Bioorganic & Medicinal Chemistry, Vol. 1, No. 1, pp. 1-17, 1993
Printed in Great Britain

0968-0896/93 $6.00 + .00

REVIEW ARTICLE

The Enzymes Involved in Biosynthesis of Penicillin and Cephalosporin; Their Structure and Function

R. D. G. Cooper
Lilly Research Laboratories, Eli Lilly & Co., Indianapolis, IN 46285

Abstract—The biosynthetic pathway resulting in the penicillins and cephalosporins contains two Fe^{2+} oxidase enzymes which are responsible for the conversion of α-aminoadipoyl-L-cysteinyl-D-valine into isopenicillin N and penicillin N into deacetoxycephalosporin C. We will discuss the studies delineating the ligand binding of these enzymes and present a possible secondary structure.

The true genesis of this project goes back over 300 million years, for it was in that evolutionary timeframe [the Palaeozoic era] that the genes responsible for the biosynthesis of penicillins and cephalosporins first became assembled in the filamentous fungi. Nature continued the synthesis of these secondary metabolites undisturbed until Fleming's discovery of penicillin in 1928 which initiated one of the most remarkable sagas in medical history. In the U.S.A. it began on a Saturday afternoon in 1942 when Anne Miller received one of the first injections of this new miracle drug. By Sunday, her temperature had fallen from 41°C to normal and her streptococcal infection had subsided.[1] Although the initial discovery of penicillin happened in 1928 it was not until the mid 40s and early 50s that the structure of penicillin was solved at Oxford University and other laboratories.[2] A few years later Prof. Edward Abraham at Oxford discovered the related structure cephalosporin C. These initial discoveries have resulted in the development of many new and effective antibiotics, resulting in the great advances in the clinical management of infectious disease that have occurred over the intervening five decades.

In 1985 a group of scientists at Lilly combined resources with the research groups of Prof. J. Baldwin and Sir Edward Abraham at Oxford University in an attempt to determine how nature was able to construct such a series of structurally unusual secondary metabolites. Perhaps the solution to this puzzle could give valuable lessons to the synthetic organic chemist, perhaps we could also make use of these genes in a very practical manner to modify the pathway to produce different products or to increase the efficiency of the penicillin and cephalosporin fermentations. This review is an overview of the progress made at the Lilly Research Laboratories during the last five years.

What was known about the biosynthetic pathway when we initiated this programme is shown in Fig. 1.[3]

The first enzyme, ACV synthetase, assembles three separate amino acids into the tripeptide ACV.[4] The second enzyme, isopenicillin N synthase [IPNS] is a non-heme iron oxidase[5] which efficiently constructs the 4,5 fused ring system of isopenicillin N.

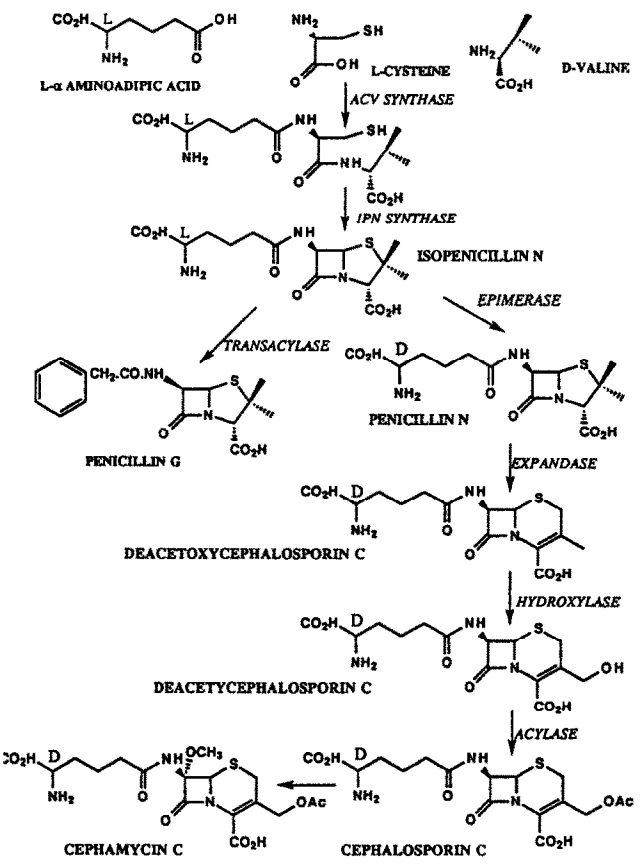


Figure 1 Biosynthetic scheme for penicillin and cephalosporin.

In *C. acremonium* and in the *Streptomyces sp.*, the next enzyme in the sequence is an epimerase,[6] which converts the L-α-aminoadipic acid side chain to the D-isomer. This new compound, penicillin N, is now a substrate for the next enzyme in the pathway, expandase, which modifies the 5-membered ring system of penicillin N to the 6-

membered ring system of deacetoxycephalosporin C.[7] A subsequent second step hydroxylates the allylic methyl group resulting in deacetylcephalosporin C. These three oxidative enzymes, IPNS, expandase, and hydroxylase are the most mechanistically interesting enzymes in this pathway. In *Penicillium chrysogenum* and *Aspergillus nidulans* the pathway diverges from that which produces cephalosporin C, for these fungi contain a transacylase enzyme which removes the L-α-aminoadipic acid from isopenicillin N to give either the nucleus 6-APA or an acylated derivative of the nucleus depending on whether a suitable carboxylic acid is added to the fermentation.

The two general approaches to the isolation of the genes responsible for a biosynthetic pathway are the use either of reverse genetics or of complementation cloning. The first method which involves the isolation of the protein that is responsible for a particular function in the pathway was the method which was used to dissect the cephalosporin pathway. Six years ago a small amount of a protein was isolated from the cephalosporin fermentation by Prof. Abraham. This protein was responsible for the IPNS activity and produced an N-terminal amino acid sequence of 23 amino acids.[8] After constructing appropriate synthetic oligomers and probing the gene library constructed from *C.acremonium,* Samson *et al.* identified a section of DNA to which this N-terminal sequence corresponded. After extensive DNA sequencing both upstream and downstream from this original sequence, the open reading frame responsible for this gene product was identified using the Bibb algorithm.[9] The GC content in the third position of the triplet codons is significantly higher in an ORF compared to the other two positions and it is this increase in GC content which readily identifies the position of an ORF. Continuation of DNA sequencing in both directions can also result in the identification of further open reading frames. Although this does not give any information as to the function of the gene products of these ORFs, this can be a valuable method in obtaining other genes in the pathway when the biosynthetic genes are closely clustered, especially when the proteins are highly unstable or difficult to purify. Lilly scientists have located other genes in this pathway by exactly this method. Using the IPNS gene of *C. acremonium* as a heterologous hybridization probe we then investigated *Penicillium chrysogenum*,[10,11] *Aspergillus nidulans*,[12,13] and using synthetic oligomers, *Streptomyces lipmanii*,[12] a prokaryotic source of the beta-lactam antibiotics, was investigated. Other workers using this information have more recently obtained the sequences of IPNS from *Streptomyces clavuligerus*,[14] *Streptomyces griseus*,[15] *Streptomyces jumojinensis*,[16] *Nocardia lactamdurans*,[17] and the gram negative bacteria *Flavobacterium sp.*[18] Thus there are now nine separate sequences available for a protein having this same function. These sequences show a reasonably high degree of homology, the discussion of which will be later in this review. IPNS is a non-heme containing ferrous iron oxidase, having a requirement for ferrous iron, dioxygen, ascorbic acid and the substrate isopenicillin N. Because of this requirement for ferrous iron and the observation that all the sequences so far obtained had two conserved cysteine residues, it was speculated that the thiol groups of these cysteines could be involved in the binding of the protein to the metal centre.

Table 1. Kinetic parameters of IPNS and IPNS mutants

Enzymes	Vmax *unit/mg*	kcat *s-1*	Km *mM*	kcat/Km *s-1M-1x10-3*
Wild Types				
Natural	5.0	3.2	0.37	8.6
Recombinant	7.2	4.6	0.56	8.2
Mutants				
Cys 255-Ser	4.8	3.1	0 78	4.0
Cys 106-Ser	2.7	1.7	7.80	0.2
Cys 106,255-Ser	2.7	1.7	7.82	0.2

Using site specific mutagenesis Samson *et al.* proceeded to systematically change individually the cysteines to serines and both cysteines to serines.[19] The results of assaying the three resulting enzymes for their ability to synthesize isopenicillin N from the ACV tripeptide is shown in Table 1.[20] All the derivative proteins are functional, converting the tripeptide into isopenicillin N, showing that neither cysteine is an absolute requirement for catalytic activity. From examination of the kinetics, determined by initial rate measurements, it was found that changing the cysteines causes a small decrease in V_{max} [33% for Cys.–Ser.255 and 63% for Cys.–Ser.106], and the enzymes with mutations at cysteine 106 with or without changes at 255 show an elevation in K_m by 14 fold and a reduction in k_{cat}/K_m of 40 fold. Further work on the delineation of the role of the two cysteines involved reaction with cysteine specific reagents.[20] It was found that the intact protein was resistant to modification with iodoacetic acid, reaction occurring only when the protein was unfolded, indicating that the cysteines were located internally. Use of the alternate cysteine modifying reagent *N*-ethylmaleimide succeeded in modifying the thiol groups in the folded protein, however this reagent preferentially reacted with Cys.-106. These results are consistent with Cys.-106 being involved in substrate binding and being more available and that both cysteines are located internally, with Cys.-255 possibly only involved in the internal architecture of the protein.

To probe the spacial and binding characteristics of the active site of this enzyme both Oxford chemists and chemists at the Lilly Research Laboratories have synthesized many substrate analogues.[21,22] The types of structural variations synthesized are shown in Fig. 2. From the ability of some of these substrates to serve as β-lactam precursors it was possible to reach the following conclusions concerning the size, shape and binding site requirements of the active site.

(a) Concerning the side chain; the carboxyl group, its configuration and its specific distance from the cysteine thiol were required, however the amino group was not critical.

(b) Methyl groups could be accommodated at the alpha and beta proR positions of the cysteine.

(c) The beta proS proton and the thiol group were essential.

(d) In the valine part of the substrate, variation in the methyl groups could be accepted as long as their steric requirement was small. It is highly likely then that there are critical basic amino acids i.e. lysine and arginine, in the enzymes that are responsible for binding of the substrate carboxylic acid groups.

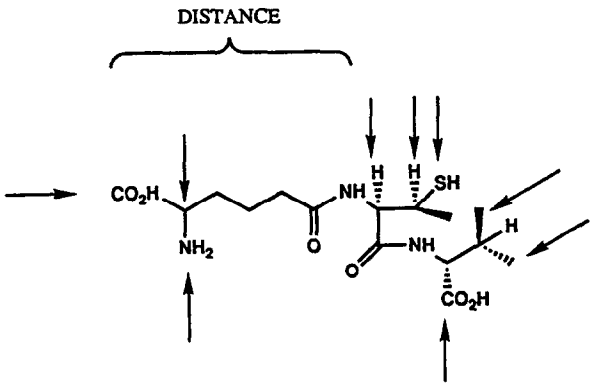


Figure 2 Structural probes of active site of IPNS. Variations on L-α-amino-adipoyl-L-cysteinyl-D-valine.

We then turned our attention to the structure of the metal centre in the enzyme. Both the wild type enzyme and the recombinant mutants were found to require one equivalent of ferrous iron for complete activity[19,23] supporting the contention that there was one active site containing one metal atom [Fig. 3]. Oxygen is a diradical, and on entering the electrostatic field of the ferrous centre undergoes a rehybridization, the degeneracy of the $1\pi g$ level is removed and both oxygen atoms of the oxygen molecule become sp^2 hybridized. One of the sp^2 hybrid orbitals of each atom participates in mutual σ bonding. Dioxygen binding to Fe^{2+} is caused by σ interaction of an sp^2 lone pair with a z orbital of the d^2sp^3 hybridized metal atom. Retrodative π bonding from iron to oxygen results in a net negative charge at the ligand. This is reported to result in an end on ligand binding with an Fe–O=O angle of 120–135°.[24,25] The degree of electron transfer from the iron to oxygen is subtly controlled by the ligands attached to the metal and the extent of transfer is one factor contributing to whether the enzyme functions as a reversible oxygen binder or as an oxidase. To further understand the structure of the active site Chen and his co-workers used Mossbauer spectroscopy [in collaboration with Prof. Munck, Carnegie-Mellon University] and electron spin resonance spectroscopy [in collaboration with Prof. Lipscolm, University of Minnesota]. To observe a Mossbauer spectrum it was first necessary to reconstitute the apoenzyme with the Fe 57 isotope. In Fig. 4 the top curve shows the spectrum of the active enzyme without substrate, the second curve with substrate, and the bottom curve is the resultant curve obtained by subtraction of the two previous curves. This latter curve [curve C in Fig. 4] then represents the spectrum of the enzyme containing substrate. The curve shows a single doublet having a splitting of ΔE_q=2.7 mm/s, and isomer shift, δ=1.30.[23]

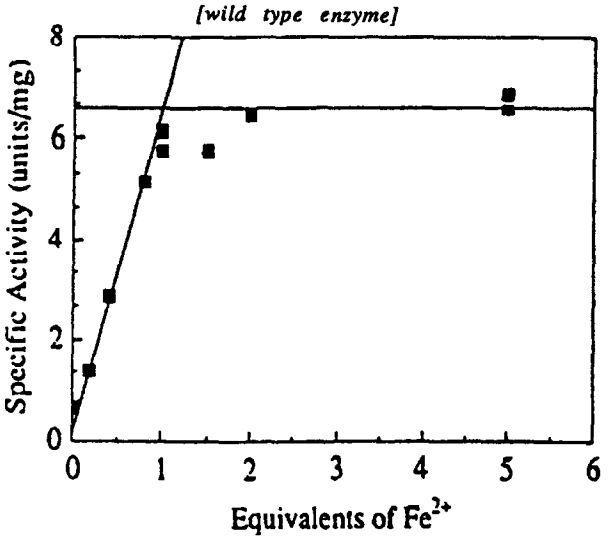


Figure 3 Iron stoichiometry of IPNS.

These results are typical for a high spin Fe^{2+} and are suggestive of a penta- or hexacoordinate sphere of nitrogen and/or oxygen surrounding the iron. The change upon addition of the substrate showed that the substrate is perturbing the environment of the iron. The isomer shift is a measure of the d-electron density at the iron nucleus which decreases on delocalization. The change in the value of δ is explained by assuming that the thiolate group of the substrate i.e. the cysteine SH group has entered the coordinate environment of the iron and is therefore a ligand of the metal centre.

High spin ferrous iron is typically EPR silent and no signal is observed for either the enzyme [containing ferrous iron] with or without substrate. An EPR signal appears on the addition of nitric oxide which complexes with the iron. The EPR spectrum obtained by this technique is shown in Fig. 5 [in collaboration with Prof. J. D. Lipscolm, University of Minnesota]. The top spectrum shows the result of Fe^{2+}–IPNS under anaerobic conditions on the addition of nitric oxide. The addition of ACV substrate causes the changes shown in the bottom spectrum in Fig. 5. This latter curve could be obtained independent of the order of addition of the ACV and NO. If this experiment is done in the presence of O^{17} water there is a resultant hyperfine broadening of both the peaks at 4.22 and 3.81. This perturbation suggests that the substrate binds to the metal centre and as the resulting spectrum can be obtained irrespective of the order of addition then the binding of NO and ACV is to different binding sites. The hyperfine line broadening indicates that water is also a ligand for the metal centre. The addition of NO to the anaerobic IPNS–ACV complex resulted in new absorptions in the visible spectrum at 508 and 720 nm[23] [Fig. 6]. These peaks are also obtained independent of the order of addition. This is indicative of a charge transfer complex between the ligand and the metal. On repeating this experiment with the Cys.–Ser. mutants previously described, the same changes

to the visible spectrum were observed, suggesting that the cysteines contained in the native protein were not involved in binding to the iron.[26]

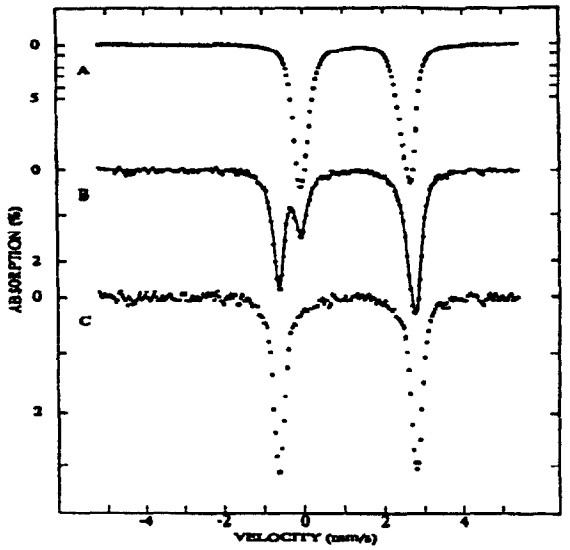


Figure 4 Anaerobic Mossbauer spectroscopy of IPNS–ACV complex.

It is possible to replace the iron metal centre with other metals, Cu^{2+} and Co^{2+}.[27] Although neither of these protein derivatives exhibit any function, the natural enzymic activity was recovered from both of these proteins by their subsequent exposure to excess ferrous iron, thus indicating that in replacing the metal no irreversible structural changes to the protein occurred. It appears that IPNS is capable of reversibly binding the surrogate metals without having the capability to function as an oxidase.

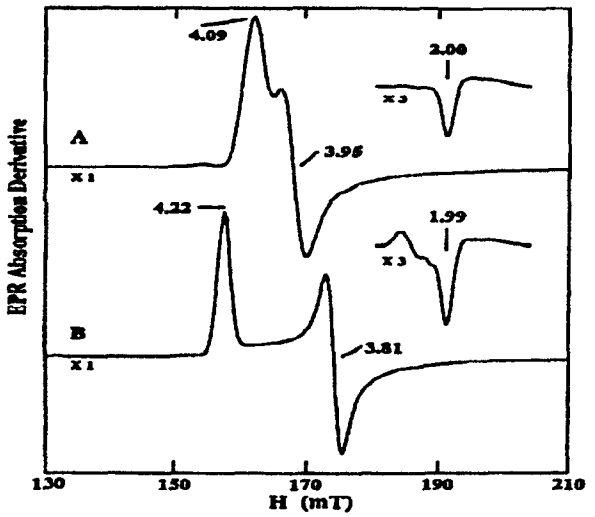


Figure 5 Electron spin resonance spectra of IPNS–NO complex.

The EPR spectrum of the Cu–IPNS complex showed a significant change on the addition of the ACV substrate [Fig. 7]. These effects were consistent with thiolate binding of the ACV to the Cu^{2+}, in an analogous manner to that of thiolate binding to other type 2 copper proteins. The EPR parameters also suggested that the ACV binding was at an equatorial site.

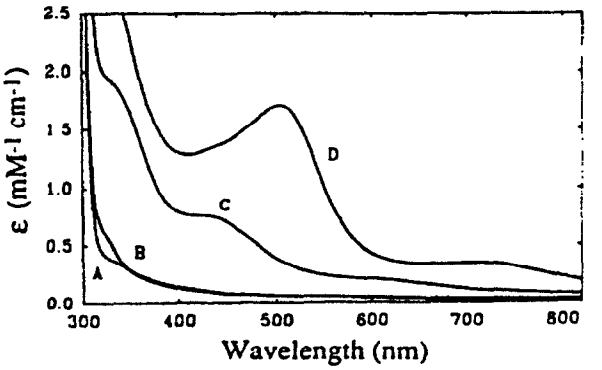


Figure 6 Optical spectra of IPNS–ACV–NO complexes [where A=enzyme alone; B=enzyme+ACV; C=enzyme+ACV+NO; D=enzyme+ACV+NO [under conditions of saturation]].

Significant spectral changes on addition of ACV were observed in the UV spectrum. There was a shift of the d–d transition to 664 nm and the appearance of an intense new peak at 385 nm, the latter was characteristic of a thiol–Cu charge transfer interaction [Fig. 8].

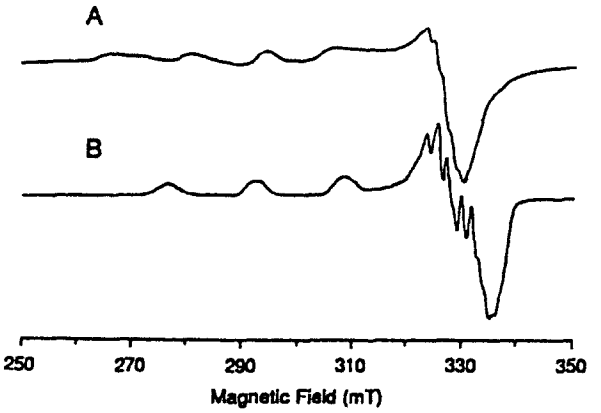


Figure 7 ESR spectrum of Cu–IPNS complex [where A=Cu+IPNS; B=Cu+IPNS+ACV].

The NMR spectrum of Fe–IPNS [in collaboration with Prof. L. Que, University of Minnesota] shows a peak at 65ppm(3H) which washes out with D_2O.[27] This peak is split on addition of ACV into two peaks at 66ppm and 56ppm in the ratio of 2:1. Based on the chemical shifts of Fe^{2+}–imidazole complexes these two peaks could be assigned to the NH protons of iron ligated histidines existing in two different environments. The cobalt containing enzyme showed a very similar effect.

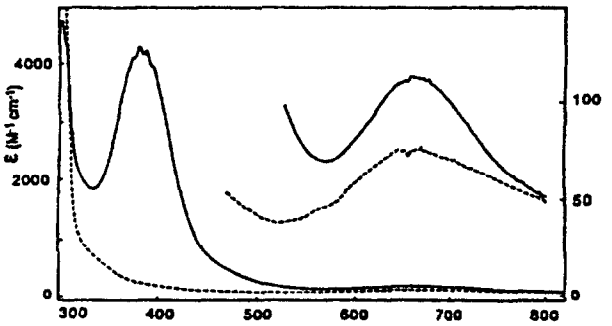


Figure 8 UV spectrum of Cu–IPNS complex [where dotted line=Cu+IPNS; solid line=Cu+IPNS+ACV].

These results indicated that three histidines are involved in the binding site of IPNS and that addition of substrate substantially affects the metal site. Further NMR studies[28] on the Co complex has identified signals which based on NOE studies were attributed to the presence of a coordinated aspartyl residue. Although equivalent signals were observed in the Fe complex they were insufficiently resolved to conduct the NOE analysis. Furthermore, on binding of NO these signals disappeared, indicating either a change in configuration of the aspartyl carboxyl configuration or displacement from the metal by the NO. The imidazole NH resonance of one of the histidines also disappeared suggesting that this ligand may become detached from the active site. Thus it appears that some of the ligands involved in binding to the metal site could vary during the catalytic process. Other studies using electron spin echo modulation on the Cu complex[29] again indicated the presence of two equatorially coordinated histidines together with a water molecule which is displaced from the metal by binding of ACV.

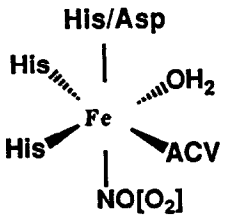


Figure 9 Model for the active site of IPNS.

Thus, it would appear that the ligand environment surrounding the iron consists of three histidines [or two histidines and an aspartic acid], two in equatorial sites and one in an axial environment, the substrate thiol group, water, and oxygen [Fig.9].

The next enzyme in this biosynthetic cascade that we will be discussing is expandase/hydroxylase. This also requires ferrous iron and oxygen, but has the distinction from IPNS in that α-ketoglutarate is an additional required cofactor. This enzyme requires penicillin N as substrate and converts it into deacetylcephalosporin C [see Fig. 1]. All attempts to separate the two functions of expandase and hydroxylase into two separate proteins in *C. acremonium* were unsuccessful,[30] however separation of activities could be accomplished in *S. clavuligerus*.[31] This enzyme having dual activities was purified by Yeh *et al.* from *C. acremonium*.[32] Attempts to obtain N-terminal amino acid sequence from this protein were unsuccessful, apparently the protein was blocked at the N-terminus. Tryptic digestion of the enzyme, separation of the fragments, and sequencing some of these fragments gave several internal sequences of the protein. Construction of appropriate DNA probes and probing a DNA library prepared from *C. acremonium* identified a fragment containing the open reading frame for this protein from which the amino acid sequence for expandase/hydroxylase was obtained. Cloning and expression of this enzyme in *E. coli* by Yeh *et al.* produced a protein which on assay was found indeed to contain both activities, thus proving for the first time that this enzyme was bifunctional. The *S. clavuligerus* enzymes, both expandase and hydroxylase were isolated by Miller and co-workers in a similar manner and the amino acid sequence of both of these enzymes obtained. They were then cloned and expressed in *E. coli.* [33,34] These three enzymes were very similar, they were all of approximately the same size, and had high homology in their amino acid sequences. A comparison of their properties is shown in Table 2.[35,36]

Table 2. Catalytic and kinetic properties of *C. acremonium* expandase/hydroxylase *S. clavuligerus* expandase and hydroxylase

Property	*C. acremonium*		*S. clavuligerus*	
	Expandase/Hydroxylase		Expandase	Hydroxylase
Optimal pH	7.5-7.8	7.0-7.5	7.4	7.0-7.4
Optimal Temperature (°C)	26-34	34-38	36	28-31
Requirement for α-KG, Fe^{2+} and O_2	Yes	Yes	Yes	Yes
Stimulation by				
DIT	Great	Great	Great	Great
Ascorbate	Great	Great	Great	Slight
ATP	Moderate	Slight	None	Moderate
Inhibition by r-HMB, DTNB or NEM	Strong	Strong	Strong	Strong
K_m (μM) for				
Penicillin N	29	--[b]	35	--
DAOC	--	18	--	25
α-KG	22	20	22	14
K_s (μM) for Fe^{2+}	6	13	4	30
V_{max} (μmol/min/mg protein)	0.4-0.8	0.1-0.3	0.4	0.5

[a]From native or/and recombinant enzymes.
[b]Not determined.

It was observed that there is no appreciable difference in the binding of the expandase substrate [penicillin N] or the hydroxylase substrate [deacetoxycephalosporin C] with either the prokaryotic enzymes or the eukaryotic enzyme. The values for the cofactors, iron and α-ketoglutarate, were also similar. When the pure enzymes became available from the cloning of the genes into *E. coli* ,[37,38] it was discovered that all of these three enzymes were in fact bifunctional, they just differed in their relative rates. Shown in Fig. 10 is a comparison of the data for the primary activity for each enzyme, the primary activity being given the arbitrary value of 100%. It also shows another interesting substrate, the exomethylene-cephalosporin C. This double bond isomer of deacetoxy-cephalosporin C which had previously been found to be an effective substrate for expandase/hydroxylase[39] can now be seen to be a substrate of all these enzymes, although a much more effective one of the hydroxylase.[40]

The data shown in Fig. 10 indicate that in all of these enzymes we are dealing with only one active site. As the sequences possess good homology to each other, there must be subtle changes in the tertiary structure which are controlling the relative rates of the two separate activities of these enzymes.

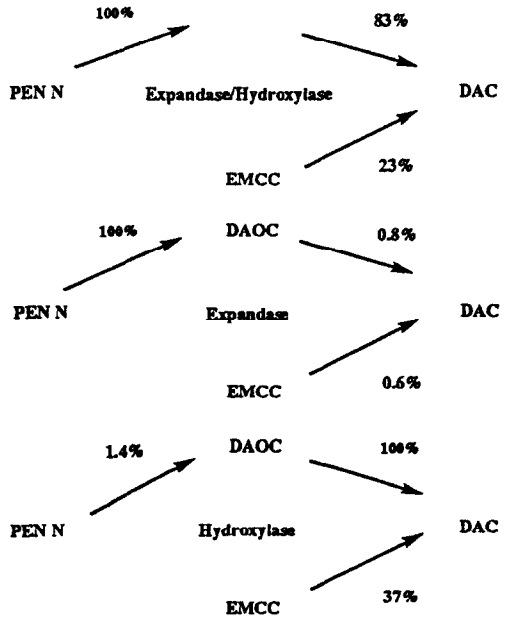
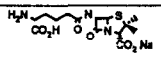
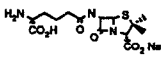
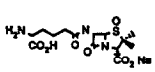
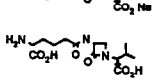
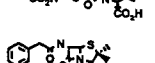
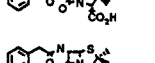
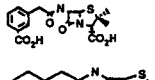
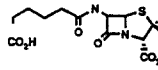
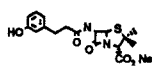


Figure 10 Relative reactivities of pen N with expandase/hydroxylase, expandase, and hydroxylase [where EMCC=exomethylene cephalosporin C; DAOC=deacetoxycephalosporin C; DAC= deacetylcephalosporin C].

Both Oxford scientists and Lilly scientists have probed the active site of expandase using different potential substrates. Some of the results of investigating the expandase activity of *C. acremonium* and *S. clavuligerus* enzymes are shown in Table 3.[35,41] The data generated for the expandase/ hydroxylase from *C. acremonium* indicate that the length of the side chain, the carboxyl group and the configuration of the α-amino group are important.

Table 3. Substrate specificity for ring-expansion of *C. acremonium* expandase/hydroxylase and *S. clavuligerus* expandase

Compound Tested (at 280 μM)	Structure	Enzyme Activity (%) Exp/Hyd	Expandase
Penicillin N		100[a]	100[b]
Isopenicillin N		0[c]	0
Penicillin N Sulfoxide		0	0
Dethio-penicillin N		0	0
Penicillin G		0	0
m-Carboxy-Penicillin G		50	?
6-Glutarylamido-penicillanic Acid		3	0
6-(m-Hydroxyphenylpro-pionamido)-penicillanic Acid		0	0
Adipamido-penicillanic Acid		40	?

[a]100%=0.32 nmol DAOC+DAC formed/min.
[b]100%=0.33 nmol DAOC formed/min.
[c]Not detectable (less than 0.5% of relative product formation) by HPLC.

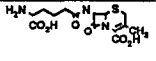
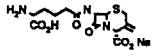
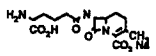
As the 7-adipoyl cephalosporin C is a substrate and the L-α-aminoadipic derivative is not, then the amino group of the L-α-aminoadipic acid configuration is not accommodated either sterically or electronically by the enzyme. Suprisingly the dethiopenicillin N is not a substrate even though it contains the D-aminoadipic side chain. We had anticipated that this compound would hydroxylate at the methyl group. We have also investigated several compounds as inhibitors of the natural substrate, the results are shown in Table 4. Not suprisingly those compounds already found to be substrates are also competitive inhibitors of the enzyme i.e. *m*-carboxypenicillin G and exomethylenecephalosporin C. The dethiopenicillin N, penicillin N sulfone, and deacetylcephalosporin C are effective inhibitors suggesting that they fit the active site but do not undergo chemistry at the active site. As cephalosporin C is not an inhibitor the presence of the extra acyl group apparently has a steric interaction with the enzyme.

Table 4. Inhibition of ring-expansion of *C. acremonium* expandase/hydroxylase and *S. clavuligerus* expandase

Compound Tested	Enzyme Inhibition (%) Exp/Hyd		Expandase	
Ratio To 100 μM Penicillin N	1	10	1	10
Penicillin N Sulfoxide	0	100	23	30
Dethio-Penicillin N	9	78	96	100
m-CO_2H-Penicillin G	12	81	15	39
Exomethylene ceph C	0	90	64	100
Deacetylceph C	14	55	26	27
Cephalosporin C	0	0	0	9

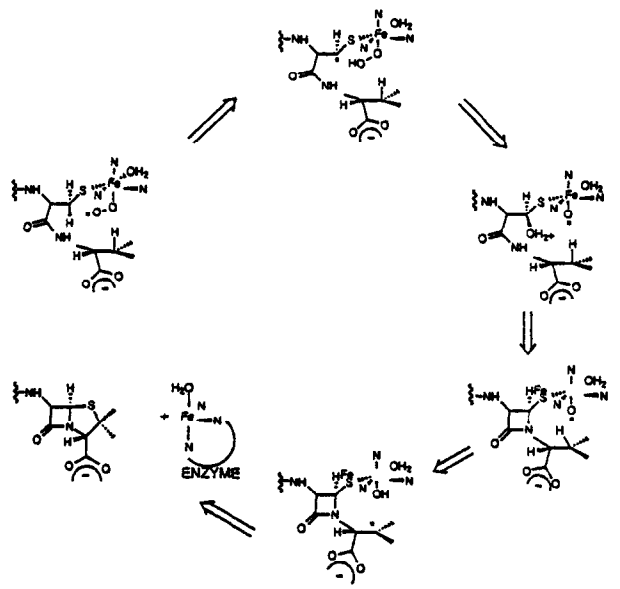
The results of probing the active site of the *S. clavuligerus* hydroxylase and the hydroxylase function of the expandase/hydroxylase from *C. acremonium* with alternate substrates are shown in Table 5. Again we see that the exomethylene isomer is an effective substrate for the hydroxylation function, however the 1-carba analogue is also a good substrate of the *C. acremonium* expandase/ hydroxylase proving that for the hydroxylation reaction the presence of the sulfur atom is not required.[42]

Table 5. Substrate specificity for 3'-hydroxylation of *C. acremonium* expandase/hydroxylase and *S. clavuligerus* hydroxylase

Compound tested (at 280 μM)		Exp/Hyd	Hydroxylase
Deacetoxy ceph C		100[a]	100[b]
Exomethylene ceph C		56	51
Carba-DAOC		33	?

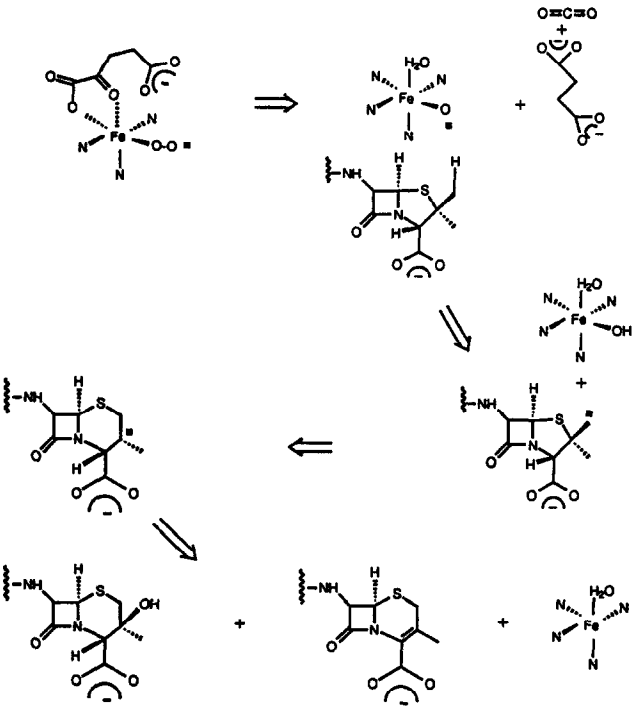
[a]100%=0.29 nmol DAC formed/min.
[b]100%=0.43 nmol DAC formed/min.

Our proposals concerning the mechanism of formation of penicillin from the tripeptide and its subsequent conversion to the cephalosporin ring system are shown in Scheme I. The active site of the enzyme contains the ferrous atom bound to water and the tripeptide substrate. The substrate also appears to be bound through the valine carboxyl group and the L-aminoadipic acid carboxyl group. Approach of one molecule of dioxygen results in a species in which the



Scheme I. Proposed mechanism of the enzymatic synthesis of penicillin.

oxygen is bound to the iron with the distal oxygen atom having an unpaired electron. This highly reactive species removes a hydrogen from the position on the substrate α to the sulfur atom. The resultant stabilized secondary radical is converted to the secondary alcohol by transfer of a hydroxyl radical from the iron to the carbon atom. Possibly after protonation, ring closure of the four membered ring takes place by a nucleophilic displacement using the amide nitrogen. The iron species now exists as a ferryl species and still retains an unpaired electron on the remaining oxygen atom. This is also a powerful oxidizing agent and can abstract the tertiary hydrogen from the β carbon of the valine.



Scheme II. Proposed mechanism of the enzymatic synthesis of cephalosporin from penicillin.

This radical displaces the sulfur atom from the iron centre with final closure of the second ring and liberation of the isopenicillin N product and reconstitution of the active catalytic site on the iron. A variation on this theme explaining in more detail the stereochemical results obtained with different substrates has recently been published.[43,21]

There is considerable similarity between IPNS and expandase/hydroxylases, with the major differences being the requirement for α-ketoglutarate and the existence of a different configuration of the α-aminoadipic acid in the substrate. We do not know the ligands around the iron but taking the IPNS active site as a model we can speculate that we have two of the IPNS ligands replaced by two ligands of the α-ketoglutarate, possibly those due to the thiol group that was required for IPNS but is not present in the expandase substrate and the water molecule or the aspartyl residue. Our proposed mechanism is shown in Scheme II. The reaction of the iron with dioxygen results in the same initial adduct as previously hypothesized with IPNS which then reacts with α-ketoglutarate resulting in CO_2, succinic acid, and the ferryl oxygen species. This initial reaction in the IPNS mechanism in which the first ring is formed giving the ferryl species may now be accomplished by using α-ketoglutarate, thus explaining the need for this cofactor. Removal of hydrogen from the β-methyl group of the penicillin substrate results in a primary radical which is stabilized by orbital overlap from the lone pair electrons on the sulfur atom. A ring expansion to the thermodynamically more stable six-membered cepham ring then occurs forming the tertiary radical species. This reorganization of a radical on the β-methyl group has chemical precedence as generation of the radical from the β-bromomethyl penicillin causes a rapid rearrangement to the six membered ring.[39] This is a rearrangement that can occur through either a one or two electron mechanism with great facility. The alternate mechanism of fragmentation of the primary radical to the thiol radical followed by readdition to the double bond resulting in the same tertiary radical has also been demonstrated to have partial chemical precedence.[44] The enzymatically generated tertiary radical on losing a hydrogen radical gives the observed product, deacetoxycephalosporin C. Alternatively, hydroxyl transfer can occur from the metal centre to the substrate, occurring from the same side of the substrate as the initial hydrogen abstraction yielding the 3β-hydroxycepham which has been observed as a minor product [2%] of the enzymatic reaction.[40] The remaining step is that of an allylic hydroxylation of the methyl group to give deacetylcephalosporin C. This is a reaction that has not been studied, but as it is carried out by either the same or very similar enzymes we presume that it is operating at the metal centre in a manner similar to that of the proposed expandase mechanism.

Although we have generated a large amount of data concerning the structure of the metal centre, we have not been successful in determining the secondary or tertiary structures of these enzymes. They have proven recalcitrant to crystallization. One advantage that we do possess with

these enzymes is that we have available the sequences of IPNS enzymes from nine different bacterial and fungal sources. This gives advantages in our attempts to determine the secondary structure of these enzymes, for it is possible to conclude that any differences in sequence are not relevent to function, areas of homology could be important, and sequence identity of potential binding sites could be very important. The degree of homology between all of these sequences is shown in Table 6. There is a consistent degree of sequence homology between the fungal sources of IPNS, i.e.*C. acremonium, P. chrysogenum,* and *A. nidulans,* of 72–79%, and also a high degree between the prokaryotic sources, *S. clavuligerus, S. lipmanii, S. griseus, S. jumonjinensis* and *Nocardia lactamdurans* of 70–80%. However the degree of homology between the fungal and the streptomyces proteins drops to a level of 54–60%. The one example of a gram negative bacterial source, *flavobacterium sp.* exhibits a similar degree of homology of 55–60% to both of the other source groups.

Table 6. Homology of IPNS sequences

	C.a	*A.n*	*S.c*	*S.l*	*S.g*	*N.l*	*S.j*	*Fl.*
P.crysogenum	76	78	54	53	55	56	58	54
C.acremonium		72	55	56	54	57	61	57
A.nidulans			57	56	55	58	59	56
S.clavuligerus				70	70	73	80	59
S.lipmanii					73	70	70	56
S.griseus						72	74	59
N.lactamdurans							76	57
S.jumjimensis								60

[where *P. chrysogenum=penicillium chrysogenum; C.a=Cephalosporium acremonium; A.n=Aspergillus nidulans; S.c=S.clavuligerus; S.l=S.lipmanii; S.g=S.griseus; N.l=Nocardio lactamdurans; S.j=S.jumonjinensis; and Fl=flavobacterium sp.*]

The percentage identities between the expandase and hydroxylase enzymes and between the IPNS and expandase enzymes are shown in Tables 7 and 8. The degree of identity between the prokaryotic and fungal sources of the enzymes is the same in both the IPNS enzymes and the expandase/hydroxylases, an observation that has evolutionary importance. What is also intriguing is the similarity between IPNS and expandase, they have a 24% identity and a 45% similarity, raising the possibility that because of this similarity of sequence and similarity of function, there is also a corresponding similarity in secondary structure.

Table 7. Homology between expandase/hydroxylase, expandase and hydroxylase

	C.a .EXP	*S.c.* .EXP	*S.c.* .HYD
S.c. EXP	56.6		
S.c. HYDR	54.6[71]	58.8[74]	

[where *C.a.ex p=C. acremonium* expandase/hydroxylase; *S.c.exp=S. clavuligerus* expandase, and *S.c.hyd* =S. [illegible] hydroxylase. The numbers are % age identity, those in [ ] are [illegible]]

Table 8. Homology between IPNS and expandase/hydroxylase and expandase and hydroxylase

	C.a. EXP.	*S.c.* EXP.	*S.c.* HYDR.
C.a. IPNS	18.4 [42]		
S.c. IPNS		24 [45]	25 [44]

[where *C.a.* IPNS=IPNS from *C. acremonium, S.c.* IPNS=IPNS from *S. clavuligerus.*]

In an attempt to generate a secondary structure of these enzymes, we concentrated on the IPNS enzyme for which we had the nine separate sequences. There are some generalized rules of protein structure that should be considered at the outset, namely that:

(1) Homologous enzymes that catalyze the same metabolic reaction in different organisms will adopt the same folding topology.
(2) The protein folding code is degenerate, highly diverse sequences can encode for the same topology.
(3) Insertions or deletions occur mainly in surface loops.
(4) Residues that are involved in catalysis are highly conserved.
(5) Variability of residues in contact with the substrate is low.
(6) Residues that perform equivalent structural and functional roles are usually conserved.
(7) Hydropathic character of the core is generally preserved, buried positions will only tolerate conservative substitution.
(8) High variability of sequence is a characteristic of surface loops.
(9) Local maxima of hydrophobicity is associated with buried β-sheets.
(10) Chain flexibility maxima correlate with surface chain segments

The basic strategy in attempting to determine the secondary structure of IPNS involved obtaining an alignment of the nine sequences. The guidelines that we established in order to achieve the optimal alignment were [a] attempt to maximally align the histidines, some of which have been implicated in binding to the iron, [b] maximally align the acidic and basic groups which are potential binding sites for the substrate, iron, and α-ketoglutarate, [c] align the proline and glycine residues which are important structural residues mainly found in turns and which operate as both α-helix and β-sheet breakers, and [d] determine areas of hydrophobicity and hydrophilicity. The optimal alignment achieved is shown in Fig. 11. There were found to be six histidines common to all nine sequences [shown in red], therefore establishing with a high probability that the two or three iron binding histidines were located among these sites. In addition there were six common basic amino acids [lysine/arginine, shown in green] and seven common acidic groups [aspartic acid/glutamic acid, shown in blue]. It is highly probable that the substrate binding sites are located amongst these few positions. The next procedure was to attempt to determine the basic structural architecture of these enzymes. As more protein structures are being solved it is becoming apparent that there is a far greater conservation in structural topology than in sequence similarity, in fact there are several examples of sequences having insignificant sequence similarity and yet have the same topology. The corollary of this is that if proteins

have related function and a reasonable sequence homology then it is highly probable that they have the same topology. Initially we attempted to determine to which structural type of protein these enzymes belonged, for proteins have been classified into four different architectural types,α [mostly helical and little or no β structure], β [mostly β-sheet with little or no helical content], α + β [α-helical and β-structures that segregate into different domains], and α/β [approximately alternating α and β domains].[45]

Table 9. Statistical analysis of the amino acid composition of IPNS and expandase/hydroxylase

AMINO-ACID	IPNS Average	Exp/Hyd Average	RANGE	α-Proteins	β-Proteins	α+βProteins	α/β Proteins
ALA	8.1	9.3	11.7-6.0	11.6	7.3	9.3	8.3
ARG	5.4	5.8	3.6-7.8	2.2	2.4	4.1	3.4
ASP	7.0	6.6	8.6-6.0	6.7	4.4	5.9	5.6
ASN	4.5	3.5	6.0-1.6	4.0	5.0	6.4	4.2
CYS	1.0	1.1	1.9-0.6	0.9	2.7	3.9	1.5
GLU	6.5	7.1	7.8-5.3	5.5	3.1	4.6	5.9
GLN	3.7	2.5	5.2-1.8	2.7	4.4	3.9	2.0
GLY	6.5	6.3	8.1-4.8	8.1	10.7	9.1	8.7
HIS	3.8	2.2	5.5-1.2	4.5	1.8	1.7	2.5
ILE	4.1	3.1	5.3-1.8	3.7	4.3	4.9	5.5
LEU	8.0	7.2	8.8-6.9	9.0	6.4	5.8	7.8
LYS	4.1	4.2	6.9-1.6	10.2	4.1	5.9	7.4
MET	2.1	2.3	3.6-0.9	2.0	0.6	1.3	2.1
PHE	5.1	5.5	6.5-3.1	5.0	3.1	2.8	3.6
PRO	6.5	5.8	7.3-5.5	3.4	4.6	3.8	4.3
SER	6.0	6.7	7.1-4.6	5.0	12.3	6.7	7.5
THR	6.0	6.3	8.8-3.8	4.9	9.1	6.2	5.5
TYR	4.9	4.0	6.1-2.7	2.6	4.0	5.7	3.0
TRP	1.2	1.2	2.8-0.6	1.3	1.6	1.6	1.7
VAL	6.5	7.6	9.0-4.7	6.8	8.2	6.5	8.7

An algorithm was developed based on a statistical analysis of the amino acid compositions,[46] see Table 9, which was claimed capable of assigning proteins to their proper conformational class with an accuracy of 80%. This is a study which shows the advantage of having available multiple sequences. The results of this technique when applied to these sequences showed considerable differences in the serine, lysine, and alanine values compared to representative values found for a series of proteins of known classes and would argue against both the all α- and the all β-protein conformations. The lower cysteine content would indicate that these proteins belong to the α/β class rather than the α+β class. Interestingly there is considerable similarity in the average values for both the IPNS sequences and the EXP/HYDROX. sequences indicating that both series of enzymes have the same overall topology.

An alternate statistical method of predicting the structural class was recently developed by Zhang and Chou,[47] the results of this approach are shown in Table 10. From the values shown in this table, the probability that these enzymes exist in either an α+β or α/β is 70–100% with an overall average of 89%. There is also a 2:1 probability that they belong to the α/β type of protein rather than the α+β, a conclusion which is in agreement with the previous method. The results of this analysis on the expandase/hydroxylase enzymes is considerably less clearcut, whilst there is an average weighting in favour of either an α+β or an α/β, it is impossible to distinguish between these two topologies.

Table 10.

Protein Class / Enzyme Source	ALPHA	BETA	ALPHA+BETA	ALPHA / BETA
C.a.	11	0	28	61
P.c.	23	0	13	63
A.n	29	0	46	25
S.c.	0	3	48	49
S.j.	0	0	19	80
N.l.	0	4	0	96
S.l.	0	11	25	65
S.g.	0	0	24	76
Fl.	9	0	70	20
Average	8	2	30	59
C.a. exp/hydrox.	0	0	49	49
S.c. exp.	0	39	33	27
S.c. hydrox.	2	31	34	32

These values are expressed in probability percentages.

Although a computer homology search has not identified any other related sequences, it was able to identify an interesting small sequence of ca. 6 amino acids which has high similarity in all of the twelve enzymes and which is repeated throughout their full sequences several times at regular intervals. Furthermore this small sequence was found to be highly homologous to a segment in the carboxy terminal end of the β-chain of human insulin which exists as a β-sheet. There is a high retention of identity or similarity in these segments which are highly hydrophobic, suggesting that they could constitute the hydrophobic core of the protein. The location of these short sections are shown in Fig. 11 in colour coded form [magenta]. In this scheme the cysteine 106 which has been implicated in substrate binding to the enzyme is located in one of the β sheets in the interior hydrophobic region.

We then analyzed the sequences for areas of hydrophobicity, surface preference, and flexibility.[48] The areas of high hydrophobicity are expected to be associated with highly conserved interior segments of the enzymes, while the flexibility and surface preference areas are expected to be areas of low homology and associated with the exterior areas of the proteins. There was found to be considerable consistency between the individual sequences in these areas. This is shown in Fig. 12 and it is immediately apparent that these domains are generally alternating throughout the sequences, highly indicative of the α/β type of architecture. We now applied a secondary structure prediction algorithm to these aligned sequences in an attempt to locate areas of definitive secondary structure, namely α-helix, β-sheet, and turns. The method chosen was "Combine", a recent version of the pioneering GOR procedure[49] in which an improvement in the prediction accuracy to 65.5% is claimed. This method was applied to each sequence separately and the initial predictions are shown in Fig.13. There is a considerable similarity between the predictions for the individual sequences. There are some obvious areas of strong secondary structure indications and several other areas where the predictions are made with less confidence. For instance, using the sequence numbering of the *Penicillium* IPNS there are strong predictions of α-helix at positions 23–36, 52–80, 140–180, and β-sheets at 41–46, 221–225, 250–254. There are weaker indications of α-helix at 200–220, and 300–

320, and β-sheet at 90–94, 183–190, 272–276. Again the periodicity of the secondary structure is observed, with the overall impression of alternating α-helix and β-sheet units. This is again highly indicative of an α/β barrel type of structure.

The three expandase/hydroxylase enzymes possessed a high similarity of sequence to each other, but a somewhat lower one [42–45%] to the equivalent IPNS enzymes. We were able to obtain a significant alignment of these sequences with the judicious use of several gaps. This alignment is shown in Fig. 14, where we have eliminated several of the IPNS sequences for simplicity, however all the consensus amino acids highlighted are also homologous in all of the IPNS sequences which are not shown. We were able to align two of the histidines in all twelve proteins, those at positions 250 and 311 in Fig. 14, thus these would have the highest probability as being the iron binding ligands. Analysis of the surface preferences also generally coincide with the gaps that have been incorporated to achieve the alignment. This is consistent with the fact that if these proteins are topologically similar then differences would be expected to be in the surface loops, and areas of high similarity are in the catalytically important interior regions. These are shown in Fig. 14 [in magenta] and there is good similarity overall with the IPNS sequences, with a few differences in the region 240–270.

In summary this analysis predicts that the secondary structure of both the IPNS and the EXP/HYDROX. enzymes is closely related and is an α/β barrel type of structure. It also predicts the location of the two histidine iron-binding sites. This type of topology which can exist in either an eight stranded parallel arrangement of alternating helix and sheet or certain variations on this theme in which either α-helix or β–sheet is missing, is a folding motif which is finding increasing occurrence. The sequences of known examples of this topology have little sequence homology and they have been found to be able to perform a diverse series of enzymatic reactions indicating that this type of structure is a particularly stable type of folding motif. In our utilization of these conclusions we should note that active sites of this motif always exist in the C-terminal part of the β-sheet and in the β–α loops. This gives direction in attempting to ascertain the critical substrate binding sites of these enzymes, however, a true picture of these mechanistically fascinating enzymes will have to await the X-ray structure analysis.

It was observed that the sequence similarity between the prokaryotic and eukaryotic beta-lactam IPNS genes is high, especially when compared to other related prokaryotic and eukaryotic genes.[50] This has resulted in some interesting proposals by Skatrud *et al.* concerning the evolutionary assemblage of these genes, see Fig. 15. The homology in the IPNS genes was ca. 55%, and this same value was found for the expandase genes.[50,51] In the prokaryotic organisms these genes were found to be tightly clustered, being physically linked within 20–30 KD.[52] This clustering was also found in the *P. chrysogenum* and *A.nidulans*,[53–55] indicating that all of these genes were transferred in one event. However, in *C. acremonium* it was found by Skatrud and Queener using molecular karyotyping analysis that the IPNS gene was on chromosome VI and the epimerase and expandase genes were located on chromosome II.[56] It was hypothesized that the genes were in two clusters, thus allowing the location and identification of the ACV synthase gene next to the IPNS gene. Based on a constant rate of evolutionary change, this resulted in the proposal that a horizontal transfer from the prokaryotes to the eukaryotes occurred about 370 million years ago, well after the prokaryotic/eukaryotic split which happened 1.8 billion years ago.

About 70 million years after this original transfer, an ancestor common to both *P. chrysogenum* and *A. nidulans* diverged from *C. acremonium*. It was hypothesized that in this eukaryotic precursor the block of biosynthetic genes was split into two chromosomes and when the transfer into *P. chrysogenum* occurred only the one containing the ACV synthase and the IPNS genes was transferred. This proposal also explained the absence of these beta-lactam genes in the yeasts, as the evolutionary split of the yeasts from the fungi occurred before the horizontal transfer from the prokaryotic organisms. The homology of the genes from the gram negative bacteria, *flavobacteria sp., is* the same to both the fungal and bacterial sources, indicating that these genes horizontally transferred into this gram negative bacterium long after the gram positive–gram negative split, and at approximately the same time period that they transferred into the eukaryotic organisms. Assuming that this theory is correct, there still exists the anomaly concerning the direction of transcription, for in *S.clavuligerus* the genes are transcribed in the same direction whereas in the fungi they are divergently transcribed. Thus for unknown reasons gene rearrangements must have occurred while also maintaining tight clustering.

We now have available to us the genes responsible for ACV synthetase, IPNS, racemase, expandase, and hydroxylase, from both eukaryotic and prokaryotic sources. We have recently been investigating the utilisation of these genes. Skatrud *et al.* have inserted the IPNS gene into a production strain of *C. acremonium* when an increase in the levels of IPNS was observed showing that the copy of the gene had been increased.[57] However no increase in antibiotic production was observed, indicating that in this particular strain this enzyme was not rate limiting. We had available a production strain of *C. acremonium* which contained residual penicillin N at the termination of the fermentation, an ideal case in which to test this technology. The expandase/hydroxylase gene was inserted by Queener *et al.* into this strain to give an additional copy of the gene.[57] The level of expandase activity was increased two fold and more excitingly the level of cephalosporin C also increased whilst the level of penicillin N decreased. Thus we are now able to manipulate the production levels of secondary metabolites by increasing the copy numbers of the genes which are rate determining.

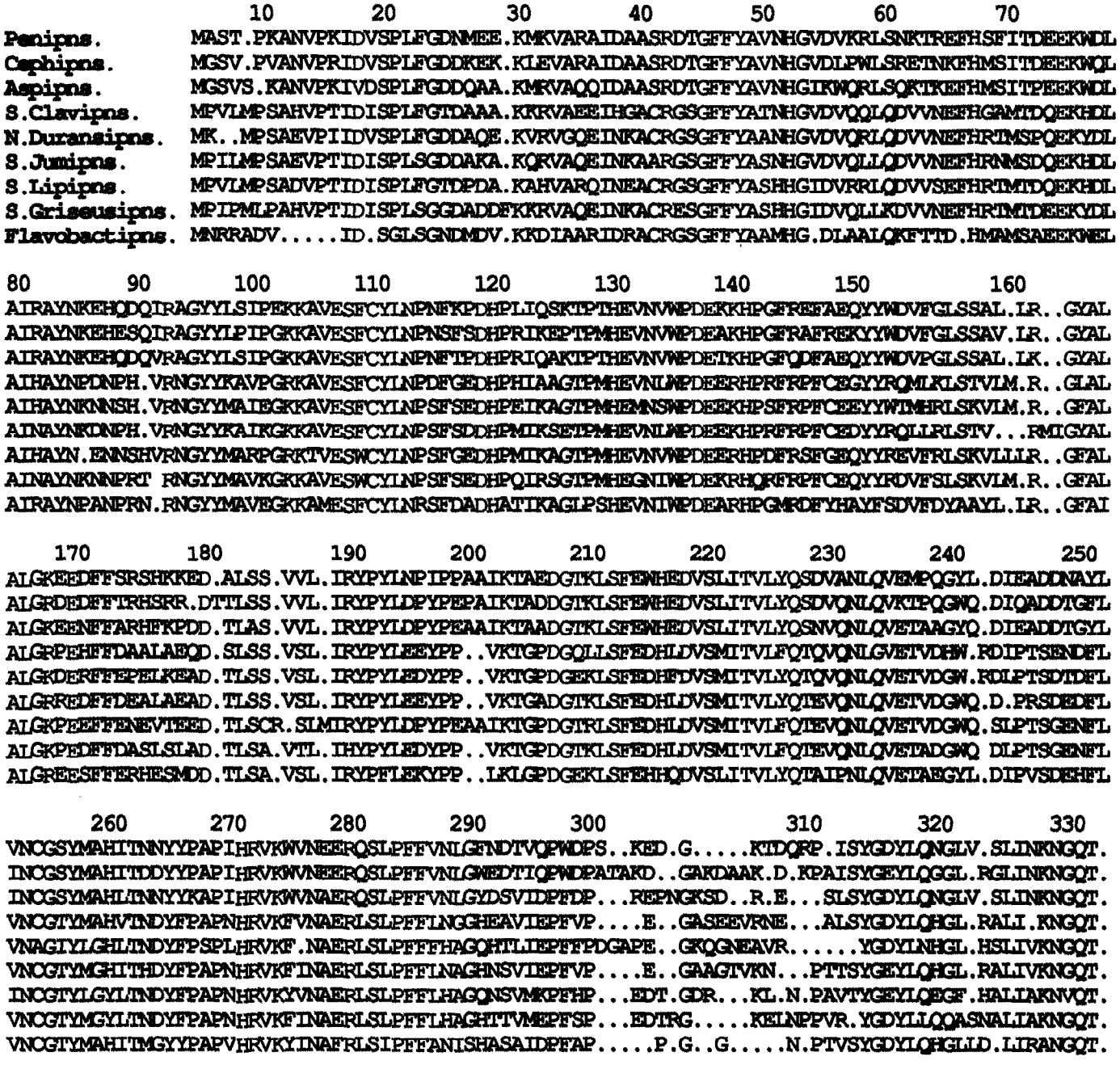


Figure 11 IPNS sequences–Areas of homology.

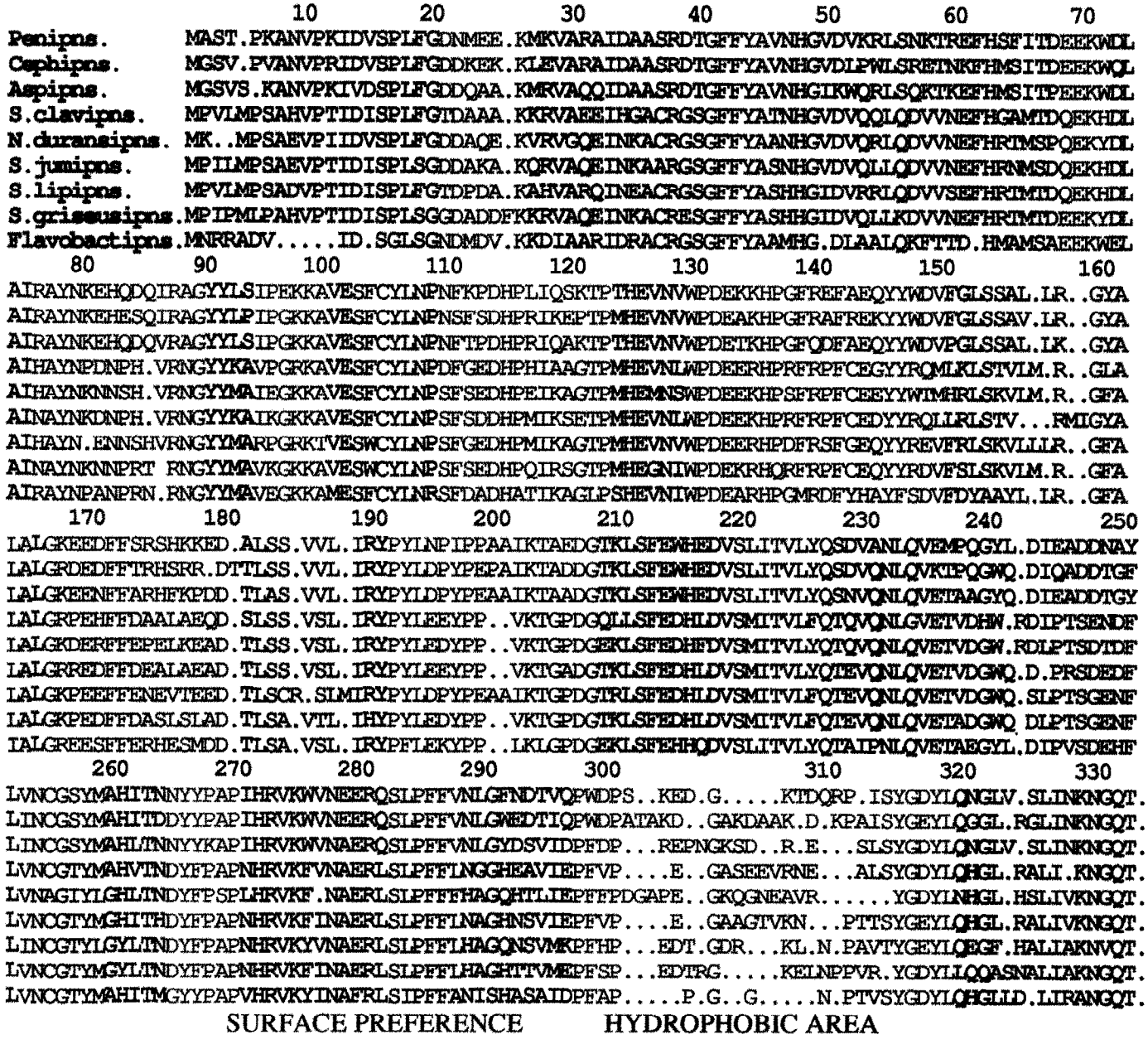


Figure 12 IPNS sequences–Areas of surface preference and hydrophobicity.

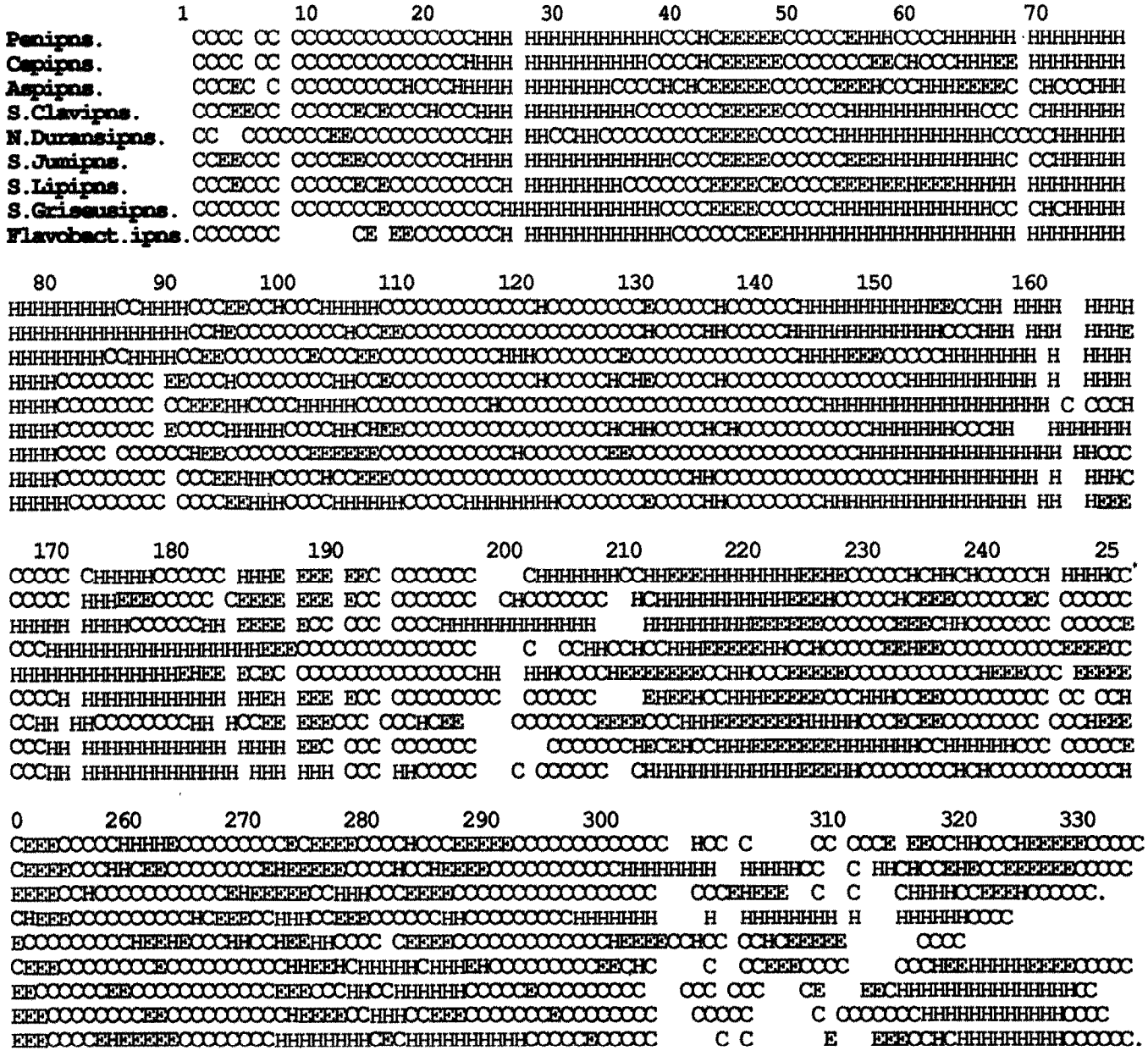


Figure 13 IPNS sequences–Predicted secondary structure.

Figure 14 IPNS and expandase sequences–Alignment and hydrophobic homology.

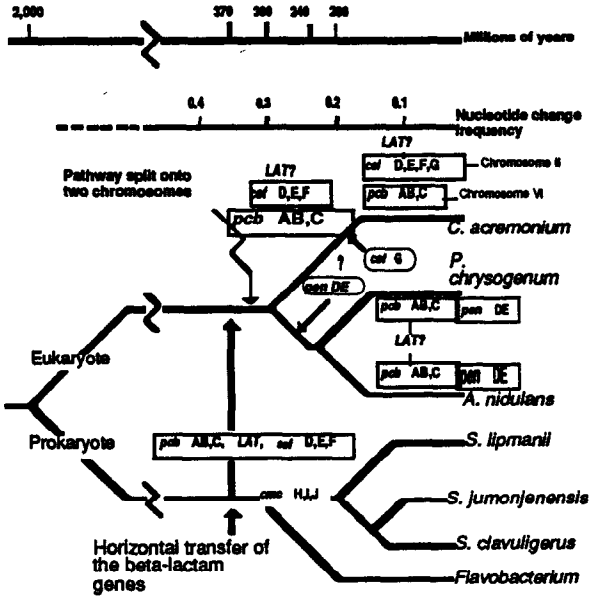


Figure 15 Phylogenic Tree.

Finally, an experiment which gives us a view into the future is the successful insertion by Queener and co-workers of the genetic material containing the genes for both epimerase and expandase into a strain of *Penicillium*.[58,59] *P. chrysogenum* does not contain these genes as the pathway diverges after IPNS, when a transacylase enzyme [or enzymes] converts isopenicillin N into 6-APA and from thence to a precursed penicillin [see Fig.1]. The genetic material that was inserted came from *S.clavuligerus*, a prokaryotic organism. Strains of *P.chrysogenum* which now produce deacetoxy-cephalosporin C although still producing high levels of penicillin were obtained . To change the direction of the biosynthetic pathway such that the *P. chrysogenum* would produce deacetoxycephalosporin C to the exclusion of 6-APA, the function of the transacylase gene would have to either be removed or deactivated.

We have now modified the direction of a biosynthetic pathway in a eukaryotic organism by the insertion of genes from a prokaryotic source, a process which Nature accomplished 370 million years ago.

Note Added In Proof

The sequence for the enzyme from *Nocardia lactamdurans* having ring expandase activity has just been published,[60] however there was no mention as to whether this enzyme had a dual function. This sequence was found to align with the sequence alignments shown in Fig. 14 in both the histidines at 250 and 311 and the conserved stretches of hydrophobicity maxima.

Acknowledgements

It gives me great pleasure to acknowledge the scientific contributions of a superb team of scientists at the Lilly Research Laboratories; Molecular Biology and Genetics—Tom Ingolia, Paul Skatrud, Jim Miller, Steve Queener, Mat Tobin, Steve Kovacevic; Protein isolation—W.-K. Yeh, Joe Dotzlaf; Enzymology—Victor Chen, Chuck Frolick; Chemistry—George Huffman; and the stimulating discussions with Prof. Jack Baldwin, Sir Edward Abraham and their talented colleagues at Oxford University.

References

1. E. Kiester, Jr, *The Smithsonian,* 173–187, 1990.

2. For a review of these early studies see; *The Chemistry of Penicillin*, p.144, Eds. H. T. Clarke, J. R. Johnson & R. Robinson; Academic Press, 1949.

3. There is a considerable body of work carried out in many laboratories which has resulted in our present understanding of this problem. It is beyond the scope of this brief review to be able to do justice to the many important contributions from these laboratories.

4. G. Banko, A. L. Demain & S. Wolfe, *J. Amer. Chem. Soc.* **109,** 2858–2860, 1987.

5. In the general class of oxidoreductase enzymes, where oxygen is the acceptor, it is correct to classify the enzyme as an oxidase, see; *Enzyme Nomenclature*, p.7, Academic Press (1984).

6. S. E. Jensen, D. W. Westlake, & S. Wolfe, *Can. J. Microbiol.* **29**, 1526–1531, 1983.

7. M. Kohsaka & A. L. Demain; *Biochem. Biophys. Res. Commun.* **70**, 465–473, 1976.

8. S. M. Samson, R. Belagaje, D. T. Blankenship, J. L. Chapman, D. Perry, P. L. Skatrud, R. M. Van Frank, E. P. Abraham, J. E. Baldwin, S. W. Queener & T. D. Ingolia, *Nature* **318**, 191–194, 1985.

9. M. J. Bibb, P. R. Findley & M. W. Johnson, *Gene* **48**, 157–166, 1984.

10. L. G. Carr, P. L. Skatrud, M. E. Scheetz, S. W. Queener & T. D. Ingolia, *Gene* **48**, 257–266, 1986.

11. J. L. Barredo, J. M. Cantoral, E. Alvarez, B. Diez & J. F. Martin, *Mol. Gen. Genet.* **216**, 91–98, 1989.

12. B. J. Weigel, S . G. Burgett, V. J. Chen, P. K. Skatrud, C. A. Frolik, S. W. Queener & T. D. Ingolia, *J. Bacteriol.* **170**, 3817–3826, 1988.

13. D. Ramon, L. Carramolino, C. Patino, F. Sanchez & M. Penalva, *Gene* **57**, 171–181, 1987.

14. B. K. Leskiw, Y. Aharonowitz, M. Mevarech, S. Wolfe, L. C. Vining, D. W. S. Westlake & S. E. Jensen, *Gene [Amst.]* **62**, 187–196, 1988.

15. M. Garcia-Dominguez, P. Liras & J. F. Martin, *Antimicrobial Agents & Chemotherapy* **35**, 44–52, 1991.

16. D. Shiffman, M. Mevarech, S. E. Jensen, G. Cohen & Y. Aharonowitz, *Mol. Gen. Genet.* **214**, 562–569, 1988.

17. J. M. Castro, P. Liras, J. Cortes & J. F. Martin, *Applied Microbiology and Biotechnology* **21**, 32–40, 1985.

18. D. Shiffman, G. Cohen, Y. Aharonowitz, H. Palissa, H. von Dohren, H. Kleinkauf & M. Mevarech, *Nucleic Acids Res.* **18**, 660, 1990.

19. S. M. Samson, J. L. Capman, R. Belagaje, S. W. Queener & T. D. Ingolia, *Proc. Natl Acad. Sci. U.S.A.* **84**, 5705–5709, 1987.

20. A. Kriauciunas, C. A. Frolik, T. C. Hassell, P. L. Skatrud, M. G. Johnson, N. L. Holbrook & V. J. Chen, *J. Biol. Chem.* **266**, 11779–11788, 1991.

21. J. E. Baldwin & M. Bradley, *Chem. Rev.* **90**, 1079–1088, 1990.

22. G. W. Huffman, P. D. Gesellchen, J. R. Turner, R. B. Rothenberger, J. L. Chapman & S. W. Queener, *J Med. Chem.* **35**, 1897–1914, 1992.

23. V. .J. Chen, A. M. Orville, M. R. Harpel, C. A. Frolik, K. Surerus, E. Munck & J. D. Lipscomb, *J. Biol. Chem.* **264**, 21677–21681, 1989.

24. Bo. G. Malmstrom, *Ann. Rev. Biochem.* **51**, 21–59, 1982.

25. H. M. Hanauske–Abel & V. Gunzler, *J. Theor. Biol.* **94**, 421–455, 1982.

26. V. J. Chen & J. D. Lipscomb, *unpublished results.*

27. L.-J. Ming, L. Que, A. Kriauciunas C. A. Frolik & V. J. Chen, *Inorg.Chem.* **29**, 1111–1112, 1990.

28. L.-J. Ming, L. Que, A. Kriauciunas, C. A. Frolik & V. J. Chen, *Biochemistry* **30**, 11653–11659, 1991.

29. F. Jiang, J. Piesach, L.-J. Ming, L. Que, Jr & V. J. Chen, *Biochemistry* **30**, 11437–11445, 1991.

30. A. Scheidegger, M. T. Kuenzi & J. Nuesch, *J. Antibiot.* **37**, 522–531, 1984.

31. S. E. Jensen, D. W. S. Westlake & S. Wolfe, *J. Antibiot.* **38**, 263–265, 1985.

32. J. E. Dotzlaf & W.-K Yeh, *J. Bacteriol.* **169**, 1611–1618, 1987.

33. S. Kovacevic, B. J. Weigel, M. B. Tobin, T. D. Ingolia & J. R. Miller, *J. Bacteriol.* **171**, 754–760, 1989.

34. S. Kovacevic & J. R. Miller, *J. Bacteriol.* **173**, 398–400, 1991.

35. W.-K. Yeh & J. E. Dotzlaf, *Symposium "50 Yrs of Penicillin"*, Berlin, 1990.

36. J. E. Dotzlaf & W.-K. Yeh, *J. Biol. Chem.* **264**, 10219–10227, 1989.

37. S. M. Samson, J. E. Dotzlaf, M. L. Slitz, G. W. Becker, R. M. Van Frank, L. E. Veal, W. K. Yeh, J. R. Miller, S. W. Queener & T. D. Ingolia, *Bio/Technology*, **5**, 1207–1214, 1987.

38. B. J. Baker, J. E. Dotzlaf & W.-K. Yeh, *J. Biol. Chem.* **266**, 5087-5093, 1991.

39. J. E. Baldwin, R. M. Adlington, R. T. Aplin, N. P. Crouch, G. Knight & C. J. Schofield, *J. Chem. Soc., Chem. Commun.* 1651–1654, 1987.

40. J. E. Baldwin, R. M. Adlington, T. W. Wong, E. Lee & C. J. Schofield, *J. Chem. Soc., Chem. Commun.* 104, 106, 1987.

41. J. E. Baldwin, R. M. Adlington, M. J. C. Crabbe, G. Knight, T. Nomoto, C. J. Schofield & H.-H. Ting, *Tetrahedron* **43**, 3009–3014, 1987.

42. J. E. Baldwin, K. C. Goh, M. E. Wood, C. J. Schofield, R. D. G. Cooper & G. W. Huffman, *Bioorg. Med. Chem. Lett.* **1**, 421–424, 1991.

43. J. E. Baldwin, M. Bradley, N. J. Turner, R. M. Adlington, A. R. Pitt & A. E. Derome, *Tetrahedron* **47**, 8223–8242, 1991.

44. Y. Maki & M. Sako, *Tetrahedron Lett.* 4291–4294, 1976.

45. M. Levitt & C. Chothia, Nature **261**, 552–558, 1976.

46. P. Y. Chou, *Prediction of Protein Structure and the Principles of Protein Conformation*, pp. 587–599, Ed. G. D. Fasman, Plenum Press, 1989.

47. C.-T. Zhang & K.-C. Chou, *Protein Sci.* **1**, 401–408, 1992. I would like to acknowledge the contribution of David Smith, Lilly Research Laboratories, for developing the computer programme to allow this application to be run on the Cray 2 supercomputer at the Lilly Research Laboratories.

48. The computer algorithm used for the hydrophobicity measurements was that of Hopp & Woods; *Proc. Natl Acad. Sci. U.S.A.* **78**, 3824–3828, 1981; and that used for the surface probability was from Ermini, *J. Virol.* **55**, 836–839, 1985. These programmes were part of the GCG Sequence Analysis Software Package developed at the University of Wisconsin; Genetics Computer Group, 575 Science Dr., Madison, WI 53711. U.S.A.

49. V. Biou, J. F. Gibrat, J. M. Levin, B. Robson & J. Garnier, *Protein Engng* **2**, 185–191, 1988. I would like to acknowledge the generosity of Dr B.Robson of Proteus Int. for supplying a copy of this programme.

50. P. L. Skatrud, A. J. Tietz, T. D. Ingolia, C. A. Cantwell, D. L. Fisher, J. L. Chapman & S. W. Queener, *Bio/Technology* **7**, 477–484, 1989.

51. P. L. Skatrud, J. Hoskins, J. S.Wood, M. B . Tobin, J. R. Miller, S. Kovacevic, C. A. Cantwell & S. W. Queener, *Proceedings of the American Mycology Meeting*, Madison, Wisconsin, USA, 1990.

52. J. R. Miller & T. D. Ingolia, *Mol. Microbiol.* **3**, 689–695, 1989.

53. P. L. Skatrud, J. Hoskins, M. B . Tobin, J. R. Miller, J. Wood, S. Kovacevic & S .Queener, Molecular genetics of β–lactam antibiotics. In *Applied Genetics of Fungi*, pp.145–159, Eds J. F. Perbody, C. E. Caten, J. E. Ogden & J. W. Bennett, Cambridge University Press (1991).

54. A. W. Smith, M. Ramsden, M. J. Dobson, S. Harford & J. F. Peberdy, *Bio/Technology* **8,** 237–240, 1990.

55. M. B. Tobin, M. D. Fleming, P. L. Skatrud & J. R. Miller, *J. Bacteriol.* **172,** 5908–5914, 1990.

56. P. L. Skatrud & S. W. Queener, *Gene* **78,** 331–338, 1989.

57. P. L. Skatrud, A. J. Tietz, T. D. Ingolia, C. A. Cantwell, D. L. Fisher, J. L. Chapman & S . W. Queener, *Bio/Technology* **7,** 477–485, 1989.

58. S. Kovacevic, M. B. Tobin & J. R. Miller, *J. Bacteriol.* **172,** 3952–3958, 1990.

59. C. Cantwell, R. Beckmann, P. Whiteman, S . W. Queener & E. P. Abraham, *Proc. Royal Soc. Lond., Ser. B* **248,** 283–289, 1992.

60. J. J. R. Coque, J. F. Martin & P. Liras, *Mol. Gen. Genet.* **236,** 453–458, 1993.

Biographical Sketch: R. D. G. Cooper

Born; Eastbourne, England. BSc at Imperial College, London University. PhD at Imperial College, London University, studying under Prof. B. C. L. Weedon–Thesis Title "Synthesis of Capsorubin and Capsanthin, some Naturally Occurring Carotenoids". Post Doc. at Imperial College with Prof. D. H. R. Barton on the synthesis of Tetracycline. Two years visiting National Academy Of Science, Fellowship with the U. S. Government. Presently holds the position of Research Advisor at Eli Lilly & Co., working primarily in the area of infectious disease.