene, that produce α CRM which is more susceptible to acid precipitation at pH 4. A second group, mapping in a cluster at the end of the *trpA* gene, produces a substantially more heat labile α CRM. Binding studies with these mutants has shown that each has a similar affinity for β_2, which is equivalent to normal α. A mutation of α, therefore, can render the catalytic site inactive without having a significant effect on the β interaction site.

Despite this early work suggesting that interactive and catalytic sites are separate, the development of more sensitive techniques to assay the effect of single and double mutations on α tertiary structure has demonstrated an area of overlap between interactive and catalytic regions (Murphy and Mills, 1968). These authors suggested that the mutations provoked only local conformational changes since quantitative immunochemical studies (with reaction to anti-α) failed to demonstrate whole protein surface changes. Half (10/21) the single and double mutants, however, had reduced β_2 binding abilities as evidenced by low association constants. Reduced affinity for β_2 was observed for A487 (Leu-177 → Arg), A46 (Gly-211 → Glu), A46PR9 (Gly-211 →Val) and A187 (Gly-211→Val; Gly-213 → Val) amongst others. The following mutants, however, had affinities similar to wild type: A3 (Glu-49 → Val), A33 (Glu-49 → Met), A446 (Tyr-175 → Cys), A223 (Thr-183 → Ile), A23 (Gly-211 → Arg), A58 (Gly-234 → Asp), A78 (Gly-234 → Cys) and A46PR8 (Tyr-175 → Cys; Gly-211 → Glu).

This study offered a new perspective on $\alpha\beta$ interaction and the effects of mutations at key sites. Position 211 appears to be involved in interaction in addition to catalysis. For example, changes at α211 (in A46, A46PR9 and A187), or at spatially close residues (in A187 and A487) have an effect on the strength of $\alpha\beta$ interaction.

Milton *et al.* (1986) described altered α-proteins generated by saturation mutagenesis with reduced (Tyr-4 $\rightarrow$ Cys) or no intrinsic enzyme activity (Phe-22 $\rightarrow$ Ser, Ala-47 $\rightarrow$ Thr/Ala-59 $\rightarrow$ Thr, Pro-53 $\rightarrow$ His, Pro-53 $\rightarrow$ His/Leu-58 $\rightarrow$ Gln, Phe-54 $\rightarrow$ Ile and Asp-60 $\rightarrow$ Asn). Two subclasses were demonstrated, comprising those able to activate β_2 and those not (*Table 2*). The reduction/abolition of activity can be assumed to be due to changes at or around the active site. Changes affecting both catalysis and the ability to stimulate β_2 can be considered to be located in the proposed area of overlap between active and interaction sites.

It is not surprising that an area of overlap has been demonstrated. Hyde *et al.* (1988) described the channelling of indole from the α active site to that of β via the indole tunnel. The tunnel begins at the α active site and must pass through the interactive region before reaching the β active site. We can therefore envisage areas where changes would affect catalysis and interaction.

Nevertheless, changes have been observed which inhibit catalytic activity, may affect the strength of interaction, yet do not abolish interaction. This suggests that there are also key interactive residues that remain unchanged in many of the mutation studies. Furthermore, as $\alpha\beta$ interaction occurs over a relatively large area (1190 $Å^2$ of α and 1110 $Å^2$ of β; Hyde *et al.*, 1988), in comparison to that of the lysozyme–anti-lysozyme reported surface of interaction (750 $Å^2$, Sherriff *et al.*, 1987; and 690 $Å^2$, Amit *et al.*, 1986), residues at the overlap may give a small, but observable interaction effect.

Mutational studies, therefore, have identified residues affecting the strength of interaction through effects on association constants and β_2 enhancement. It is probable that these changes act at residues involved in modulation of the strength of interaction and not in direct physical contact with β_2.

It is interesting to note that data derived in the early 1960s showed that non-functional α-mutations map at the ends of the *trpA* gene (Yanofsky, Helinski and Maling, 1961). The middle of the polypeptide would appear unrelated to catalytic function and, despite the subsequent demonstration of overlapping regions having effects on both catalysis and $\alpha\beta$ interaction (described above), we can still envisage α incorporating separate regions solely involved in interaction or catalysis. The central region could be predicted to constitute the β interactive site. With the advent of recent X-ray crystallographic studies (Hyde *et al.*, 1988), which have identified the regions contacting β_2 to be part of the central region, this inference derived from early mutational studies is confirmed.

RESIDUES IMPLICATED IN α FUNCTION

Random mutagenesis, second-site mutagenesis, chemical modification and

X-ray crystallography have identified a number of potentially key α-residues. In many cases site-directed mutagenesis (SDM) has been used to further study the roles of such residues. Positions having catalytic, interactive and structural roles are discussed.

Catalytic residues

α Glu-49. Initial studies (Yanofsky, 1967; Yanofsky and Crawford, 1972; Yanofsky and Horn, 1972; Murgola and Yanofsky, 1974c; Yutani *et al.*, 1984) identified seven mis-sense mutations at α49, suggesting a role in catalytic function. To thoroughly evaluate Glu-49 in terms of structure and function, Yutani *et al.* (1987) used SDM at α49 to complete the set of variants possible at this position. All 20 α49 variants formed an $\alpha\beta_2\alpha$ complex, full enhancement of β_2 was obtained and similar association constants were observed. Thus, none of these changes significantly modulate interaction and α49 would not appear to be in the locale of the interaction site.

The 19 mutant αs, however, are totally devoid of activity in the reactions normally catalysed at the α active site. The finding that even the conservative Glu-49 to Asp substitution is inactive is especially significant, being strong evidence that Glu-49 is an essential catalytic residue. X-ray crystallography results agree with an active-centre location for Glu-49.

α Asp-60. Nagata, Hyde and Miles (1989), using the *Salmonella typhimurium* enzyme, studied the role of Asp-60, a key residue described by Milton *et al.* (1986). X-ray crystallography has located the carboxyl group of Asp-60 in close proximity to the N-atom of the substrate analogue indole 3-propanolphosphate (IPP) at the active site of α. The suggestion, therefore, is that Asp-60 is involved in catalysis. Replacement by asparagine, alanine or tyrosine results in a loss of activity, although binding of IPP to the mutants occurs and elicits activity and spectral effects observed in wild-type α. Asp-60 appears not to affect substrate binding, but only catalysis.

X-ray crystallography has identified a weak electron density at Asp-60, believed to reflect mobility of the side-chain (Nagata, Hyde and Miles, 1989). Substitution with glutamic acid by SDM, resulted in a partially active mutant. α60 is a highly conserved residue (*see* pp. 264–282) with aspartic acid present in all but *Caulobacter crescentus*, where glutamic acid is preferred (*see Figure 5*). It has been suggested that α60 has limited flexibility, with a need only for the carboxyl group, and may, therefore, be a less rigid catalytic residue than Glu-49.

Other putative catalytic sites. Milton *et al.* (1986) isolated a double mutant with Thr-47/Thr-59, which is catalytically inactive, but interactive. It would be interesting to know whether one of these mutations alone could cause inactivation, or whether inhibition is due to a cumulative effect. An additional double mutant, with His-53/Gln-58, is both inactive and non-interactive. Clearly, it would be interesting to know whether one change

caused both the effects, or whether singly the mutations affected catalysis or interaction. At present these aspects await clarification.

α Active site. The active site for the aldolysis of InGP was located using X-ray crystallography of the enzyme–IPP complex (Hyde *et al.*, 1988). The active site is formed from side-chains of residues near the C-termini of the eight β-strands (*see also* pp. 251–255). These regions contain many residues that are intergenerically conserved and for which substitution gives rise to inactive mutants. The indole moiety of the substrate appears to contact Phe-22, Leu-100, Tyr-102, Leu-127, Ala-129, Ile-153, Tyr-175 and Leu-177. The phosphate side-chain of the substrate appears to contact the Gly-211–Gly-213 region and the Gly-234–Ser-235 region. Conformational changes on substrate binding occur only at the phosphate-binding area. Glu-49 is in the hydrophobic core of the enzyme near the scissile bond and Asp-60 is near the nitrogen atom of indole.

Interactive residues

Milton *et al.* (1986) isolated a group of mutants that were non-interactive, being unable to activate $β_2$ but retaining their intrinsic enzyme activity. These mutants were Pro-28 → Leu, Ser-33 → Leu, Gly-44 → Ser and Gly-51 → Asp. Initially it was hoped that these mutants specified part of the interaction site; however, further evidence has since tempered this view. X-ray crystallography has identified areas of the α-protein surface that make important contacts with $β_2$ (Hyde *et al.*, 1988). The residues at α28, α33 and α44, although at or near the protein surface, are not in any of those contact areas. The residue α51, however, is located at the surface in an area thought to contact β and is therefore the only putative non-interactive mutant of Milton *et al.* (1986) that does not contradict the three-dimensional structure. Instability of the α-protein from the non-interactive mutants (*Table 2*), in conjunction with the phenotype of site-specific changes at α33, α44 and α51 (Milton *et al.*, 1986), combine to suggest a structural rather than functional role for these residues, and their effects are therefore likely to be secondary. It should be noted, however, that Lim, Smith-Somerville and Hardman (1989) in developing a technique to purify the α Leu-33 mutant from insoluble cell debris, used the activation of $β_2$ in the β-reaction to assay for the protein. This is in contradiction to the original observations of Milton *et al.* (1986) which considered the Leu-33 mutant to be non-interactive. No further clarification of this discrepancy is currently available.

In our laboratory we have used the inherent infidelity of the *Taq* polymerase in the polymerase chain reaction (PCR) to randomly generate mutations in the *E. coli trpA* gene (Swift, Kuhn and Stewart, manuscript in preparation). *TrppoL*-controlled expression of plasmid-encoded, PCR-derived *trpA* in an *E. coli trpA*$^-$ mutant has permitted phenotypic screening and the selection of non-interactive α-mutants.

The non-interactive α-phenotype was expected to impart a reduced capacity for tryptophan biosynthesis on any mutant clone, evidenced by slow

growth on a minimal medium. Furthermore, the inability to enhance the β_2-subunit was expected to reduce growth rate on minimal medium containing limiting levels of indole. As discussed earlier when outlining the development of selection techniques for mutant αs, the important co-repressor effect of FMT was introduced. Putative non-interactive αs were identified by their inability to grow on a minimal medium containing high levels of FMT (with low levels of indole being supplied by the unenhanced α-subunit). Plasmid DNA isolated from these mutants was used to transform a mutant *E. coli* carrying both a deletion of part of the *trpA* gene and a point mutation in the *trpB* gene which renders β_2 inactive unless interacting with α (further defined on pp.255–262). Four transformants, unable to make tryptophan, were deemed to produce a non-interactive α-subunit on the basis that interaction would have restored prototrophy. Plasmid sequencing of *trpA* from these clones identified the changes involved as Lys-109 → Arg, Cys-118 → Arg, Val-126 → Glu and Val-128 → Asp.

In conjunction with X-ray crystallography (*see Figure 4*) and intergeneric comparison (*see Figure 5*) it is possible to further describe these mutants. Lys-109 is located in the loop between β-strand 3 and α-helix 3 in a region making important contacts with the β-subunit (Hyde *et al.*, 1988). α109 is not, however, conserved between species, with arginine incumbent in some cases. It is possible, therefore, that the αβ interaction effects of the α Lys-109 → Arg change are limited to *E. coli*.

Cys-118, as we will mention again later, appears to perform a key structural role in the α-subunit. It is located in α-helix 3, away from the αβ interaction surface (Hyde *et al.*, 1988) and conserved as cysteine in the Enterobacteriaceae studied, *N. crassa* and *Lactobacillus casei* (*see Figure 5*). It is probable that introduction of the bulky charged arginine at α118 distorts the conformation at the nearby β interaction site. The mutation appears, therefore, not to be at the actual binding site, but at some secondary position.

Both α126 and α128 are located in the loop between β-strand 4 and α-helix 4, a region identified as making important contacts with the β-subunit (Hyde *et al.*, 1988). Furthermore, intergeneric comparisons show both residues to be conserved as a valine or similar aliphatic side-chain amino acid, suggesting an important functional role. The effect of these mutants is likly to be at the site of direct interaction between α- and β-subunits, with the introduction of a charged residue disrupting important hydrophobic interactions involved in subunit interaction. Evidence implicating hydrophobic interactions in subunit interaction has been provided by analytical ultracentrifugation studies (DiCamelli, Balbinder and Lebowitz, 1973; DiCamelli and Balbinder, 1976).

Key structural residues with a role in conformation but not directly implicated in catalysis or interaction

α Gly-211 and Gly-234. Henning and Yanofsky (1962) described two previously isolated α CRMs, A23 and A46, with alterations extremely close to one another on the genetic map. The A23 mutant α was shown to be heat

labile, whereas A46 α was heat stable (Maling and Yanofsky, 1961). Peptide fingerprinting identified the A23 α-trypsin/chymotrypsin peptide harbouring the change. Sequencing of the relevant peptide from A23 and the analogous A46 peptide identified a change at the same residue, α211. The wild-type glycine had been replaced by arginine in A23, and glutamic acid in A46 (Yanofsky, Helinski and Maling, 1961). These independent mutagenic events inactivate α, but provide the α CRMs with different heat stabilities. This suggests that α211 may be a key structural or stability residue in the area of the active site and may not have a key catalytic function.

Subsequent reversion studies with A23 and A46 (Henning and Yanofsky, 1962) showed that any one of four different amino acids can functionally occupy the α211 position. In addition to the wild-type glycine, alanine and serine replacements at α211 appear as full revertants. A substitution with valine at α211 gave a partial revertant with slight activity in the αβ-reaction.

Mutational and reversion studies have told us much about position 211 in the α-chain. α Gly-211 is a key position, being one of the few sites at which mutation completely abolishes α catalytic activity. Substitution of α Gly-211 with similar small, uncharged residues gives full activity. Substitution with larger, non-polar residues, such as valine, however, yields only a partially active α, suggesting that these residues interfere with some spatial requirement. Charged groups appear to have a profound effect on the local area of α211, with substitutions to arginine$^+$ and glutamic acid$^-$ resulting in inactive α-chains. These early studies confirm the usefulness of primary-site reversion in elucidating the structural constraints on a particular position, once that position has been identified by mutagenesis.

Reversion of amber α211 codons (Murgola and Yanofsky, 1974a) to form complete α-chains able to interact with β_2, identified functional and non-functional residues that fitted the original scheme as regards the importance of size and charge. Non-polar leucine was partially active, charged lysine and glutamine were inactive, and bulky residues with aromatic side-chains, tyrosine and tryptophan were also inactive. The seven α211 residues shown to have activity were arranged in an order according to their *in vivo* activity. The activity range is glycine ≥ alanine > serine > threonine > isoleucine > valine > asparagine.

Parallel studies to those on α211, have also been carried out on α234. Murgola and Hijazi (1983) used the indole/FMT technique (previously described) to increase the number of mis-sense mutations known at α234; a site identified by Yanofsky, Ito and Horn (1966) and Murgola (1981) as a crucial residue. Glycine (wild type) and alanine were found to be active, but serine, leucine, cysteine, glutamic acid, glutamine, aspartic acid, arginine, lysine, tryptophan and tyrosine were inactive. The α234 glycine therefore is not as tolerant to change as α211.

An interesting perspective on the situation at α211 and α234 was given by Tucker, Murgola and Hijazi (1989). Using reversion from a nonsense codon to identify mis-sense revertants able to stimulate β_2, they isolated two novel deletion mutants. The first, derived from an α211 nonsense mutant, had a deletion of 21 base pairs, extending from the third nucleotide of α209 to the

second of α216. This deletion removed seven codons, including α211, and changed α209 from leucine to phenylalanine. Although unable to support growth on minimal medium, this mutant α was able to fully activate β_2. The second, derived from an α234 nonsense mutant, had a 3 bp deletion removing the stop codon at α234. The resultant strain grew as well as wild type on indole/FMT medium, indicating a strong interaction with β_2. The poor, but definite, growth on minimal medium of the revertant shows that catalytic ability was retained even though a previously suspected key residue had been deleted.

It seems, therefore, that α211 is not directly involved in interaction and α234 is not directly involved in interaction or catalysis. Mutation at these sites, however, appears to disrupt important structural motifs. Furthermore, it appears that spatial requirements are influential, especially at α234, where deletion of the residue has less effect on activity than the introduction of a large or charged molecule.

α Pro-28. It is interesting to note the effect of the amino acid present at α28, described by Milton *et al.* (1986). Wild-type proline is fully active, interactive and stable. Mutant serine, although active and interactive, is unstable. Mutant leucine, although stable and active in the α-reaction, is non-interactive. α28 therefore appears to be a key position for stability and possibly interaction, although the interaction effect may be a result of secondary conformational changes.

α Cysteine residues. Early chemical modification studies with the *E. coli* α-subunit suggested that the three cysteines had functional roles. The inhibition observed was due to substitution of the sulphydryl groups with a bulky charged group derived from the sulphydryl reagent. Hardman and Yanofsky (1965) and Malkinson and Hardman (1969) found that the three sulphydryl residues reacted to different extents with various reagents at different pH values. This was interpreted as suggesting that the three groups were situated in the same area of native α, since the reaction with one molecule of a bulky reagent stearically restricted subsequent reactions.

Freedberg and Hardman (1971) demonstrated the similar locations of Cys-81 and Cys-118 by cross-linking with the bifunctional sulphydryl reagent *N,N′-bis*(maleimidomethyl)ether. Hardman and Hardman (1971) observed that the binding of InGP protected Cys-81 and Cys-118 from modification by *N*-ethylmaleimide and *N,N′-bis*(maleimidomethyl)ether, and concluded that the cysteine residues were at the substrate-binding site. Nevertheless, because of the bulk of the maleimide ester, the potential for steric effects or conformational distortion being directly responsible for the apparent role of the cysteine residues was not excluded.

Ahmed *et al.* (1988) continued to investigate the roles of these cysteine residues in the α-subunit from *S. typhimurium* by replacement with serine at α81, 118 and 154. The serine-substituted mutants are useful for heavy-metal binding which can help in X-ray crystallographic studies. It was important,

therefore, to ascertain whether the individual cysteine → serine changes had any effect on the α-protein.

Although the Ser-154 mutant proved unstable and was not studied, the Ser-81 and Ser-118 mutants were purified and fully active in tryptophan synthase reactions. Mutant complexes dissociated more readily and were less heat stable, however, and consequently structural roles were proposed for the α cysteines.

Our own work has added to the evidence suggesting a structural role for α Cys-118. As previously mentioned, we have identified a cysteine → arginine change at α118, producing an α which, while retaining intrinsic catalytic function, is unable to interact with β_2. The amino acid substitution with arginine introduces a bulky, charged group which is likely to have more drastic effects on the local conformation than a change to serine. X-ray crystallography (Hyde *et al.*, 1988) locates the α118 position in α-helix 3 (*see Figure 4*) away from any sites of direct contact between α and β, indicating that any effects on interaction are secondary, and probably the result of conformational changes at the interaction site induced by introduction of the bulky, charged arginine group. Position α118 in other organisms studied is invariably hydrophobic and, although phenylalanine is present on two occasions (*see Figure 5*) a dislike for large, charged residues is seen.

The disagreement between the results of chemical modification and SDM demonstrates the limited use of chemical modification techniques in probing structure–function relationships. The results from both sets of studies do tell us that alterations in the vicinity of functional residues can have effects on activity, even though the altered residue is not directly involved. Steric hindrance effects and local conformational changes are believed responsible for these effects. We must therefore bear in mind that mutational changes, identified as affecting the αβ interaction, may in fact be at one of these secondary sites.

STRENGTH OF INTERACTION

We have mentioned previously the concept of residues that, while not being involved in the physical binding of β, do affect the strength of interaction.

The effect of substitution of α211 on an α22 phenylalanine → leucine mutant

The involvement of α211 in αβ interaction, initially thought not to occur, was only detected due to the development of more sensitive techniques (Murphy and Mills, 1968). Changes at and around α211 do not inhibit interaction, but do affect the strength of interaction, with a reduction in the association constant being measured. It is possible, therefore, that this area of α is involved in the modulation of the strength of αβ interaction, once the subunits have recognized and bound each other.

The cumulative effects of mutation at α211 and α22, which may inhibit catalysis depending on the nature of the residue at α211, have been shown to have a similar effect on the ability to enhance β_2 activity (*Figure 2*).

Examination of colony sizes (a qualitative measure of growth rate) for the double mutant auxotrophs on minimal medium supplemented with low levels of indole and the co-repressor 5-methyltryptophan (FMT), suggests that mutants may differ in their ability to activate β_2 (Murgola and Yanofsky, 1974b). Theory predicts that FMT limits the expression of β_2 by acting as a co-repressor, thus low β_2 activity cannot be compensated for by high levels of expression, and activation by α is required for good growth. The reduction in colony size, down the α211 activity range, suggests that the more drastic the change at α211 the greater the effect, both on catalysis and the ability to activate β_2.

α Arg-179

Kawasaki *et al.* (1987) targeted Arg-179 on the basis of chemical modification studies with phenylglyoxal (Eun and Miles, 1984), which had indicated a crucial role in the binding of InGP for an arginine residue. Arg-179 was chosen because of its proximity to previously determined key sites at α175 and α177. The work was performed on the *S. typhimurium* system.

Analysis of the mutant suggested that Arg-179 played a role in the reciprocal transmission of substrate-induced conformational changes which occur between native α- and β-subunits in the $\alpha\beta_2\alpha$ complex. The mutant complex differed from wild type in the nature and extent of conformational changes transmitted between heterologous subunits, induced by ligand binding.

It appears, therefore, that α179 is involved in the modulation of the $\alpha\beta$ interaction. The exact role of α179 will be discussed on p. 255.

Intraprotein interaction

Yanofsky *et al.* (1961) have described an A46 (Gly-211 → Glu) partial revertant, A46PR8. Peptide fingerprinting showed that glutamic acid was retained at the α211 mutant site but a second change, identified as Tyr-175 → Cys, was detected in another peptide. The second change appeared to compensate for the first. Surprisingly, the single mutant containing α Cys-175

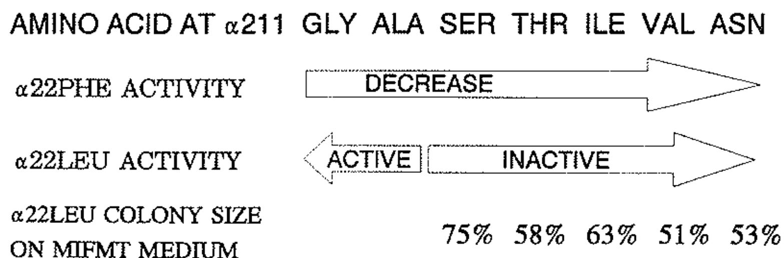


Figure 2. The effect of changes at α22 and α211. The effect of amino acid substitution at α211 (wild-type glycine) on activity with wild-type α22 phenylalanine and mutant α22 leucine is shown. The effect on the ability to enhance β_2 activity of α22 leucine mutants with inactivating changes at α211, measured by a reduction in colony size (given as a percentage of the wild-type growth on a minimal medium) on a minimal medium supplemented with indole and 5-methyl tryptophan (MIFMT), is also shown.

and wild-type Gly-211, was defective. These findings were considered to reflect some functionally significant relationship between these two specific regions of α; an intraprotein interaction.

Yanofsky, Horn and Thorpe (1964) have provided further proof of intraprotein interaction between the regions around α175 and α211. Moreover, they have demonstrated that the cumulative effects of individual mutations reducing α activity, could collectively inactivate the α catalytic function. The double mutant A187 has glycine at α positions 211 and 213 replaced by valine. Single mutants with valine substituted at either position were partially active. It seems that spatial requirements in this region can functionally accommodate one valine, but not two. A second-site revertant of A187 with Leu-177 → Arg was partially active, further demonstrating intraprotein interaction.

To identify further residues with structural proximity to α211, mutagenesis of a Gly → Ser α211 revertant (which provides a modified α only slightly less active than wild type) was performed (Yanofsky and Horn, 1972). An inactive double mutant was isolated (A218) with the second change, Phe → Leu, at a hitherto undetected position, α22.

As mentioned previously, Murgola and Yanofsky (1974b) have demonstrated that the single mutant α Leu-22 is partially active. Given that α Ser-211 and α Leu-22 are both partly active in single mutant αs, it is the accumulated effects of both mutations in a single α-chain that lead to inactivation. Furthermore, the seven α211 residues previously shown to produce α-mutants with an *in vivo* activity range of glycine ⩾ alanine > serine > threonine > isoleucine > valine > asparagine, formed double mutants with Leu-22. The mutants in the range from serine to asparagine were inactive, whereas alanine and glycine were active. Again, the cumulative effects of both mutations appear responsible for inactivation.

α Tyr-175

X-ray crystallography data have located Tyr-175 in the α active site, with the phenolic hydroxide group in close proximity to the side-chain of IPP (Hyde *et al.*, 1988). Although previous mutagenic studies had suggested a critical role for α175, functional replacement with phenylalanine had ruled out a role for the phenolic hydroxyl (Nagata, Hyde and Miles, 1989). Furthermore, replacement of α Tyr-175 with phenylalanine (functional) or cysteine (non-functional) does not affect IPP binding.

Since α Tyr-175 is not involved in catalysis or substrate binding, its effect seems to be to distort the proper geometry of the active site. Thus, while substitutions are detectable as providing a mutant phenotype, its effect is secondary to the actual catalytic and binding residues.

Second-site reversion studies have observed intraprotein interaction between α175 and α211 (Yanofsky, 1967). The restoration of α activity in the second-site revertant is due to restoration of the proper active-site geometry which the single mutation distorts.

α FOLDING UNITS

Intraprotein interactions have been further characterized by folding studies which have described two independent folding units within the single-domain α-subunit.

Complementation studies

Jackson and Yanofsky (1969) demonstrated complementation for the recovery of α catalytic activity between inactive mutant α-subunits in artificial dimers formed following renaturation in 6 M urea. These results prompted them to suggest the interaction between two α-domains. Complementation depends on the unfolding of the mutant α-chains and the subsequent formation of a heterologous dimer during refolding. Furthermore, only heterologous dimers formed from mutant proteins having amino acid substitutions at opposite ends of the α-chain regain enzyme activity. Refinement of the technique allowed fragments of α, generated by either cyanogen bromide cleavage or from nonsense mutants, to complement intact mutant αs, where overlap of the site of amino acid change was observed.

The presence of the β_2-subunit in the denaturation/renaturation reaction inhibited the formation of α dimers. It was suggested that $\alpha\beta$ interaction blocked $\alpha\alpha$ formation, or that the presence of β during renaturation promoted a more rapid and accurate folding of α.

Limited proteolysis yields an active complex

In 1978 Miles and Higgins demonstrated that limited tryptic proteolysis of $\alpha\beta_2\alpha$ yields an active complex, containing nicked-α (α'), of the formula $\alpha'\beta_2\alpha'$. α' was resolved into α–1 and α–2 components using gel filtration in 6 M urea (Higgins, Fairwell and Miles, 1979). The cleavage site was identified at α Arg-188 using N-terminal protein sequencing. Circular dichroism (CD) studies demonstrated that the independently renatured fragments retained an ordered tertiary structure. Neither fragment alone had catalytic activity nor the ability to bind either the catalytic inhibitor indole 3-propanolphosphate (IPP) or the holo-β_2 subunit. Interestingly, upon mixing α–1 and α–2, catalytic and binding properties were restored, leading to the suggestion that α–1 and α–2 were independently folding units that must have interacted to form a functional enzyme.

Folding studies on the native α-subunit

Initial analysis of unfolding and refolding transitions (Crisanti and Matthews, 1981; Matthews and Crisanti, 1981), using UV difference spectroscopy at 286 nm, CD measurements at 222 nm and urea gradient gel electrophoresis, indicated a folding process more complex than a direct native to unfolded transition and supportive of a transition involving one or more partially folded forms.

Miles, Yutani and Ogasahara (1982) compared the effect of guanidine hydrochloride (GuHCl) on the unfolding of α, α′, α–1 and α–2. CD measurements and the susceptibility of tryosyl residues to chemical modification were used to determine the degree of 'foldedness', with the results revealing a stepwise unfolding. First, (at lower GuHCl concentrations) α–2, or the equivalent unit in the intact α-subunit, unfolded. Subsequently (at higher GuHCl concentrations) α–1, or the corresponding unit, unfolded. The ^{1}H-NMR studies of Iwahashi *et al.* (1983) supported this stepwise domain-unfolding hypothesis.

Notwithstanding that folding of the nascent polypeptide chain probably begins early during translation (Tsou, 1988), studies on the refolding of complete chains also describe α folding as the stepwise folding of two units. The prediction that the N-terminal α–1 unit folds first is in agreement with the co-translational folding model (Tsou, 1988).

Refolding kinetics have suggested that the rate-limiting step for the folding of α–1 is a proline isomerization. The subsequent α–2 folding, to give native α, appears more complex and involves a pH-dependent step (Crisanti and Matthews, 1981). Both steps are considered in more detail in the next two sections.

Folding studies with α-mutants

Urea-induced unfolding and refolding studies on mutant α-subunits (Matthews *et al.*, 1983; Beasty *et al.*, 1986; Hurle, Tweedy and Matthews, 1986) have given a clearer understanding of the α–2 folding step and the concomitant production of native α.

Matthews *et al.* (1983) working with the non-functional α Gly-211 → Glu mutant, observed the pH-dependent refolding transition to be composed of both a pH-dependent (twice as slow in the mutant) and a pH-independent phase (rate as wild type). The pH-independent step showed kinetics consistent with proline isomerization and appeared to be the rate-limiting step in the folding of the α–2 unit. The pH-dependent step follows this proline isomerization and consists of some additional folding and/or folding-unit-association steps.

Beasty *et al.* (1986) examined the final step in more detail using the non-functional mutants α Phe-22 → Leu, α Glu-49 → Met, α Gly-234 → Asp and α Gly-234 → Lys. With the exception of α Lys-234, all the mutants had an effect on the kinetics of the final α folding. As residues 211 and 234 in the α–2 unit and residues 22 and 49 in the α–1 unit affect this final step, it was concluded that unit association or some other molecule-wide phenomenon was involved, and not simply folding of α–2. These results are consistent with second-site mutagenesis studies implying intraprotein interactions involving these α-residues. It was suggested, therefore, that these residues interact at the interface of the two folding units.

Hurle, Tweedy and Matthews (1986) have demonstrated a further interaction that adds to the stability of the α-protein between α175 and α211. The interaction is present in the wild-type α with Tyr-175 and Gly-211 and the

second-site revertant α with Cys-175 and Glu-211, but not in either of the inactive single mutants. Extrapolation of this stability effect to second-site reversion studies of αβ interaction, should help locate interactive residues on complementary subunits.

Hurle *et al.* (1987) further investigated the folding-unit-association step by looking at the effects of viscosity, pH and temperature on urea-induced unfolding and refolding kinetics. An increase in solution viscosity, effected by the addition of sucrose to 0·58 M, indicated that unit association was not solely dependent upon diffusion: viscosity effects were not as pronounced as expected for a simple diffusion-limiting process. pH studies indicated that the protonation of a single residue (pK 8·8 (native), pK 7·1 (intermediate)) occurred, suggesting that a key ionizable residue is important in unit association. The energy change involved in protonation is in the range observed for electrostatic interactions. The activation energy for unfolding was calculated at near zero, in the range expected for the disruption of hydrophobic or elctrostatic interactions.

Proline isomerization

Urea-induced unfolding followed by rapid refolding after a known period of time, was used by Hurle and Matthews (1987) to study the proposed proline isomerization steps (*see also* above). Results for the rate-limiting step in α–2 unit folding were consistent with the proline isomerization theory. However, results for the analogous step in the α–1 unit were more rapid than expected for proline isomerization and have yet to be precisely defined in structural terms. The results did, however, demonstrate another unfolded species, the folding of which was not limited by either isomerization or the domain association.

The folding of the α-subunit is summarized in *Figure 3*.

SECONDARY AND TERTIARY STRUCTURE STUDIES DESCRIBE α AS AN EIGHTFOLD α/β-BARREL

In all of the previous discussions we have deliberately used the term α folding unit rather than α folding domain, which was employed in the original literature. We have done this in order to limit confusion in contempory nomenclature; however, at this point it is necessary to explain the basis for using the term folding unit in preference to domain.

Results from folding studies on both wild-type and mutant α-subunit have been interpreted as consistent with the folding of a two-domain protein. Secondary and tertiary structure studies (Crawford, Niermann and Kirschner, 1987; Hyde *et al.*, 1988) described below, however, contradict a two-domain structure for the α-subunit. In the light of this recent work it is necessary to slightly modify the original interpretations derived from folding data. The data itself remains consistent with the latest folding model, all that is required is the marginal adjustment of considering folding units within a single-domain structure.

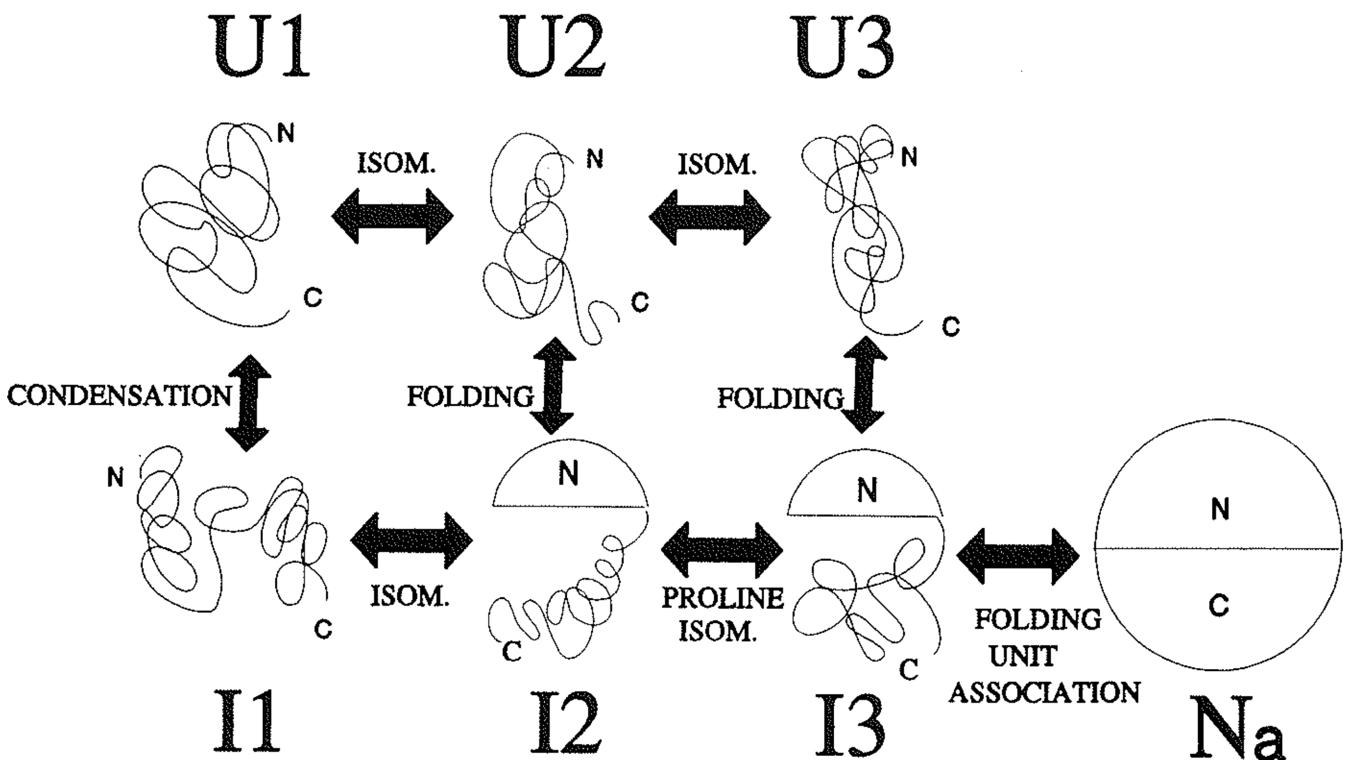


Figure 3. Diagrammatic representation of the proposed folding model for the α-subunit of tryptophan synthase from *Escherichia coli*. Na, I and U represent native, intermediate and unfolded forms, respectively. Amino (N) and carboxy (C) termini/folding units are indicated. Isomerization is abbreviated as ISOM. (Adapted from Tweedy *et al.*, 1990.)

Crawford, Niermann and Kirschner (1987) predicted the secondary structure (α-helices, β-strands, loops and random coils) for 10 α-sequences all possessing some primary structure homology. The conserved secondary structure observed, of eight repeated β-strand-loop-α-helix units with an extra N-terminal α-helix, suggested a tertiary structure of an eightfold α/β-barrel; a single-domain structure which has been observed in other proteins. The data also suggested that both the catalytic site and the β binding site were at the end of the barrel formed by the C-termini of the β-strands. The three-dimensional structure of α determined by Hyde *et al.* (1988) using X-ray crystallography of the *S. typhimurium* tryptophan synthase complex, confirmed the eightfold α/β-barrel conformation. The structure was described as a barrel with the eight β-strand–loop–α-helix units packed as staves, with unit 8 adjacent to unit 1 due to the twisting of the molecule. N-terminal α-helix 0, preceeding unit 1, acts to cap the bottom of the barrel opposite the active site (*Figure 4*). The structure actually resembles a canon more than a barrel due to the staves being at about an angle of 35° to the vertical. Loops occur between the strand and helix of each stave. Of particular interest are loop 2, highly conserved and large at 26 residues; loop 6, which contains the

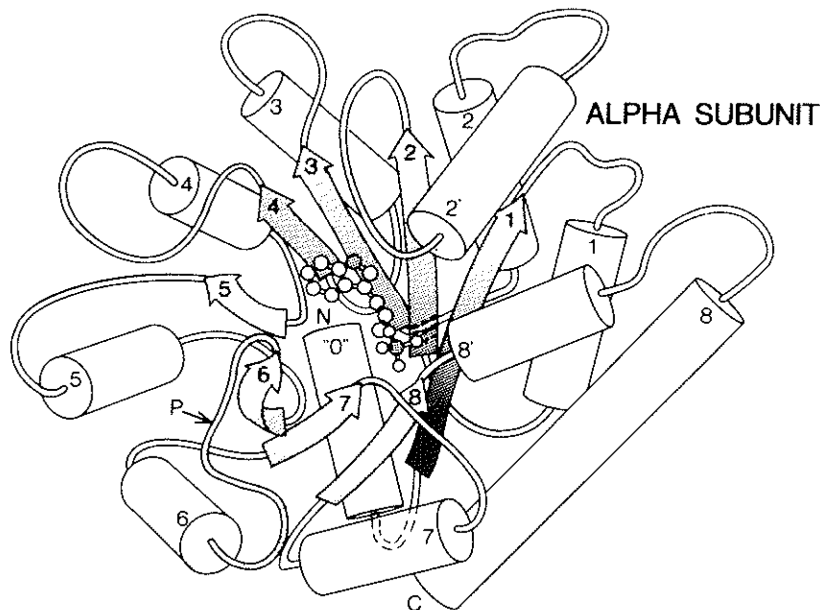


Figure 4. Schematic view of the eightfold α/β-barrel structure of the tryptophan synthase α-subunit. β-strands are shown as flattened arrows with arrowheads at their carboxy-termini; α-helices are represented as cylinders and are labelled on their amino-termini. Additional α-helices 2′ and 8′ are located above the active site and helix 0 is located below on the opposite side of the subunit. The active site is indicated by the position of bound inhibitor, indole propanolphosphate, shown in a ball-and-stick representation (the phosphorous and indole ring nitrogen are shaded). The loops following β-strands 2,3,4 and 5 in the α-subunit make important contacts with the β-subunit. N and C mark the amino- and carboxy-termini of the polypeptide chain, and P marks the site of limited proteolysis (Arg-188) producing α–1 and α–2. The schematic drawing of the α-subunit was reproduced by kind permission from Hyde *et al.* (1988).

trypsin proteolysis site Arg-188 separating α–1 and α–2; and loops 3 and 4, which form much of the interface with β.

Despite the considerable amount of evidence previously considered consistent with two folding domains, it would appear that only one α-domain exists. It seems that the partial barrel structures are more stable than previously predicted, and folding proceeds from the initial folding of the N-terminal part of the barrel (up to β-strand 6) to the folding of the C-terminal part of the barrel (from α-helix 6) and ending with a fitting together of the two folded termini involving intramolecular interactions and the molecule-wide conformational changes previously mentioned. In the light of X-ray crystallography data, Tweedy *et al.* (1990) described the late docking of β-strands 6 (amino folding unit) and 7 (carboxy folding unit) in the folding of the single α-domain.

Combining mutagenesis, folding and three-dimensional structure studies, we may be able to build a better picture of intraprotein interaction in the α-subunit. First, without mentioning specific residues, we can imagine the α polypeptide chain beginning to fold co-translationally. The N-terminal α–1 unit folds during translation and then, post-translationally, the C-terminal α–2 unit folds to give the complete α-domain structure.

Into this model we can re-introduce those specific residues thought to be involved in intraprotein interaction between the α folding units. We will mention Gly-211 and Gly-213, Tyr-175, Leu-177, Phe-22 and Arg-179.

Tweedy *et al.* (1990) have investigated the effect of changing the residue at position 211 on folding unit association and dissociation. Replacement of the wild-type glycine with serine or valine has little effect on the re-folding and unfolding kinetics involved with association and dissociation of the folding units. A replacement with tryptophan only significantly affected the dissociation. In contrast, glutamic acid and arginine replacements have significant and opposing effects on both association and dissociation kinetics, with folding decelerated by the incorporation of a negative charge and accelerated by the introduction of a positive charge. The insignificant effect of replacement with aspartic acid defied explanation. Long-range electrostatic interactions between residues in the opposing folding units are thought to be responsible for the severe folding effects, with short-range steric effects and hydrogen bonding playing lesser roles.

The effect of mutant Glu-211 was shown by Nagata, Hyde and Miles (1989) to be due to a disruption of the proper geometry of the α active site. Second-site reversion effects of the α Tyr-175 → Cys change were demonstrated to restore the proper geometry at the active site, using computer graphics modelling of the α active site with X-ray crystallography coordinates.

α175 is in β-strand 6 of the amino folding unit and α211 is in β-strand 7 of the carboxy folding unit. As previously mentioned, β-strands 6 and 7 dock late in folding, in a step reflecting the intraprotein interactions of folding unit association. It thus appears that a proper geometry of folding units in the native α is important for formation of the active site.

Earlier discussions have identified functionally important intraprotein interactions of α211 with α22, α177 and α213. It is likely that these effects are

due to alteration of the conformational geometry of residues at the folding unit interface. The reduced interactive ability exhibited by some of those mutants described earlier may also be a result of the altered geometry at the folding unit interface.

Arg-179 is in the region of the folding unit interaction surface. It is tempting to speculate that the effects on reciprocal transmission of ligand-binding effects seen after mutation to leucine at α179, are due to the inability of α to undergo the relevant conformational changes because of a defective intramolecular interaction between folding units of the α-domain.

The β-subunit

Genetic and biochemical investigations have made important contributions in studies characterizing the functional properties of the tryptophan synthase β-subunit. In a similar manner to the selection of novel α-mutants, the manipulation of screening techniques has allowed characterization of β-mutant types beyond the initial mutagenic studies described earlier.

Crawford and Johnson (1963), for example, recognized that in the original studies many β-mutants remained prototrophic (and hence undetected) because interaction with the wild-type α-subunit could enhance β catalytic activity. By using UV-irradiation of an αCRM^- nonsense mutant, A2, they were able to identify novel β-mutants, some of which have a bearing on the study of αβ interaction. This, and subsequent work, will be discussed later in this section.

Other mutational and chemical modification studies will be discussed that identify sites which modulate subunit interaction and residues having a potentially key catalytic function. The use of SDM to probe further the roles of residues identified by such chemical modification are also described.

Folding and three-dimensional structure studies have identified a two-domain structure for β and identified aspects of co-factor binding, dimerization and interaction with α, which will also be reviewed.

CHARACTERIZATION OF β-MUTANT TYPES ISOLATED USING AN IMPROVED SELECTION TECHNIQUE

As indicated above, Crawford and Johnson (1963) identified a number of double mutants with a nonsense mutation in *trpA* providing for an α CRM^- phenotype and a mutation in *trpB* preventing growth on a minimal medium supplemented with indole.

Further classification of these mutants was obtained on the basis of accumulation products, the ability to spontaneously revert to prototrophy, β CRM production and *in vitro* enzyme assays (Crawford and Johnson, 1963, 1964).

Class I mutants produce no demonstrable β CRM or β activity in any of the tryptophan synthase reactions. Subsequent studies (Crawford *et al.*, 1970) identified the majority of class I mutants as nonsense or frameshift.

Class IIa mutants produce a β CRM able to enhance the intrinsic enzyme

activity of the α-subunit. These mutants are analogous to the β CRM formers described previously (*see* pp. 235–236).

Class IIb mutants also produce a β CRM, but are inactive in all the tryptophan synthase reactions. It appears that interactive and catalytic sites are sufficiently close for the respective functions to be abolished by a single amino acid change. Nevertheless, the existence of class IIa mutants means that there cannot be a total overlap between interactive and catalytic sites on the β-subunit. Mutations affecting only interactive residues should, therefore, be discernible.

Class III mutants produce a labile β_2 which, although possessing some activity in the *in vitro* β-reaction, becomes rapidly inactivated in the cell.

Mutant classes IV , V and VI possess less functionally impaired β-subunits. In an α CRM^- background these βs are catalytically inert, but if associated with α in the tryptophan synthase complex, catalysis of all three tryptophan synthase reactions is possible (although not necessarily without some impairment).

Class IV mutants give prototrophs accumulating large amounts of indole as well as some InG. The accumulation of indole suggests the ratio of α:β activity is higher than the normal 1:1, and that the catalytic function of the β-subunit is significantly impaired. β interaction with α *in vitro* is markedly stimulated by the presence of ammonium ions at 1·4 M. Hatanaka *et al.* (1962) have found that ammonium ions favourably alter the conformation of β for catalysis. The class IV mutational event appears to produce an inactive conformation which is partly restored by interaction with α and further restored by the presence of ammonium ions.

Class V mutants are also partially prototrophic when associated with α, accumulating large amounts of InG but not indole. These mutants are similar to class IV mutants, having a reduced tryptophan biosynthetic capacity, but the class V β is less impaired in the αβ-reaction.

Class VI mutants are described as enzymatically inert with β alone, but sufficiently repaired by interaction with α not to accumulate any biosynthetic intermediates. As mentioned previously, the class VI mutant has considerable potential for studies of αβ interaction. It is possible, for example, to detect α-subunits with defective interaction by their inability to effect functional repair of the class VI β (*see* pp. 242–243).

All the above classification was determined prior to the discovery of nonsense suppressors (Brenner, Stretton and Kaplan, 1965). Re-evaluation in the light of this discovery, and with it the possibility that class IV, V and VI mutants could reflect simultaneous suppression of *trpB* and *trpA* ochre mutations, revealed that approximately one-third of the mutants originally classified as repairable mis-sense types on the basis of reversion studies alone, were found to be nonsense mutants (Crawford *et al.*, 1970). Of the *trpB* mis-sense mutations which remained, recombination tests allowed fine structure mapping. Intragenic clustering of both these reclassified β-mutants and additional newly generated nitrosoguanidine mutants, was observed (Crawford *et al.* 1970). The results were used to indicate, first, the position of residues intimately and obligatorily involved in the active site of the

β_2-subunit and, secondly, those sites where modification produces a structural distortion that can be restored by interaction with α. Residues of the first type appeared primarily in two areas of the front third of *trpB*. Residues of the second type were similarly found in those areas, and also occurred in areas of the final third of *trpB*. The central third of *trpB* contained few mis-sense mutations, although nonsense mutations were found. The situation is, therefore, analogous to that of the *trpA* gene and α-subunit, and may indicate that the central portion of the polypeptide chain constitutes at least part of the interaction site for the α-subunit.

β-MUTATIONS AFFECTING THE MODULATION OF SUBUNIT INTERACTION

Gibson, Gibson and Yanofsky (1961) have described the mutant *E. coli* 7-4 which produces a β CRM defective in β catalytic function, but able to interact with α. The binding of ligands PLP and L-serine to the β-subunit stimulate the α catalytic function in the tryptophan synthase complex *in vitro*. This stimulation is not observed in wild-type $\alpha\beta_2\alpha$ and it appears that the exploration of $\alpha\beta$ subunit interaction can be facilitated by the removal of the β catalytic function. Following some of the arguments presented for the α-subunit, these results additionally suggest that the catalytic site and the interaction site of the β-subunit are at least partially separate.

Mosteller, Goldstein and Nishimoto (1977) have described non-functional β-mutants B9578 and B9611 which have an increased affinity for α, but enhance activity in the α-reaction to a lesser extent. The strength of the interaction appears to be functionally important. In earlier discussions of the α-subunit, we have mentioned cases where changes in the strength of interaction have been observed. Only in this specific case of the β-subunit, however, has any functional effect been detected, and it is conceivable that important $\alpha\beta$ interface residues govern the strength of interaction.

LOCALIZATION OF FUNCTIONAL β-RESIDUES

Mis-sense mutants

Cotton and Crawford (1972) identified the amino acid change responsible for a class IIa mis-sense mutation, B244. Amino acid sequence analysis of cyanogen bromide, tryptic and chymotryptic fragments specified a Lys-384 → Asn change. The role played by the lysine residue is still unknown (the lysine residue forming a Schiff base with PLP in the resting enzyme has been located elsewhere on the protein chain).

Biochemical studies and site-directed mutagenesis

Chemical modification studies have suggested roles for His-86 (or possibly 82) in the removal of the α proton from L-serine (Miles, Bauerle and Ahmed, 1987); Lys-87 in the formation of a Schiff base with PLP (Fluri *et al.*, 1971;

Higgins, Miles and Fairwell, 1980); Arg-148 in the binding of the substrate, L-serine (Tanizawa and Miles, 1983); and Cys-230, which has been located in the vicinity of the PLP-binding site (Miles, 1970; Higgins and Miles, 1978; Miles and Higgins, 1980).

However, the chemical modification techniques used in these studies have intrinsic difficulties (Miles *et al.*, 1989). Specificity and stoichiometry may not be absolute, and the introduction of charged or bulky groups may cause inhibition of activity by steric hindrance or conformational distortion as well as by modifying the chemistry of a functional group. In accordance with the above, Miles *et al.* (1989) employed SDM to establish the precise importance of those residues previously implicated as having functional significance by chemical modification.

β *His-86.* His-86 was thought to be a catalytic residue for proton transfer. Kinetic and stereochemical evidence suggested that a histidyl residue catalysed the removal of the α proton from L-serine and any subsequent intramolecular proton transfer. Photo-oxidation sensitized by PLP has been shown to destroy an essential histidyl residue located near the PLP-binding site (Miles, 1974; Miles and Kumagai, 1974). Enzyme inactivation of β_2 with the acylating reagent, diethylpyrocarbonate, and regeneration by hydroxylamine treatment have also implicated histidyl residue(s) at the active site (Miles and Kumagai, 1974). Despite the above, no definite proof has been obtained for the proposal that an essential histidyl residue exists and catalyses the removal of the α proton from L-serine.

Replacement with lysine at β86 produced a mutant β_2 able to bind PLP, α and exhibit partial catalytic activity. The reduction in activity (to 20–30% wild type) may reflect conformational distortion at the active site. The fact that activity is still observed, however, indicates that His-86 is not an obligatory requirement for catalysis or substrate binding. This observation is in keeping with the three-dimensional structure determined by X-ray crystallography, which places His-86 too far from the L-serine-binding site to be involved in catalysis (Hyde *et al.*, 1988).

β *Lys-87.* Fluri *et al.* (1971) used sodium borohydride to covalently attach PLP to the ε-amino group of the lysyl residue forming a Schiff base with the co-factor in the resting enzyme. Higgins, Miles and Fairwell (1980) identified the lysyl group as Lys-87.

Replacement of Lys-87 with threonine yields a mutant β having no detectable activity in the reactions catalysed at the β active site. Interaction with α is not affected, however, and PLP binds the enzyme through a Schiff base. Spectral analysis of mutant enzyme–substrate intermediates implies that the β-reaction is blocked before the removal of the α proton from L-serine. Lys-87 rather than His-86 may, therefore, fill this catalytic role.

β *Arg-148.* Tanizawa and Miles (1983) showed the holo-β_2-subunit to be protected from inactivation with the arginine-specific dicarbyl reagent, phenylglyoxal, by the binding of substrate, L-serine. As no protection effect was

observed for apo-β_2 and phenylglyoxal-modified holo-β_2 could not bind L-serine, it was concluded that an essential arginine residue was involved. Use of ^{14}C-labelled phenylglyoxal identified Arg-148 as the modified residue.

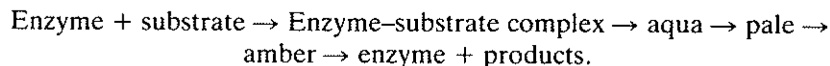
Substitution to give Gly-148 produces an $\alpha\beta_2\alpha$ complex having low but significant activity (Miles *et al.*, 1989). The non-essential role is consistent with the determined three-dimensional structure (Hyde *et al.*, 1988).

β *Cysteine residues.* Cys-230 appears to be at the β active centre, as the binding of PLP protects β_2 from modification by *N*-ethylmaleimide and 5,5′-dithiobis(2-dinitrobenzoic acid). The single sulphydryl residue reacting with holo-β_2 was identified as Cys-170 (Miles, 1970; Miles and Higgins, 1980). Nitrothiocyanobenzoic acid was used to selectively modify Cys-230 by replacement of the sulphydryl group with a cyanide group. The loss of activity, regenerated by treatment with 2-mercaptoethanol, suggested that the sulphydryl group is essential (Higgins and Miles, 1978).

Despite this, SDM results have assigned Cys-170 and Cys-230 as non-essential (Miles *et al.*, 1989). Substitutions giving Ser-170, Ser-230 or Ala-230 had no significant effects on β activity. It appears, therefore, that replacement of the sulphydryl group at β230 by hydroxyl (serine) or methyl (alanine) groups can be functionally accommodated, but substitution with the slightly larger cyanide group (nitrothiocyanobenzoic acid modification) cannot.

THE FOLDING AND THREE-DIMENSIONAL STRUCTURE OF THE β_2-SUBUNIT

In comparison to the α-subunit, a study of the folding and native structure of the β_2-subunit is more complex, particularly given dimer formation and pyridoxal phosphate (PLP) co-factor binding. In the following section we will outline the current model of native β_2 and indicate how that structure is derived. The distinct UV and visible properties of the PLP moiety in β-reaction enzyme(β_2 or $\alpha\beta_2\alpha$)–substrate intermediates has allowed a comprehensive molecular description of the reaction pathway (Goldberg and Baldwin, 1967; Goldberg, York and Stryer, 1968; York, 1972; Miles, 1979; Lane and Kirschner, 1983a, 1983b; Drewe and Dunn, 1985, 1986). Thus the 'aqua', 'pale' and 'amber' catalytic intermediates have been defined in the β-reaction pathway.


Enzyme + substrate → Enzyme–substrate complex → aqua → pale → amber → enzyme + products.

Limited proteolysis and domains

Limited trypsin proteolysis of β_2 results in a nicked β_2 product, β_2' (Högberg-Raibaud and Goldberg, 1977a). Unlike the analogous α-fragments, however, the β-fragments, termed F-1 and F-2, cannot interact to produce a fully active protein. Interaction between F-1 and F-2 does occur, producing a dimer which binds co-factor PLP, substrates L-serine and indole, and gives rise to 'aqua complex'; however, catalysis is not completed, β-reaction L-serine

deaminase activity is not observed and β_2' is unable to bind the α-subunit. Limited proteolysis does not disrupt $\beta\beta$ interaction, substrate binding or the initial catalytic processes. It does disrupt the $\alpha\beta$ interaction and prevents the full catalytic reaction.

Further characterization of F-1 and F-2 (Högberg-Raibaud and Goldberg 1977a, b; Crawford, Decastel and Goldberg, 1978) supports F-1 (29 kDa) to be the N-terminal folding domain, having β Thr-2 at its N-terminal and β Lys-272 at its C-terminal end. F-2 (12 kDa) is co-ordinate with the C-terminal folding domain, with β Lys-283 at the N-terminal and β Ile-397 at the C-terminal end. Trypsin cleaves β at three sites, Lys-272, Arg-275 and Lys-283 (Ahmed *et al.*, 1986) removing 11 residues from a proposed hinge or connector region. The loss of some material had previously been predicted by analytical ultracentrifugation studies (Högberg-Raibaud and Goldberg, 1977a) which detected the presence of small peptides in the β tryptic digest and showed the sum of F-1 and F-2 molecular masses to be lower than that of native β monomers (45 kDa).

Roles suggested for the connector include holding the F-1 and F-2 domains in the correct orientation for interaction with α and catalysis, and for the physical binding of α. The role of the connector and the folding domains will be discussed further in the following sections.

Folding of β_2

F-1 and F-2 are proposed as independent folding domains. Separation using gel filtration in 6 M urea, followed by removal of the denaturant, results in a spontaneous folding of the isolated fragments. Aromatic residue fluorescent emission spectra and CD spectra approximate these fragments to the corresponding domains in native β_2. Mixing F-1 and F-2 produces β_2', a protein indistinguishable from β_2 using the above optical techniques.

The kinetics of folding (Zetina and Goldberg, 1982; Blond and Goldberg, 1986), folding domain association (Zetina and Goldberg, 1982) and β-monomer dimerization (Blond and Goldberg, 1987), have been studied after GuHCl-induced unfolding for β_2, β_2', F-1 and F-2. The results obtained have helped elucidate a six-step renaturation sequence for native β_2, and the following model was proposed by Blond and Goldberg (1986).

Step 1 reflects the folding of the N-terminal F-1 domain, characterized by the burial of β Trp-177 into a very hydrophobic environment and the alignment of Trp-177 and Lys-87 into adjacent positions in a native conformation, as evidenced by fluorescent energy transfer from Trp-177 to a fluorescent probe linked to Lys-87 (Blond and Goldberg, 1986).

Step 2 is observed by fluorescent energy transfer from Trp-177 to a fluorescent probe linked to Cys-170 and involves a local conformational change in this area (Blond and Goldberg, 1985). Interaction of the F-1 domain with the F-2 domain is required for this change, as it only occurs in the presence of F-2. The transition may involve conformation adjustments seen as differences in the immunoreactivity of F-1 on association with F-2 (Zakin *et al.*, 1980).

Step 3 is a rate-limiting isomerization of the monomer, which is required before dimerization can occur (Blond and Goldberg, 1985).

Step 4 is the rapid dimerization of the monomers (Blond and Goldberg, 1985).

Step 5 is an isomerization of the dimer, leading to the formation of the PLP-binding site, involving interactions between the folding domains (detailed below).

Step 6 is a slow isomerization involving the F-1 domain and completes the renaturation of β_2; this step also occurs during the refolding of isolated F-1 fragments (Zetina and Goldberg, 1982).

The model proposed above demonstrates the existence of constraints exerted by intraprotein interactions involved in the folding process. Conformational changes are seen after each step towards the native conformation, to accommodate the new interactions.

Pyridoxal phosphate (PLP) binding

The concerted mechanism (Monod, Wyman and Changeux, 1965) best describes the results of kinetic studies of PLP binding to β_2 and its proteolysis products. Tschopp and Kirschner (1980) concluded that apo-β_2 must spontaneously undergo structural transitions between the high-affinity R state and the low-affinity T state; the isomerization between the two being a relatively slow reaction. Binding of α in αapo-$\beta_2\alpha$ fixes apo-β_2 in the R state, with α acting as an allosteric activator with respect to PLP binding, fixing the binding site in a rigid state optimally arranged for rapid, high-affinity, single-step binding. β_2' appears frozen in the T state, the T to R isomerization not being possible, so it seems that the connector region is essential for this conformational change to occur, perhaps involving intraprotein interactions and domain-alignment changes. It should be noted that cleavage of β at Arg-275 by the endopeptidase arg-C has the same effect as removal of the connector (Ahmed *et al.*, 1986). It is β cleavage, therefore, rather than the removal of essential residues, that prevents T to R isomerization.

The mechanism of PLP binding to apo-β_2 and apo-$\alpha\beta_2\alpha$ has been investigated using rapid-mixing techniques, with the reaction being monitored through absorption and fluorescence changes associated with the bound PLP moiety (Bartholmes, Balk and Kirschner, 1980). A three-step process was described, involving the formation of a non-covalent complex which isomerizes to an enzymically inactive internal aldinine (covalent complex) followed by formation of the holo-β_2. The two binding sites on β_2 appear to bind PLP independently in $\alpha\beta_2\alpha$; however, a co-operative binding was observed for β_2. The difference in PLP binding to $\alpha\beta_2\alpha$ and β_2 is consistent with the concerted mechanism, where apo-β_2 is in the low-affinity T state and the co-operative binding is due to conformational changes giving the high-affinity R state.

Lane (1983) and Chaffotte and Goldberg (1984) have demonstrated that β Trp-177 is more accessible to the solvent in holo-β_2 than apo-β_2, apo-β_2' and holo-β_2', using fluorescence quenching of β177 tryptophan by acrylamide. The phenomenon is thought to be due to a real change in the conformation of

the β_2-subunit, reflecting some allosteric transition induced by PLP binding to intact β. In agreement with a co-factor-induced conformational change increasing the compactness of β_2, the apo-β_2 has been reported to be less stable (Zetina and Goldberg, 1980) and more flexible than holo-β_2 (Chaffotte and Goldberg, 1983). In support of the claim that β_2' is frozen in the low-affinity T state, Chaffotte and Goldberg (1984) determined that the fluorescence quenching constants for apo-β_2, apo-β_2' and holo-β_2' are the same. Thus, isomerization to the R state involves a conformational change in the enzyme for which the connector is vital but only as a tether to allow repositioning of the F-1 and F-2 domains in space.

STRUCTURAL STUDIES

X-ray crystallography studies of the *Salmonella typhimurium* enzyme (Hyde *et al.*, 1988) verified the two-domain structure of the β-monomer. Both domains have a α-helix–β-sheet tertiary structure, with PLP binding at the domain interface. The amino domain (β Met-1–Gln-204), and the carboxy domain (β Arg-205–Ile-397), are not equivalent to the proteolysis fragments, F-1 (β Met-1–Lys-272) and F-2 (β Lys-283–Ile-397). Furthermore, a portion of the amino-terminal half of the β polypeptide chain (β Pro-55–His-85) forms part of the structural carboxy domain.

X-ray crystallography has provided information on protein–protein interactions involving β. A portion of the carboxy-terminal half of β (β Lys-260–Gly-310), having no defined secondary structure and not forming any part of the helix–strand–helix domain structure, makes several contacts with α, and thus may be involved in the αβ interaction. ββ interaction occurs over a broad, nearly flat surface, through which a dyad axis of symmetry passes; part of the amino domain of each monomer interacts with part of the complementary monomer's carboxy domain. Hydrogen bonding, hydrophobic interactions, van der Waals interactions and ion-pair interactions add to the stability of the ββ contact surface in a tight association of approximate area 1440 $Å^2$ (Hyde *et al.*, 1988).

α AND β_2 INTERACTION

Once α and β_2 attain their native structures, interaction may occur. Recent work (Jaenicke, 1991; Landry and Gierasch, 1991) suggests that the *in vivo* process may be different to the *in vitro* process studied. The folding of individual subunits and their subsequent interaction may begin co-translationally and involve additional protein helpers which ensure correct subunit folding and complex assembly.

Early work

Creighton and Yanofsky (1966) demonstrated the reversible binding of α to two identical and independent sites on β_2, with each combined site contributing the same enzyme activity. The reluctance of α to form dimers and β to

exist as a monomer, coupled with analytical ultracentrifuge evidence for a complex of the size of $\alpha\beta_2$ when β_2 is in excess, indicates a sequential addition of α to β_2 in the association reaction.

Hydrophobic interaction

As previously mentioned, analytical ultracentrifugation studies have demonstrated that hydrophobic bonding plays an important role in $\alpha\beta$ interaction in homologous and heterologous *E. coli* and *S. typhimurium* tryptophan synthase complexes (DiCamelli, Balbinder and Lebowitz, 1973; DiCamelli and Balbinder, 1976). Furthermore, the involvement of ionic bonding was suggested by the interference of monovalent and divalent cations in the $\alpha\beta$ interaction (DiCamelli, Balbinder and Lebowitz, 1973). In another study, Bartholmes and Teuscher (1979) observed a moderate temperature dependence in the binding of α to apo-β_2 indicative of a hydrophobic interaction. They also demonstrated that α binds more strongly to the holo-β_2 subunit than to apo-β_2 which suggests that conformational changes induced by PLP binding, involving the arrangement of β folding domains, affects the interaction site in addition to the catalytic site.

Adding α to β_2

The addition of α to β_2 and $\alpha\beta_2$ was studied using stopped-flow techniques (Lane, Paul and Kirschner, 1984). In both cases addition proceeded through the formation of an initial $\alpha\beta$ protomer which slowly isomerized to the native state. Negative co-operativity was seen, with the weaker binding of the second α possibly being due to steric hindrance. The slow isomerization of both subunits includes conformational changes to a catalytically more effective (high-efficiency) state involving intraprotein interactions and changes in folding unit or domain alignment. Interaction-specific changes in the conformations of α and β have also been observed using small-angle X-ray scattering (Wilhelm *et al.*, 1982), fluorescence energy transfer (Lane and Kirschner, 1983c) and calorimetry (Wiesinger, Bartholmes and Hinz, 1979). Although the binding of substrates to the individual active sites stabilizes the high-efficiency conformation (by reducing the reverse isomerization rate) it does not induce the production of the high-efficiency state.

Quaternary structure

Small-angle neutron scattering (Ibel *et al.*, 1985) involving ^{2}H-labelling of α folding units, α–1 and α–2, and β_2 has further contributed to knowledge of the enzyme complex quarternary structure. Individual α-subunits were shown to be completely separated (thus the tryptophan synthase formula $\alpha\beta_2\alpha$ supercedes the $\alpha_2\beta_2$ formula first adopted (Creighton and Yanofsky, 1966)) with its two folding units intimately juxtaposed. Conformation changes observed on $\alpha\beta$ interaction were seen to occur only in the β_2-subunit, where compactness

increased. These observations argue against steric hindrance as a mechanism to explain the negative co-operativity of α binding.

X-ray crystallography (Hyde *et al.*, 1988) has given the most detailed picture so far of the gross quarternary structure. As detailed earlier, the four peptide chains are arranged linearly in an αββα order, forming a complex 150 Å long. α and β active sites are about 25 Å apart and connected by a tunnel, of diameter matching that of indole, believed to facilitate diffusion between active sites, preventing loss of indole to the solvent. As would be expected, in order to allow formation of the indole tunnel, part of the α active site is at the interface with β. This supports those observations seen in mutagenesis studies, where catalytic residues affect interaction. αβ Interaction occurs over a relatively large area, with 1190 $Å^2$ of α and 1110 $Å^2$ of β buried at the interface. From this we can anticipate that a number of residues are solely involved in αβ interaction.

Intergeneric comparison

Tryptophan synthase has become an archetypal example of gene structure–function operation and has been studied more extensively and in a greater variety of organisms than any other enzyme. Initial studies aimed at understanding the genetics and biochemistry of a variety of medically and commercially important organisms have produced information relating to gene arrangements, gene regulation and protein primary structure sequence homology. Primary structure conservation within tryptophan biosynthetic enzymes has been used primarily to determine interspecies evolutionary divergence. It has also been used, however, to compare the generally more thermostable enzymes from thermophilic bacteria with those from mesophilic bacteria in an attempt to understand thermostability. The interspecies conservation of proposed catalytic residues is concordant with essential function.

In parallel with the considerable work on *E. coli* described previously, the work comparing tryptophan biosynthesis in different organisms encompasses protein tertiary structure analysis (immunological cross-reactivity and subunit interchange), protein primary structure studies (N-terminal sequencing of purified proteins, superceded by DNA sequencing of cloned genes), and gene arrangement and cloning studies (including expression of cloned foreign genes in host organisms). The overall picture presented by this work describes generally conserved protein structures with striking inflexibility at key residues and the retention of some interspecies tertiary structure homology. Nevertheless, there has been a considerable divergence in gene arrangements, gene–enzyme function and specification, and the regulation of tryptophan biosynthesis.

In the following section, using *E. coli* as the reference organism, we compare and contrast tryptophan synthase from a wide diversity of sources, paying special attention to protein interaction considerations and the limitations in the use of such data to develop generalized views of protein–protein interaction.

TERTIARY STRUCTURE ASSAYS

Immunological studies

Three principle methodologies have utilized antisera against tryptophan synthase α- and β_2-subunits.

1. Enzyme neutralization tests involving determination of the ratio (homologous/heterologous antigen) of the number of Enzyme Units neutralized by a given amount of antiserum, made to the *E. coli* subunit. Results are expressed as a neutralization efficiency.
2. Microcomplement fixation and determination of the index of dissimilarity, involving determination of the ratio of antiserum concentrations required with homologous (*E. coli*) and cross-reacting antigens to give identical levels of complement fixation. Complement fixation (Wasserman and Levine, 1961) measures the extent of antigen–antibody interaction by quantitatively measuring residual complement activity. (That is, antigen + antibody → complex which fixes complement. Any complement remaining is measured and, being inversely proportional to the amount of antigen–antibody interaction, it is assumed to reflect tertiary structure relatedness. Results are expressed as an index of dissimilarity.)
3. Precipitin formation using the technique of Ouchterlony (1949).

The Enterobacteriaceae have been studied in greatest detail using techniques (1) and (2), as summarized in *Table 3.* Immunodiffusion experiments, technique (3), have been limited; Balbinder (1964) showed small differences between both subunits of *Serratia marcesens* and *E. coli*, but failed to demonstrate differences between *E. coli* and *S. typhimurium* β_2-subunits. Maurer and Crawford (1971a) demonstrated some cross-reactivity between anti-*E. coli* β- and β_2-subunits of *Pseudomonas putida*.

Work on non-enteric bacteria and eukaryotes is scarce, with information existing only for *P. putida, Bacillus subtilis, Anabaena variabilis, Chlorella ellipsoidae* and *N. crassa* (*Table 4*; Sakaguchi, 1970; Hoch and Crawford, 1973). Technique (1), therefore, appears the more sensitive test for detecting distant evolutionary relationships, whereas technique (2) appears more useful for the investigation of intergroup relationships.

The relative sensitivities of techniques (1) and (2) further suggest that (1)

Table 3. Immuno-crossreactivity of Enterobacteriaceae tryptophan synthase subunits (data from Murphy and Mills, 1969; Rocha, Crawford and Mills, 1972; Reyes and Rocha, 1977)

Organism	Index of dissimilarity		Neutralization efficiency	
	α	β	α	β
Escherichia coli	1·0	1·0	1·0	1·0
Shigella dysenteriae	1·1	1·0	1·0	1·02
Salmonella typhimurium	2·2	1·4	1·25	1·49
Klebsiella aerogenes	2·8	1·2	1·39	1·64
Serratia marcescens	8·4	1·8	1·89	2·38
Proteus vulgaris	ND*	1·0	ND	ND
Erwinia carotovora	ND	1·7	ND	ND

* ND, not determined.

Table 4. Immuno-crossreactivity of non-Enterobacteriaceae tryptophan synthase subunits (data from Sakaguchi, 1970; Maurer and Crawford, 1971a; Hoch, 1973a)

Protein source	Antisera source	Protein	Neutralization (efficiency)*
Pseudomonas putida	*Escherichia coli*	β_2	weak
Shigella dysenteriae	*Escherichia coli*	β_2	weak (1·5%)
Chlorella ellipsoidae	*Salmonella typhimurium*	β_2	none
Anabaena variabilis	*Salmonella typhimurium*	β_2	weak (2·9%)
Esherichia coli	*Salmonella typhimurium*	β_2	strong
Aspergillus niger	*Salmonella typhimurium*	TS†	none
Saccharomyces	*Salmonella typhimurium*	TS	none
Escherichia coli	*Pseudomonas putida*	β_2	none
Bacillus subtilis	*Pseudomonas putida*	β_2	weak (3·5%)
Chlorella ellipsoidae	*Neurospora crassa*	β_2	weak (0·05%)
Anabaena varabilis	*Neurospora crassa*	β_2	none
Escherichia coli	*Neurospora crassa*	β_2	none
Aspergillus niger	*Neurospora crassa*	TS	yes
Saccharomyces	*Neurospora crassa*	TS	yes

* Data presented as a percentage rather than as an index because intergeneric cross-reactivity is weak.
† TS, tryptophan synthase.

reflects similarities at functionally important structures (e.g. substrate-binding sites at the active centre), and (2) reflects less rigid, whole-protein similarities.

Looking at the immunological data as a measure of tertiary-structure relatedness, it is possible to conclude from a compilation of all studies that the β_2-subunit is more conserved than α, and to present reasoned functional arguments for this finding. For example, while β has binding sites for α, β, indole, L-serine and pyridoxal phosphate and forms the majority of the indole tunnel, α only binds β, indole and InGP. β could be considered, therefore, to have more functional regions to conserve. Furthermore, studies on the quaternary structure of tryptophan synthase have indicated that β is on the inside and α on the outside of the $\alpha\beta_2\alpha$ complex (Wilhelm *et al.*, 1982; Lane and Kirschner, 1983c; Ibel *et al.*, 1985; Hyde *et al.*, 1988). It is generally accepted that internal residues are less tolerant to change than external, again supporting a greater conservation of the β-subunit.

Table 4 shows that the β-subunits of *P. putida, B. subtilis* and *S. typhimurium* are related to that of *E. coli*, and since the β of *A. variabilis* is related to that from *S. typhimurium*, it too is probably related to the *E. coli* β. The data available are, however, incomplete. It would be interesting to examine relatedness throughout the groups of organisms studied by determination of cross-reactivity between antisera raised against all purified subunits and their heterologous equivalents.

Subunit interchange

Potentially a formidable source of protein interaction information, subunit interchange data are unfortunately in considerable need of extension. Nevertheless, those successful interchanges that have been demonstrated are encouraging as they suggest a conservation of the key structural residues at the protein subunit interaction sites. Unsuccessful attempts, especially in view of studies by Riverin and Drapeau (1976), cannot be confidently trusted for

reasons that will be discussed below. The basic ability of heterologous subunits to form a tryptophan synthase complex *in vitro* is shown in *Table 5*.

Support for the basic precept that the β-subunit is more conserved in tertiary structure than α (with *E. coli* as reference) can be obtained from *Table 5*. It can be seen that the *E. coli* α-subunit, and indeed the α-subunit from certain other Enterobacteriaceae, forms a wider range of heterologous tryptophan synthase complexes than does the *E. coli* β-subunit, or the β-subunit of the other Enterobacteriaceae. The predomination of + in the top right of *Table 5*, and – in the bottom left, is evidence of this. The implication from this is that the binding site for *E. coli* α is conserved on a wider range of β-subunits than is the *E. coli* β-binding site on the α-subunit. Unfortunately, insufficient subunit interchange data are available, thus allowing only this rather diffuse view of the real situation.

The data in *Table 5* only hint at the knowledge we would like to obtain from subunit interchange experiments. The majority of the data was generated in the 1970s using *in vitro* stimulation of β by α in the easily assayed β-reaction of tryptophan synthase. The bulk of information for Enterobacteriaceae and *B. subtilis* was obtained using this technique. Unfortunately, as the β-reaction is only an indirect measure of subunit interaction, it is open to misinterpretation. Riverin and Drapeau (1976) working with *P. mirabilis* tryptophan synthase subunits, demonstrated the limitations of using this technique. Heterologous complexes of *P. mirabilis* α with *E. coli* β_2 and *P. mirabilis* β_2 with *E. coli*, *S. typhimurium* and *S. marcescens* αs were all found to be effective in the physiologically important αβ-reaction (formation of tryptophan from InGP and L-serine), but gave either reduced or no enhancement of the partial reactions. The indication must be, therefore, that in this case evolutionary pressure has maintained the αβ interaction to conserve the physiologically important active site arrangement.

Coupling this conclusion with that already inferred from antisera neutral-

Table 5. *In vitro* subunit interchange in tryptophan synthase to give a heterologous complex (data from Balbinder, 1964; Murphy and Mills, 1969; Sakaguchi, 1970; Maurer and Crawford, 1971a; Rocha, Crawford and Mills, 1972; Hoch, 1973a, b; Hoch and Crawford, 1973; Riverin and Drapeau, 1976)

	β											
α	Eç*	St	Sm	Sd	Ka	Pm	Pp	Bs	Bp	Ba	Av	Ce
Escherichia coli	+	+	$+^R$	+	+	+	–	$+^R$	N	N	$+^R$	$+^R$
Salmonella typhimurium	+	+	N	N	N	+	N	N	N	N	$+^R$	$+^R$
Serratia marcescens	+	+	+	N	N	+	N	N	N	N	$+^R$	N
Shigella dysenteriae	+	+	N	+	N	N	N	N	N	N	N	N
Klebsiella aerogenes	+	+	N	N	+	N	N	N	N	N	N	N
Proteus mirabilis	+	N	N	N	N	+	N	N	N	N	N	N
Pseudomonas putida	–	N	N	N	N	N	+	$+^R$	N	N	N	N
Bacillus subtilis	–	N	N	N	N	N	–	+	N	N	N	N
Bacillus pumilis	N	N	N	N	N	N	N	+	+	N	N	N
Bacillus alvei	N	N	N	N	N	N	N	–	N	+	N	N
Anabaena variabilis	–	–	–	N	N	N	N	N	N	N	+	+
Chlorella ellipsoidae	–	–	N	N	N	N	N	N	N	N	+	+

* Column headers are arranged in the same order as the rows, but abbreviated with generic and specific initials.

+, Complementation; $+^R$, complementation reduced; –, no complementation; N, no data available.

ization experiments, we can confidently conclude that the $\alpha\beta_2\alpha$ active site tertiary structure is highly conserved. From this we may identify two further points. First, it suggests that a lack of activity in a partial reaction, e.g. indole to tryptophan, may not necessarily mean a lack of specific subunit binding to give a catalytically viable enzyme. Secondly, as we have previously discussed, evidence for an overlap of catalytic and interactive sites is substantial. The α active site is thought to be intimately associated with certain αβ interaction sites (Hyde *et al.*, 1988); those αβ interaction sites involved in the formation of the catalytic site, therefore, may also be highly conserved.

Measurements of the strength of subunit interaction have provided additional support for a number of interspecies subunit interactions (*Table 6*). Murphy and Mills (1969) studied the association of αs from a range of Enterobacteriaceae with β_2s from *E. coli* and *S. typhimurium* using the technique of Creighton and Yanofsky (1966). Rocha, Crawford and Mills (1972) performed similar experiments with β_2s from a range of Enterobacteriaceae with α from *E. coli*. The association constant, K_a, determined for the complexes tested are shown in *Table 6*:

$$K = [\alpha\beta]/[\alpha]\cdot[\beta] = K_a = AB/(A_t - AB)\cdot(B_t - AB)$$

where

- $[\alpha]$ = concentration of free α-subunits;
- $[\beta]$ = concentration of free half β_2-subunits;
- $[\alpha\beta]$ = concentration of αβ couplets;
- AB = enzyme activity of complex present;
- A_t = enzyme activity which would be observed if the total α present were saturated with β_2;
- B_t = enzyme activity which would be observed if the total β_2 present were saturated with α.

DiCamelli and Balbinder (1976) studied the association of α- and β-subunits from *E. coli* and *S. typhimurium* in homologous and heterologous combinations using sucrose density gradient centrifugation. The homologous combinations showed similar sedimentation patterns: increasing the r.p.m. to

Table 6. Association constants for homologous and heterologous tryptophan synthase complexes (data from Murphy and Mills, 1969; Rocha, Crawford and Mills, 1972)

α source	β_2	K_a/1/(enzyme units ml^{-1})
E.coli	*E.coli*	4·0
E.coli	*S.typhimurium*	1·0
E.coli	*S.dysenteriae*	1·7
E.coli	*K.aerogenes*	0·88
E.coli	*S.marcescens*	0·33
S.typhimurium	*E.coli*	0·72
S.typhimurium	*S.typhimurium*	2·7
S.dysenteriae	*E.coli*	2·8
S.dysenteriae	*S.typhimurium*	0·44
K.aerogenes	*E.coli*	7·1
K.aerogenes	*S.typhimurium*	1·6
S.marcescens	*E.coli*	3·2
S.marcescens	*S.typhimurium*	0·56

50 000 from 39 000 at 5°C dissociated the complex, whereas a similar increase at 20°C did not. The *E. coli* α–*S. typhimurium* β combination showed a weaker association, with the subunits remaining dissociated at 50 000 r.p.m./20°C. The reciprocal *E. coli* β–*S. typhimurium* α combination showed a stronger association, with the subunits remaining associated at 50 000 r.p.m./5°C. These results are in agreement with an earlier study of the interaction between the α- and β_2-subunits of *E. coli* (DiCamelli, Balbinder and Lebowitz, 1973) which indicated the αβ interaction to be of a hydrophobic nature.

The presence of residues responsible for modulating the strength of subunit interaction, previously discussed in biochemical terms, is further supported by these results. It would appear that a certain degree of αβ interaction is important and maintained, at least between *E. coli* and *S. typhimurium*, and possibly through evolution. Thus a change at an interactive residue in one subunit may be compensated for by an evolutionary change in the other. A hypothetical αβ interaction surface, comprising highly conserved key residues responsible for the maintenance of recognition and binding sites and less highly conserved, but complementary, regions responsible for controlling the strength of interaction, can be imagined. However, it should be recognized that, between *E. coli* and *S. typhimurium*, amino acid changes, which possibly include such less highly conserved regions capable of modulating the strength of interaction in either heterologous or hybrid complexes, have no observable effect on catalytic activity (DiCamelli and Balbinder, 1976; Schneider, Nichols and Yanofsky, 1981).

Any future extension of existing work to give a comprehensive study of subunit interchange, should include:

1. Measurements of both partial and full reaction enzyme activities.
2. An indication of subunit interaction and the strength of this interaction, e.g. using analytical ultracentrifugation.
3. *In vivo* complementation studies, where a subunit from one organism is expressed in a mutant of another organism having the complementary subunit (*see* pp. 269–270 and p. 271).

Gene cloning

The selection of *trp* gene clones from various gene libraries, or from genes mobilized using the R68 variants of the wide host-range P1 group plasmid (Hedges, Jacob and Crawford, 1977), have employed two strategies. The first relies on the ability of cloned *trp* genes to complement chromosomal mutations, thus allowing transformants to grow on a minimal medium.

The complementation strategy has a number of potential complications. Expression of foreign DNA in host organisms can present a number of problems. For example, the native promoter may function poorly, or not at all, or codon usage differences may contribute to reduced expression. Expression at low levels may not be enough to support growth without exogenous tryptophan, and so the correct transformant would slip through

the screen. Even if properly expressed, inherent instabilities or inactivity at temperatures appropriate for host growth may prevent selective identification of *trp*$^+$ clones. Furthermore, if the formation of heterologous enzyme complexes is required (anthranilate and tryptophan synthase), any subunit incompatibilities will lead to a reduced biosynthetic potential and may result in the inability to select the desired clone.

The *trp* mutant hosts are mostly *E. coli* derivatives, as used in the cloning of *Bacillus stearothermophilus trpBA* (Ishiwata *et al.*, 1989), *Brevibacterium lactofermentum trpE* (Matsui, Miwa and Sano, 1987), *Caulobacter crescentus trpF* (Winkler *et al.*, 1984), *Clostridium thermocellum trpEG* (Sato *et al.*, 1989), *Lactobacillus casei trpDCFBA* (Natori, Kano and Imamoto, 1990), *Methanococcus voltae trpFBA* (Sibold and Henriquet, 1988), *Streptomyces griseus trpCBA* (Rivero-Lezcano *et al.*, 1990) and *Thermus thermophilus trpEG* (Sato *et al.*, 1988). In cases where problems with an *E. coli* host were anticipated or encountered, other hosts were used. For example, in the cloning of *Rhizobium meliloti trpE, trpDC* and *trpFBA* regions, *Rhizobium leguminosarum* and *Pseudomonas aeruginosa* mutants were employed (Berringer, Hoggan and Johnston, 1978; Johnston, Bibb and Berringer, 1978). Mutant strains of the donor species were used in other cases, for example in the cloning of *B. lactofermentum trpEGDCBA* (Matsui, Miwa and Sano, 1987), *C. crescentus trpFBA* (Ross and Winkler, 1988a), and *T. thermophilus trpBA* (Koyama and Furukawa, 1990). The details of each individual cloning are dependent on the organisms used, and will not be mentioned.

The second selection strategy relies on the DNA sequence homology observed between *trp* genes of closely related organisms, noted first for *Klebsiella aerogenes, S. typhimurium* and *E. coli* by Nichols, Blumenberg and Yanofsky (1981). Cloned DNA from one organism has been used as a hybridization probe for the *trp* genes in transformants. Band, Shimotsu and Henner (1984) used cloned *Bacillus pumilis* DNA to identify *B. subtilis trp* genes, and Crawford and Eberly (1989) used cloned *P. aeruginosa* DNA to identify *P. putida trp* genes.

For the cloning of fungal *trp* genes, both of the above strategies have been used. Carbon *et al.* (1977) used complementation of both yeast and *E. coli* mutants in the selection of clones containing the *TRP*-5 gene from *S. cerevisiae*. The *N. crassa TRP*-3 gene clone was identified from a cosmid bank by its ability to complement an *N. crassa* mutant (Vollmer and Yanofsky, 1986). Most recently, the *TRP*-1 gene encoding tryptophan synthase in the basidiomycete *Coprinus cinereus* was identified using the *TRP*-5 tryptophan synthase gene of *S. cerevisiae* as a hybridization probe (Skrzynia *et al.*, 1989).

Once a piece of DNA containing a complementing *trp* gene has been specified, it may be used as a hybridization probe to identify chromosomal fragments containing other *trp* genes linked in an operon (Matsui, Miwa and Sano, 1987). The extent of *trp* genes present on a cloned fragment can be estimated by the range of *trp* mutants complemented.

In vivo *subunit interchange*

The tryptophan synthase genes from a wide range of bacteria have now been cloned. *In vivo* subunit interchange simply involves the complementation of tryptophan synthase mutations on the host chromosome (usually *E. coli*) with the equivalent foreign subunit introduced on a plasmid. Complementation has been used as a technique for the selection of plasmids containing *trp* genes from gene libraries, and for the verification of a *trp* phenotype in cloned DNA fragments hybridizing to *trp* probes. Unfortunately, studies were not, on the whole, addressing protein interaction and thus the information relating to protein interaction is limited.

Factors affecting subunit interchange in vivo

Three possible *in vivo* complementation phenotypes are observable: good growth, poor growth and no growth. A consideration of factors other than protein interactions that may effect growth in *in vivo* subunit interchange experiments is important before analysis of the results obtained. First, descriptions of the growth are problematical, with an enigmatic 'yes' or 'no' result often published for complementation tests, despite the fact that dissociated subunits may still have sufficient intrinsic activity to support growth. Secondly, as mentioned in the section describing *trp* gene cloning, if the gene products are poorly expressed, unstable or inactive in the host organisms, growth will suffer. Thirdly, uncoupling the expression and regulation of the tryptophan synthase subunits may imbalance relative intracellular amounts, with an overproduction of the plasmid-encoded protein probable. Excess of either subunit may enhance the *in vivo* interaction between heterologous pairs. Interpretation of *in vivo* complementation is, therefore, difficult.

Nevertheless, the current knowledge available for *in vivo* subunit exchange is presented in *Table 7* and the source for these data, including problems in interpretation, is described below.

Expression problems. Expression difficulties were observed with *C. crescentus* DNA in both *E. coli* and *P. putida*; and *R. meliloti* DNA in *E. coli*. Expression from the *trpFBA* operon of *C. crescentus*, in either host, was only obtained with the *tet* promoter. In *E. coli*, only the *trpF* gene was functionally expressed, with no expression of *trpBA* seen. Sequence analysis (Ross and Winkler, 1988a) has determined the codon usage in *C. crescentus* to be dramatically different to *E. coli*, including a GTG *trpB* start codon. The inability to effectively recognize this unusual initiation site, other codon usage differences and the probable translational coupling between *trpB* and *trpA* may account for the poor expression in *E. coli*. Even in *P. putida* the *trpBA* gene products do not perform well. The *C. crescentus* tryptophan synthase is unable to function properly within the host tryptophan biosynthetic pathway, with the addition of exogenous indole being required. The inability to utilize

Table 7. *In vivo* subunit interchange

Plasmid-encoded genes	Host genotype	Growth¶	Promoter	Inference	Reference
B.stearothermophilus trpBA	*E.coli trpBA*⁻	good	original*	foreign TS** responsible for growth	Ishiwata *et al.* (1989)
B.stearothermophilus trpBA	*E.coli trpB*⁻	good	original	foreign TS responsible for growth	Ishiwata *et al.* (1989)
B.stearothermophilus trpB	*E.coli trpB*⁻	good	original	heterologous complex	Ishiwata *et al.* (1989)
B.subtilis trpBA	*E.coli trpA*⁻	yes	original	foreign TS responsible for growth	Henner, Band and Shimotsu (1984)
B.lactofermentum trpBA	*E.coli trpB*⁻	good	original	foreign TS responsible for growth	Matsui, Miwa and Sano (1987)
B.lactofermentum trpBA	*E.coli trpA*⁻	good	original	foreign TS responsible for growth	Matsui, Miwa and Sano (1987)
B.lactofermentum trpB	*E.coli trpB*⁻	good	original	heterologous complex	Matsui, Miwa and Sano (1987)
C.crescentus trpBA	*E.coli trpB*⁻	none	original	no expression due to promoter	Winkler *et al.* (1984)
C.crescentus trpBA	*E.coli trpA*⁻	none	original	no expression due to promoter	Winkler *et al.* (1984)
C.crescentus trpBA	*E.coli trpB*⁻	none	*tet*†	faulty *trpBA* expression	Winkler *et al.* (1984)
C.crescentus trpBA	*E.coli trpA*⁻	none	*tet*	faulty *trpBA* expression	Winkler *et al.* (1984)
C.crescentus trpBA	*P.putida trpB*⁻	none	original	no expression due to promoter	Winkler *et al.* (1984)
C.crescentus trpBA	*P.putida trpA*⁻	none	original	no expression due to promoter	Winkler *et al.* (1984)
C.crescentus trpBA	*P.putida trpB*⁻	+ indole	*tet*	foreign TS partially active	Winkler *et al.* (1984)
C.crescentus trpBA	*P.putida trpA*⁻	yes	*tet*	foreign TS responsible for growth	Winkler *et al.* (1984)
L.casei trpBA	*E.coli trpB*⁻	yes	*lac*‡	foreign TS responsible for growth	Natori, Kano and Imamoto (1990)
L.casei trpB	*E.coli trpB*⁻	no	*lac*	no complementation	Natori, Kano and Imanoto (1990)
L.casei trpBA	*E.coli trpA*⁻	yes	*lac*	foreign TS responsible for growth	Natori, Kano and Imamoto (1990)
L.casei trpA	*E.coli trpA*⁻	no	*lac*	no complementation	Natori, Kano and Imamoto (1990)
M.voltae trpBA	*E.coli trpA*⁻	good	original	foreign TS responsible for growth	Sibold and Henriquet (1988)
M.voltae trpA	*E.coli trpA*⁻	good	original	heterologous complementation	Sibold and Henriquet (1988)
P.aeruginosa trpA§	*E.coli trpA*⁻	poor	*lac*	activities complement, no/poor complex	Hadero and Crawford (1986)
R.meliloti trpBA	*E.coli trpA*⁻	none	original	no expression	Johnston, Bibb and Berringer (1978)
R.meliloti trpBA mutant	*E.coli trpA*⁻	good	?	foreign TS responsible for growth	Johnston, Bibb and Berringer (1978)
R.meliloti trpBA	*P.aeruginosa trpB*⁻	good	original	foreign TS responsible for growth	Johnston, Bibb and Berringer (1978)
R.meliloti trpBA	*P.aeruginosa trpA*⁻	good	original	foreign TS responsible for growth	Johnston, Bibb and Berringer (1978)
S.griseus trpBA	*E.coli trpBA*⁻	good	*tet*	foreign TS responsible for growth	Rivero-Leczano *et al.* (1990)
S.griseus trpBA	*E.coli trpB*⁻	good	*tet*	foreign TS responsible for growth	Rivero-Leczano *et al.* (1990)
S.griseus trpBA	*E.coli trpA*⁻	good	*tet*	foreign TS responsible for growth	Rivero-Leczano *et al.* (1990)
T.thermophilus trpBA	*E.coli trpB*⁻	yes	*lac*	foreign TS responsible for growth	Koyama and Furukawa (1990)
T.thermophilus trpBA	*E.coli trpA*⁻	yes	*lac*	foreign TS responsible for growth	Koyama and Furukawa (1990)

* Original = the indigenous promoter for the operon or gene, or an internal promoter.
** TS, tryptophan synthase complex.
† *tet*, the tetracycline resistance gene promoter from pBR322.
‡ *lac*, the lactose operon promoter in the pUC vector series.
§ The cloned *trpA* from *P.aeruginosa* has an altered N-terminus. The original (Met) Ser Arg residues are replaced by a short segment of *lacα* from pUC8: (Met) Thr Met Ile Thr Asn Ser Arg Gly Ser Val Asp.
¶ Growth is described as published; thus, 'yes' means growth occurred, 'good' indicates a wild-type growth, and 'poor' describes a sub-wild-type growth.

pathway InGP suggests that the problem lies with the α-subunit. The $β_2$-subunit appears unaffected, as indole can be utilized.

The cloned *R. meliloti trpBA* system originally gave no functional expression in *E. coli* (Johnston, Bibb and Berringer, 1978). A mutation on the plasmid at a site upstream of the structural genes, however, allowed complementation of the corresponding host *trp* mutations, implying that expression problems determine the original inactivity.

Functional expression of foreign tryptophan synthase complexes in host mutants. Numerous heterologous tryptophan synthase systems have been shown to complement host *trpBA*⁻, *trpB*⁻ and *trpA*⁻ mutations. Subunit interchange with the formation of a heterologous tryptophan synthase complex is not expected, however, as translational coupling and co-translation favour the formation of the homologous foreign gene product enzyme complex. In the *E. coli* host, complementation of this type has been demonstrated with cloned tryptophan synthase systems from *B. stearothermophilus, B. subtilis* and *B. lactofermentum* without the need for a recombinant promoter. Similar results were seen in the *P. aeruginosa* host with the *R. meliloti* tryptophan synthase. Further complemention studies in *E. coli* have required the use of the *lac* promoter with *L. casei* and *T. thermophilus trpBA*, and the *tet* promoter with *S. griseus trpBA*.

Functional heterologous complexes. Functional heterologous complexes have been demonstrated for *E. coli* with *B. stearothermophilus* β, *E. coli* α with *B. lactofermentum* β and *M. voltae* α with *E. coli* β. The interactions imply that recognition and interaction sites in the above subunits are conserved and the results are thus of some functional importance.

The *M. voltae* case is interesting as although free α will complement *trpA*⁻ *E. coli* mutants, even *trpBA* from the *Methanococcus* will not complement *trpB*⁻ *E. coli* mutants. The non-functional nature of *M. voltae* β in *E. coli* is again thought to be an expression problem since there is good primary structure agreement with the β consensus sequence. By contrast, complementation of the *E. coli* α mutation appears to result from the formation of a heterologous complex.

Tn5 mutagenesis of the cloned *M. voltae* operon, however, has resulted in the selection of a mutant with no detectable complementation. An insertion was located upstream of *trpA*, around the *trpBA* junction and, taken in isolation, could argue against interaction between the heterologous α- and β-subunits. To the above information must be added the fact that the *trpA*⁻ *E. coli* mutant expressing *M. voltae* α grows poorly on minimal medium. Considered with other complementation evidence, the poor growth observed after Tn5 mutagenesis is best explained by effects at the level of α expression. The danger of making conclusions with incomplete data is, therefore, clear.

Non-complementing crosses. Three cases have been demonstrated where heterologous subunits have been unable to form a functional complex.

Combinations of *L. casei* α with *E. coli* $β_2$ and *E. coli* α with *L. casei* $β_2$

were described as *trp*⁻, although slow-growing clones with dissociated subunits may not have been left to grow long enough for detection. Obviously, in these cases heterologous subunit interaction to form a fully active complex does not occur. That a homologous active complex for *L. casei trpBA* can complement *E. coli* has been shown, however, with complementation of both *trpB*⁻ and *trpA*⁻ mutants (Natori, Kano and Imamoto, 1990).

A *P. aeruginosa* α with *E. coli* β_2 (*trpA*⁻) was noticeably slower growing, and accumulated large amounts of InG, indicating a tryptophan synthase deficiency (Hadero and Crawford, 1986). In this case an altered N-terminus for the recombinant α may be responsible, but a more likely explanation is the non-interaction of α and β (supported by *in vitro* subunit interaction work with the closely related *P. putida, Table 5*).

Evidence for non-interacting intergeneric tryptophan synthase subunits implies that evolutionary changes at key site(s) has occurred. It is sensible to predict, therefore, that these key sites will only be fully conserved in complementing organisms.

It is probable that the formation of a functional heterologous complex between the subunits mentioned above is impossible either because of a recognition failure in the cell cytosol, or the inability to functionally interact once recognition has occurred.

PROTEIN PRIMARY STRUCTURE STUDIES

Protein sequencing

The first indication of an intergeneric conservation of tryptophan synthase primary structure was given by Creighton *et al.* (1966). Tryptic digests of α from *E. coli* K-12, *E. coli* B, *Shigella dysenteriae, S. typhimurium* and *K. aerogenes* were shown to be similar. The *E. coli* and *S. dysenteriae* patterns were identical, while the others showed common peptide regions.

Amino acid sequencing in the 1970s allowed a comparison of the first 50 N-terminal residues of α-subunits from *E. coli, S. typhimurium, S. dysenteriae* (Li and Yanofsky, 1972), *K. aerogenes* (Li and Yanofsky, 1973), *P. putida* (Crawford and Yanofsky, 1971) and *B. subtilis* (Li and Hoch, 1974). In this limited comparison of only 15–20% of the protein, a noticeable amount of intergeneric conservation was seen. The observation became more significant when considered together with the finding that most of the important α-residues (as determined by mutagenesis studies; Yanofsky and Horn, 1972) were located at the ends of the polypeptide chain.

The sequence of the tryptic peptide containing the highly conserved pyridoxal phosphate binding site was first determined in *E. coli* (Fluri *et al.*, 1971), *P. putida* (Maurer and Crawford, 1971b), and *S. marcescens* (Rocha, Deeley and Crawford, 1979). The first 100 residues of *E. coli* and *S. typhimurium* β_2-subunits were sequenced shortly after (Higgins, Miles and Fairwell, 1980) and showed a considerable conservation between these two enteric bacteria. The development of DNA sequencing techniques, replaced

protein sequencing efforts to determine the rest of the primary structures.

Protein sequences derived from DNA sequences

A move to cloning *trp* structural and regulatory regions has been evident in recent years, especially with the recognition of their commercial and model system potential. It is a simple procedure to deduce the amino acid sequence from DNA sequences. Moreover, deduced sequences can be compared with separately determined amino acid sequences to ensure fidelity. DNA sequencing, however, does not identify post-translational modifications (more important in eukaryotes) such as the simple removal of the N-terminal f-Met, glycosylation, N-terminal substitution or other more drastic proteolytic processing events, and in many cases the assignment of the start codon is not easy.

Nevertheless, DNA sequencing has identified the prevalence of overlapping genes in *trp* gene clusters where the gene products form a complex (*Table 8* shows the *trpBA* overlapping regions). A strong case for translational coupling in these cases has been presented, and will be discussed below.

A comparison of sequences for both tryptophan synthase subunits, for a range of organisms, is presented in *Figure 5*, and will also be discussed later.

DNA sequencing of fungal tryptophan synthase genes has confirmed the fusion protein theory in the αβ order. The nature and role of the connector region has also been investigated.

Protein primary structure conservation

Mutational shift during evolution, with the nucleotide composition changing to fit the G+C composition of the given organism, and its preferred amino acids and codon usages, has not left much nucleotide homology. Homology can, however, be found in protein sequences, although in some cases alignment of polypeptide chains has required the insertion of pseudo-residues at some places, a potential source of error.

Of the 397 β-residues, 91 are fully conserved throughout the species studied, and 50 are partially conserved, having residues of the same type at

Table 8. Overlapping *trpBA* gene regions

Organism	Overlap	Sequence*
E.coli	1 bp	CGAGGGGAAATCTGATG
P.aeruginosa	4 bp	CAACAGGAGTCGAAAGCATGAGCCGC
P.putida	4 bp	GCCCAGGAGAAACAGGCATGAGCCGT
C.crescentus	12 bp separation	ATCTAGGACCGACCTCGCGTGACCAAA
B.subtilis	8 bp	AGAAGAGGTGAAACGCCATGTTTAAATTGG
B.stearothermophilus	20 bp	AAAGAAGGTGAAGATGTTGCTGCTATCCGTTAAGTTA
L.casei	14 bp	AAATACAAAGGAGTCGATGTTGATGAGTAAGTTA
B.lactofermentum	4 bp	AAATGAAGGACAACCGATGAGCCGTT
T.thermophilus	4 bp	TTCTGGGAGGGGAGCTGTGACCACCC
M.voltae	37 bp separation	TAAAATTAATTAAAATTAATTAAAATATCGAGGAATTTAAATGAAA

* Shine–Dalgarno sequences (Shine and Dalgarno, 1974), stop and start codon are underlined. Overlaps are in italics.

Figure 5. Intergeneric conservation of tryptophan synthase proteins. Single-letter abbreviations are used for amino acid residues, with hyphens representing gaps or pseudo-residues and asterisks at the carboxy terminus. The abbreviations and references for the different organisms are: Ec, *Escherichia coli* (Yanofsky *et al.*, 1981); St, *Salmonella typhimurium* (Crawford, Nichols and Yanofsky, 1980); Ka, *Klebsiella aerogenes* (Nichols, Blumenberg and Yanofsky, 1981); Vp, *Vibrio parahemolyticus* (Crawford, Niermann and Kirschner, 1987); Pa, *Psudomonas aeruginosa* (Hadero and Crawford, 1986); Pp, *Pseudomonas putida* (Crawford and Eberly, 1989); Cc, *Caulobacter crescentus* (Ross and Winkler, 1988a); Bs, *Bacillus subtilis* (Henner, Band and Shimotsu, 1984); Bt, *Bacillus stearothermophilus* (Ishiwata *et al.*, 1989); Lc, *Lactobacillus casei* (Natori, Kano and Imamoto, 1990); Bl, *Brevibacterium lactofermentum* (Matsui, Sano and Ohtsubo, 1986); Tt, *Thermus thermophilus* (Koyama and Furukawa, 1990); Mv, *Methanococcus voltae* (Sibold and Henriquet, 1988); Sc, *Saccharomyces cerevisiae* (Zalkin and Yanofsky, 1982); Nc, *Neurospora crassa* (Burns and Yanofsky, 1989); Ci, *Coprinus cinereus* (Skrzynia *et al.*, 1989). Numbering is of the *E. coli* subunits. The sequences are aligned to maximize similarities. Conserved residues are in bold type; *, all residues at the position are identical; ○, all residues at the position are of a similar type; 1, all but one of the residues at the position are of a similar type; 2, all but two of the residues are of a similar type. The similarity groupings are I=L=M=V; D=E; F=Y=W; K=R; N=Q; S=T.

The α-subunit primary sequence

```
                       1 o1    *    2           122 o*o*1loo**olo*2 o1 2   2            1   o      *o 12
   -------MERYESLFAQLKERKEGAFVPFVTLGDPGIEQSLKIIDTLIEA-GADALELGIPFSDPLADGPTIQNATLRAFAAGVTPAQCFEMLALIRQKHPTIPIGLLM
   -------MERYENLFAQLNDRREGAFVPFVTLGDPGIRQSLKIIDTLIDA-GADALELGVPFSDPLADGPTIQNANLRAFAAGVTPAQCFEMLALIREKHPTIPIGLLM
 , -------MERYETLFAQLKHRQEGAFVPFVTLGDPGPEQSLKIIDALIEG-GADALELGIPFSDPLADGPTIQGAALRAFAAGVTPAQCFEMLAAIRQKHPTIPIGLLM
.p -------MSRYEKHFARLHERNOGAFVPFVTVCDPNAEQSYKIHETLVES-GADALELGIPFSDPLADGPTIQGANIRALDSGATPDICFEQIGKIRAKYPDLPIGLLM
Pa -------MSRLQTRFAQLKQENRAALVTFVTAGDPDYASSLEILKGLPGA-GADVIELGMPFTDPMADGPAIQLANIRALEGGQTLARTLQHVREFRSGDSETPLVLMG
Pp -------MSRLEQRFAELKAEGRSALVTFVTAGDPGYDASLQILKGLPAA-GADVIELGMPFTDPMADGVAIQLATLRALEAGQTLAKTLQHVREFRVDHHTTPIVLMG
Cc -----MYKDRIDRRFAALKAENRAGFVTYVHAGDPDAATTLSVLKGLPAA-GADLIELGFPFSDPMAEGPTIQRAAQRGLKSGHTLQGTLDLVAAFREGDQDTPIILMG
Bs ------------MFKLDLQPSEKLFIPFITAGDPVPEVSIELAKSLQKA-GATALELGVAYSDPLADGPVIQRASKRALDQGMNIVKAIELGGEMKKNGVNIPIILFT
Bt --------------MLLLSVNPPLFIPFIVAGDPSPEVTVDLALALEEA-GADLLELGVPYSDPLADGPTIQRAAARALAGNMTLPKAIHLVAEMRKKGVTIPIILFT
Lc ----------MLHSKLADVFKNHKVFIPFIVADDPDFETTVKNVVALAKG-GADIVELGIPFSDPVADGPVIQAADLRAFAANVRTKTVFDIVEAARKETA-VPIVFLT
Bl ------MSRYDDLFGDASTRSGEGAFVPFIMLSDPSPEEAFQIISTAIER-GADALELGVPFSDPVADGPTVAESHLRALDGGATVDSALEQIKRVRAAYPEVPIGMLI
Tt --------MTTLEAFAKARSEGRAALIPYLTAGFPSREGFLQ-AVEEVLP-YADLLEIGLPYSDPLGDGPVIQRASELALRKGMSVQGALELVREVRAL-TEKPLFLMT
Mv MKNLENLEKDLKMDLKKDLKKEKPILVSFLVSGDPNIEATLKFMNALDEYCGV--IELGIPFSDPIADGSTIQEANVRSLSNGYKIHQSFDVLREFRKFSDT-PVVLMT
Sc ------MSEQLRQTFANAKKENRNALVTFMTAGYPTVKDTVPILKGFQDG-GVDIIELGMPFSDPIADGPTIQLSNTVALQHGVTLPQTLEMVSQARNEGVTVPIILMG
Nc -------MEGIKQTFQRCKAQNRAALVTYVTAGFPEPEQTPDILLAMEKG-GADVIELGVPFTDPIADGPTIQTANTIALQHGVTLQSTLQMVRDARQRGLKAPVMLMG
Ci ----------MDCVLTRSRGPGCDCLRCFRDRRYPKKEDTVPVLLALQAG-GADIIELGIPFSDPIADGPVIQEANTVALKNDIDYPTVLGQIREARQQGLTAPVLLMG
              1        10        20        30        40        50        60        70        80        90
   * * 1  1 o 2       2    oo *1* 2             o  o o 1      o 1       o 1 o*   *2o1          1   o 1        1
Ec YAMLVFNKGIDEFYAQCEKVGVDSVLVADVPVEESAPFRQAALRHNVAPIFICPPNADDDLLRQIASYGRGYTYLLSRAGVTGA-ENRAALPLNHLVAKLKEYN-AAPP
St YAMLVFNNGIDAFYARCEQVGVDSVLVADVPVEESAPFRQAALRHNIAPIFICPPNADDDLLRQVASYGRGYTYLLSRSGVTGA-ENRGALPLHHLIEKLKEYH-AAPA
Ka YAMLVFSPGIDAFYAQCARVGVDSVLVADVPVEESAPFRQAAMRHNIAPIFICPPNADDDLLRQIASYGRGYTYLLSRAGVTGA-ENRAALPLNHLVEKLAEYH-AAPP
Vp YAMLVYSRGIESFYERCAKAGIDSVLIADVPTNESAEFVAAAEKFGIHPIFIAPPTASDETLKQVSELGGGYTYLLSRAGVTGA-ETKANMPVDHMLEKLNQF-DAPPA
Pa YFNPIHHYGVERFIAEAKEVGVDGLIVVDLPPEHNEDLCHPAQAAGLDFIRLTTPTTGDQRLPTVLEGSSGFVYYVSVAGVTGA-NAATLEHVEEAVARLRRHT-DLPI
Pp YYNPIHRFGVEKFVAEAKQAGVDGLIIVDLPPEHDAELATPAQAAGIDFIRLTTPTTDDARLPRVLERSSGFVYYVSVAGVTGA-GSATTEHVTEAIARLRRHT-DLPI
Cc YLNPVLHKGFETFAALAAKAGVDGLIIVDCPPEEADPLSDALEAEGIALIRLAAPTTDDARLPHVVRRTSGFVYYVSVAGVTGV-KSADAADVAPAVERLRKAS-GLPV
Bs YYNPVLQLNKEYFFALLRENHIDGLLVPDLPLEESNSLQEECKSHEVTYISLVAPTS-ESRLKTIIEQAEGFVICVSSLGVTGV-RNEFNSSVYPFIRTVKNLS-TVPV
Bt YYNPVLQLGEESFFALARENGANGVLIPDLPFEESGPLRELGERFDLPLISLVAPTS-KQRIERIASVAQGFLYCVSSLGVTGM-RETLPESLGDFVSEVKRHS-RVPV
Lc YLNIVFKYGYDAFLKRCADLNVAGLVIPDLPYESRDEIVPIAEKYGIDIIPLITPTS-GHRIEKIAKSASGFIYVVSSMGITGE-RDEFFAGLKALVAEIKQYT-NVPT
Bl YGNVPFTRGLDRFYQEFAEAGADSILLPDVPVREGAPFSAAAG---IDPIYIAPANASEKTLEGVSAASKGYIYAISRDGVTGTERESSTDGLSAVVDNIKKF-DGAPI
Tt YLNPVLAWGPERFFGLFKQAGATGVILPDLPPDEDPGLVRLAQEIGLETVFLLAPTSTDARIATVVRHATGFSYAVSVTGVTGM-RERLPEEVKDLVRRIKART-ALPV
Mv YYNPIYKRGIENFVIQALEAGANGLIIVDLPLDEAEQYRAICKKHDMGTVFLVAPNTPDERLMYSDEASTLPLYVISTFGITGA-RGSFEKHTFEFIARAKNLCDKNKL
Sc YYNPILHYGEERFIQDAAKAGANGFIIVDLPPEEALKVRNYINDNGLSLIPLVAPSTTDERLELLSHIADSFVYVVSRMGTTGV-QSSVASDLDELISRVRKYTKDTPL
Nc YYNPLLSYGEERLLNDCKEAGVNGFIIVDLPPEEAVSFRQLCTRGGLSYVPLIAPATSDARMRVLCQLADSFIYVVSRQGVTGA-SGTLNANLPELLARVKKYSGNKPA
Ci YYNPMLAYGEDKAIQDAAEAGANGFIMVDLPPEEAIAFRQKCAASNLSYVPLIAPSTTLKRIQFLASIADSFIYVVSKMGGSSA-NVAVNEELPTILSRIREYT-HVPL
             110       120       130       140       150       160       170       180       190       200
     1**o   2            1 o  ** 11  1
Ec LQGFGISAPDQVKAAIDAGAAGAI-SGSAIVKIIEQHINE----PEKMLAALKVFVQPMKAATRS*
St LQGFGISSPEQVSAAVRAGAAGAI-SGSAIVKIIEKNLAS----PKQMLAELRSFVSAMKAASRA*
Ka LQGFGISAPEQVSAAIDAGAAGAI-SGSAIVKIIERHLDE----PQTMLDELKAFVQSLKAATKTA*
Pa GIGFGIRSAEHA-AAVARLADGVV-VGSALIDRIAKARDH----AQAVKDVLALCGELAEGVRHAR*
Pp SVGFGIRTPEQA-ANIARLADGVV-VGSALVDKIAQATSA----DQAVHDVLSLCSALAEGVRGARR*
Cc AVGFGIRTPDQA-AAVAKVADAAV-VGSALVDEIESAAQL----DEDVTEKVLLKASELAKAVRSARVGTK*
Bs AVGFGISNREQV-IKMHEISDGVV-VGSALVRKIEELKD------RLISAETRNQALQEFEDYAMAFSGLYSLN*
 * AVGFGISTPEQV-AMLKEVCDGVV-IGSALVQKVEQLGE------RLLAPEEKEAAIAEFAAYARSLAAPLHAPCSLR*
   AIGFGIHTPEQAQTHAGIAD-GVI-IGSAIVDLVAKEK-------QQAPAAIEKFTKQIRVAVDAKKQISVK*
   .LGFGISSPQHVADAIAAGASGAI-TGSAITKIIASHCEGEHPNPSTIRDMDGLKKDLTEFISATEGSDQEGLGL*
    VGFGVSG--KATAAQAAVADGVV-VGSALVRALEEG--------RSLAPLLQEIRQGLQRLEANPGLKESSKKPLS*
    VGFGISNGEHAEKIIENGADGVI-VGSAFVDIIKEYGDS----NETIYKLKELARELSEGIEKGYVKYNEKNKY*
    .VGFGVSTREHF-QSVGSVADGVV-IGSKIVTLCGDAPEG-----KRYDVAKEYVQGILNGAKHKVLSKDE...
   .AVGFGVSTHDHF-TQVGAIADGVV-VGSHIITTIQKAAKG-----EEVKAVQEYCSYLCGRNFEQSAHELN...
   .AVEFGVATRDQFNYVADAGADGVVVIGSRIVNAIKAAGEG-----QVPQFVENYCREVSGKGEPSRVRSPG...
    210       220       230        240           250       260
```

Fig. 5 (contd)

The β-subunit primary sequence

```
                     11 *** oo 2 *   o  o          22              o 2*1  *     o1         2 oo2**oo* * * ** *o o
Ec -----------MTTLLNPYFGEFGGMYVPQILMPALRQLEEAFVSAQKDPEFQAQFNDLLKNYAGRPTALTKCQNIT----AGTNTTLYLKREDLLHGGAHKTNQVL
St -----------MTTLLNPYFGEFGGMYVPQILMPALNQLEEAFVRAQKDPEFQAQFADLLKNYAGRPTALTKCQNIT----AGTATTLYLKREDLLHGGAHKTNQVL
Pa ----MT-S-YRNGPDAKGLFGGFGGQYVAETLMPLILDLAREYEKAKDDPAFQEELAYFQRDYVGRPSPLYFAERLT---EHCGGAKIYLKREELNHTGAHKINNCI
Pp ----MTQSQYRPGPDANGLFGSFGGRYVAETLMPLVLDLAREYEAAKADPKFLEELAYFQRDYIGRPNPLYFAERLT---EHCGGAKIFFKREELNHTGAHKVNNCI
Cc MNAPAKPNDYSAYPDAEGRFGGFGGRYVAETLMPLVLDLGKAYADAKADPEFQAQLKSYDTHYAGRPSPLYFAERLT---EHPGGAKIYFKRDELNHTGSHKINNAL
Bs ---------MYPYPNEIGRYGDFGGKFVPETLMQPLDEIQTAFKQIKDDPAFREEYYKLLKDYSGRPTALTYADRVT---EYLGGAKIYLKREDLNHTGSHKINNAL
Bt ---------MERVPNEHGRFGDFGGKFVPETLMLPLEEIEAELDKALADESFKQEYIRILQHYSGRPTPLTFAPNLT---RQLGGAKMYLKREDLNHTGAHKINNAI
Lc  MKTLNETTQQSTRAGRYGKDFGGQYIPETLMTELELVTKAFNDLKDNPEFKAELNDLLVNYANRPSLLYYAKNMT---EDLGGAKIYLKREDLNHTGAHKINNVI
Bl ---MTEKENLGGSTLLPAYFGEFGGQFVAESLLPALDQLEKAFVDATNSPEFREELGGYLRDYLGRPTPLTECSNLPLAGEGKGFARIFLKREDLVHGGAHKTNQVI
Tt ----MLTLPDFPLPDARGRFGPYGGRYVPETLIPALEELEAAYREAKKDPAFLEELDHYLRQFAGRPTPLYHAKRLS---EYWGGAQVFLKREDLLHTGAHKINNTL
Nv -------MKCNTKCDKNGYFGEFGGQYIPEVLKPAVEELKEAYKELKDDEDFQNELAYYLKHYAGRETPLYYAKNLT---EKLGGAKIYLKREDLLHGGAHKTNNTI
Sc ......LDEFDENHKEPIRFGDFGGQYVPEALHACLRELEKGFDEAVADPTFWEDFKSLYS-YIGRPSSLHKAERLT---EHCQGAQIWLKREDLNHTGSHKINNAL
Nc ......AQLAALEGTIPKRFGEFGGQYVPEALMDCLSELEEGFNKIKDDPAFWEEYRSYYP-WNGRPGQLHKAERLT---EYAGGANIWLKREDLNHTGSHKINNAL
Ci ...............LPARFGQFGGQYVPESLVDALAELEEAHKSAIEDPAFWEEVRSLYT-YSNRPSNLYLAENLT---EA-GGANIWLKREDLNHTGSHKINNAL
                   1         10        20        30        40        50        60            70        80        90
   1* ***  o**  o*******2**12 1   *  *1 1 oo**  *2 **  1o2*o o1*o   o 2*  *2 2*o**  1111 *        *oo*o  **1*o
Ec GQALLAKRMGKTEIIAETGAGQHGVASALASALLGLKCRIYMGAKDVERQSPNVFRMRLMGAE--VIPVHSGSATLKDACNEALRDWSGSYETAHYMLGTAAGPHPY
St GQALLAKRMGKSEIIAETGAGQHGVASALASALLGLKCRIYMGAKDVERQSPNVFRMRLMGAE--VIPVHSGSATLKDACNEALRDWSGSYETAHYMLGTAAGPHPY
Pa GQILLARRMGKKRIIAETGAGMHGLATATVAARFGLQCVIYMGTTDIDRQQANVFRMKLLGAE--VIPVTAGTGTLKDAMNEALRDWVTNVDSTFYLIGTVAGPHPY
Pp GQVLLAKRMGKKRIIAETGAGMHGLATATVAARFGLQCVIYMGTTDIDRQQANVFRMKLLGAE--IVPVTAGTGTLKDAMNEALRDWVTNVEDTFYLIGTVAGPHPY
Cc GQILLAMRMGKTRIIAETGAGQHGVATATVCARFGLPCVVYMGATDVERQKPNVFRMNLLGAE--VRPVSSGTGTLKDAMNEAMRDWVTNVHDTYYLIGTAAGPHPY
Bs GQALLAKKMGKTKIIAETGAGQHGVAPATVAAKFGFSCTVFMGEEDVARQSLNVFRMKLLGAE--VVPVTSGNMTLKDATNEAIRYWVQHCEDHFYMIGSVVGPHPY
Bt GQALLAKRMGKKKLIAETGAGQHGVAAATVAAHFGMDCIVFMGEEDIKRQELNVFRMKLLGAE--VVPVSSGNRTLKDATNEAIRYWVAHCDDHFYMIGSVVGPHPY
Lc GQALLAKHLGKKRLIAETGAGQHGVATATIAALMGMDCEIFMGKEDTDRQKLNVTRMELLGAK--VHPVTSGSMVLKDAVNATLQEWASPSDDTFYVLGSAVGPAPF
Bl GQVLLAKRMGKTRIIAETGAGQHGTATALACALMGLECVVYMGAKDVARQQPNVYRMQLEGAK--VIPVESGSGTLRDAVNEALRDWTATFHESHYLLGTRAGPHPF
Tt GQALLARRMGKRRVIAETGAGQHGVSVATVAALFGLECVVYMGEEDVRRQALNVFRMKLLGAE--VRPVAAGSRTLKDATNEAIRDWITNVRTTFYILGSVVGPHPY
Nv GQALLAKKMGKTRIIAETGAGQHGVGTSHAGALFGLETEIFMGRVDTERQQPNVARMKLLGAK--VTPVDTGSKVLKDAVNEAMRNWPATPENTHYLLGTVMGPHPY
Sc AQVLLAKRLGKKNVIAETGAGQHGVATATACAKFGLTCTVFMGAEDVRRQALNVFRMRILGAK--VIAVTNGTKTLRDATSEAFRFWVTNLKTTYYVVGSAIGPHPY
Nc GQLLLARRLGKKKIIAETGAGQHGVATATVCAKFGMECTVFMGAEDVRRQALNVFRMKLLGAK--VVAVEAGSRTLRDAVNEALRYWVVNLADTHYIIGSAIGPHPF
Ci GQILLAKRIGKTRIIAETGAGQHGVATATVCAKFGLECVIYMGAEDVRRQALKIFRIEMLGGKAWVIPVHSGSCTLKDAVNEAMRDWVTHLSTTHYLVGSAIGPHPF
            100       110       120       130       140       150         160       170       180       190
   * o*o *1 o*2 * o *       22o** oo*1o*****o1*2*  o                  1 oo*2*    *1 1  1 *2o  *  *o 2*    2o2   **o
Ec PTIVREFQRMIGEETKAQILERE-GRLPDAVIACVGGGSNAIGMFADFI---------NETNVGLIGVEPGGHGIETGEHGAPLKHGRVGIYFGMKAPMMQTEDGQI
St PTIVREFQRMIGEETKAQILDKE-GRLPDAVIACVGGGSNAIGMFADFI---------NETNVGLIGVEPGGHGIETGEHGAPLKHGRVGIYFGMKAPMMQTADGQI
Pa PAMVRDFQAVIGKETREQLAEKE-GRLPDSLVACIGGGSNAMGLFHEFL---------EEPSVQIIGVEAGGHGVHTDKHAASLNGGVPGVLHGNRTFLLQDADGQI
Pp PAMVRDFQSIIGKETRAQLQEKE-GRLPDSLVACIGGGSNAMGLFHPFL---------DDAGVQIVGVEAAGHGIDTGKHAASLNGGVPGVLHGNRTYLLQDEDGQI
Cc PVMVRDFQSVIGAEAREQILEME-GRLPDAVVACIGGGSNAIGLFHPFL---------GDEGVRLIGVEAAGHGVSTDKHAASLTGGRPGVLHGNRTYLLQDDDGQI
Bs PQVVREFQKMIGEEAKDQLKRIE-GTMPDRVVACVGGGSNAMGMFQAFL----------NEDVELIGAEAAGKGIDTPLHAATISKGTVGVIHGSLTYLIQDEFGQI
Bt PKMVREFQRIIGDEAKEQFLACE-GKLPDVIVACVGGGSNAIGMFYPFL----------QDDVRLVGVEAAGKGIDTPYHAATITKGTKGVIHGAMTYLLQDEYGQI
Lc PEMVKHFQSVISTESKQQLQAKE-AQLPDMVVACVGGGSNAIGSFAAYI---------DDPSVQLVGVEAAGKGADTDRTAATIERGSVGIFHGMKSLFMQNEDGQI
Bl PTIVREFHKVISEEAKAQMLERT-GKLPDVVVACVGGGSNAIGMFADFI---------DDEGVELVGAEPAGEGLDSGEHGATITNGQIGILHGTRSYLMRHSDGQV
Tt PMMVRDFQSVIGEEVKRQSLELF-GRLPDALIAAVGGGSNAIGLFAPFAY-------LPEGRPKLIGVEAAGEGLSTGRHAASIGAGKRGVLHGSYMYLLYDHDGQI
Nv PTMVRDFQSVIGKEVKKQIMEQEE-RLPDYLVACIGGGSNAMGLFHPFLSNNISTGNDDAKNVKMIGIEAAGKGLNTSLHGASITKGEKGVLHGMLSYFLQDEDGQI
Sc PTLVRTFQSVIGKETKEQFAAMNNGKLPDAVVACVGGGSNSTGMFSPFE---------HDTSVKLLGVEAGGDGVDTKFHSATLTAGRPGVFHGVKTYVLQDSDGQV
Nc PTIVRTFQSVIGNETKQQMLEKR-GKLPDAVVACVGGGSNAVGMFYPFS---------NDPSVKLLGVEAGGDGVDTPRHSATLTAGSKGVLHGVRTYILQNQYGQI
Ci PTIVRDFQKVIGEEIKAQLKEVR-GKLPDVVVACVGGGSNAIGTFYDFI---------PDKSVRLVGVEAGGEGIDGHKESATLSMGQPGVLHGVRTYILQDKAGQI
      200       210       220       230       240              250       260       270       280       290
    2  *o*****o*2o*1          1  1    *1  12  2       ***o11112 12o                     oo1 o1******o  1   1
Ec EESYSISAGLDFPSVGPQHAYLNSTGRADYVSITDDEALEAFKTLCLHEGIIPALESSHALAHALKMMREHPD--KEQLLVVNLSGRGDKDIFTVHDILKARGEI*
St EESYSISAGLDFPSVGPQHAYLNSIGRADYVSITDDEALEAFKTLCRHEGIIPALESSHALAHALKMMREQPE--KEQLLVVNLSGRGDKDIFTVHDILKARGEI*
Pa IDAHSISAGLDYPGIGPEHAWLHDIGRVEYTSITDDEALEAFHTCCRLEGIIPALESSHALAEVFKRAPSLP---KEHIMVVNLSGRGDKDMQTVMHHM-QQESKA*
Pp TDAHSISAGLDYPGIGPEHAYLHEVKRVEYVSITDDEALDAFHATCRLEGIIPALESSHALAEAIKRAPKLP---KDHLMVVCLSGRGDKDMQTVMNHMAAQEKQA*
Cc IDAHSISAGLDYPGIGPQHSFLHDIGRAEYVSTTDTEALEAFKLCSTLEGIIPALEPAHALARVGEIAQELG---KGKIVVMNLCGRGDKDIFTVAEALGRKI*
Bs IEPYSISAGLDYPGIGPEHAYLHKSGRVTYDSITDEEAVDALKLLSEKEGILPALESAHALAKAFKLAKGMD---RGQLILVCLSGRGDKDVNTLMNVLEEEVKRHV*
Bt VEPYSISAGLDYPGVGPEHAYLASIGRVRYESVTDEEAVAAFRLLAQTEGIIPAIESAHAVAKAVELAQSMS---PDETVLICLSGRGDKDVQTMMRHLGAKEGEDVA
Lc DPVYSISAGLDYPGVGPEHAALAQEGRAQYVGITDDEAVEAFTYIAKQEGIVAAIESCHAIAYVEKIAPQMA---KDQIIICTLSGRGDKDVASIAKYKGVDVDE*
Bl EESYSISAGLDYPGVGHSTHTCTPPARTTLVSPTPKPSKESSS-LARYEGIIPRTGILTRVRLRLKRAKTAEEEGQWLTILVSLSGRGDKDVDHRAGTLEENPELILK
Tt TPAHSVSAGLDYPGVGPEHSYYADAGVAEYASVTDEEALEGFKLLARLEGIIPALESAHAIAYAAKVVPEMD---KDQVVVINLSGRGDKDVYEVMRLLGGEL*
Nv EEAYSISAGLDYPGIGPEHAYLHNLGRVQYASATDKQALKAFMELTRTEGIIPALESSHAIAYAIENAGNMD---KDDIMVINLSGRGDKDLNTVINAVHKLGC*
Sc HDTHSVSAGLDYPGVGPELAYWKSTGRAQFIAATDAQALLGFKLLSQLEGIIPALESSHAVYGACELAKTMK---PDQHLVINISGRGDKDVQSVAEVLPKLGPKIGWA
Nc EDTHSISAGLDYPGVGPELSNWKDTERAKFVAATDAQAFEGFRLMSQLEGIIPALESSHGIWGALELAKTHK---PDEDVVICLSGRGDKDVQSVADELPIIGPKIGW
Ci IETHSISAGLDYPGVGPEHAWLKDSGRADYVVCTDEDALRGFRMLTQKEGIIPALESSHAIWEGVKIAKSLP---KDKDIVICLSGRGDKDVEQISELLPKWADKLDW
          300       310       320       330       340       350       360        370       380       390
Ec
St
Pa
Pp
Cc
Bs
Bt AIR*
Lc
Bl DNR*
Tt
Nv
Sc DLRFEEDPS
Nc DLRF
Ci VSSNAIPSK
```

given positions in all of the organisms studied. Thirty-seven are partially conserved in all but one of the organisms studied, and 28 are partially conserved in all but two. In the α-subunit, 17 are fully conserved and 23, 24 and 12 are partially conserved in all, all but one and all but two, respectively. The numbers are drawn from *Figure 5*, with *E. coli* as the reference organism.

Despite many of the reservations previously discussed for early studies on subunit conservation, the first and most striking observation from primary sequence data is that, as previously described, β appears more highly conserved than α. Importantly, there are certain residues within both subunits that show almost universal conservation and it is these that may be of greatest interest. To resist change through evolution from the proposed common ancestor to archaebacteria, eubacteria and fungi, these residues must have some crucial role to play. Indeed, genetic and biochemical studies of the *E. coli* subunits have identified residues believed to play crucial roles, as has been discussed above. Alterations (chemical or mutational) at such residues leads to alterations in enzyme activity. In the α-chain these residues include Phe-22, Pro-28, Ser-33, Gly-44, Glu-49, Gly-51, Pro-53, Phe-54, Asp-60, Tyr-175, Leu-177, Thr-183, Gly-211, Gly-213, Gly-234 and Ser-235 (Milton *et al.*, 1986; Miles, Bauerle and Ahmed, 1987; Yutani *et al.*, 1987; Hyde *et al.*, 1988). In the β-chain these residues include His-82 or His-86, Lys-87, Gly-116, Arg-148, Cys-230, Gly-281 and Lys-382 (Cotton and Crawford, 1972; Miles, Bauerle and Ahmed, 1987; Zhao and Somerville, 1987).

The majority of these key residues are conserved either fully or partially but do not account for all of the conserved residues. Looking at the α-subunit, it is striking that, despite extensive structure–function studies, no mis-sense mutants have been found in the central third of the protein. This is particularly interesting given that the three-dimensional model of the *S. typhimurium* tryptophan synthase complex apportions most of the interactive surface of α to structural motifs within the central third of the polypeptide chain (*see* pp. 251–255). It may be that mutations at the α interactive surface, although preventing subunit association and mutual enhancement, do not prevent growth on a minimal medium. Over-expression of individual subunits will therefore allow a growth response in the absence of tryptophan (or indole) and thus such alterations will not be detected by penicillin selection screens for *trp*$^-$ mutants.

Examination of the β-subunit reveals even more conserved residues of unknown function consistent with the more complex nature of β. The highly conserved pyridoxal phosphate binding region (*see* β-residues 77–94 in *Figure 5*) demonstrates a functional area which tolerates little change. The region is conserved between other PLP-containing enzymes as demonstrated by Parsot (1987) when describing the similarities between the tryptophan synthase β-subunit of *E. coli, S. typhimurium* and *B. subtilis* and the threonine synthase protein of *B. subtilis*. Nevertheless, we must note the danger of close comparisons. *Shigella dysenteriae* is a closer taxonomic relation of *E. coli* than *S. typhimurium* but is virtually a tryptophan auxotroph, despite showing a large amount of homology in both regulatory regions and structural genes. There are two mutations in the promoter giving a tenfold reduction in

strength, and two changes in anthranilate synthase virtually rendering the enzyme inactive. As a result, *S. dysenteriae* produces low, nearly constitutive, levels of pathway enzymes capable of salvaging pathway intermediates and susceptible to reactivation to a fully active regulated state by a few specific point mutations (Manson and Yanofsky, 1976).

Interspecies differences in the conservation of certain residues may also be of importance. Subunit interchange experiments have been described where no complementation has occurred. The non-interactive nature may be due to changes at residues conserved between organisms capable of complementation. We might therefore expect non-interactive *E. coli* mutants to map at residues not, for example, conserved in *L. casei* or *P. aeruginosa*.

Overlapping trpBA *genes, translational coupling and co-translation*

An overlapping arrangement of genes can have regulatory importance at the level of expression, and implications at the level of protein–protein interaction (Normark *et al.*, 1983). It has been suggested that if termination of translation of a gene is very close to or even overlaps the translation initiation site for the next gene, expression of these genes can be coupled at the translational level.

Translational coupling is a commonly observed means of translational control where efficient translation of a downstream coding region of a polycistronic mRNA is partially or completely dependent upon the prior translation of the adjacent upstream coding region. It was first demonstrated by Oppenheim and Yanofsky (1980) with the *trpED* system of *E. coli*, where a 1 bp overlap is observed. A *trpE* nonsense mutation, not significantly affecting the levels of the *trpCBA* gene products, significantly decreased expression of *trpD*. Suppression of the *trpE* mutation elevated *trpD* expression in a similar proportion to *trpE*. Translation of *trpE* was therefore required for the optimal expression of *trpD*.

Das and Yanofsky (1984) reported the phenomenon for the tryptophan synthase *trpBA* gene pair of *E. coli* (1 bp overlap). Deletion mutagenesis studies demonstrated that the *trpA* Shine–Dalgarno sequence (within *trpB*) and/or the location of the *trpB* stop codon influenced *trpA* expression. The introduction of a stop codon, for *trpB* 48 nucleotides prior to the *trpA* start, decreased α levels to about 25% of normal. Deletions altering the Shine–Dalgarno sequence reduced α levels to 1% of normal. Further studies (Das and Yanofsky, 1989) demonstrated that the proximity of the *trpB* stop codon to the *trpA* start codon markedly influences α expression. Indeed, when *trpB* translation proceeded in the wild-type reading frame and terminated at the normal *trpB* stop codon, α levels were elevated over the levels observed in plasmid constructs when translation stopped before or after the natural *trpB* stop codon.

The common occurrence of translation coupling (Normark *et al.*, 1983) has facilitated the examination of a number of systems and the proposal of molecular explanations. The first model (favoured for the *E. coli trp* system) assumes upstream translation of the mRNA transcript by the 70S ribosome,

disrupts mRNA secondary structure, unmasking otherwise hidden ribosome binding sites and start codons, allowing initiation of translation. A second model assumes that no secondary structure masking occurs. The initiation site is thought to be inefficient, binding free 30S ribosome subunits poorly. The 30S subunit of an incoming 70S ribosome is thought to be able to initiate efficiently, however, perhaps because its 16S rRNA has already paired the Shine–Dalgarno sequence.

Work on translational coupling is still in the early stages. Many systems with genes containing juxtaposed stop–start codons exist. Investigation of these systems may identify as yet undiscovered features of translational coupling and help our understanding of the phenomenon.

As to a function for translational coupling, it seems a good mechanism to ensure the synthesis of subunit proteins at a constant molar ratio from polycistronic mRNA.

Examination of the tryptophan synthase genes from the range of organisms studied to date shows an overlap of the *trpBA* genes in numerous cases (*Table 8*), suggesting that translational coupling may occur here also. Furthermore, and although possibly coincidental, it is interesting to note that the two organisms identified so far as not having an overlap of the *trpBA* genes (*C. crescentus* and *M. voltae*) also have, as yet, failed to demonstrate any significant controls of *trp* expression at the transcriptional level.

There is also the additional possibility, especially with genes having larger overlaps, that co-translation of the subunit genes is required to form a functional complex. This, for example, may be seen with the anthranilate synthase system of *E. coli* (Kuhn, Pabst and Somerville, 1972). We can imagine a case where the partial translation of *trpA* yields a peptide folding to give a tertiary structure which recognizes and interacts with the β-subunit that has just been made. The absence of the β-subunit from the close proximity of the translation of *trpA*, as would occur in an uncoupled system, would allow complete folding of the α-subunit, potentially masking the β interaction site. In cases such as the *in vivo* subunit interchange between *L. casei* and *E. coli* described above, it is not inconceivable that the failure of complementation is due to the need for co-translation of the *L. casei* subunits. Disruption of translational coupling, leading to poor *L. casei trpA* expression, may also account for the failure of the *L. casei trpA* gene to complement the *E. coli trpA*$^-$ mutant.

Eukaryotic and eubacterial tryptophan synthase fusions

Early genetic and biochemical evidence suggested that in fungi the *trpB* and *trpA* coding regions are fused. The single tryptophan synthase protein dimerizes to form the tryptophan synthase complex (Bonner, DeMoss and Mills, 1965; Manney, 1968; Matchett and DeMoss, 1975; Dettwiler and Kirschner, 1979). The fusion itself was not surprising, as evolutionary pressure would be expected to select rearrangements producing a monocistronic coding region. The interesting point, however, is that the genetic elements specifying the α- and β-domains appear in the reverse order of that

found in bacteria (Manney, 1968; Hütter, Niederberger and DeMoss, 1986). DNA sequencing of the cloned genes from ascomycetes *N. crassa* (Burns and Yanofsky, 1989) and *S. cerevisiae* (Zalkin and Yanofsky, 1982) and the basidiomycete *C. cinereus* (Skrzynia *et al.*, 1989), has verified this orientation of the respective genetic elements.

In several bacterial species, where the genes overlap and are ordered *trpBA*, insertion of a single nucleotide at the *trpB* stop codon could easily fuse the two polypeptides in the βα order. The question, therefore, is how have fungi chosen the more complex fusion route to an αβ order, and what are the functional reasons?

Comparison of the deduced primary protein tryptophan synthase sequences for *N. crassa* and *S. cerevisiae*, with the sequences deduced for bacteria (Zalkin and Yanofsky, 1982; Burns and Yanofsky, 1989), reveal that the fused fungal polypeptides homologous to the bacterial α- and β-subunits are joined by a connector region. The connectors are approximately 50 (Burns and Yanofsky, 1989) to 70 (Zalkin and Yanofsky, 1982) amino acid residues long and have little or no homology with the α- or β-subunits. The sections showing little homology reflect the respective ends of the bacterial subunits (areas of poor intergeneric conservation, *see Figure 5*). The central region that lacks homology is about 10 residues in length.

Studies have investigated the importance of both the length and sequence of the connector. Comparison of the connectors of *N. crassa, S. cerevisiae* and *C. cinereus* shows no significant sequence conservation. The connectors of *N. crassa,* 76 residues, *S. cerevisiae*, 67 residues, and *C. cinereus*, 39 residues, show no specific connector size requirement, although definitions of the connector boundaries vary. Within the individual species the length of the connector appears to have a crucial role (Crawford *et al.*, 1987). Deletion of an 18-residue section of the yeast connector rendered the enzyme inactive. Replacement with a 14-residue section of different sequence recovered 6% of the activity, and a spontaneous duplication of two amino acids, giving a 16-residue replacement, recovered 26% of the activity. Breaking the fusion by dividing the enzyme at the connector (Crawford *et al.*, 1987) obtained segments, albeit low in activity, able to combine and mutually enhance activities by thirtyfold, to 20% of normal, indicating that the connections are not a functional requirement.

Fusion of the *E. coli* tryptophan synthase genes (Yanofsky *et al.*, 1987; Burns *et al.*, 1990) has indicated that fusion in the αβ order is required for optimal enzyme activity. Active fusion in both orders, with and without a synthetic connector of 10 residues, have been produced. Both types of fusion proteins in the βα order had reduced activity and showed a marked tendency to aggregate and associate with insoluble cell material, which was not attributable to over-expression. The properties of the βα fusions suggest an abnormal enzyme. Fusion in the αβ order without the connector gave full activity, but also gave association with insoluble cell material, a property significantly reduced by the presence of the connector.

Evaluation of fusion results, in relation to the three-dimensional structure of the *S. typhimurium* tryptophan synthase (Burns *et al.*, 1990), have

produced some interesting observations. It is assumed that any differences between *E. coli* and *S. typhimurium* are negligible. The three-dimensional model of Hyde *et al.* (1988) indicates that in tryptophan synthase the α N-terminus and β C-terminus are separated by 50 Å, and the β N-terminus and α C-terminus by 70 Å. Theory predicts a connector of 20–30 residues to be required to allow proper folding to give a functional complex. Nature has supplied fungi with longer connectors, but the fusion experiments provided connectors somewhat too short. Some structural distortion is therefore likely, due to the constraint of the fusion. Yet the enzymes are functional, if only partially, so the active site is not radically changed and the subunit recognition sites are able to interact. Indeed, competition experiments with mutant α33 subunits have indicated that the indole tunnel is retained in the fusions (Burns *et al.*, 1990).

Conclusions

The molecular basis of inter-protein recognition and interaction remains one of the unanswered questions in molecular biology. In the opening section of this review we defined a number of requirements for model systems in the continuing study of protein–protein interaction. Tryptophan synthase meets these requirements, since a variety of techniques are effective in altering the primary structure of individual tryptophan synthase subunits, and the differential phenotype of interactive/non-interactive subunits appear sufficiently dissimilar to facilitate the development of screening methods. In addition, the three-dimensional structure of the enzyme has been solved by X-ray crystallography, providing a structural perspective for both mutational and folding data.

In this review we have described a number of residues with a putative involvement in αβ interaction; α Val-126 and Val-128, α Gly-51 and α Lys-109 have been implicated at the αβ contact region. Effects on αβ interaction have been observed with changes at α-residues Pro-28, Ser-33, Gly-44, Cys-118 and Arg-179. In addition to catalytic effects, changes at Phe-22 and Gly-211 have been shown to affect intersubunit interaction. The effect on interaction caused by residues not located at the αβ contact region is thought to be indirect, with mutation affecting the local conformation around the interaction site, and in some cases also the catalytic site.

Of all the residues mentioned above, only α Lys-109 is not conserved (at least throughout the Enterobacteriaceae studied). Evolutionary divergence has arranged the different *trp* genes in many ways on the chromosome, altered their nucleotide composition and devised numerous ways of regulating the activity of the tryptophan biosynthetic enzymes for which they code. Nevertheless, throughout the species studied the proteins produced have certain conserved, presumably key, residues.

The α- and β-residues that have so far been attributed with an essential or otherwise important role do not account for all of the conserved residues identified in sequencing studies. The goal of identifying the key descriptors of protein–protein interaction is unlikely to be achieved, even for this model

system, while the role for such residues remains unknown. It is not unreasonable to suggest that some are involved in the interaction of α and β_2; a highly important function of both tryptophan synthase subunits which gives mutual enhancement of activity, allows for the formation of the indole tunnel and facilitates communication between α and β active sites.

In order to determine the nature of the $\alpha\beta$ interaction site, a greater proportion of the residues functionally involved at the $\alpha\beta$ contact region need to be identified. Now that an effective technique for the selection of non-interactive α-mutants has been developed (Swift, Kuhn and Stewart, in preparation), the identification of further randomly generated non-interactive α-mutants should be possible. In a similar manner to the elucidation of intraprotein interactions within the α-subunit, it should be possible to expand the number of residues known to be involved in $\alpha\beta$ interaction on both subunits using second-site mutagenesis. Once key sites have been identified, the use of site-directed mutagenesis allows for the determination of functional, partially functional and non-functional replacement amino acids. Further characterization of these proteins can be accomplished with measurements of association constants and computer-aided modelling of mutant subunits using X-ray crystallography data.

Despite the wealth of knowledge obtained from 40 years of research into tryptophan synthase, we would still wish to echo Charles Yanofsky's comments of 1987, 'the end is by no means in sight!' The details of the catalytic mechanism and the key active site residues involved are still unclear and, although the gross aspects of folding and subunit association *in vitro* have been characterized, how the scheme compares with the *in vivo* situation remains unknown. Elucidation of the molecular basis of $\alpha\beta$ and $\beta\beta$ protein interactions and the key residues or structural motifs involved is in its infancy. We hope that this review will serve to encourage the way forward.

Acknowledgements

One of us (G.S.A.B.S.) is greatly indebted to Dr Jonathan Kuhn for providing a molecular insight into tryptophan synthase as a model for protein interaction. He kindled the spark of interest that led to the work presented here. During this work S.S. was funded by a University of Nottingham Studentship.

References

AHMED, S.A., FAIRWELL, T., DUNN, S., KIRSCHNER, K. AND MILES, E.W. (1986). Identification of three sites of proteolytic cleavage in the hinge region between the two domains of the β_2 subunit of tryptophan synthase of *Escherichia coli* or *Salmonella typhimurium*. *Biochemistry* **25**, 3118–3124.

AHMED, S.A., KAWASAKI, H., BAUERLE, R., MORITA, H. AND MILES, E.W. (1988). Site-directed mutagenesis of the α subunit of tryptophan synthase from *Escherichia coli*. *Biochemical and Biophysical Research Communications* **151**, 672–678.

AMIT, A.G., MARIUZZA, R.A., PHILLIPS, S.E.V. AND POLJAK, R.J. (1986). Three-dimensional structure of an antigen-antibody complex at 2·8 Å resolution. *Science* **233**, 747–753.

BAE, Y.M. AND CRAWFORD, I.P. (1990). The *Rhizobium meliloti trpE(G)* gene is regulated by attenuation, and its product, anthranilate synthase, is regulated by feedback inhibition. *Journal of Bacteriology* **172**, 3318–3327.

BAE, Y.M., HOLMGREN, E. AND CRAWFORD, I.P. (1989). *Rhizobium meliloti* anthranilate synthase gene: cloning, sequence and expression in *Escherichia coli*. *Journal of Bacteriology* **171**, 3471–3478.

BALBINDER, E. (1964). Intergeneric complementation between A and B components of bacterial tryptophan synthases. *Biochemical and Biophysical Research Communications* **17**, 770–774.

BAND, L., SHIMOTSU, H. AND HENNER, D.J. (1984). Nucleotide sequence of the *Bacillus subtilis trpE* and *trpD* genes. *Gene* **27**, 55–65.

BARTHOLMES, P. AND TEUSCHER, B. (1979). Cooperative binding of α subunits to the apo-β_2 subunit of tryptophan synthase from *Escherichia coli*. *European Journal of Biochemistry* **95**, 323–326.

BARTHOLMES, P., BALK, H. AND KIRSCHNER, K. (1980). Mechanism of reconstitution of the Apoβ_2 subunit and the $\alpha_2$apoβ_2 complex of tryptophan synthase with pyridoxal 5′-phosphate: kinetic studies. *Biochemistry* **19**, 4527–4533.

BEASTY, A.M., HURLE, M.R., MANZ, J.T., STACKHOUSE, T., ONUFFER, J.J. AND MATTHEWS, C.R. (1986). Effects of the phenylalanine22–leucine, glutamic acid49–methionine, glycine234–aspartic acid, and glycine234–lysine mutations on the folding and stability of the α subunit of tryptophan synthase from *Escherichia coli*. *Biochemistry* **25**, 2965–2974.

BERLYN, M.B., LAST, R.L. AND FINK, G.R. (1989). A gene encoding the tryptophan synthase β subunit of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **86**, 4604–4608.

BERRINGER, J.E., HOGGAN, S.A. AND JOHNSTON, A.W.B. (1978). Linkage mapping in *Rhizobium leguminosarum* by means of R plasmid mediated recombination. *Journal of General Microbiology* **104**, 201–207.

BLOND, S. AND GOLDBERG, M.E. (1985). Kinetics and importance of the dimerisation step in the folding pathway of the β_2 subunit of *Escherichia coli* tryptophan synthase. *Journal of Molecular Biology* **182**, 597–606.

BLOND, S. AND GOLDBERG, M.E. (1986). Kinetic characterisation of early intermediates in the folding of *E. coli* tryptophan synthase β_2 subunit. *Proteins: Structure, Function and Genetics* **1**, 247–255.

BLOND, S. AND GOLDBERG, M. (1987). Partly native epitopes are already present on early intermediates in the folding of tryptophan synthase. *Proceedings of the National Academy of Sciences, USA* **84**, 1147–1151.

BONNER, D.M., DEMOSS, J.A. AND MILLS, S.E. (1965). The evolution of an enzyme. In *Evolving Genes and Proteins* (V. Bryson and H.J. Vogel, Eds), pp.305–318. Academic Press, New York.

BRENNER, S., STRETTON, A.O.W. AND KAPLAN, S. (1965). Genetic code: the 'nonsense' triplets for chain termination and their suppression. *Nature* **206**, 994–998.

BURNS, D.M. AND YANOFSKY, C. (1989). Nucleotide sequence of the *Neurospora crassa trp-3* gene encoding tryptophan synthetase and comparison of the *trp-3* polypeptide with its homologs in *Saccharomyces cerevisiae* and *Escherichia coli*. *Journal of Biological Chemistry* **264**, 3840–3848.

BURNS, D.M., HORN, V., PALUH, J. AND YANOFSKY, C. (1990). Evolution of the tryptophan synthetase of fungi. *Journal of Biological Chemistry* **265**, 2060–2069.

BUVINGER, W.E., STONE, L.C. AND HEATH, H.E. (1981). Biochemical genetics of tryptophan synthesis in *Pseudomonas acidovorans*. *Journal of Bacteriology* **147**, 62–68.

CARBON, J., RATZKIN, B. AND CLARKE, L. AND RICHARDSON, D. (1977). In *Brookhaven Symposia in Biology*, volume 29 (H.H. Smith, Ed.), pp.277–296. Brookhaven National Laboratory, Upton, New York.

CHAFFOTTE, A.F. AND GOLDBERG, M.E. (1983). Immunological evidence for confor-

mational flexibility and its modulation by specific ligands in the β_2 subunit of *Escherichia coli* tryptophan synthase. *Biochemistry* **22**, 2708–2714.

CHAFFOTTE, A.F. AND GOLDBERG, M.E. (1984). Fluorescence-quenching studies on a conformational transition within a domain of the β_2 subunit of *Escherichia coli* tryptophan synthase. *European Journal of Biochemistry* **139**, 47–50.

CHANG, M. AND CRAWFORD, I.P. (1990). The roles of indoleglycerol phosphate and the *trpI* protein in the expression of *trpBA* from *Pseudomonas aeruginosa*. *Nucleic Acids Research* **18**, 979–988.

COHN, W. AND CRAWFORD, I.P. (1976). Regulation of enzyme synthesis in the tryptophan pathway of *Acinetobacter calcoaceticus*. *Journal of Bacteriology* **127**, 367–379.

COTTON, R.G.H. AND CRAWFORD, I.P. (1972). Tryptophan synthase β_2 subunit. Application of genetic analysis to the study of primary structure. *Journal of Biological Chemistry* **247**, 1883–1891.

CRAWFORD, I.P. (1989). Evolution of a biosynthetic pathway: The tryptophan paradigm. *Annual Reviews of Microbiology* **43**, 567–600.

CRAWFORD, I.P. AND EBERLY, L. (1989). DNA sequence of the tryptophan synthase genes of *Pseudomonas putida*. *Biochimie* **71**, 521–531.

CRAWFORD, I.P. AND JOHNSON, L.M. (1963). Mutants of *Escherichia coli* defective in the B protein of tryptophan synthetase. I Selection and classification. *Genetics* **48**, 725–736.

CRAWFORD, I.P. AND JOHNSON, L.M. (1964). Mutants of *Escherichia coli* defective in the B protein of tryptophan synthetase. II Intragenic position. *Genetics* **49**, 267–278.

CRAWFORD, I.P. AND STAUFFER, G.V. (1980). Regulation of tryptophan biosynthesis. *Annual Review of Biochemistry* **49**, 163–195.

CRAWFORD, I.P. AND YANOFSKY, C. (1971). *Pseudomonas putida* tryptophan synthetase: Partial sequence of the α subunit. *Journal of Bacteriology* **108**, 248–253.

CRAWFORD, I.P., DECASTEL, M. AND GOLDBERG, M.E. (1978). Assignment of the ends of the β-chain of *E. coli* tryptophan synthase to the F_1 and F_2 domains. *Biochemical and Biophysical Research Communications* **85**, 309–316.

CRAWFORD, I.P., NICHOLS, B.P. AND YANOFSKY, C. (1980). Nucleotide sequence of the *trpB* gene in *Escherichia coli* and *Salmonella typhimurium*. *Journal of Molecular Biology* **142**, 489–502.

CRAWFORD, I.P., NIERMANN, T. AND KIRSCHNER, K. (1987). Prediction of secondary structure by evolutionary comparison: application to the α subunit of tryptophan synthase. *Proteins: Structure, Function and Genetics* **2**, 118–129.

CRAWFORD, I.P., SIKES, S., BELSER, N.O. AND MARTINEZ, L. (1970). Mutants of *Escherichia coli* defective in the B protein of tryptophan synthetase. III Intragenic clustering. *Genetics* **65**, 201–211.

CRAWFORD, I.P., CLARKE, M., VAN CLEEMPUT, M. AND YANOFSKY, C. (1987). Crucial role of connecting region joining the two functional domains of Yeast tryptophan synthetase. *Journal of Biological Chemistry* **262**, 239–244.

CREIGHTON, T.E. AND YANOFSKY, C. (1966). Association of the α and β_2 subunits of the tryptophan synthase of *Escherichia coli*. *Journal of Biological Chemistry* **241**, 980–990.

CREIGHTON, T.E., HELINSKI, D.R., SOMERVILLE, R.L. AND YANOFSKY, C. (1966). Comparison of tryptophan synthetase α subunits of several species of *Enterobacteriaceae*. *Journal of Bacteriology* **91**, 1819–1826.

CRISANTI, M.M. AND MATTHEWS, C.R. (1981). Characterisation of the slow steps in the folding of the α subunit of tryptophan synthetase. *Biochemistry* **20**, 2700–2706.

DAS, A. AND YANOFSKY, C. (1984). A ribosome binding site sequence is necessary for efficient expression of the distal gene of a translationally-coupled gene pair. *Nucleic Acids Research* **12**, 4745–4768.

DAS, A. AND YANOFSKY, C. (1989). Restoration of a translational stop start overlap

reinstates translational coupling in a mutant *trpB'-trpA* gene pair of the *Escherichia coli* tryptophan operon. *Nucleic Acids Research* **17**, 9333–9340.

DETTWILER, M. AND KIRSCHNER, K. (1979). Tryptophan synthetase from *Saccharomyces cerevisiae* is a dimer of two polypeptide chains of M_r 76,000 each. *European Journal of Biochemistry* **102**, 159–165.

DICAMELLI, R.F. AND BALBINDER, E. (1976). The association of tryptophan synthetase subunits from *Escherichia coli* and *Salmonella typhimurium* in homologous and heterologous combinations. *Genetical Research* **27**, 323–333.

DICAMELLI, R.F., BALBINDER, E. AND LEBOWITZ, J. (1973). Pressure effects on the association of the α and β_2 subunits of the tryptophan synthetase from *Escherichia coli* and *Salmonella typhimurium. Archives of Biochemistry and Biophysics* **155**, 315–324.

DREWE, W.F., JR AND DUNN, M.F. (1985). Detection and identification of intermediates in the reaction of L-serine with *Escherichia coli* tryptophan synthase via rapid-scanning ultraviolet-visible spectroscopy. *Biochemistry* **24**, 3977–3987.

DREWE, W.F., JR AND DUNN, M.F. (1986). Characterisation of the reaction of L-serine and indole with *Escherichia coli* tryptophan synthase via rapid-scanning ultraviolet-visible spectroscopy. *Biochemistry* **25**, 2494–2501.

DUNN, M.F., AGUILAR, V., BRZOVIC, P., DREWE, W.F., JR, HOUBEN, K.F., LEJA, C.A. AND ROY, M. (1990). The tryptophan synthase bienzyme complex transfers indole between the α- and β-sites via a 25–30 Å long tunnel. *Biochemistry* **29**, 8598–8607.

ESSAR, D.W., EBERLY, L. AND CRAWFORD, I.P. (1990). Evolutionary differences in chromosomal locations of four early genes of the tryptophan pathway in fluorescent pseudomonads: DNA sequences and characterisation of *Pseudomonas putida trpE* and *trpGDC. Journal of Bacteriology* **172**, 867–883.

ESSAR, D.W., EBERLY, L., HAN, C.-Y. AND CRAWFORD, I.P. (1990). DNA sequences and characterisation of four early genes of the tryptophan pathway in *Pseudomonas aeruginosa. Journal of Bacteriology* **172**, 853–866.

EUN, H.-M. AND MILES, E.W. (1984). Reaction of phenylglyoxal and (*p*-hydroxyphenyl)glyoxal with arginines and cysteines in the α subunit of tryptophan synthase. *Biochemistry* **23**, 6484–6491.

FANKHAUSER, H., PYTHOUD, F. AND KING, P.J. (1990). A tryptophan auxotroph of *Hyoscyamus muticus* lacking tryptophan synthase activity. *Planta* **180**, 297–302.

FLURI, R., JACKSON, L.E., LEE, W.E. AND CRAWFORD, I.P. (1971). The tryptophan synthetase β_2 subunit. Primary structure of the pyridoxal peptide from the *Escherichia coli* enzyme. *Journal of Biological Chemistry* **246**, 6620–6624.

FREEDBERG, W.B. AND HARDMAN, J.K. (1971). Structural and functional roles of the cysteine residues in the α subunit of the *Escherichia coli* tryptophan synthetase. III Studies with the bifunctional sulfhydryl reagent bismaleimidomethylether. *Journal of Biological Chemistry* **246**, 1439–1448.

GIBSON, F., GIBSON, M.I. AND YANOFSKY, C. (1961). A mutational alteration of the tryptophan synthetase of *Escherichia coli. Journal of General Microbiology* **24**, 301–312.

GOLDBERG, M.E. AND BALDWIN, R.L. (1967). Interactions between the subunits of the tryptophan synthetase of *Escherichia coli*. Optical properties of an intermediate bound to the $\alpha_2\beta_2$ complex. *Biochemistry* **6**, 2113–2119.

GOLDBERG, M.E., YORK, S. AND STRYER, L. (1968). Fluorescence studies of substrate and subunit interactions of the β_2 protein of *Escherichia coli* tryptophan synthase. *Biochemistry* **7**, 3662–3667.

GRYCZAN, T., CONTENTE, S. AND DUBNAU, D. (1980). Molecular cloning of heterologous chromosomal DNA by recombination between plasmid vector and a homologous resident plasmid in *Bacillus subtilis. Molecular and General Genetics* **177**, 459–467.

HADERO, A. AND CRAWFORD, I.P. (1986). Nucleotide sequence of the genes for

tryptophan synthase in *Pseudomonas aeruginosa*. *Molecular and Biological Evolution* **3**, 191–204.

HARDMAN, J.K. AND HARDMAN, D.F. (1971). Reaction of the α subunit of the *Escherichia coli* tryptophan synthase with 1,5-difluoro-2,4-dinitrobenzene. *Journal of Biological Chemistry* **246**, 6489–6496.

HARDMAN, J.K. AND YANOFSKY, C. (1965). Studies on the active site of the A protein subunit of the *Escherichia coli* tryptophan synthetase. *Journal of Biological Chemistry* **240**, 725–732.

HATANAKA, M., WHITE, E.A., HORIBATA, K. AND CRAWFORD, I.P. (1962). A study of the catalytic properties of *Escherichia coli* tryptophan synthetase, a two component enzyme. *Archives of Biochemistry and Biophysics* **97**, 596–606.

HATHAWAY, G.M. AND CRAWFORD, I.P. (970). Studies on the association of β-chain monomers of *Escherichia coli* tryptophan synthetase. *Biochemistry* **9**, 1801–1808.

HEDGES, R.W., JACOB, A.E. AND CRAWFORD, I.P. (1977). Wide ranging plasmid bearing the *Pseudomonas aeruginosa* tryptophan synthase genes. *Nature* **267**, 283–284.

HENNER, D.J., BAND, L. AND SHIMOTSU, H. (1984). Nucleotide sequence of the *Bacillus subtilis* tryptophan operon. *Gene* **34**, 169–177.

HENNING, U. AND YANOFSKY, C. (1962). An alteration in the primary structure of a protein predicted on the basis of genetic recombination data. *Proceedings of the National Academy of Sciences, USA* **48**, 183–190.

HIGGINS, W. AND MILES, E.W. (1978). Affinity labeling of the pyridoxal phosphate binding site of the β_2 subunit of *Escherichia coli* tryptophan synthase. *Journal of Biological Chemistry* **253**, 3964–3967.

HIGGINS, W., FAIRWELL, T. AND MILES, E.W. (1979). An active proteolytic derivative of the α subunit of tryptophan synthetase. Identification of the site of cleavage and characterisation of the fragments. *Biochemistry* **18**, 4827–4835.

HIGGINS, W., MILES, E.W. AND FAIRWELL, T. (1980). Location of three active site residues in the NH_2-terminal sequence of the β_2 subunit of tryptophan synthase from *Escherichia coli*. *Journal of Biological Chemistry* **255**, 512–517.

HOCH, S.O. (1973a). Tryptophan synthetase from *Bacillus subtilis*. Purification and characterisation of the β_2 component. *Journal of Biological Chemistry* **248**, 2992–2999.

HOCH, S.O. (1973b). Tryptophan synthetase from *Bacillus subtilis*. Purification and characterisation of the α component. *Journal of Biological Chemistry* **248**, 2999–3003.

HOCH, S.O. AND CRAWFORD, I.P. (1973). Enzymes of the tryptophan pathway in three *Bacillus* species. *Journal of Bacteriology* **116**, 685–693.

HÖGBERG-RAIBAUD, A. AND GOLDBERG, M.E. (1977a). Isolation and characterisation of independently folding regions of the β chain of *Escherichia coli* tryptophan synthetase. *Biochemistry* **16**, 4014–4020.

HÖGBERG-RAIBAUD, A. AND GOLDBERG, M.E. (1977b). Preparation and characterisation of a modified form of β_2 subunit of *Escherichia coli* tryptophan synthetase suitable for investigating protein folding. *Proceedings of the National Academy of Sciences, USA* **74**, 442–446.

HURLE, M.R. AND MATTHEWS, C.R. (1987). Proline isomerisation and the slow folding reactions of the α subunit of tryptophan synthase from *Escherichia coli*. *Biochimica et Biophysica Acta* **913**, 179–184.

HURLE, M.R., TWEEDY, N.B. AND MATTHEWS, C.R. (1986). Synergism in folding of a double mutant of the α subunit of tryptophan synthase. *Biochemistry* **25**, 6356–6360.

HURLE, M.R., MICHELOTTI, G.A., CRISANTI, M.M. AND MATTHEWS, C.R. (1987). Characterisation of a slow folding reaction for the α subunit of tryptophan synthase. *Proteins: Structure, Function and Genetics* **2**, 54–63.

HÜTTER, R. AND NIEDERBERGER, P. (1983). Biochemical pathways and mechanisms.

Nitrogen, amino acid and carbon metabolism. *Biotechnology Advances* **1**, 179–191.

HÜTTER, R., NIEDERBERGER, P. AND DEMOSS, J.A. (1986). Tryptophan biosyntheteic genes in eukaryotic microorganisms. *Annual Review of Microbiology* **40**, 55–77.

HYDE, C.C. AND MILES, E.W. (1990). The tryptophan synthase multienzyme complex: Exploring structure–function relationships with X-ray crystallography and mutagenesis. *Biotechnology* **8**, 27–31.

HYDE, C.C., AHMED, S.A., PADLAN, E.A., MILES, E.W. AND DAVIES, D.R. (1988). Three dimensional structure of the tryptophan synthase $\alpha_2\beta_2$ multienzyme complex from *Salmonella typhimurium*. *Journal of Biological Chemistry* **263**, 17857–17871.

IBEL, K., MAY, R.P., KIRSCHNER, K., LANE, A.N., SZADKOWSKI, H., DAUVERGNE, M.-T. AND ZULAUF, M. (1985). The domain structure of tryptophan synthase. *European Journal of Biochemistry* **151**, 505–514.

ISHIWATA, K., YOSHINO, S., IWAMORI, S., SUZUKI, T. AND MAKIGUCHI, N. (1989). Cloning and sequencing of *Bacillus stearothermophilus* tryptophan synthase genes. *Agricultural and Biological Chemistry* **53**, 2941–2948.

IWAHASHI, H., YUTANI, K., OGASAHARA, K., TSUNSAWA, S., KYOGOKU, Y. AND SUGINO, Y. (1983). NMR evidence for the stepwise unfolding of the two domains of tryptophan synthase α subunit. *Biochemica et Biophysica Acta* **744**, 189–192.

JACKSON, D.A. AND YANOFSKY, C. (1969). Restoration of enzymic activity by complementation *in vitro* between mutant α subunits of tryptophan synthetase and between mutant subunits and fragments of the α subunit. *Journal of Biological Chemistry* **244**, 4539–4546.

JAENICKE, R. (1991). Protein folding: local structures, domains, and subunit assemblies. *Biochemistry* **30**, 3147–3161.

JOHNSTON, A.W.B., BIBB, M.J. AND BERRINGER, J.E. (1978). Tryptophan genes in *Rhizobium* – their organisation and their transfer to other bacterial genera. *Molecular and General Genetics* **165**, 323–330.

KAPLAN, J.B., GONCHAROFF, P., SEIBOLD, A.M. AND NICHOLS, B.P. (1984). Nucleotide sequence of the *Acinetobacter calcoaceticus trpGDC* gene cluster. *Molecular Biology and Evolution* **1**, 456–472.

KAWASAKI, H., BAUERLE, R., ZON, G., AHMED, S.A. AND MILES, E.W. (1987). Site specific mutagenesis of the α subunit of tryptophan synthase from *Salmonella typhimurium*. *Journal of Biological Chemistry* **262**, 10678–10683.

KOYAMA, Y. AND FURUKAWA, K. (1990). Cloning and sequence analysis of tryptophan synthetase genes of an extreme thermophile, *Thermus thermophilus* HB27: Plasmid transfer from replica-plated *Escherichia coli* recombinant colonies to competent *T. thermophilus* cells. *Journal of Bacteriology* **172**, 3490–3495.

KUHN, J.C., PABST, M.J. AND SOMERVILLE, R.L. (1972). Mutant strains of *Escherichia coli* K12 exhibiting enhanced sensitivity to 5-methyltryptophan. *Journal of Bacteriology* **112**, 93–101.

KURODA, M.I., HENNING, D. AND YANOFSKY, C. (1988). Cis-acting sites in the transcript of the *Bacillus subtilis trp* operon regulate expression of the operon. *Journal of Bacteriology* **170**, 3080–3088.

KURODA, M.I., SHIMOTSU, H., HENNER, D.J. AND YANOFSKY, C. (1986). Regulatory elements common to the *Bacillus pumilis* and *Bacillus subtilis trp* operon. *Journal of Bacteriology* **167**, 792–798.

LANDRY, S.J. AND GIERASCH, L.M. (1991). Recognition of nascent polypeptides for targeting and folding. *Trends in Biochemical Sciences* **16**, 159–163.

LANE, A.N. (1983). The accessibility of the conformation states of the β_2 subunit of tryptophan synthase studied by fluorescence quenching. *European Journal of Biochemistry* **133**, 531–538.

LANE, A.N. AND KIRSCHNER, K. (1983a). The mechanism of binding of L-serine to

tryptophan synthase from *Escherichia coli*. *European Journal of Biochemistry* **129**, 561–570.

LANE, A.N. AND KIRSCHNER, K. (1983b). The catalytic mechanism of tryptophan synthase from *Escherichia coli*. Kinetics of the reaction of indole with the enzyme–L-serine complexes. *European Journal of Biochemistry* **129**, 571–582.

LANE, A.N. AND KIRSCHNER, K. (1983c). The quarternary structure of tryptophan synthase from *Escherichia coli*. Fluoresence and hydrodynamic studies. *European Journal of Biochemistry* **129**, 675–684.

LANE, A.N., PAUL, C.H. AND KIRSCHNER, K. (1984). The mechanism of self assembly of the multienzyme complex tryptophan synthase from *Escherichia coli*. *EMBO Journal* **3**, 279–287.

LERNER, P. AND YANOFSKY, C. (1957). An immunological study of mutants of *Escherichia coli* lacking the enzyme tryptophan synthase. *Journal of Bacteriology* **74**, 494–501.

LI, S.-L. AND YANOFSKY, C. (1972). Amino acid sequence of the tryptophan synthetase α chain of *Bacillus subtilis*. *Journal of Bacteriology* **118**, 187–191.

LI, S.-L. AND YANOFSKY, C. (1972). Amino acid sequence of fifty residues from the amino termini of the tryptophan synthetase α chains of several enterobacteria. *Journal of Biological Chemistry* **247**, 1031–1037.

LI, S.-L. AND YANOFSKY, C. (1973). Amino acid sequence studies with the tryptophan synthetase α chain of *Aerobacter aerogenes*. *Journal of Biological Chemistry* **248**, 1837–1843.

LIM, W.K., SMITH-SOMERVILLE, H.E. AND HARDMAN, J.K. (1989). Solubilisation and renaturation of overexpressed aggregates of mutant tryptophan synthase α subunits. *Applied and Environmental Microbiology* **55**, 1106–1111.

MALING, B.D. AND YANOFSKY, C. (1961). The properties of altered proteins from mutants bearing one or two lesions in the same gene. *Proceedings of the National Academy of Sciences, USA* **47**, 551–566.

MALKINSON, A.M. AND HARDMAN, J.K. (1969). Structural and functional roles of the cysteine residues in the α subunit of the *Escherichia coli* tryptophan synthetase. 1. Structural roles and reactivity of the cysteine residues. *Biochemistry* **8**, 2769–2776.

MANCH, J.N. AND CRAWFORD, I.P. (1981). Ordering tryptophan synthase genes of *Pseudomonas aeruginosa* by cloning in *Escherichia coli*. *Journal of Bacteriology* **146**, 102–107.

MANCH, J.N. AND CRAWFORD, I.P. (1982). Genetic evidence for a positive acting regulatory factor mediating induction in the tryptophan pathway of *Pseudomonas aeruginosa*. *Journal of Molecular Biology* **156**, 67–77.

MANNEY, T.R. (1968). Evidence for chain termination by super-suppressible mutants in yeast. *Genetics* **60**, 719–733.

MANNEY, T.R. (1970). Physiological advantage of the mechanism of the tryptophan synthase reaction. *Journal of Bacteriology* **102**, 483–488.

MANSON, M.D. AND YANOFSKY, C. (1976). Naturally occuring sites within the *Shigella dysenteriae* tryptophan operon severely limit tryptophan biosynthesis. *Journal of Bacteriology* **126**, 668–678.

MATCHETT, W.H. AND DEMOSS, J.A. (1975). The subunit structure of tryptophan synthase from *Neurospora crassa*. *Journal of Biological Chemistry* **250**, 2941–2946.

MATSUI, K. AND SANO, K. (1987). Structure and function of the *trp* operon control regions of *Brevibacterium lactofermentum*, a glutamic acid producing bacterium. *Gene* **53**, 191–200.

MATSUI, K., MIWA, K. AND SANO, K. (1987). Cloning of tryptophan genes of *Brevibacterium lactofermentum* a glutamic acid-producing bacterium. *Agricultural and Biological Chemistry* **51**, 823–828.

MATSUI, K., SANO, K. AND OHTSUBO, E. (1986). Complete nucleotide sequence and

deduced amino acid sequences of the *Brevibacterium lactofermentum* tryptophan operon. *Nucleic Acids Research* **14**, 10113–10114.

MATTHEWS, C.R. AND CRISANTI, M.M. (1981). Urea induced unfolding of the α subunit of tryptophan synthase: Evidence for a multistate process. *Biochemistry* **20**, 784–792.

MATTHEWS, C.R., CRISANTI, M.M., MANZ, J.T. AND GEPNER, G.L. (1983). Effect of a single amino acid substitution on the folding of the α subunit of tryptophan synthase. *Biochemistry* **22**, 1445–1452.

MAURER, R. AND CRAWFORD, I.P. (1971a). Properties and subunit structure of the B component of *Pseudomonas putida* tryptophan synthetase. *Archives of Biochemistry and Biophysics* **144**, 193–203.

MAURER, R. AND CRAWFORD, I.P. (1971b). Tryptophan synthetase β_2 subunit. Primary structure of the pyridoxal peptide from the *Pseudomonas putida* enzyme. *Journal of Biological Chemistry* **246**, 6625–6630.

MEADE, H. (1985). Cloning of *ArgG* from *Streptomyces*: loss of gene in *Arg*$^-$ mutants of *S. cattleya*. *Bio/Technology* **3**, 917–918.

MILES, E.W. (1970). The B protein of *Escherichia coli* tryptophan synthetase. I. Effects of sulfhydryl modification on enzymatic activities and subunit interaction. *Journal of Biological Chemistry* **245**, 6016–6025.

MILES, E.W. (1974). Evidence that the essential, photosensitive histidyl residue in the β_2 subunit of tryptophan synthetase is in the pyridoxal peptide. *Biochemical and Biophysical Research Communications* **57**, 849–856.

MILES, E.W. (1979). Tryptophan synthetase. *Advances in Enzymology* **49**, 127–186.

MILES, E.W. (1986). Pyridoxal phosphate enzymes catalyzing β-elimination and β-replacement reactions. In *Vitamin B_6 Pyridoxal Phosphate Part B. Chemical, Biochemical and Medical Aspects*, (D. Dolphin, R. Poulson and O. Avramovic, Eds), pp.253–310. Wiley, New York.

MILES, E.W. AND HIGGINS, W. (1978). An active $\alpha'_2\beta_2$ derivative of tryptophan synthetase formed by limited proteolysis. *Journal of Biological Chemistry* **253**, 6266–6269.

MILES, E.W. AND HIGGINS, W. (1980). Location of the reactive sulfhydryl residues in the primary sequence of tryptophan synthase of *Escherichia coli*. *Biochemical and Biophysical Research Communications* **93**, 1152–1159.

MILES, E.W. AND KUMAGAI, H. (1974). Modification of essential histidyl residues of the β_2 subunit of tryptophan synthetase by photo-oxidation in the presence of pyridoxal 5′-phosphate and L-serine and by diethyl pyrocarbonate. *Journal of Biological Chemistry* **249**, 2843–2851.

MILES, E.W., BAUERLE, R. AND AHMED, S.A. (1987). Tryptophan synthase from *Escherichia coli* and *Salmonella typhimurium*. *Methods in Enzymology* **142**, 398–414.

MILES, E.W., YUTANI, K. AND OGASAHARA, K. (1982). Guanidine hydrochloride induced unfolding of the α subunit of tryptophan synthase and of the two proteolytic fragments: Evidence for stepwise unfolding of the two α domains. *Biochemistry* **21**, 2586–2592.

MILES, E.W., KAWASAKI, H., AHMED, S.A., MORITA, H. AND NAGATA, S. (1989). The β subunit of tryptophan synthase. Clarification of the roles of histidine 86, lysine 87, arginine 148, cysteine 170 and cysteine 230. *Journal of Biological Chemistry* **264**, 6280–6287.

MILTON, D.L., NAPIER, M.L., MYERS, R.M. AND HARDMAN, J.K. (1986). *In vitro* mutagenesis and overexpression of the *Escherichia coli trpA* gene and the partial characterisation of the resultant tryptophan synthase mutant α-subunits. *Journal of Biological Chemistry* **261**, 16604–16615.

MONOD, J., WYMAN, J. AND CHANGEUX, J.P. (1965). On the nature of allosteric transitions: a plausible model. *Journal of Molecular Biology* **12**, 88–118.

MOSTELLER, R.D., GOLDSTEIN, R.V. AND NISHIMOTO, K.R. (1977). Interactions of

tryptophan synthetase subunits in *Escherichia coli* containing mutationally altered β_2 subunits. *Journal of Biological Chemistry* **252**, 4527–4532.

MURGOLA, E.J. (1981). Restricted wobble in UGA codon recognition by glycine tRNA suppressors of UGG. *Journal of Molecular Biology* **149**, 1–13.

MURGOLA, E.J. (1985). tRNA, suppression, and the code. *Annual Review of Genetics* **19**, 57–80.

MURGOLA, E.J. AND HIJAZI, K.A. (1983). Selection for new codons corresponding to position 234 of the tryptophan synthetase alpha chain of *Escherichia coli*. *Molecular and General Genetics* **191**, 132–137.

MURGOLA, E.J. AND YANOFSKY, C. (1974a). Selection for new amino acids at position 211 of the tryptophan synthetase α chain of *Escherichia coli*. *Journal of Molecular Biology* **86**, 775–784.

MURGOLA, E.J. AND YANOFSKY, C. (1974b). Structural interactions between amino acid residues at positions 22 and 211 in the tryptophan synthetase alpha chain of *Escherichia coli*. *Journal of Bacteriology* **117**, 444–448.

MURGOLA, E.J. AND YANOFSKY, C. (1974c). Suppression of glutamic acid codons by mutant glycine transfer ribonucleic acid. *Journal of Bacteriology* **117**, 439–443.

MURPHY, T.M. AND MILLS, S.E. (1968). Immunochemical comparisons of mutant and wild-type α-subunits of tryptophan synthetase. *Archives of Biochemistry and Biophysics* **127**, 7–16.

MURPHY, T.M. AND MILLS, S.E. (1969). Immunochemical and enzymatic comparisons of the tryptophan synthase α subunits from five species of *Enterobacteriaceae*. *Journal of Bacteriology* **97**, 1310–1320.

MURRY-BRELIER, A. AND GOLDBERG, M.E. (1988). Kinetics of appearance of an early immunoreactive species during the refolding of acid denatured *Escherichia coli* tryptophan synthase β_2 subunit. *Biochemistry* **27**, 7633–7640.

MYERS, R.M., LERMAN, L.S. AND MANIATIS, T. (1985). A general method for saturation mutagenesis of cloned DNA fragments. *Science* **229**, 242–247.

NAGATA, S., HYDE, C.C. AND MILES, E.W. (1989). The α subunit of tryptophan synthase. Evidence that aspartic acid 60 is a catalytic residue and that the double alteration of residues 175 and 211 in a second-site revertant restores the proper geometry of the substrate binding site. *Journal of Biological Chemistry* **264**, 6288–6296.

NATORI, Y., KANO, Y. AND IMAMOTO, F. (1990). Nucleotide sequences and genomic constitution of five tryptophan genes of *Lactobacillus casei*. *Journal of Biochemistry* **107**, 248–255.

NICHOLS, B.P., BLUMENBERG, M. AND YANOFSKY, C. (1981). Comparison of the nucleotide sequence of *trpA* and the sequences immediately beyond the *trp* operon of *Klebsiella aerogenes*, *Salmonella typhimurium*, and *Escherichia coli*. *Nucleic Acids Research* **9**, 1745–1755.

NORMARK, S., BERGSTRÖM, S., EDLUND, T., GRUNDSTRÖM, T., JAURIN, B., LINDBERG, F.P. AND OLSEN , O. (1983). Overlapping genes. *Annual Review of Genetics* **17**, 499–525.

OPPENHEIM, D.S. AND YANOFSKY, C. (1980). Translational coupling during expression of the tryptophan operon of *E. coli*. *Genetics* **95**, 785–795.

OUCHTERLONY, O. (1949). Antigen–antibody reactions in gels. *Archiv för Kemi, Mineralogi och Geologi* **B26**, 1–9.

PARSOT, C. (1987). A common origin for enzymes involved in the terminal step of the threonine and tryptophan biosynthetic pathways. *Proceedings of the National Academy of Sciences, USA* **84**, 5207–5210.

REYES, G.R. AND ROCHA, V. (1977). Immunochemical comparison of phosphoribosylanthranilate isomerase-indoleglycerol phosphate synthetase among the *Enterobacteriaceae*. *Journal of Bacteriology* **129**, 1448–1456.

RIVAS, M.V., JARVIS, E.D. AND RUDNER, R. (1990). The structure of the *trpE*, *trpD* and 5′*trpC* genes of *Bacillus pumilis*. *Gene* **87**, 71–78.

RIVERIN, M. AND DRAPEAU, G.R. (1976). Purification and properties of the $\alpha_2\beta_2$

complex of tryptophan synthetase of *Proteus mirabilis*. *Journal of Biological Chemistry* **251**, 3875–3880.

RIVERO-LEZCANO, O., ANGUITA-CASILLO, J., LOPEZ-NIETO, J. AND NAHARRO-CARRASCO, G. (1990). Cloning and expression in *Escherichia coli* of tryptophan genes from *Streptomyces griseus* IMRU3570. *FEMS Microbiology Letters* **68**, 201–206.

ROCHA, V., CRAWFORD, I.P. AND MILLS, S.E. (1972). Comparative immunological and enzymatic study of the tryptophan synthetase β_2 subunit in the *Enterobacteriaceae*. *Journal of Bacteriology* **111**, 163–168.

ROCHA, V., DEELEY, M. AND CRAWFORD, I.P. (1979). Conservation of primary structure of the pyridoxal peptide of *Escherichia coli* and *Serratia marcescens* tryptophan synthase β_2 protein. *Journal of Bacteriology* **137**, 700–703.

ROSS, C.M. AND WINKLER, M.E. (1988a). Structure of the *Caulobacter crescentus trpFBA* operon. *Journal of Bacteriology* **170**, 757–768.

ROSS, C.M. AND WINKLER, M.E. (1988b). Regulation of tryptophan biosynthesis in *Caulobacter crescentus*. *Journal of Bacteriology* **170**, 769–774.

SAKAGUCHI, K. (1970). The similarity of tryptophan synthetases of *Anabaena variabilis* and *Chlorella ellipsoidae* with that of bacteria. *Biochimica et Biophysica Acta* **220**, 580–593.

SATO, S., NAKADA, Y., SHIGENORI, K. AND TANAKA, T. (1988). Molecular cloning and nucleotide sequence of *Thermus thermophilus* HB8 *trpE* and *trpG*. *Biochimica et Biophysica Acta* **950**, 307–312.

SATO, S., NAKADA, Y., HON-NAMI, K., YASUI, K. AND AKIKO, S. (1989). Molecular cloning and the nucleotide sequence of *Clostridium thermocellum trpE* gene. *Journal of Biochemistry* **105**, 362–366.

SAWULA, R.V. AND CRAWFORD, I.P. (1972). Mapping of the tryptophan genes of *Acinetobacter calcoaceticus* by transformation. *Journal of Bacteriology* **112**, 797–805.

SCHNEIDER, W.P., NICHOLS, B.P. AND YANOFSKY, C. (1981). Procedure for production of hybrid genes and proteins and its use in assessing significance of amino acid differences in homologous tryptophan synthetase α polypeptides. *Proceedings of the National Academy of Sciences, USA* **78**, 2169–2173.

SHERRIFF, S., SILVERTON, E.W., PADLAN, E.A., COHEN, G.H., SMITH-GILL, S.J., FINZEL, B.C. AND DAVIES, D.R. (1987). Three-dimensional structure of an antibody–antigen complex. *Proceedings of the National Academy of Sciences, USA* **84**, 8075–8079.

SHIMOTSU, H., KURODA, M.I., YANOFSKY, C. AND HENNER, D.J. (1986). Novel form of transcription attenuation regulates expression of the *Bacillus subtilis* tryptophan operon. *Journal of Bacteriology* **166**, 461–471.

SHINE, J. AND DALGARNO, L. (1974). The 3′ terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense tripets and ribosome binding sites. *Proceedings of the National Academy of Sciences, USA* **71**, 1342–1346.

SIBOLD, L. AND HENRIQUET, M. (1988). Cloning of the *trp* genes from archaebacterium *Methanococcus voltae*: Nucleotide sequence of the *trpBA* genes. *Molecular and General Genetics* **214**, 439–450.

SKRZYNIA, C., BINNINGER, D.M., ALSPAUGH, J.A. AND PUKKILA, P.J. (1989). Molecular characterisation of *TRP1*, a gene coding for tryptophan synthetase in the basidiomycete *Coprinus cinereus*. *Gene* **81**, 73–82.

SMITHERS, C.M. AND ENGEL, P.P. (1974). Gene–enzyme relations of tryptophan mutants in *Streptomyces coelicolor*. *Genetics* **78**, 799–808.

SOMERVILLE, R.L. (1988). The *trp* promoter of *Escherichia coli* and its exploitation in the design of efficient protein purification systems. *Biotechnology and Genetic Engineering Reviews* **6**, 1–41.

TANIZAWA, K. AND MILES, E.W. (1983). L-serine binds to arginine 148 of the β_2 subunit of *Escherichia coli* tryptophan synthase (EC 4.2.1.20). *Biochemistry* **22**, 3594–3603.

TSCHOPP, J. AND KIRSCHNER, K. (1980). Kinetics of cooperative ligand binding to the apoβ_2 subunit of tryptophan synthase and its modulation by the α subunit. *Biochemistry* **19**, 4521–4527.

TSOU, C.-L. (1988). Folding of the nascent peptide chain into a biologically active protein. *Biochemistry* **27**, 1813–1817.

TUCKER, S.D., MURGOLA, E.J. AND HIJAZI, K.A. (1989). Reversion of *trpA* nonsense mutations by deletion of the chain termination codons. *Biochimie* **71**, 721–728.

TURNER, V.R. AND MATCHETT, W.H. (1968). Alteration of tryptophan-mediated regulation in *Neurospora crassa* by indoleglycerol phosphate. *Journal of Biological Sciences* **258**, 3571–3575.

TWEEDY, N.B., HURLE, M.R., CHRUNYK, B.A. AND MATTHEWS, C.R. (1990). Multiple replacements at position 211 in the α subunit of tryptophan synthase as a probe of the folding unit association reaction. *Biochemistry* **29**, 1539–1545.

VALLIER, H. AND WELKER, N.E. (1990). Genetic map of the *Bacillus stearothermophilus* NUB36 chromosome. *Journal of Bacteriology* **172**, 793–801.

VOLLMER, S.J. AND YANOFSKY, C. (1986). Efficient cloning of genes of *Neurospora crassa*. *Proceedings of the National Academy of Sciences, USA* **83**, 4869–4873.

WASSERMAN, E. AND LEVINE, L. (1961). Quantitative microcomplement fixation and its use in the study of antigenic structure by specific antigen–antibody inhibition. *Journal of Immunology* **87**, 290–295.

WEGMAN, J. AND CRAWFORD, I.P. (1968). Tryptophan synthetic pathway and its regulation in *Chromobacterium violaceum*. *Journal of Bacteriology* **95**, 2325–2335.

WIESINGER, H., BARTHOLMES, P. AND HINZ, H.J. (1979). Subunit interaction in tryptophan synthase of *Escherichia coli*: Calorimetric studies on association of α and β_2 subunits. *Biochemistry* **18**, 1979–1984.

WILHELM, P., PILZ, I., LANE, A.N. AND KIRSCHNER, K. (1982). Small-angle X-ray scattering of tryptophan synthase from *Escherichia coli* and its α and β_2 subunits. *European Journal of Biochemistry* **129**, 51–56.

WINKLER, M.E., SCHOENLEIN, P.V., ROSS, C.M., BARRETT, J.T. AND ELY, B. (1984). Genetic and physical analyses of *Caulobacter crescentus trp* genes. *Journal of Bacteriology* **160**, 279–287.

WOESE, C.R. (1987). Bacterial Evolution. *Microbiological Reviews* **51**, 221–271.

YANOFSKY, C. (1952). Tryptophan desmolase of *Neurospora* – Partial purification and properties. *Journal of Biological Chemistry* **194**, 279–286.

YANOFSKY, C. (1967). Gene structure and protein structure. *The Harvey Lectures* **61**, 145–168.

YANOFSKY, C. (1981). Attenuation in the control of expression of bacterial operons. *Nature* **289**, 751–758.

YANOFSKY, C. (1984). Comparison of regulatory and structural regions of genes of tryptophan metabolism. *Molecular Biology and Evolution* **1**, 143–163.

YANOFSKY, C. (1987). Tryptophan synthetase: Its charmed history. *Bioessays* **6**, 133–137.

YANOFSKY, C. AND BONNER, D.M. (1955). Gene interaction in tryptophan synthetase formation. *Genetics* **40**, 761–769.

YANOFSKY, C. AND CRAWFORD, I.P. (1959). The effects of deletions, point mutations, reversion and suppressor mutations on the two components of the tryptophan synthetase of *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA* **45**, 1016–1026.

YANOFSKY, C. AND CRAWFORD, I.P. (1972). Tryptophan synthetase. In *The Enzymes* (P.D. Boyer, Ed.), 3rd edition, volume 7, pp.1–31. Academic Press, Orlando.

YANOFSKY, C. AND HORN, V. (1972). Tryptophan synthetase α chain positions affected by mutations near ends of the genetic map of *trpA* of *Escherichia coli*. *Journal of Biological Chemistry* **247**, 4494–4498.

YANOFSKY, C., HELINSKI, D.R. AND MALING, B.D. (1961). The effects of mutation

on the composition and properties of the A protein of *Escherichia coli* tryptophan synthetase. *Cold Spring Harbor Symposia on Quantitative Biology* **26**, 11–24.

YANOFSKY, C., HORN, V. AND THORPE, D. (1964). Protein structure relationships revealed by mutational analysis. *Science* **146**, 1593–1594.

YANOFSKY, C., ITO, J. AND HORN, V. (1966). Amino acid replacements and the genetic code. *Cold Spring Harbor Symposia on Quantitative Biology* **31**, 151–162.

YANOFSKY, C., PLATT, T., CRAWFORD, I.P., NICHOLS, B.P., CHRISTIE, G.E., HOROWITZ, H., VAN CLEEMPUT, M. AND WU, A.M. (1981). The complete nucleotide sequence of the tryptophan operon of *Escherichia coli. Nucleic Acids Research* **9**, 6647–6668.

YANOFSKY, C., PALUH, J.L., VAN CLEEMPUT, M. AND HORN, V. (1987). Fusion of *trpB* and *trpA* of *Escherichia coli* yields a partially active tryptophan synthetase polypeptide. *Journal of Biological Chemistry* **262**, 11584–11590.

YORK, S.S. (1972). Kinetic spectroscopic studies of substrate and subunit interactions of tryptophan synthetase. *Biochemistry* **11**, 2733–2740.

YUTANI, K., OGASAHARA, K., AOKI, K., KAKUNO, T. AND SUGINO, Y. (1984). Effect of amino acid residues on conformational stability in eight mutant proteins variously substituted at a unique position of the tryptophan synthase α-subunit. *Journal of Biological Chemistry* **259**, 14076–14081.

YUTANI, K., OGASAHARA, K., TSUJITA, T., KANEMUTO, K., MATSUMOTO, M., TANAKA, S., MIYASHITA, T., MATSUSHIRA, A., SUGINO, Y. AND MILES, E.W. (1987). Tryptophan synthase α subunit glutamic acid 49 is essential for activity. *Journal of Biological Chemistry* **262**, 13429–13433.

ZAKIN, M.M., BOULOT, G. AND GOLDBERG, M.E. (1980). Immunochemical study of the β-chain of *Escherichia coli* tryptophan synthase and its proteolytic fragments. *European Journal of Immunology* **10**, 16–21.

ZALKIN, H. AND YANOFSKY, C. (1982). Yeast gene *TRP5*: Structure, function, regulation. *Journal of Biological Chemistry* **257**, 1491–1500.

ZETINA, C.R. AND GOLDBERG, M.E. (1980). Reversible unfolding of the β_2 subunit of *Escherichia coli* tryptophan synthetase and its proteolytic fragments. *Journal of Molecular Biology* **137**, 401–414.

ZETINA, C.R. AND GOLDBERG, M.E. (1982). Kinetics of renaturation and self-assembly of intermediates on the pathway of folding of the β_2 subunit of *Escherichia coli* tryptophan synthetase. *Journal of Molecular Biology* **157**, 133–148.

ZHAO, G.-P. AND SOMERVILLE, R.L. (1987). Genetic, physiological and biochemical analysis of a missense mutation in the hinge region of the tryptophan synthase β_2 subunit. *Federation Proceedings* **46**, 1938.