Structural Studies of Two Enzymes of Pantothenate Biosynthesis in Escherichia Coli

Florian Johannes Schmitzberger

A thesis submitted for the degree of Doctor of Philosophy

Sidney Sussex College
Cambridge - United Kingdom
2004

Declaration

This dissertation is a summary of the research carried out in the Department of Biochemistry, University of Cambridge, between October 2000 and October 2003, under the supervision of Professor Sir Thomas L. Blundell. The work described in the dissertation is my own, unless otherwise stated in the text. It has not, either in part, or as a whole, been submitted for a degree, diploma, or other qualification at any other university.

Florian J. Schmitzberger
Department of Biochemistry
University of Cambridge
February 2004

Contents

List of Figures

List of Tables

Acknowledgements

Many people have contributed to enable this work. First and foremost I want to express my deep gratitude to my supervisor, Tom Blundell, for providing me with the opportunity to carry out my Ph.D. research training in his laboratory. I am greatly indebted to him for his excellent guidance, his continuous support throughout my scientific studies, and for the many enthusiastic scientific discussions. I have been fortunate that I was able to learn and work in the stimulating environment of his laboratory, with so many motivated and talented people of diverse cultural and scientific backgrounds.

I would like to thank Dimitri Chirgadze who was teaching me macromolecular crystallography, and assisted greatly in my progress. Particularly, I appreciate him for the many hours he spent with me at the synchroton, for solving numerous computer problems, and for always being happy to give a helping hand when needed. Michael Webb and Michael Witty provided a lot of support with protein purification at the early stage of my project. Tsuyoshi Inoue, a visiting post-doctorial researcher from Japan, supervised me during the first six month of my Ph.D. studies, and taught me several useful crystallographic skills. Mairi Kilkenny and Carina Lobley joined the project when I entered my second year and brought a lot of fresh momentum and impetus to it. I would like to thank them very much for their collaboration in the last two years. Frank von Delft,

although far away in San Diego, was always quick to reply to inquiring e-mails. I am grateful to Chris Abell and Alison Smith for teaching me how (not) to write scientific articles, and for often being so very patient with me. Loïc Bertrand joined the project during my last year and I appreciated to work with him. Many thanks especially for the useful LaTeX 2_ε commands, without which this thesis would have taken even longer to write.

I am very much indebted to Martyn Symmons. He has been a great reviewer for my publications and for this thesis. But more importantly, he always was interested to discuss scientific problems. I would like to thank my Ph.D. companion Mark DePristo for rewarding scientific discussions, and for the numerous nights out at Hughes Hall College. Andy Doré took the time to teach me molecular cloning. Kenji Mizuguchi was at any time there to help me with bioinformatics-related problems. Many thanks go to Louise Birch, Graham Eliff, Paul de Bakker, Nic Harmer, Adrian Harrington, Marko Hyvönen, Diana Matak-Vinkovic, Adrian Saldanha, Dave Story, Mladen Vinkovic, and Christine Witchell, who, in many ways, helped me in the laboratory. Ben Luisi is acknowledged for the help with X-ray diffraction data collection at the synchroton.

Certainly, I would like to thank *all* group members of the Tom Blundell and the Ben Luisi laboratories for creating such a good working environment, and for assisting me in one or the other way during the time of my Ph.D. studies.

Monique Gangloff, Maria Marcaida, and Liisa van Vliet are thanked for the countless dinner evenings in Cambridge, for the great food, the fun, and for creating such a pleasant and comfortable atmosphere outside the laboratory.

This work, surely, would not have been possible without the tremendous help of my parents, Mathilde and Johann, and my sister Barbara. I am immensely grateful to them for creating the necessary background, for

always supporting me, and my interests, and for giving me so much freedom. My appreciations also go to Eva Schober, for more love and support than I could have ever hoped for, and Paul Schanda, who has been a great friend throughout all those years.

At last, I would like to acknowledge funding for my research from the Cambridge European Trust, research-support studentships of the University of Vienna, the Kepler-International-studentship fund, as well as financial support from Sidney Sussex College.

Abstract

Pantothenate (vitamin B_5), which is the invariable metabolic precursor to coenzyme A, is synthesized from L-aspartate and α-ketoisovalerate in a converging four-step process in bacteria. Here, structural studies of two enzymes of pantothenate biosynthesis in *Escherichia coli*, L-aspartate-α-decarboxylase and ketopantoate hydroxymethyltransferase, are described.

Ketopantoate hydroxymethyltransferase catalyzes the transfer of a hydroxymethyl group on to α-ketoisovalerate, assisted by the cofactor 5,10-methylene-5,6,7,8-tetrahydrofolate. In order to determine the mode of cofactor binding to the protein, ketopantoate hydroxymethyltransferase was crystallized in the presence of two 5,10-methylene-5,6,7,8-tetrahydrofolate analogues and α-ketoisovalerate. X-ray diffraction patterns, collected on the in-house X-ray diffraction data collection facility, extended to $\sim$4.0 Å. Unit cell dimensions derived from these diffraction patterns indicate an asymmetric unit with one decameric enzyme.

A detailed comparative structural analysis of the fold of ketopantoate hydroxymethyltransferase was carried out. Based on this investigation it was possible to assign the enzyme to the phosphoenolpyruvate/pyruvate enzyme superfamily. Furthermore, similarities in the mode of ligand binding to the catalytic magnesium, as well as differences in the mechanisms between the enzymes within this superfamily could be delineated.

In common with a small, but widely distributed, group of mechanistically-related enzymes, L-aspartate-α-decarboxylase is translated as an inactive pro-enzyme, which self-processes at a specific site. In this process of intra-molecular protein maturation a covalently bound pyruvoyl cofactor is formed. A fast purification system for eight L-aspartate-α-decarboxylase mutants was established that allows production of large amounts of enzyme. In order to gain insights into the molecular mechanism of self-processing, crystallographic studies were carried out. Several of the purified mutants have been crystallized. X-ray diffraction data from glycine 24 to serine and serine 25 to threonine mutants were collected, to a maximum resolution of 1.26 Å. The respective crystal structures were solved by molecular replacement. Along with the structures of an unprocessed, native precursor form of L-aspartate-α-decarboxylase and a serine 25 to alanine mutation, the structure models were refined and evaluated, and the models deposited in the Protein Data Bank.

Analysis of these four structures together with four other L-aspartate-α-decarboxylase mutant structures revealed specific conformational constraints on the self-processing mechanism. Threonine 57 and a water molecule could be identified as catalytic elements, most likely essential for acid-base catalysis, and stabilization of the oxyoxazolidine intermediate in the self-processing reaction. A molecular mechanism for self-processing in L-aspartate-α-decarboxylase, largely based on the threonine 57 and a water molecule, is proposed. The differences in the structures of the cleavage site of the serine 25 to alanine and serine 25 to threonine mutants, relative to the structure of the unprocessed native precursor, suggest that molecular models of the cleavage site and mechanisms, based solely on serine to alanine and serine to threonine mutants, may lead to erroneous interpretations of the mechanism. On comparison with other self-processing systems, particularly, glycosylasparaginase, remarkable parallels in the structural features of the environment of the cleavage site were identified.

List of original scientific publications

List of publications and conference proceedings the work, described in this thesis, has contributed to or resulted in.

- Lobley C.M.C., Schmitzberger F., Kilkenny M.L., Whitney H., Ottenhof H.H., Chakauya E., Webb M.E., Birch L.M., Tuck K.L., Abell C., Smith A.G., and Blundell T.L. (2003) Structural insights into the evolution of the pantothenate biosynthesis pathway. *Biochemical Society Transactions*, **31**:563-571.

- Schmitzberger F., Smith A.G., Abell C., and Blundell T.L. (2003) Comparative analysis of the *Escherichia coli* ketopantoate hydroxymethyltransferase crystal structure confirms that it is a member of the $(\beta\alpha)_8$ phosphoenolpyruvate/pyruvate superfamily. *Journal of Bacteriology*, **185**:4163-71.

- von Delft F., Inoue T., Saldanha S.A., Ottenhof H.H., Schmitzberger F., Birch L.M., Dhanaraj V., Witty M., Smith A.G., Blundell T.L., and Abell C. (2003) Structure of *E. coli* ketopantoate hydroxymethyltransferase complexed with ketopantoate and Mg^{2+}, solved by locating 160 selenomethionine sites. *Structure*, **11**:985-96.

- Schmitzberger F., Kilkenny M.L., Lobley C.M.C., Webb M.E., Vinkovic M., Matak-Vinkovic D., Witty M., Chirgadze D.Y., Smith A.G., Abell C., and Blundell T.L. (2003) Structural constraints on protein self-processing in L-aspartate-α-decarboxylase. *The EMBO Journal*, **22**:6193-6204.

- Ottenhof H.H., Ashurst J.L., Whitney H.M., Saldanha S.A., Schmitzberger F., Gweon H.S., Blundell T.L., Abell C., and Smith A.G. (2004) Organisation of the pantothenate (vitamin B_5) biosynthesis pathway in higher plants. *The Plant Journal*, **37**:61-72.

List of abbreviations

For amino acids, the standard one and three-letter abbreviation-code is used. For latin names of the organism, the first name will be abbreviated after the first mentioning in the text and in figures.

α-kiva	α-ketoisovalerate
ADC	L-aspartate-α-decarboxylase
AdometDC	*S*-adenosylmethionine decarboxylase
AMP	adenosine monophosphate
Å	Ångström units
APS	ammonium persulphate
ATP	adenosine triphosphate
ArgDC	arginine decarboxylase
A24a-S25b	L-aspartate-α-decarboxylase mutant with an alanine insertion before serine 25
CDP	cysteine diphosphate
CTP	cysteine triphosphate
DDGA	2-dehydro-3-deoxy-galactarate aldolase
DNA	desoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacidic acid

5,10-meTHF	N_5,N_{10}-methylene-5,6,7,8-tetrahydrofolate
G24S	glycine 24 to alanine mutant of L-aspartate-α-decarboxylase
H-bond	hydrogen bond
H11A	histidine 11 to alanine mutant of L-aspartate-α-decarboxylase
Hepes	N'-2-hydroxyethylpiperazine N-2-ethanesulphonic acid
HisDC	histidine decarboxylase
ICL	isocitrate lyase
IPTG	isopropyl-β-D-thiogalactopyranoside
K9Q	lysine 9 to glutamine mutant of L-aspartate-α-decarboxylase
KPHMT	ketopantoate hydroxymethyltransferase
KPR	ketopanoate reductase
LB medium	Luria-Bertani medium
OD_{600}	optical density at 600 nm
NADPH	nicotinamide adenine dinucleotide phosphate
Ntn	N-terminal nucleophile
PAGE	polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PEG	polyethylene glycol
PEP	phosphoenolpyruvate
PEPC	PEP carboxylase
PEPM	PEP mutase
PK	pyruvate kinase
PPDK	pyruvate phosphate dikinase
PPi	pyrophosphate
pro-ADC	unprocessed L-aspartate-α-decarboxylase enzyme
PS	pantothenate synthetase
R54K	arginine 54 to lysine mutant of L-aspartate-α-decarboxylase
R.m.s.d.	root mean square deviation
RNA	ribonucleic acid
R.p.m.	rounds per minute

SDS	sodium dodecyl sulfate
S25A	serine 25 to alanine mutant of L-aspartate-α-decarboxylase
S25a-A25b	L-aspartate-α-decarboxylase mutant with an alanine insertion after serine 25
S25T	serine 25 to threonine mutant of L-aspartate-α-decarboxylase
S25C	serine 25 to cysteine mutant of L-aspartate-α-decarboxylase
SRS	synchrotron radiation source
(6S)-5-foTHF	(6*S*)-5-formyl-5,6,7,8-tetrahydrofolic acid
(6S)-5-meTHF	(6*S*)-5-methyl-5,6,7,8-tetrahydrofolic acid
TEMED	N,N,N',N'-tetramethylethylenediamine
THF	tetrahydrofolic acid
TIM	triosephosphate isomerase
TLS	translation, libration, screw
Tricine	tris-hydroxymethylmethylglycine
Tris	tris-hydroxymethylaminomethane
Y58F	tyrosine 58 to phenylalanine mutant of L-aspartate-α-decarboxylase
UV	ultraviolett
v/v	volume/volume
v	version
w/v	weight/volume

Part I

Introduction

1 Small molecule metabolic pathways

1.1 Biological metabolism

All living biological organisms require a continuous input of free energy to sustain their own and their genes' existence on this planet. In order to accomplish the variety of tasks that the biological cells of these organisms have to carry out, the free energy has to be converted and distributed. This occurs in the form of chemical molecules within and between the cells in specific ways (see Stryer, 1995, pg. 91-100, for an introduction). Both the transport and the conversion of free energy are predominantly facilitated by specific proteins, and enzymes, which are the most common biological catalysts. The accompanying chemical processes are coordinated in space and time, often in highly integrated and specific networks of (bio)chemical reactions. These biological processes are collectively known as *metabolism*.

Metabolism can be viewed as being schematically organized in interconnected functional units, which are called pathways (see Schilling *et al.*, 2000; Schilling and Palsson, 1998, for a mathematical definition). The partition of the metabolism into distinct pathways is somewhat arbitrary (Ger-

rard *et al.*, 2001), in that it is often difficult to define a beginning and an end of a pathway. As a consequence, the definition of a pathway is not straightforward and several may exist. However, a pathway can, generally, be defined as a series of coupled and consecutive biochemical reactions, the enzymes of which are expressed together (Schuster *et al.*, 2000; Selkov *et al.*, 1998). The degradative processes breaking down more complex molecules to simpler compounds, to release free energy, are referred to as catabolic metabolism, whereas those consuming free energy to build more complex molecules from simpler ones are defined as anabolic metabolism.

More than two thousand different chemical reactions take place in as simple an organism as *Escherichia coli* (Karp *et al.*, 2002c). Whereas the complexity of metabolism may appear overwhelmingly vast, closer investigation reveals that metabolism contains many common motifs (Rison and Thornton, 2002; Teichmann *et al.*, 2001b). Consequently, although the overall number of biochemical reactions in metabolism is large, the number of unique reactions is relatively small (Stryer, 1995, p.443). It appears that biochemical pathways require a small number of different enzyme types to accomplish the chain of chemical transformations essential for life (Tsoka and Ouzounis, 2001). Likewise, related mechanisms of regulation can be observed in many different metabolic pathways. Activation and inhibition of genes and proteins by biochemical molecules, and post-translational modifications of RNA and proteins are the main mechanisms for regulation of metabolism.

1.2 Small molecule metabolism

Metabolism can be schematically divided into primary and secondary metabolism. Primary metabolism forms the core, and includes the sum of the essential processes that keep every cell alive. By and large, primary me-

tabolism is shared by all cells within an organism. Although it is modified by losses, substitutions and innovations during evolution (Dandekar *et al.*, 1999; Huynen *et al.*, 1999; Makarova *et al.*, 1999), it has remained largely conserved in all living eukaryotic and prokaryotic organisms. In contrast, secondary metabolism is usually highly organism- as well as cell-specific. In general, it involves more complex molecules, which are not directly essential for the function of the cell but for the survival of the organism. Examples of secondary metabolites are antibiotics and toxins. An elementary part, predominantly of the primary biological metabolism, is formed by the *small molecule* metabolic pathways. These refer to the metabolism involving small, non-polymeric, predominantly organic, chemical compounds as substrates. Examples of small molecules are glucose or pyruvate, but the definition does not include biopolymers, such as proteins, DNA, or RNA (for an approximate partition of the small molecule metabolism see [1]).

Small molecule metabolism is composed of a complex network of enzymes, substrates, and cofactors, the specific characteristics of which can differ between organisms, particularly between eukaryotes and prokaryotes. Traditionally, the metabolic pathways of the prokaryotic organism *E. coli* have been among the most thoroughly studied and consequently the most popular ones. Almost all of the enzymes in small molecule metabolism have been characterized, and their positions in a pathway verified, experimentally (Riley, 1998)[2]. Additionally, the entire *E. coli* genome has been sequenced (Blattner *et al.*, 1997), and several aspects of the global properties of the organism's metabolic map, such as the frequency of multifunctional enzymes (∼16%), have been determined (Ouzounis and Karp, 2000; Simeonidis *et al.*, 2003; Tsoka and Ouzounis, 2001).

The EcoCyc database[3] (Karp *et al.*, 2002a) provides comprehensive, and

[1] http://www.genome.ou.edu/act/small_molecule_metabolism.html
[2] http://genprotec.mbl.edu/
[3] http://www.ecocyc.org/

presumably close to complete, information on the 106 small molecule metabolic pathways in *E. coli* and the corresponding 581 enzymes as well as their genes. Several other databases are available, such as the PATHWAY database from KEGG[4] (Kanehisa *et al.*, 2002) and the WIT database[5] (Overbeek *et al.*, 2000), which pertain not only to *E. coli*, but to several other organisms, mostly microorganisms and pathogenic bacteria. Although attempts are under way to catalogue similar information for higher, eukaryotic organisms, such as *Homo sapiens*[6] (Karp *et al.*, 2002b), comprehensive information for these is not available yet. However, in a recent comparison of the small molecule metabolism, using the KEGG database, it was shown that 48 pathways, or a core of 271 enzymes have been conserved between *E. coli* and *Saccharomyces cerevisiae* (Jardine *et al.*, 2002). Consequently, more than half of all proteins involved in small molecule metabolism appear to have remained conserved, since the separation of prokaryotes and eukaryotes, and carry out common reactions. In contrast, in the part of the small molecule metabolism that varies, very different extensions to it have been made in the two organisms.

1.3 General characteristics of the small molecule metabolic pathways

A remarkable molecular plasticity (i.e. the same biochemical reactions being catalyzed by different enzymes) and functional versatility (i.e. the association of enzymes with distinct biochemical reactions and pathways) (Ouzounis and Karp, 2000) exist in some pathways, even as central as the glycolysis pathway (Huynen *et al.*, 1999). These features can not only be observed in between organisms, but also within a single species (Dandekar *et al.*, 1999; Tsoka and Ouzounis, 2001). Alternative routes of sub-

[4] http://gpath.ym.edu.tw:8090/kegg/kegg2.html
[5] http://wit.mcs.anl.gov/WIT/
[6] http://metacyc.org/

strate flux and multiple functions of (iso)enzymes within the same organism, as well as varying enzyme patterns in different organisms can be found more frequently than expected. These features imply that some pathways are more stable and relatively immune to failures that could result from malfunctioning of some of the enzymes.

Almost half of the *E. coli* proteins involved in small molecule metabolism are composed of a single protein domain (Teichmann *et al.*, 2001a), which, following the convention in the SCOP database[7] and 3Dee database[8], is the basic structural and evolutionary unit of the protein. Insights into the overall organization of the small molecule metabolic pathways were gained by studying the distribution of proteins, classified into evolutionarily related protein-(super)families, according to their amino acid sequence similarity and/or structural homology, in the small molecule metabolism (Teichmann *et al.*, 2001b). A relatively small repertoire of 213 protein-domain families constitute 90% of the enzymes in the small molecule metabolism (Teichmann *et al.*, 2001a). Interestingly, it was found that *E. coli* enzymes belonging to the same protein-family are widely distributed among different pathways. Protein homologues were also observed to be twice as likely to be found in different pathways than in the same pathway (Teichmann *et al.*, 2001b). In contrast, homology between enzymes within the same small molecule metabolic pathway was generally rare. However, if observed, homologous proteins tend to conserve their chemical reaction/catalytic mechanism and/or cofactor binding rather than similar substrate recognition (Teichmann *et al.*, 2001b). Another strong correlation appears to exist between the pathway distance, as defined by the number of steps by which two enzymes are separated in metabolism, and the gene interval (Rison *et al.*, 2002). Genes, which were found to be localized closer to each other on the *E. coli* chromosome, tend to encode enzymes catalyzing biochemical reactions that are separated by a smaller number of steps in metabolism. Several blocks of co-regulated

[7] http://scop.mrc-lmb.cam.ac.uk/scop/
[8] http://www.compbio.dundee.ac.uk/3Dee/

genes, similar to operons, of enzymes with a small pathway distance were observed. These enzymes also seem to have a higher likelihood of being functionally, but not necessarily evolutionarily related (Rison *et al.*, 2002).

1.4 Evolution of the small molecule metabolic pathways

E. coli is a representative of the descendants of the metabolically competent last common ancestor. The universal presence of many proteins of the primary metabolism in the cell suggests that much of the basic protein repertoire was developed in organisms very much simpler than any known at present. Apart from the *ab initio* invention of a set of different domains, which most likely occurred at an early stage of the development in life, the mechanisms that generate the protein repertoire in metabolism are dominated by gene duplication, divergence, and recombination (Babbitt and Gerlt, 2000; Chothia *et al.*, 2003). In the context of pathway evolution, the recruitment of proteins [or domains] across pathways is a very common phenomenon and explains the absence of regularity in the pattern of proteins, found in metabolic pathways (Teichmann *et al.*, 2001b).

New enzymes are more likely to have evolved from enzymes belonging to the same enzyme family or class, than from the same pathway. This implies that an evolutionary mechanism, known as patchwork approach, predominates, where enzymes are widely recruited from different pathways with no particular order or bias in respect to their position in the metabolic network (Copley, 2000). Several examples of homology within the metabolic neighbourhood however indicate that most likely this is not the only mechanism, and that retrograde evolutionary mechanisms are also active. The patchwork mechanism proposed by Jensen (Jensen, 1976), implies chemistry-driven evolution, whereas Horowitz theory of retro-

gade evolution mechanism (Horowitz, 1945), in which enzymes within a pathway are related to each other, would seem to be substrate-driven. Overall, it appears that both pathway evolution mechanisms occur in concert (see Alves *et al.*, 2002), although enzymes seem to be mainly selected for their mechanism and function.

2 Pantothenate and Coenzyme A biosynthesis

2.1 Pantothenate biosynthesis

Pantothenate is a central and vital metabolic compound in all organisms (reviewed by Tahiliani and Beinlich (1991)). It constitutes the invariant presursor of 4'-phosphopantetheine, the moiety of coenzyme A and the acyl carrier protein. However, pantothenate is synthesized in bacteria, fungi, and plants only; animals have lost the corresponding biosynthetic pathway. As such, pantothenate is classified as a vitamin (vitamin B_5). Consequently, enzymes of the pathway are considered attractive targets for the development of selectively toxic antimicrobial, fungicidal, and herbicidal drugs.

The pantothenate pathway is an anabolic, primary, small molecule metabolic pathway. It is reasonable to assume that the pantothenate pathway evolved rather early in the development of life on Earth, and experiments support the notion that pantothenate may have existed in the prebiotic soup (Keefe *et al.*, 1995). Whereas the biochemical steps in *E. coli* and *Salmonella typhimurium* in the biosynthesis of pantothenate have es-

sentially been known for a longer time (Maas, 1952a), only more recently have detailed characterization and investigation of the enzymes been pursued.

In *E. coli* the pantothenate pathway comprises four enzymatic steps (Figure 2.1). R-pantothenate is synthesised from α-ketoisovalerate (α-kiva) and L-aspartate in a converging four-step process (Brown and Williamson, 1982). The transfer of the C_{11} group of N_5,N_{10}-methylene tetrahydrofolate (5,10-meTHF) on to α-kiva to form ketopantoate (4-hydroxy-3,3,-dimethyl-2-oxobutanoate), by ketopantoate hydroxymethyltransferase (EC 2.1.2.11, KPHMT), is traditionally referred to as the first committed reaction (Teller *et al.*, 1976). α-kiva itself is predominantly supplied as an intermediate of pyruvate-dependent L-valine and L-leucine biosynthesis (Figure 2.1). However, in *E. coli* only a comparatively small fraction of $\sim$10% of the total flux of α-kiva through this pathway is consumed by conversion into ketopantoate (Cronan *et al.*, 1982). Subsequent to the first step, ketopantoate reductase (EC 1.1.1.169, KPR) reduces ketopantoate to pantoate, using NADPH as a cofactor (Elischewski *et al.*, 1999; Frodyma and Downs, 1998a,b; King *et al.*, 1974; Wilken *et al.*, 1975; Zheng and Blanchard, 2000a,b, 2003). In addition, another gene product has been implicated with ketopantoate processing. The primary reaction of acetohydroxy acid isomeroreductase (Biou *et al.*, 1997) (see Figure 2.1), encoded by the *ilvC* gene, is the reduction of α-acetolactate and α-aceto-α-hydroybutyrate to α,β-dihydroxyisovalerate and α,β-dihydroxy-β-methylvalerate, respectively in the above mentioned biosyntheses of L-valine and L-leucine. This enzyme was shown to rescue KPR auxotrophs with low, but significant, KPR activity (Primerano and Burns, 1983). Concomitant with the reduction of ketopantoate, in a separate branch of the pathway, L-aspartate-α-decarboxylase (EC 4.1.1.11, ADC) converts L-aspartate to β-alanine (Cronan, 1980). β-alanine production uses less than 2% of the total amount of cellular L-aspartate in *E. coli* (Cronan *et al.*, 1982). In the final step of the pantothenate pathway, β-alanine is coupled to the ATP activated C_1 of L-pantoate with an amide bond by pantothenate synthetase (EC 6.3.2.1, PS)

yielding R-pantothenate (Figure 2.1) (Maas, 1952b; Miyatake *et al.*, 1976, 1978, 1979; Zheng and Blanchard, 2001).

The cellular synthesis of β-alanine appears to be the limiting factor in pantothenate production in *E. coli*, implying the accumulation of free pantoate (Cronan, 1980; Dusch *et al.*, 1999). Nevertheless, in *S. typhimurium* a mutation in the promoter gene sequence for *panB* resulted in overexpression of KPHMT and led to an increase of the pantothenate level (Rubio and Downs, 2002). However, both *E. coli* and *Bacillus subtilis* overproduce and excrete significant amounts of pantothenate (Baigori *et al.*, 1991; Cronan *et al.*, 1982; Jackowski and Rock, 1981). The excess, which in *E. coli* can even reach fifteen-fold requirement levels, is exported efficiently (Cronan *et al.*, 1982; Jackowski and Rock, 1981).

The abundance of pantothenate in these organisms indicates that the pathway is not particularly tightly regulated or controlled. A possible explanation for this could be that the metabolic cost of regulating the pathway could exceed the cost of overproduction. In this respect, it is interesting that extracellular uptake of pantothenate also occurs. In *E. coli* this was shown to be facilitated by pantothenate permease, an inner membrane protein encoded by the *panF* gene (Jackowski and Alix, 1990; Vallari and Rock, 1985). This pantothenate carrier concentrates the vitamin by a sodium co-transport mechanism, and overexpression of the *panF* gene resulted in elevated intracellular levels of pantothenate. A pantothenate transporter importing extracellular pantothenate was also identified in *S. cerevisiae* (Stolz and Sauer, 1999) and appears to exist in animals (Barbarat and Podevin, 1986; Lopaschuk *et al.*, 1987; Wang *et al.*, 1999). A different H^+-coupled pantothenate transport was identified in the malaria parasite *Plasmodium falciparum* (Saliba and Kirk, 2001).

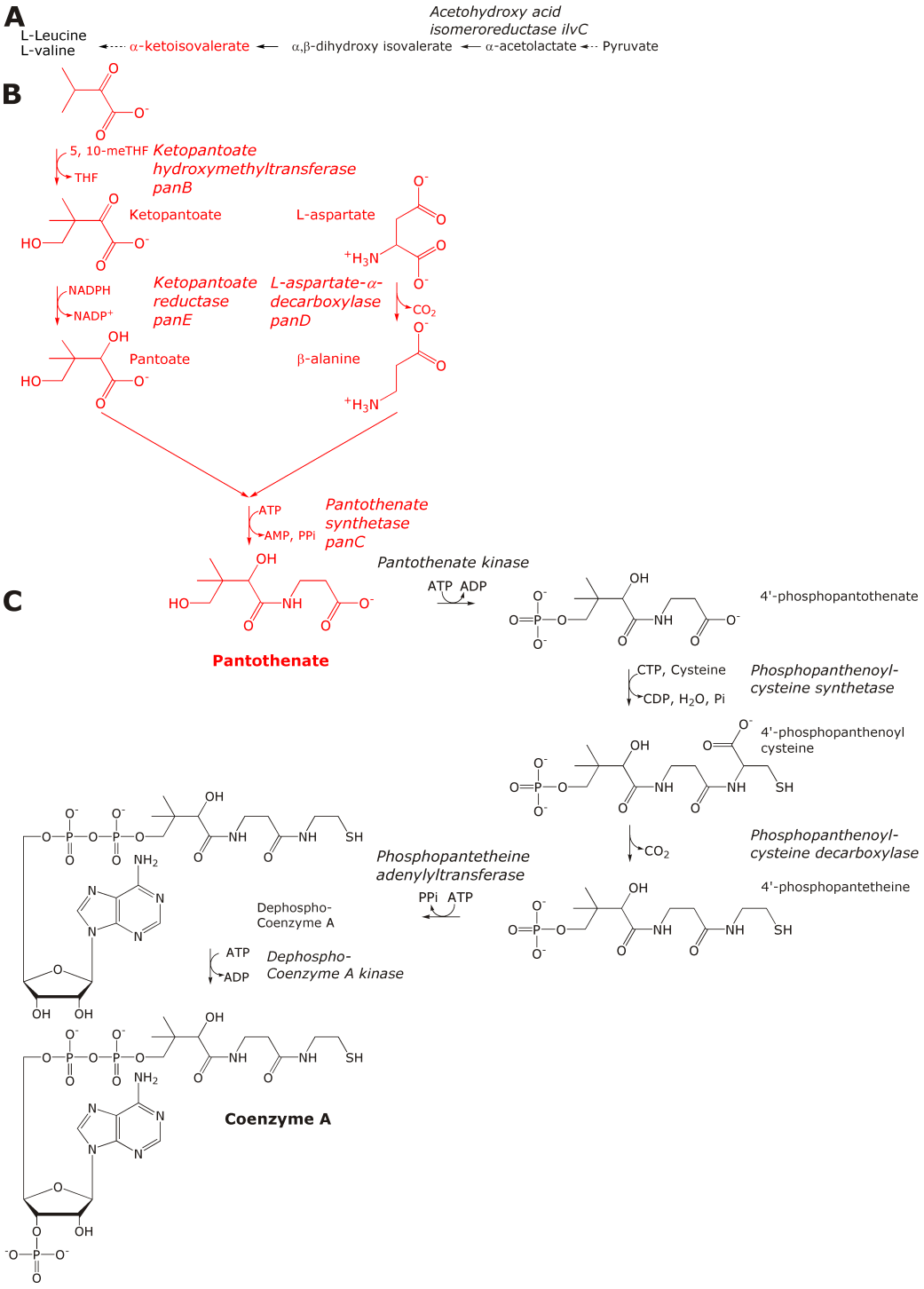

Figure 2.1 Schematic representation of pantothenate and coenzyme A biosynthesis. The names of the enzymes and gene designation are shown in italic font; substrates, products, and cofactors and their turnover are shown in normal font. Dashed arrows in **A** indicate several enzymatic steps. (**A**) α-kiva as an intermediate of L-valine and L-leucine biosynthesis. (**B**) The biosynthesis of pantothenate (colour, red). (**C**) The conversion of pantothenate into coenzyme A.

2.1.1 Organization, characteristics, and occurrence of the *pan* genes

As mentioned in Section 2.1, the pantothenate pathway is not present in animals, and, consequently, *H. sapiens*. However, enzymatic activities or relevant genes have been found in eubacteria as well as eukaryotes, such as fungi and plants.

The genes, and their organization, are characterized best in *E. coli* and *S. typhimurium* (Jackowski, 1996) of which the nomenclature *panB*, *panC*, *panD*, and *panE*, corresponding to KPHMT, PS, ADC, and KPR, respectively has been derived. In *E. coli* the *panB*, *panD*, and *panC* genes were shown to be co-localised (Cronan, 1980), in this clockwise order (Cronan *et al.*, 1982; Merkel and Nichols, 1996). The organization in the genome suggests a, possibly ancestral, operon, which has been shown to be located at 3 min on the genetic map. In *B. subtilis* and *Corynebacterium glutamicum* an identical *panBC* gene cluster exists (Sahm and Eggeling, 1999; Sorokin *et al.*, 1996), and in the former *panD* is also co-localised with this cluster. The *panD* gene of *Corynebacterium glutamicum* has also been cloned, but was not observed to be localised close to the other *pan* genes (Dusch *et al.*, 1999). Cloning of the *panD* gene from *Ralstonia eutropha* has also been reported (Hoppensack *et al.*, 1999). *panE* was identified, cloned and characterized rather late, compared with *panB*, *panC*, and *panD*, possibly because of the existence of *ilvC* (Section 2.1) and confusing reports about the enzymatic function of *apbA* in *S. typhimurium*. The latter gene eventually turned out to be identical to *panE* (Frodyma and Downs, 1998b). It maps at 10 min at both the *E. coli* and *S. typhimurium* genome (Frodyma and Downs, 1998a). The *panE* gene product has furthermore been purified and partially characterized from *Pseudomonas maltophilia 845* (Shimizu *et al.*, 1988) and *Nocardia asteriodes* (Kataoka *et al.*, 1992). In *C. glutamicum* the *ilvC* gene product alone appears to account for ketopantoate reductase activity (Merkamm *et al.*, 2003). However, in general, the key residues of the

sequences of the *pan* genes seem to be well conserved among bacteria (see amino acid sequence alignments in Section 11.3 and Section 13.3).

Pantothenate auxotrophs of *Aspergillus nidulans* were first observed by Kaefer (1958). The corresponding mutants were then identified and genetically localized as *panB* (Rever, 1965) and *panC* (Bal *et al.*, 1977). More recently, *panB* has been cloned and overexpressed from *A. nidulans* by Kurtov *et al.* (1999). Further known fungal genes include *panB* and *panC* from *S. cerevisiae* and *Schizosaccharomyces pombe* as well as *panB* from *Emericella nidulans*. In addition, the *panE* product was reported to have been purified and partially characterized from *S. cerevisiae* (King *et al.*, 1974; King and Wilken, 1972).

Evidence for the pathway in plants was provided by Jones *et al.* (1994) and subsequently the *panB* and *panC* genes from *Arabidopsis thaliana* were cloned and characterized (Jones *et al.*, 1993; Ottenhof *et al.*, 2004). In addition, the *panC* gene has been cloned and the gene product characterized from both *Lotus japonicus* and *Oryza sativum* (Genschel *et al.*, 1999). Indications for the existence of enzymatic activities in the pantothenate pathway are also known from *Datura innoxia* (Savage *et al.*, 1979) and *Pisum sativum* (Jones *et al.*, 1994). However, the identification of *panE* from plants has remained elusive so far, despite extensive database searches of fully sequenced genomes (Ottenhof *et al.*, 2004).

Similarly, so far, no evidence for a *panD* gene in cells of eukaryotic organisms has been found. This may not that surprising, since it is known that there are two known alternative ways of supplying β-alanine (Campbell, 1957; Canellakis, 1956; Fink *et al.*, 1953; Rendina and Coon, 1957). It is well established that in fungi and plants β-alanine can be physiologically derived in cells by the reductive degradation of pyrimidines, such as uracil (Slotnick, 1956). Although evidence remains inconclusive, current knowledge (Ottenhof *et al.*, 2004) suggests that the *panD* gene probably is completely absent from eukaryotes.

2.2 Coenzyme A biosynthesis

Coenzyme A is invariably bio-synthesized from pantothenate, constitutes the principal and universal carrier of acyl groups (Stryer, 1995), central to anabolic (e.g. fatty acid synthesis) and catabolic (e.g. glycolysis) metabolism. It features a terminal thiol group (Figure 2.1) that readily forms labile, and consequently reactive, thioester bonds with a variety of acyl groups, such as acetyl or malonyl, thereby activating these for biological Claisen reactions. Coenzyme A is critical to cell growth and function. In particular, the 4'-phosphopantetheine group of coenyzme A is important in fatty acid biosynthesis and degradation, in the biosynthesis of polyketides as well as in nonribosomal peptide biosynthesis (Kleinkauf and von Dohren, 1995; Kleinkauf and Von Dohren, 1996). 4'-phosphopantetheine itself constitutes an essential prosthetic group in a number of enzyme systems including acyl carrier protein components of bacterial and eukaryotic fatty acid synthases. Metabolically, coenzyme A is designated as a cofactor. In general, it is estimated that ~4% of all enzymes use coenzyme A or 4'-phosphopantetheine-containing compounds as cofactors in their reactions (Webb, 1992).

As first shown by Brown (1959) and Abiko (1975), coenzyme A is biochemically synthesized in five enzymatic steps from pantothenate (Figure 2.1). In contrast to the biosynthesis of the latter, the biosynthetic steps from pantothenate to coenzyme A do exist in higher eukaryotic organisms and animals, such as *H. sapiens*. Based on this fact, the whole biosynthesis of coenzyme A in bacteria can therefore schematically be divided into two phases (Figure 2.1), the synthesis of pantothentate and its subsequent conversion into coenzyme A (reviewed by Begley *et al.* (2001) and Jackowski (1996)). However, the coenzyme A pathway appears to be highly conserved throughout all taxons.

The continuation of the pantothenate pathway into the coenzyme A

pathway in *E. coli* begins with the phosphorylation of the hydroxyl group of pantothenate by pantothenate kinase resulting in the formation of 4′-phosphopantothenate (Figure 2.1) (Calder *et al.*, 1999; Rock *et al.*, 2000; Song and Jackowski, 1992a,b, 1994; Yun *et al.*, 2000). Subsequently, 4′-phosphopantothenate is ligated to L-cysteine by phosphopantothenate cysteine synthetase to produce 4′-phosphopantothenoylcysteine (Kupke, 2002; Strauss *et al.*, 2001), the cysteine moiety of which becomes decarboxylated by phosphopantothenoylcysteine decarboxylase (EC 4.1.1.36) to 4′-phosphopantetheine (Kupke, 2001; Kupke *et al.*, 2001; Strauss and Begley, 2001). In *E. coli*, the enzymatic acitivities of the latter two proteins are located in one bi-functional enzyme, the N-terminal domain carrying out the decarboxylation (Kupke *et al.*, 2000; Spitzer *et al.*, 1988) and the C-terminal domain the ligation to the cysteine. In the penultimate step of coenzyme A biosynthesis, phosphopantetheine adenylyltransferase catalyzes the reversible transfer of an adenylyl group from ATP to 4′-phosphopantetheine forming dephospho-coenzyme A (Geerlof *et al.*, 1999; Izard and Blackwell, 2000; Izard and Geerlof, 1999). Finally, the ribose 3′-hydroxyl group of dephospho-coenzyme A is phosphorylated by dephospho-coenzyme A kinase (Figure 2.1) to form coenzyme A (Mishra *et al.*, 2001; O'Toole *et al.*, 2003).

A genetic footprinting study by Gerdes *et al.* (2002) confirmed the indispensable importance of the coenzyme A biosynthesis pathway for metabolism. Genes encoding the enzymes for the five catalytic steps in the coenzyme A pathway have been identified and characterized in *E. coli* (Kupke *et al.*, 2001), *H. sapiens* (Aghajanian and Worrall, 2002; Daugherty *et al.*, 2002; Zhyvoloup *et al.*, 2002), and *A. thaliana* (Kupke, 2002). However, these and other studies observed significant differences, both in the gene sequence and in the organization of the gene products of the enzymes in the coenzyme A pathway between prokaryotic and eukaryotic organisms. Cloning and analysis of the *A. nidulans* (Calder *et al.*, 1999) and two *Mus musculus* (Rock *et al.*, 2000) pantothenate kinases showed that neither enzyme bore significant resemblance in primary sequence to the prokaryotic

enzyme. Phosphopantothenate cysteine ligase and phosphopantothenoylcysteine decarboxylase are two separate enzymes in eukaryotes that have little sequence similarity to the respective enzymes in bacteria (Manoj *et al.*, 2003). Whereas in bacteria, such as *E. coli* and *Corynebacterium ammoniagenes* (Martin and Drueckhammer, 1993), separate enzymes exist for the last two biosynthetic steps in coenzyme A synthesis, in higher eukaryotic organisms, such as mammals and flies, both phosphopantetheine adenylyltransferase and dephospho-coenzyme A kinase reactions are catalyzed by a single bi-functional enzyme. This enzyme is designated coenzyme A synthase (Aghajanian and Worrall, 2002; Daugherty *et al.*, 2002; Hoagland and Novelli, 1954; Suzuki *et al.*, 1967; Worrall *et al.*, 1985; Worrall and Tubbs, 1983; Zhyvoloup *et al.*, 2003). In *S. cerevisiae*, coenzyme A synthase is part of a multi-enzyme complex (Bucovaz *et al.*, 1997). As implied above, in general, sequence similarity between the orthologous enzymes in eukaryotes and prokaryotes of the coenzyme A pathway is low.

All of the *E. coli* enzymes catalyzing reactions in coenzyme A biosynthesis have been functionally characterized. Correlation of these results with structural information, although not exclusively from *E. coli* nor from any single organism, is also possible for these enzymes. Crystal structures have been determined for *E. coli* pantothenate kinase (Yun *et al.*, 2000), phosphopantothenate cysteine ligase from *H. sapiens* (Manoj *et al.*, 2003), phosphopantothenoylcysteine decarboxylase from *A. thaliana* (Albert *et al.*, 2000), phosphopantetheine adenylyltransferase from *E. coli* (Izard, 2003; Izard and Blackwell, 2000; Izard and Geerlof, 1999), and dephospho-coenzyme A kinase from both *E. coli* (O'Toole *et al.*, 2003) and *Haemophilus influenzae* (Obmolova *et al.*, 2001).

2.2.1 Regulation of coenzyme A biosynthesis

In contrast to the pantothenate pathway, which does not appear to be particularly regulated, regulation of the coenzyme A biosynthesis is more tightly controlled.

Metabolic experiments in *E. coli* (Jackowski and Rock, 1981) and of myocytes of *Rattus norvegicus* (Robishaw *et al.*, 1982; Robishaw and Neely, 1984, 1985) showed that the utilization rather than the supply of pantothenate controls the rate of coenzyme A biosynthesis. Enhanced expression of pantothenate kinase in *M. musculus* cells eliminated the intracellular pantothenate pool and triggered a thirteen-fold increase in intra-cellular coenzyme A content (Rock *et al.*, 2000). This is consistent with reports about the positive correlation of elevated tissue levels of pantothenate kinase and increased coenzyme A concentration (Fisher *et al.*, 1985; Halvorsen, 1983; Skrede and Halvorsen, 1979). Overexpression of the pantothenate kinase gene also lead to increased pantothenate uptake (Rock *et al.*, 2000). These and other experiments (Jackowski, 1996; Robishaw *et al.*, 1982; Robishaw and Neely, 1985; Rock *et al.*, 2003) have shown that the catalysis by pantothenate kinase is the most important regulatory step, controlling the flux through the coenzyme A biosynthesis pathway in both bacteria and animals. Thereby, non-esterified as well as esterified coenzyme A are acting as a feedback inhibitor of pantothenate kinase by competitive binding to the ATP binding site (Song and Jackowski, 1994; Vallari *et al.*, 1987; Yun *et al.*, 2000). However, bacteria appear far less sensitive to inhibition by acetyl-coenzyme A than to coenzyme A (Rock *et al.*, 2000). In contrast, in eukaryotic pantothenate kinases, acetyl-coenzyme A is the more potent inhibitor (Fisher *et al.*, 1985; Halvorsen and Skrede, 1982) and non-esterified coenzyme A even seems to be able to stimulate pantothenate kinase activity (Calder *et al.*, 1999), although earlier studies reported inhibition by coenzyme A (Falk and Guerra, 1993; Fisher *et al.*, 1985; Karasawa *et al.*, 1972; Kirschbaum *et al.*, 1990; Robishaw and Neely, 1985). The contradic-

tory reports may be due to the existence of several different pantothenate kinase isoforms and different expression patterns in different tissues.

A potential second, although far less significant, regulation step, constitutes the catalysis by phosphopantetheine adenylyltransferase. This was indicated by the accumulation of 4'-phosphopantetheine (Jackowski and Rock, 1984). Regulatory control at this step may also involve the reutilization of 4'-phosphopantetheine, which can arise by the turnover of 4'-phosphopantetheine of the acyl carrier protein or the cleavage of coenzyme A (Jackowski and Rock, 1984; Vallari and Jackowski, 1988). However, Cronan *et al.* (1982) and Frodyma *et al.* (2000) showed that the level of coenzyme A could be significantly, but not proportionally, elevated in pantothenate auxotrophs of *E. coli* by supplementation of pantothenate in the growth medium. Supplementation with β-alanine in the former study and also by Jackowski and Rock (1981) resulted in a similar increase of the intracellular coenzyme A level, whereas supplementation with either pantoate or ketopantoate was less effective. Somewhat contrasting this, results from (Rubio and Downs, 2002) indicated that elevated levels of ketopantoate hydroxymethyltransferase lead to physiologically significant increased coenzyme A concentration in *S. typhimurium*. Similarly, Lopaschuk *et al.* (1986) showed that increasing the exogenous pantothenate supply to heart tissue elevated cellular coenzyme A content.

Older studies reported that mitochondria can transport coenzyme A into the matrix, implying that all coenzyme A biosynthetic enzymes are cytosolic (Neuburger *et al.*, 1984; Tahiliani, 1989, 1991; Tahiliani and Beinlich, 1991; Tahiliani and Neely, 1987). However, because mitochondria and peroxisomes contain the highest concentration of coenzyme A and its thioesters in eukaryotes (Van Broekhoven *et al.*, 1981), it has been proposed that the last two enzymatic activities in the pathway are located inside these compartments (Skrede and Halvorsen, 1983). More recently, this has been convincingly demonstrated for mitochondria. Coenzyme A synthase from *H. sapiens* is indeed located at the mitochondrial outer

membrane, but not on peroxisomes, and the enzyme is regulated by phospholipids (Zhyvoloup *et al.*, 2003). 4'-phosphopantotheine arising from coenzyme A degradation may be expelled from these compartments into the cytosol, where it cannot be further metabolized. One isoform of pantothenate kinase is a cytosolic enzyme (Halvorsen and Skrede, 1982; Robishaw *et al.*, 1982), whereas another localizes to the mitochondria (Hortnagel *et al.*, 2003). The latter results could suggest that mitochondria do have the entire complement of coenzyme A biosynthetic enzymes. However, for the former isoform the biochemical regulators (i.e. Coenzyme A and its thioesters) are largely sequestered from pantothenate kinase, which renders the enzymes incapable of sensing a significant proportion of the total coenzyme A pool directly. In general, the compartmentalization of the coenzyme A biosynthetic pathway and its regulatory effects on the various isoforms of pantothenate kinase are not very well understood features in eukaryotes. The regulation of the coenzyme A content is complex, and likely governed by the interplay between coenzyme A biosynthesis, transport through membranes, and degradation (Ramaswamy *et al.*, 2003).

The metabolic state of the organism or the cell (Kondrup and Grunnet, 1973; Smith *et al.*, 1978; Smith and Savage, 1980), pathological states such as diabetes (Reibel *et al.*, 1981b) as well as drugs (Voltti *et al.*, 1979), can all significantly alter the cellular level of coenzyme A, and these fluctuations are reflected by concomitant changes in the level of tissue pantothenate kinase activity (Halvorsen, 1983). Cells can modulate pantothenate kinase expression to modify the coenzyme A levels in tissue in response to diet and disease. Pantothenate kinase activity in heart cells decreased in response to glucose, pyruvate, fatty acids, or insulin, and hepatic concentrations of coenzyme A increased in response to glucagon, starvation, high-fat diet, or diabetes (Kondrup and Grunnet, 1973; Reibel *et al.*, 1981a,b; Smith and Savage, 1980; Tahiliani and Beinlich, 1991; Voltti *et al.*, 1979). However, to date these regulatory effects are not fully understood.

2.2.2 Pathogenic syndromes of disturbed pantothenate metabolism

Recently, pantothenate kinase was linked to both the neurodegenerative Hallervorden-Spatz syndrome in *H. sapiens*, an autosomal recessive disorder (Cossu *et al.*, 2002; Hayflick, 2003a,b; Hayflick *et al.*, 2003), and to HARP syndrome (Ching *et al.*, 2002). Mutations in the gene for the *H. sapiens* pantothenate kinase isoform, which is located in the mitochondria, create a metabolic imbalance leading to Hallervorden-Spatz syndrome. This is characterized by iron accumulation in the basal ganglia of the brain. Inactivation of the enzyme is the causative agent of progressive childhood Hallervorden-Spatz syndrome and thus the disease was very recently renamed pantothenate kinase-associated neurodegeneration (Zhou *et al.*, 2001). Both pantothenate kinase-associated neurodegeneration and HARP disease indicate that defective expression or sequence alteration of a single pantothenate kinase isoform can selectively impact the function of specific tissues. Pantothenate kinase-associated neurodegeneration bears all the hallmarks of a mitochondrial disorder and could serve as a model for other complex neurodegenerative diseases, such as Alzheimers and Parkinson, since all exhibit pathologic accumulation of iron in the brain (Zhou *et al.*, 2001). Consequently, perturbed pantothenate metabolism could be a general feature of neurodegenerative diseases.

Additionally, mutation in the gene for pantothenate kinase was also associated with aberrant mitosis and meiosis in flies (Afshar *et al.*, 2001).

2.3 Outlook

In order to manipulate, create, or destroy certain functions in metabolism, an understanding of the general organization of metabolism as well as of

the particular biochemical reactions catalyzed by the individual enzymes is necessary. Consequently, one of the most important long term goals of genome sequencing initiatives, proteomics, and structural genomics, is to gain an understanding how the cell works and how it changes and adapts. This should enable a description of molecular function in the global context of the cell and at the whole organism level.

However, whereas large scale initiatives, such as proteomics and (structural) genomics, will be essential for the growing understanding of the metabolic universe, they will never suffice to substitute detailed investigation of individual pathways and their enzymes. Both things are, and will remain complementary. Since a limited set of biochemical reaction appears to exist (see Section 1.3), the results of detailed characterizations of biochemical reactions for individual enzymes always have the potential to be applicable to functionally and/or structurally-related, but yet uncharacterized, enzymes.

With respect to the pantothenate pathway, considerable interest from industry exists in manipulating the pathway to overproduce pantothenate (Dusch *et al.*, 1999). Processed food, animal feed, pharmaceuticals and cosmetics are often supplemented with extra vitamins, which are produced either by chemical synthesis or by fermentation. At least 4 000 000 kg of pantothenate are industrially produced annually (Vandamme, 1992). In order to improve production strains, knowledge of the complete biosynthetic pantothenate pathway is a prerequisite (see Elischewski *et al.*, 1999). Overproduction of pantothenate by engineering overexpression of some of the enzymes in the pantothenate pathway has already been accomplished by Sahm and Eggeling (1999).

Secondly, the pantothenate pathway constitutes a viable target for antimicrobial drugs. The majority of the existing antibiotics utilize a limited number of core chemical structures and target only a few cellular functions as potential targets, such as cell wall synthesis, DNA replication, tran-

scription, and translation. A growing number of antibiotic-resistant microbial pathogens present serious challenges to modern medicine. In the last decade, the emergence of both vancomycin- and methicillin-resistant bacteria has led to an urgent search for novel classes of antibiotics and new anti-microbial targets (Zhao *et al.*, 2003). Investigation into the potential of the enzymes of the pantothenate and coenzyme A biosynthesis pathway as antibacterial targets is now an active area of research. Results from Gerdes *et al.* (2002) and Zhao *et al.* (2003) indicate the value structural information of the enzymes involved in coenzyme A synthesis can have for the development of inhibitors.

The genes in the coenzyme A pathway appear to be significantly different between eukaryotes and prokaryotes, and thus the enzymes potentially can be exploited for the development of drugs. However, the complete absence of the pantothenate pathway in animals may make it more straightforward to identify potent inhibitors for the bacterial enzymes. In the following, I will introduce to the two *E. coli* enzymes of the pantothenate pathway, KPHMT and ADC, that are the focus of study of this thesis.

3 Ketopantoate hydroxymethyltransferase

3.1 Background

Kuhn and Wieland (1942) were the first to describe the synthesis of ketopantoate from α-kiva. Evidence that *E. coli* cells can synthesize ketopantoate from α-kiva, whereas certain pantoate-requiring mutants could not, suggested that α-kiva is the precursor for ketopantoate (Maas and Vogel, 1953). Shortly afterwards, McIntosh *et al.* (1957) reported the partial purification of an enzyme catalyzing this reaction, which required free formaldehyde as a substrate. However, since the Michaelis-Menten K_m values for formaldehyde and α-kiva were rather high, the physiological relevance seemed questionable. Powers and Snell (1976) and Teller *et al.* (1976) re-investigated the reaction and described the purification and biochemical characterization of a different 5,10-meTHF dependent enzyme, which they designated KPHMT, from *E. coli*. The transferase activity was found to be absent in a ketopantoate auxotroph, which together with reasonable Michaelis-Menten constants of $<$1 mM for its substrates, led them to conclude that this is the enzyme involved in pantothenate biosynthesis. Subsequently, the *panB* gene was cloned and the gene product character-

ized from *E. coli* (Jones *et al.*, 1993), *A. nidulans* (Kurtov *et al.*, 1999), and *Mycobacterium tuberculosis* (Sugantino *et al.*, 2003), and respective *panB* clones were overexpressed in *E. coli*. The gene has also been characterized in *S. typhimurium* (Downs and Roth, 1991; Rubio and Downs, 2002), and *Arabidopsis thaliana* (Ottenhof *et al.*, 2004). In the latter organism two genes, both encoding functional KPHMT isoenzymes, were found.

3.2 The transferase reaction of KPHMT

In chemical terms, the reaction catalyzed by KPHMT is a common crossed aldol addition reaction, where a general base deprotonates the carbon adjacent to the ketogroup, followed by the nucleophilic attack of the resulting double bond on a carbonyl group and C-C bond formation. The pK_a value of the α-hydrogen atom in aldehydes and ketones is usually between 19 and 21 (Vollhardt and Schore, 1995), which is far lower than that of usual C-bonded hydrogens. The relatively acidic character of aldehydes and ketones has two reasons, one is the electron-drawing effect of the positively-polarized carbon atom in the aldehyde or ketogroup. The second, and more important, is the resonance or conjugation effect after deprotonation. During deprotonation the σ-orbital of the C-H bond assumes π-orbital character and stands perpendicular to the plane of the sp^2-orbital. This way it can undergo π-orbital conjugation with the π-system of the adjacent carbonyl group. In the case of α-kiva the resulting negative charge of the deprotonation is further dispersed over the carboxylate group, which also provides two electronegative oxygen atoms (Figure 3.1).

The classical aldol addition proceeds with nucleophilic attack of the α-carbon on the π-antibonding orbital of a carbonyl bond. This is followed by C-C bond formation and protonation of the generated negative oxygen atom, which stands in the β-position to the carbonyl of the nucleophile (Figure 3.1). The electrophilic carbonyl carbon can be replaced by other

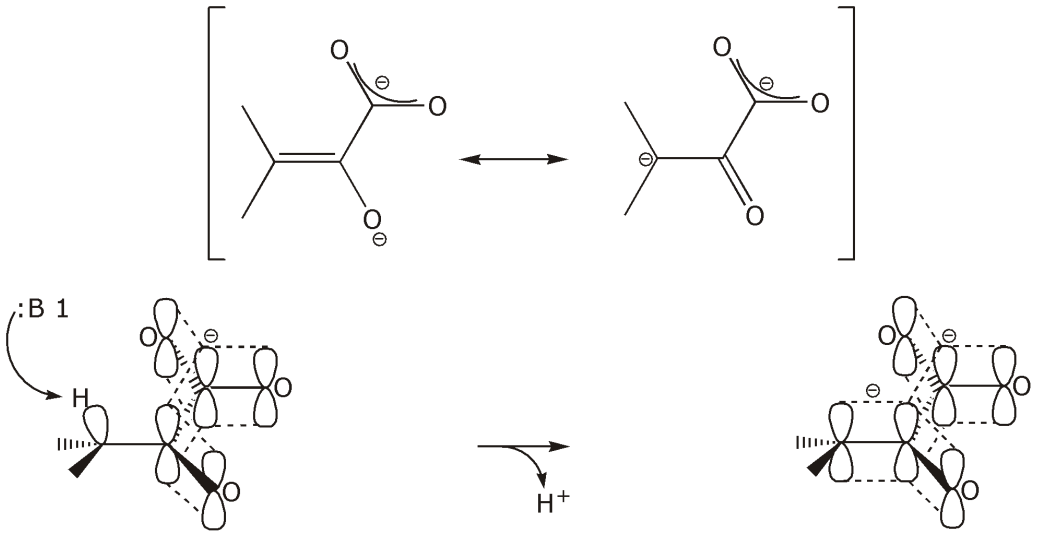

Figure 3.1 Illustration of the deprotonatation of the ketoacid α-kiva. The π-atomic orbitals are indicated to illustrate the π-conjugated system after deprotonation. Two of the most likely resonance states are shown at the top.

electrophiles, which in the case of KPHMT is the 5,10-methylene carbon of 5,10-meTHF. Kallen and Jencks (1966) reported that the reactive component of 5,10-meTHF is the iminium intermediate (Figure 3.2). With regard to the steric course of the KPHMT reaction, the addition of the carbon from 5,10-meTHF proceeds with retention of configuration, as indicated by isotopic labelling experiments (Aberhart and Russell, 1984).

In biological aldol reactions, increased stabilization of the arising negative charge is accomplished by dispersing it even further into a cationic electron sink at the active site of the enzyme (Figure 3.2). Based on the mechanism, two classes of aldolases may be distinguished (Morse and Horecker, 1968). In class I aldolases a Schiff base or imine is formed with an essential ϵ-amino group of a catalytic lysine residue in the enzyme. In contrast, class II aldolases use a divalent metal ion, such as Mg^{2+}, as an electron sink. The imine, formed in class I aldolases, can be reduced by borohydride, which results in irreversible inactivation of the enzyme. In KPHMT no such inactivation could be observed and, thus, it was function-

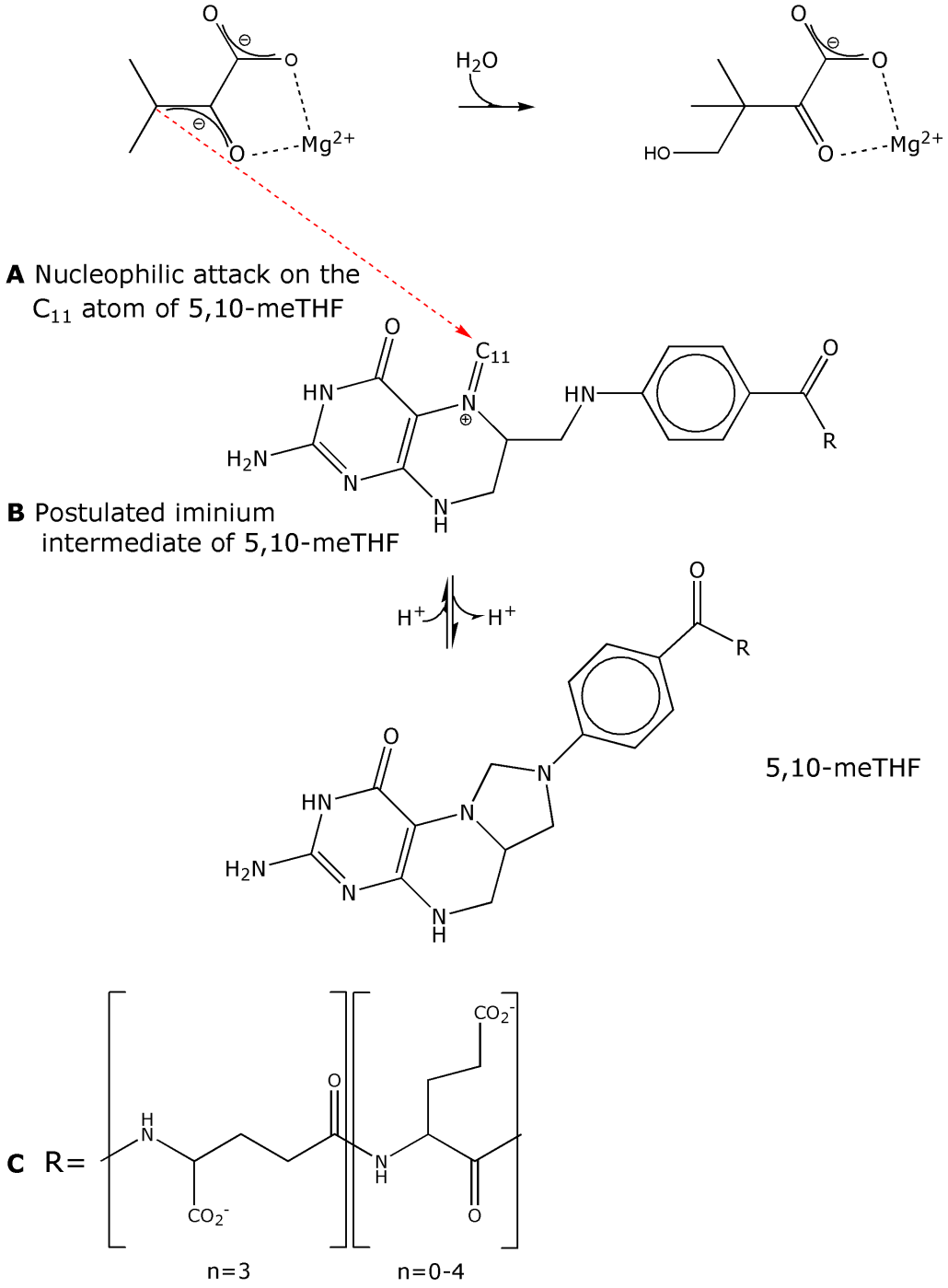

Figure 3.2 Illustration of the chemistry of the aldol reaction catalyzed by KPHMT. (**A**) The nucleophilic attack on the C_{11} atom of 5,10-meTHF. (**B**) The postulated iminium intermediate of 5,10-meTHF. (**C**) The polyglutamate tail of 5,10-meTHF.

ally assigned to the class II aldolase group (Powers and Snell, 1976). Class I aldolases are mainly found in higher order organisms whereas class II aldolases predominate in bacteria and fungi (Schurmann and Sprenger, 2001). The two classes of enzymes are considered to be evolutionarily unrelated (Flechner *et al.*, 1999).

3.3 Catalytic properties of KPHMT

E. coli KPHMT is inhibited by pantoate and pantothenate at concentrations greater than 50 µM and 500 µM, respectively. In both cases the V_{max} is decreased, the K_M is increased and the cooperativity for ketopantoate was greatly increased, whereas the enhancement of cooperativity for pantoate seemed to be less pronounced. Coenzyme A also appeared to be inhibitory above 1 mM. The data indicated feedback inhibition and possible communication between the subunits during inhibition and also catalysis. Therefore, KPHMT bears characteristics of an enzyme with multiple allosteric binding sites. KPHMT requires Mg^{2+} for activity, whereas Mn^{2+}, Co^{2+}, Ni^{2+}, and Zn^{2+} were found to be progressively less activating (Powers and Snell, 1976). Similar results have also been obtained for *M. tuberculosis* KPHMT by Sugantino *et al.* (2003).

Powers and Snell (1976) report that α-keto-β-methylvalerate, α-ketobutyrate, and α-ketovalerate can all replace α-kiva as substrate, whereas pyruvate, isovalerate, and valine proved to be inhibitors (Figure 3.3). This implies that the substitution on the C_3 is essential, but can vary. Isovalerate, 3-methyl-2-butanone, as well as D-and L-valine inhibit, and indicate the importance of the α-keto and the carboxyl group. However, substitution at the C_3 does apparently not impair the enolization reaction — i.e. abstraction of the proton — of the substrate, as demonstrated by Sugantino *et al.* (2003).

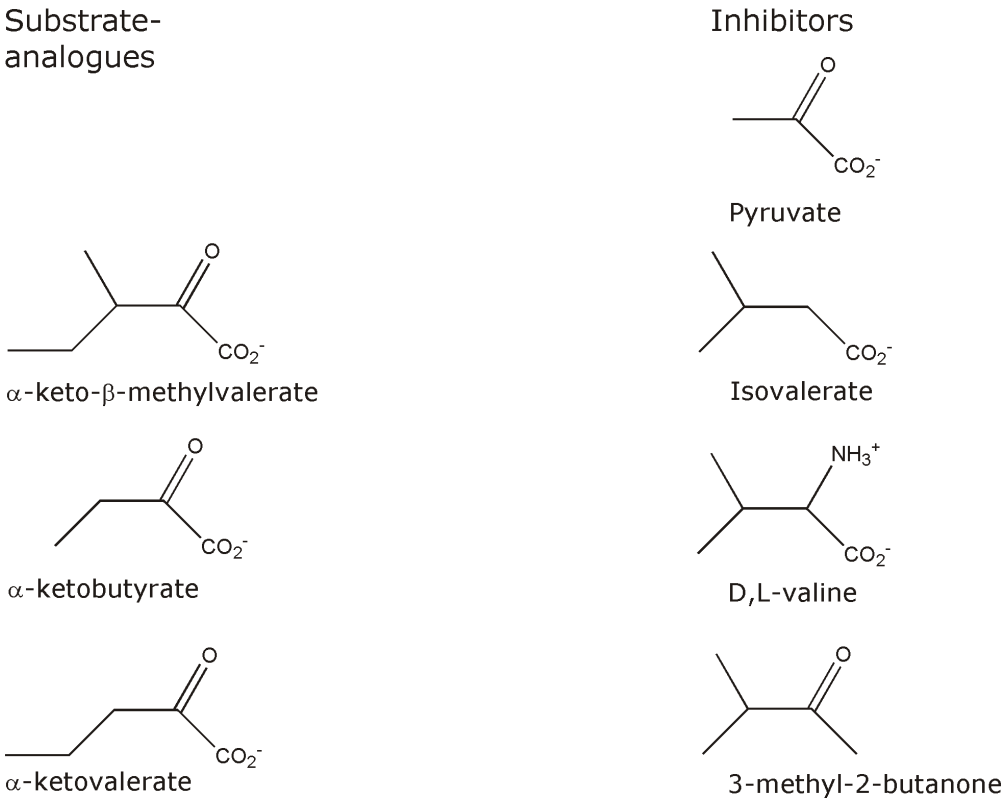

Figure 3.3 Illustration of substrate-analogues of α-kiva and inhibitors of KPHMT.

E. coli KPHMT requires 5,10-meTHF as a cofactor for the transferase reaction. Tetrahydrofolate (THF) itself is the principal one-carbon-group carrier in metabolism, capable of carrying one carbon units at the oxidation level of methanol, formaldehyde, and formate (Stover and Schirch, 1992). Only the *S*-isomer of 5,10-meTHF appears to be utilized as a cofactor by KPHMT, although the *R*-isomer did not inhibit the enzyme (Powers and Snell, 1976). THF derivatives *in vivo* differ in the number of glutamate residues that are attached to the para-aminobenzoic ring. The most common naturally occurring forms have between two and seven glutamate residues. The predominant natural conjugates found in extracts of *E. coli* cells were THF with five glutamate residues, followed by four, three, or six (Powers and Snell, 1976). All of them were effective, but 5,10-meTHF

with four or five glutamates exhibited the lowest K_Ms and thus strongest affinity for binding to the enzyme. In eukaryotes, the glutamate residues are γ-linked, connected together by their γ-carboxyl groups. In contrast, in *E. coli*, in general, only the first three glutamate residues are γ-linked and subsequent ones are α-linked (Figure 3.2) (Scarsdale *et al.*, 2000; Schirch and Strong, 1989).

3.4 Structural features of KPHMT

Based on native polyacrylamide gel electrophoresis (PAGE) and gel-filtration, Powers and Snell (1976) reported a decameric organization of *E. coli* KPHMT with a subunit molecular weight of between 25 700 and 27 000 Da. Analytical sedimentation equilibrium ultracentrifugation confirmed the decameric quaternary macromolecular assembly (von Delft, 2000), which eventually was helpful in solving the crystal structure. The recombinant protein used for this had 264 amino acid residues, corresponding to a subunit molecular weight of 28 237 Da. The crystal structure of the enzyme with the product ketopantoate (PDB-ID: 1M3U) was solved to a nominal resolution of 1.8 Å (von Delft *et al.*, 2003).

Experimental phases for structure solution were derived by multiple wavelength-anomalous dispersion of the selenomethionine derivatized protein. Since two decamers were located in the asymmetric unit, this meant that ~560 kDa of protein had to be phased. 160 of the possible 180 selenomethionine-sites were located, making it the largest selenomethionine-substructure solved to date.

The protomeric tertiary structure of the subunits is a classical $(\beta\alpha)_8$- or triosephosphate isomerase- (TIM)-like barrel fold (Figure 3.4) with some minor, but important variations. A helix, usually found between β-strands 7 and 8, is here replaced by a loop, and an additional N-terminal α-helix, α-1', preceding the usual first β-strand in the sequence is located at the base

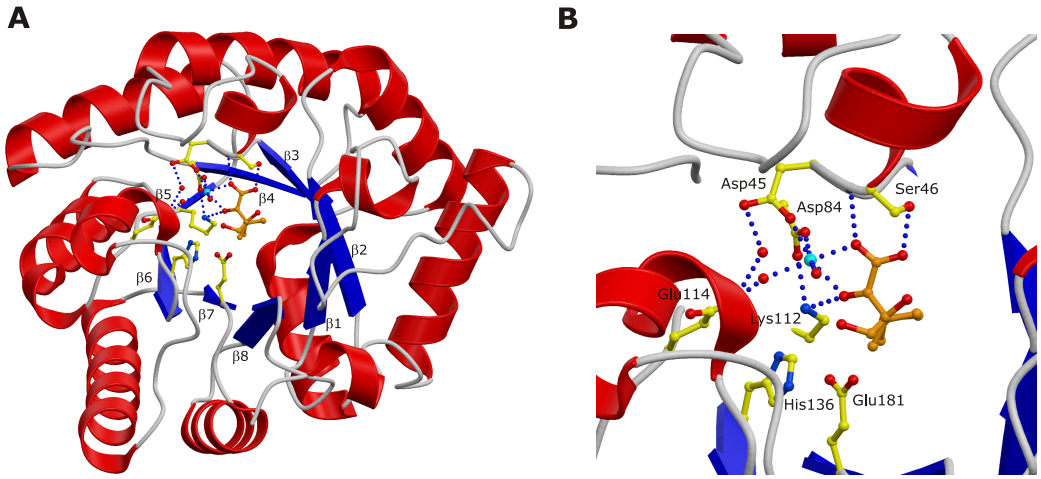

Figure 3.4 Cartoon representation of the KPHMT subunit with the active site (von Delft *et al.*, 2003). (**A**) View down into the β-barrel and the active site. The Mg^{2+} is shown in blue, coordinated by Asp45, Asp84, two water molecules, and the product ketopantoate. For this and all following figures, unless otherwise stated, the colour scheme for ball and stick representation of atoms is as follows: carbon (protein): yellow; carbon (ligand): orange; hydrogen: white; nitrogen: blue; oxygen: red; sulfur: green. (**B**) Illustration of a close up view of the active site of KPHMT, with important active site residues shown and numbered.

of the β-barrel (see Figure 3.5). The quaternary assembly of the protein is shown in Figure 3.6. Based on the analysis of both the hydrophobic and polar interactions, and the buried surface area of the subunits (von Delft *et al.*, 2003), KPHMT is best described as a pentamer of dimers.

The subunits are arranged such that the vertical axis of the $(\beta\alpha)_8$-barrel is perpendicular to the internal five-fold axis of the decamer (Figure 3.6). The extensive interface can explain the high temperature resistance of the protein (Powers and Snell, 1976). The active site of KPHMT could be reliably identified, because it contained the prosthetic Mg^{2+} with the product ketopantoate bound to it (von Delft *et al.*, 2003). The active site of KPHMT is located in a large cavity at the C-termini of the β-strands with dimensions of 20 Å in length, 10 Å in width, and 15 Å in depth. It is inaccessible from the bottom edge of the $(\beta\alpha)_8$-barrel.

The co-crystallized Mg^{2+} is bound in a distorted octahedral geometry,

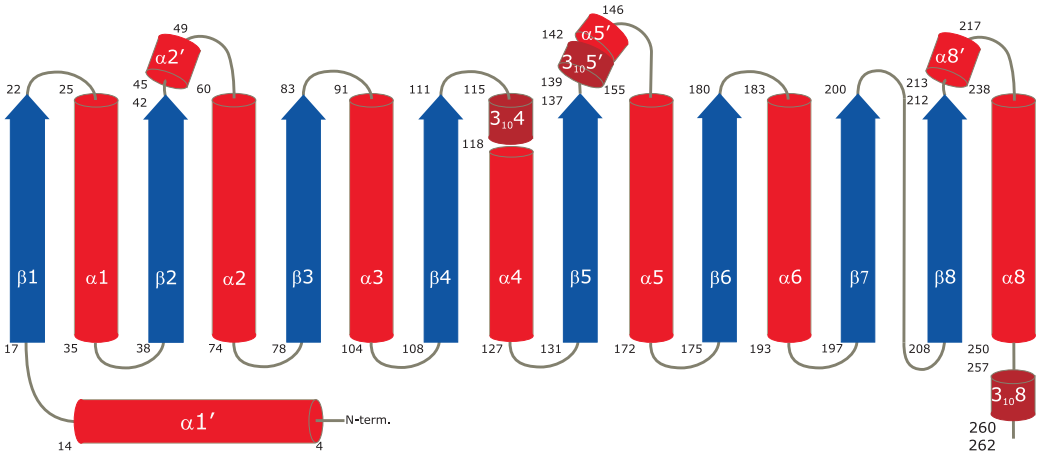

Figure 3.5 Schematic secondary structure topology representation of the KPHMT fold. For this and the following schematic figure of secondary structures the colour scheme for secondary structural elements is the same as in Figure 3.6, with the exception of 3_{10} helices which are coloured brown, instead. Residues from the *E. coli* KPHMT amino acid sequence bordering the respective secondary structural elements are shown with numbered labels. The scheme for this figure was adapted and modified from von Delft *et al.* (2003).

where residues Asp45 and Asp84 occupy axial and equatorial positions, respectively, while Glu114 coordinates through a water molecule that occupies an equatorial position (Figure 3.4). The keto- and carboxyl-groups of ketopantoate assume axial and equatorial positions to the Mg^{2+}, respectively and the remaining equatorial positions are occupied by two water molecules (see Figure 3.4 and Section 11.5).

More recently, crystal structures for the *M. tuberculosis* (PDB-ID: 1OY0) (Chaudhuri *et al.*, 2003) and *Neisseria meningitidis* KPHMT (no reference available; PDB-ID: 1O66) have also been reported. For the latter a crystal structure in complex with substrate α-kiva has also been determined (PDB-ID: 1O68).

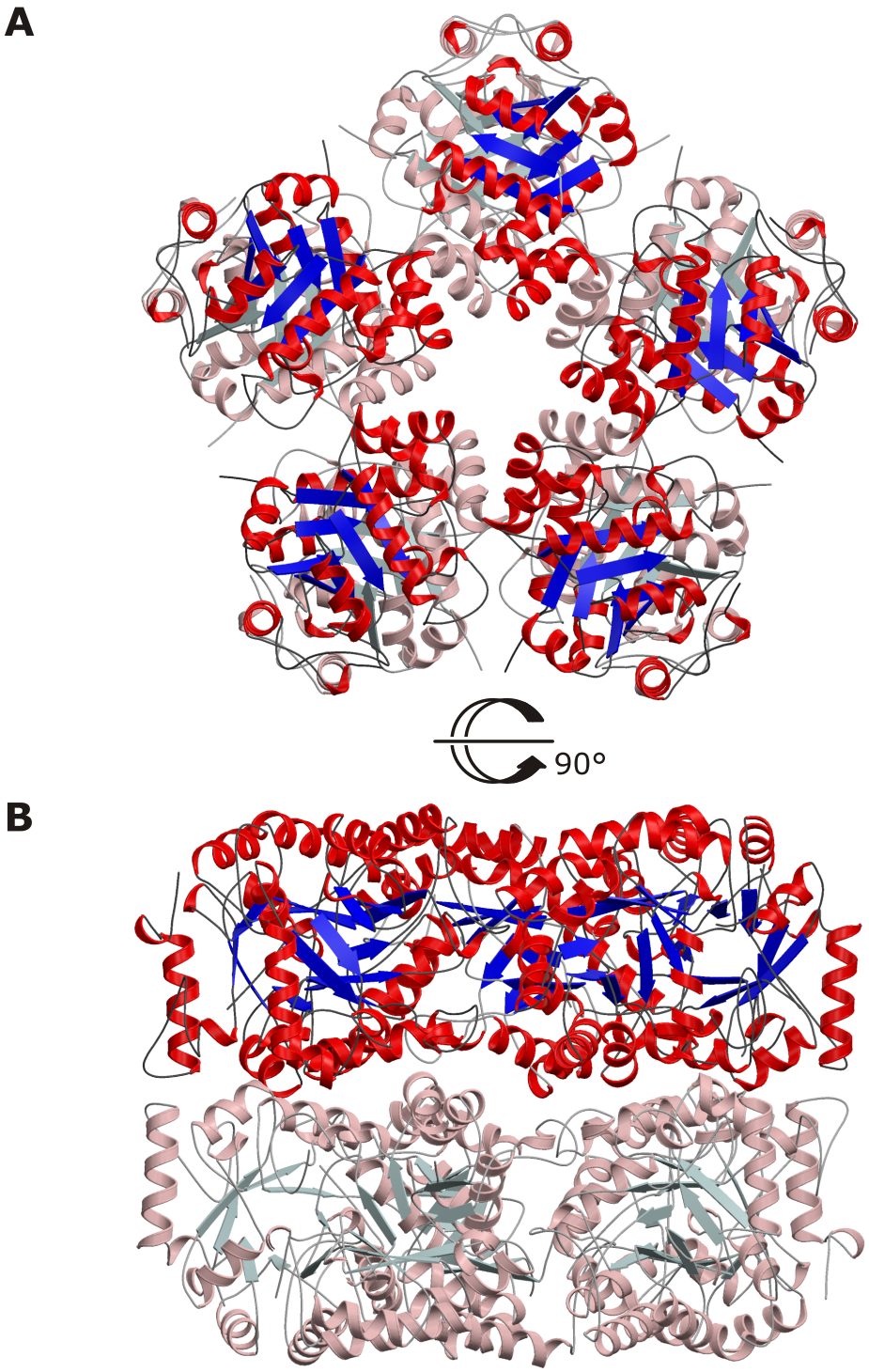

Figure 3.6 Cartoon representation of the KPHMT decamer (von Delft *et al.*, 2003). For this, and the following figures, the colour scheme is as follows, unless otherwise stated: α-helices: red; β-strands: blue; loops and coils: grey. (**A**) The KPHMT decamer viewed down the internal five-fold symmetry axis. For the subunits in the background the secondary structural elements are shown in lighter colours. (**B**) The KPHMT decamer rotated by $\sim$90° from the view in **A**. These and all following cartoon figures of proteins were prepared with MOLSCRIPT (Kraulis, 1991) and rendered in RASTER3D (Merritt and Bacon, 1997), unless stated otherwise.

3.5 Motives for studying KPHMT

Reasons for studying the pantothenate pathway in general have already been given in Section 2.3. KPHMT catalyzes the first committed reactions in this pathway and is absent in animals. As such it is a particularly attractive target for the development of pantothenate pathway inhibitors.

With the crystal structure available, the mechanistic issues can now be addressed, which are to identify the cofactor binding site, characterize key catalytic residues, and carry out a detailed comparative structural analysis. Relatively little is known about the binding of the cofactor 5,10-meTHF to enzyme in general, and about the mode of transfer of groups from THF from a structural point of view. High resolution crystal structures of KPHMT bound to 5,10-meTHF analogues could provide insight into the catalysis. In order to rationally design inhibitors against the enzyme a thorough understanding of the mechanism is required.

The potential of aldolases in synthetic organic chemistry is widely recognized as they are carrying out specific C-C bond condensations with exquisite control of stereochemistry. Their utilization has already proven useful in the synthesis of compounds for the pharmaceutical industry (Cooper *et al.*, 1996), and directed evolution has successfully been employed to engineer new activities onto existing scaffolds (Wymer *et al.*, 2001). The $(\beta\alpha)_8$-barrel fold is particularly well-suited for this strategy, because the functional and structural features are well separated in the enzyme. The substrate specificity and the catalytic step, itself could be changed by directed evolution (Gerlt and Babbitt, 2001).

KPHMT is certainly a complex and interesting enzyme, its research reaching beyond the application-oriented aims towards an understanding of this crucial enzyme in vitamin B_5 biosynthesis from a structural and mechanistic point of view.

4 L-aspartate-α-decarboxylase

4.1 Background

Virtanen and Laine (1937) discovered that β-alanine is formed in the fermentation mash of the microorganism *Rhizobium triolii*, presumably from the L-aspartate in the culture medium. Williamson and Brown (1979) were subsequently the first to demonstrate the biosynthesis of β-alanine from L-aspartate by ADC in *E. coli*. They also described the purification and some biochemical properties of the enzyme. Independently, Cronan (1980) discovered the enzyme in extracts of *E. coli* and showed that the catalysis by L-aspartate-α-decarboxylase provides for the main route of β-alanine production in *E. coli*. β-alanine auxotrophs were deficient in ADC activity and the lesion could be mapped to the *panD* gene locus. The *panD* gene was subsequently cloned and the enzymatic properties characterized in more detail by both Smith (1988) and Ramjee *et al.* (1997). The gene has also been cloned from both *Helicobacter pylori* (Kwon *et al.*, 2002) and *M. tuberculosis* (Chopra *et al.*, 2002), and in the latter the gene product has been biochemically characterized.

4.2 Posttranslational intramolecular self-processing

4.2.1 Self-catalytic pyruvoyl formation

ADC is a particularly interesting enzyme in that it is capable of forming its own cofactor, prior to carrying out the actual catalysis. In common with a small, but widely distributed, group of mechanistically related pyruvoyl-dependent enzymes, ADC is translated as an inactive pro-enzyme, the subunit of which is also designated π-chain. This self-processes at a specific cleavage site to create a covalently bound prosthetic group, a pyruvoyl cofactor (Figure 4.1).

The latter is formed by intramolecular, non-hydrolytic serinolysis in which the side-chain oxygen of Ser25 atttacks the carbonyl carbon of Gly24. This process does not require any external auxiliary effectors. The initial step of the reaction, which is also known as an N→O acyl shift, results in the formation of an ester in processing (Figure 4.1), which was first demonstrated by isotope labelling experiments with O^{18} by Recsei *et al.* (1983a). Direct evidence for the ester intermediate in the reaction has also been provided by the crystal structures of processed, active ADC (Albert *et al.*, 1998), and human *S*-adenosylmethionine decarboxylase (AdoMetDC) His243→Ala mutant (Ekstrom *et al.*, 2001). β-elimination of the ester intermediate produces dehydroalanine, which hydrolyses to form the pyruvoyl group (Figure 4.1). In *E. coli* ADC this process yields a larger α-chain of 11 kDa with the pyruvoyl group at the amino-terminus and a smaller β-chain of 2.8 kDa with Gly24 at the carboxyl-terminus, as shown by PAGE, electrospray mass spectrometry (Ramjee *et al.*, 1997), and the ADC crystal structure (Albert *et al.*, 1998). However, both subunits remain non-covalently attached to each other to form the active, functional subunit of the enzyme.

Interestingly, purified, tetrameric *E. coli* ADC was observed to be only partially processed and SDS-PAGE showed the presence of significant amounts of the π-chain (see Section 12.2). Mass-spectrometry experiments by Ramjee *et al.* (1997) indicated the presence of three processed subunits per tetramer and one subunit with an N-terminal Ser residue at the α-chain. The initial crystal structure (Albert *et al.*, 1998) partly confirmed this by finding three processed subunits and an ester intermediate in one of the subunits. In general, self-processing in *E. coli* ADC appears to be considerably slower than in other self-processing enzymes and was observed to be faster *in vivo* than *in vitro*, which could indicate the existence of an external processing factor or a specific pH-dependence. Self-processing of the last subunit was slow enough to survive purification, crystallization, and X-ray data collection (Albert *et al.*, 1998). More recent studies on tetrameric *M. tuberculosis* ADC by Chopra *et al.* (2002) confirmed the finding of *E. coli* ADC that protein self-processing in ADC is very slow. The *M. tuberculosis* enzyme essentially shows biochemical properties and characteristics similar to *E. coli* ADC. The *E. coli* and *M. tuberculosis* enzymes exhibit 45% amino acid sequence identity, with the key residues being identical (see amino acid sequence alignment in Section 13.3).

The overall molecular mechanism of self-processing and the formation of the pyruvoyl group is assumed to be essentially the same as in other pyruvoyl-dependent enzymes (reviewed by van Poelje and Snell, 1990). Pyruvoyl formation is known from AdometDC, arginine decarboxylase (ArgDC), histidine decarboxylase (HisDC) of Gram-positive bacteria, phosphatidylserine decarboxylase, and from glycine-, proline-, and sarcosine reductases.

Lactobacillus 30a HisDC was the first enzyme extensively studied because of the formation of the pyruvoyl cofactor (Gallagher *et al.*, 1989; Huynh *et al.*, 1984a; Huynh and Snell, 1986a,b; Huynh *et al.*, 1984b; Recsei *et al.*, 1983a; Recsei and Snell, 1970, 1973, 1982; Snell and Huynh, 1986; Vanderslice *et al.*, 1988) and served as a model for the study of the others.

The same enzyme has also been characterized from *Clostridium perfringens* and *Lactobacillus buchneri* (Huynh and Snell, 1985a,b; Recsei *et al.*, 1983b; Recsei and Snell, 1985). Moreover, research has been carried out on AdometDC from *H. sapiens* (Stanley and Pegg, 1991; Stanley *et al.*, 1989, 1994; Xiong and Pegg, 1999; Xiong *et al.*, 1999), *Methanococcus jannaschii* (Kim *et al.*, 2000a), *Acanthamoeba castellani* (Hugo and Byers, 1993), *Sulfolobus solfataricus* (Porcelli *et al.*, 1993), *Catharanthus roseus* (Schroder and Schroder, 1995) and *S. cerevisiae* (Kashiwagi *et al.*, 1990). AdometDC is a critical regulatory enzyme in the polyamine (spermidine and spermine) biosynthetic pathway. Both processing and enzyme activity in the *H. sapiens* and *S. cerevisiae* forms were shown to be enhanced by putrescine (Pegg *et al.*, 1998). Pyruvoyl-dependent phosphatidylserine decarboxylase has been characterized from *E. coli* (Li and Dowhan, 1988, 1990) and *S. cerevisiae* (Trotter *et al.*, 1995; Voelker, 1997), and glycine- proline-reductases from *Clostridium sticklandii* (Bednarski *et al.*, 2001; Kabisch *et al.*, 1999; Seto, 1978).

In addition to the crystal structure of ADC (Albert *et al.*, 1998), the processed crystal structures of three pyruvoyl-dependent enzymes have been reported, namely, *Lactobacillus* 30a HisDC (Gallagher *et al.*, 1993), *H. sapiens* AdometDC (Ekstrom *et al.*, 1999) and *M. jannaschii* arginine decarboxylase (ArgDC) (Tolbert *et al.*, 2003a). Although all these enzymes have their cleavage site located in a loop region between two β-strands, only ArgDC and HisDC are homologous to each other. The absence of similarity in sequence and structure between the pyruvoyl-dependent enzymes points towards convergent evolution for the molecular mechanism of cleavage.

4.2.2 Related forms of self-processing

In self-catalyzed protein processing the initial biochemical step, common to all self-processing enzymes, is the formation of an ester or thioester (Figure 4.1). In an open system, the thermodynamic equilibrium for this step

would be expected to be on the side of the amide bond, and consequently the reaction would be unfavourable. However, chemical model systems have shown that the equilibrium in N→O is on the side of the ester, when the amine can be protonated (Iwai and Ando, 1967).

The ester formation may be additionally driven by relief of strain present in the loop region preceding the self-cleavage site. In HisDC mechanical strain appeared to be built into the Ser81-Ser82 cleavage site, which was indicated by the relatively large distance of 6 Å between the cleaved N-and C-terminal ends of the cleaved chain (Gallagher *et al.*, 1993). Likewise, unusual peptide conformations were observed by Xu *et al.* (1999) in the crystal structure of an unprocessed glycosylasparaginase mutant, another self-processing enzyme. Relaxation of this strain by the N-O acyl rearrangements was proposed to shift the amide-ester equilibrium in favour of the ester intermediate. In general, the biochemical rationale for the formation of the ester intermediate presumably lies in its greater reactivity compared with amides, making them intermediates that are more suitable for the subsequent biochemical steps in self-catalyzed reactions. In recent years, self-processing has emerged as an important post-translational modification for protein maturation that is observed in a growing number of, often evolutionary and structurally unrelated, proteins. The molecular mechanism is assumed to have strong parallels in all self-processing systems. Consequently, self-processing in ADC is thought to bear the general characteristics of intramolecular, self-catalyzed backbone rearrangements.

As already indicated above, whereas the inital N→O or N→S shift is chemically the same in all self-processing systems, the subsequent reaction, which invariably leads to cleavage of the rearranged peptide bond, can proceed in different ways. Based on the fate of the ester intermediate, these enzymes can be mechanistically classified into four main systems (reviewed by Paulus, 2000): hedgehog proteins, inteins, N-terminal nucleophilic aminohydrolases (Ntn hydrolases) and pyruvoyl-dependent

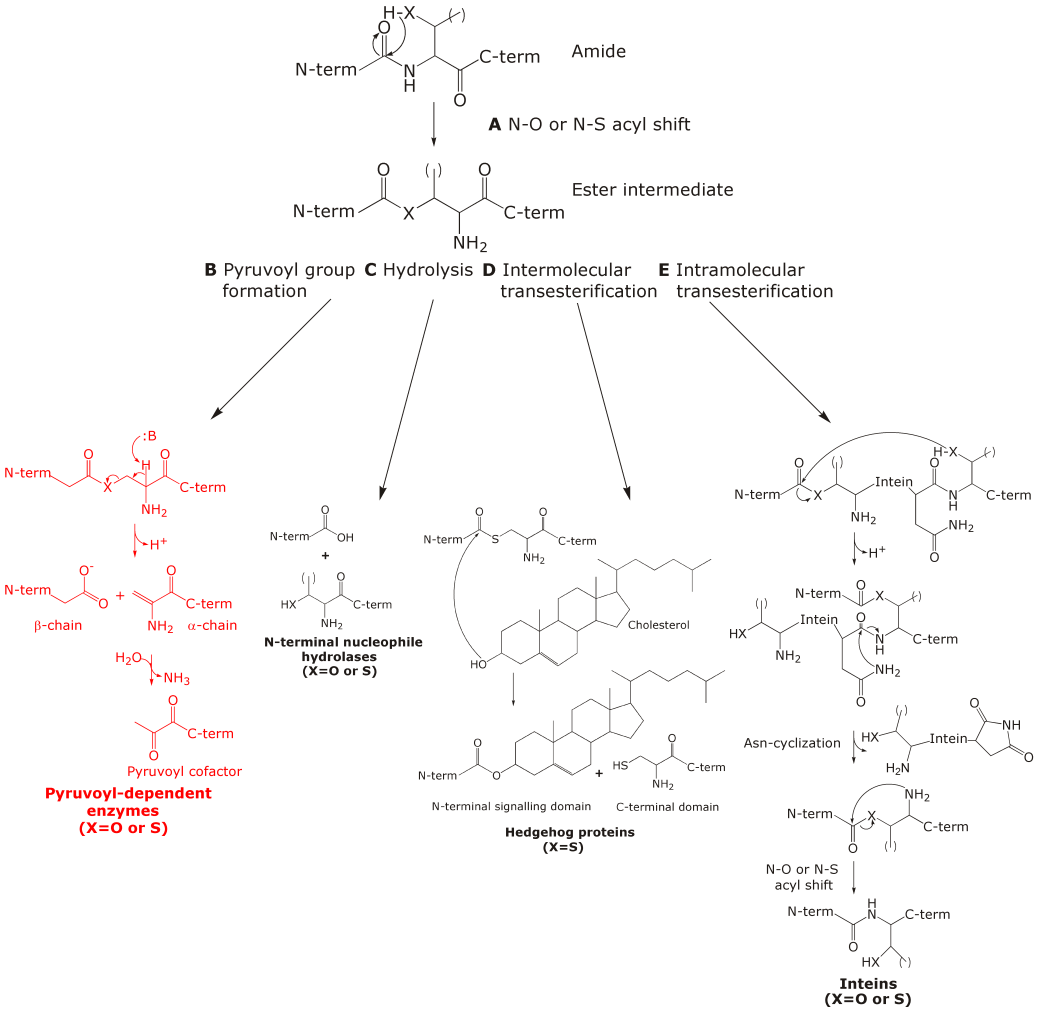

Figure 4.1 Schematic overview of the chemistry of the various forms of intramolecular protein self-processing. (**A**) The initial N→O or N→S acyl shift common to all forms of protein self-processing. (**B**) Pyruvoyl group formation, as in ADC, coloured in red. (**C**) Hydrolysis of the ester as in N-teminal nucleophile hydrolases. (**D**) Intermolecular attack of the ester by cholesterol as in the case of the hedgehog proteins. (**E**) Intramolecular transesterification as in inteins.

enzymes. An initial N→S or N→O acyl shift with invariably a Cys, Ser or Thr nucleophile, resulting in a thioester or ester intermediate, respectively, is common to all four groups. The following steps differ, but always result in breakdown of the activated (thio)ester (Figure 4.1).

In Ntn hydrolases the (thio)ester is simply hydrolyzed generating a prosthetic N-terminal Cys, Ser, or Thr at the cleaved C-terminal part, which then becomes crucial to the actual catalysis of the enzyme (Figure 4.1) (Artymiuk, 1995). One of the most prominent examples for an Ntn hydrolase is the β-subunit of the proteasome, the principal protease responsible for non-lysosomal protein degradation in most eukaryotic cells. Upon self-processing, the much smaller N-terminal part appears to act as molecular chaperone (Chen and Hochstrasser, 1996) whereas the resulting new N-terminal residue at the C-terminal part is central in carrying out the nucleophilic attack for peptide bond hydrolysis of the substrates (Chen and Hochstrasser, 1996; Groll *et al.*, 1999; Seemuller *et al.*, 1996). Other known and characterized Ntn hydrolases include glycosylasparaginase (Guan *et al.*, 1996, 1998; Ikonen *et al.*, 1991; Liu *et al.*, 1998; Riikonen *et al.*, 1995; Saarela *et al.*, 1998, 2004; Tikkanen *et al.*, 1996), glutamine phosphoribosylpyrophosphate amidotranferase (Li *et al.*, 1999a), glutaryl 7-aminocephalosporanic acid acylase (Lee *et al.*, 2000; Lee and Park, 1998), L-aminopeptidase D-Ala-esterase (Bompard-Gilles *et al.*, 2000), penicillin acylases (Martin *et al.*, 1991; Slade *et al.*, 1991), and cephalosporin acylases (Li *et al.*, 1999b).

The Ntn glycosylasparaginase is one of the most detailed studied Ntn hydrolases to date. Site-directed mutagenesis and structural studies have revealed several important aspects of self-processing. A wide spectrum of different disease-causing mutations are known in the human gene of glycosylasparaginase, several of which impair self-processing (Tikkanen *et al.*, 1996). A resulting deficiency of enzyme activity causes the lysosomal accumulation disease aspartylglycosaminuria in *H. sapiens*, which results in mental retardation (Mononen *et al.*, 1993).

Penicillin and cephalosporin acylases are industrially very important enzymes, used to hydrolyze penicillin or cephalosporin, respectively to give the β-lactam nucleus of many semi-synthetic drugs (Valle *et al.*, 1991).

Although there is hardly any sequence similarity among most of the Ntn hydrolase enzymes, all enzymes share a common, specific fold. Consequently, they appear to be evolutionary related (Artymiuk, 1995; Brannigan *et al.*, 1995), and have been classified into one superfamily in SCOP (Murzin *et al.*, 1995). The fold has been the subject of a detailed comparison by Oinonen and Rouvinen (2000). The SCOP classification enabled the crystal structure of the conserved protein MTH 1020 from *Methanobacterium thermoautotrophicum* to be assigned to this superfamily (Saridakis *et al.*, 2002). Based on sequence comparison the proteins of the peptidase family U34, of unknown function, could also be classified into the Ntn superfamily (Pei and Grishin, 2003).

In hedgehog proteins an intermolecular attack of the 3β hydroxyl group of cholesterol on the thioester occurs, resulting in covalent linkage of cholesterol to the C-terminus of the smaller N-terminal signalling domain and release of the larger C-terminal domain (Beachy *et al.*, 1997; Hammerschmidt *et al.*, 1997) (Figure 4.1). Members of the hedgehog family of secreted signalling proteins are among the primary signals generated by certain embryonic patterning centers, and are consequently central in developmental biology of both invertebrates and vertebrates. Secreted signalling proteins encoded by the hedgehog gene family induce specific patterns of differentiation in a variety of tissues (Lee *et al.*, 1994; Porter *et al.*, 1996, 1995). The N-terminal domain of the cleaved hedgehog protein, accounting for local and long-range signalling activity, remains tightly cell surface-associated, whereas the C-terminal domain diffuses freely upon secretion in embryos. The cholesterol modification of the N-terminal domain was demonstrated to be crucial for its association with the cell surface-membrane, and is essential for proper signaling function (Gallet *et al.*, 2003). Self-processing in hedgehog proteins appears to

be concentration-independent (Porter *et al.*, 1995). Genes that are similar in sequence to the self-processing domain in hedgehog proteins have been identified in several nematode genes of unknown function (Hall *et al.*, 1997).

Protein splicing in inteins was first described by Kane *et al.* (1990). Inteins are internal polypeptide sequences, between 100 and 600 amino acids in size, that are posttranslationally excised from a protein precursor (Perler *et al.*, 1997). In inteins the (thio)ester is intramolecularly attacked by a nucleophile from a spatially different region, which splices the intein, the central part of the protein, and ligates the two flanking N- and C-terminal polypeptides, termed exteins, together to form the mature and active external protein (Figure 4.1) (reviewed by Gogarten *et al.*, 2002; Liu, 2000; Noren *et al.*, 2000; Paulus, 2000). Protein splicing is distantly reminiscent of self-splicing in the ribozyme, self-cleaving RNAs that are synthesized as one polynucleotide and catalyze cleavage via an internal phosphoester transfer (Doherty and Doudna, 2000). The catalytic power rests within the intein and catalysis only requires one additional Cys, Ser, or Thr residue of the extein at the C-terminal splice junction.

To date, no apparent cellular function has been established for inteins. It is assumed that inteins are parasitic genetic elements which can persist, because they do not impair the function of the enzyme as they are being spliced. Most of the inteins contain a homing endonuclease activity which allows them to insert the intein gene into other genomes. This guarantees their mobility and also enables them to spread by horizontal gene transfer. Homing endonucleases are a class of sequence-specific double-stranded endonucleases that recognize and cleave DNA at or near the intein insertion site in homologous genes lacking the intein element (Belfort and Perlman, 1995; Belfort and Roberts, 1997). Intein endonucleases cleave the intein-minus copy of the intein gene at or near the intein-encoding site and the intein-plus copy is then used as a template to repair the intein-minus gene, converting it into an intein-containing allele. More than 130 inteins

have been identified in members of all three phylogenetic kingdoms of life and are listed in the Intein database[1] (Perler, 2002). They have been found inserted in a variety of archaeal, bacterial, and chloroplast proteins as well as vacuolar ATPases from *S. cerevisiae* (Noren *et al.*, 2000). The majority of these mature proteins are nucleotide triphosphate-utilizing enzymes, such as DNA and RNA polymerases, gyrases, and helicases (Liu, 2000; Pietrokovski, 1994).

Significant sequence similarity exists between the C-terminal part of hedgehog proteins and inteins and the C-terminal self-processing domain of hedgehog and inteins bear strong structural similarity. The conserved fold was termed HINT (HedgehogIntein) module (Dalgaard *et al.*, 1997; Hall *et al.*, 1997). In addition, a conserved Thr-X-X-His (where X is any amino acid) sequence motif in both hedgehog proteins and inteins, that is spacially close to the cleavage site, could be identified (Lee *et al.*, 1994). Together, these data indicate a common evolutionary origin of inteins and hedgehog proteins, with hedgehog proteins likely having evolved from inteins (Dalgaard *et al.*, 1997).

Structural information for the processed forms of self-processing enzymes is available for the hedgehog protein (Hall *et al.*, 1995, 1997), inteins (Ding *et al.*, 2003; Duan *et al.*, 1997), the proteasome (Lowe *et al.*, 1995), and for several other Ntn hydrolases, such as cephalosporin acylase (Kim *et al.*, 2000b), glycosylasparaginase (Guo *et al.*, 1998; Oinonen *et al.*, 1995), from glucosamine 6-phosphate synthase (Isupov *et al.*, 1996), glutamine phosphoribosylpyrophosphate (Smith *et al.*, 1994), glutaryl 7-aminocephalosporanic acid acylase (Kim *et al.*, 2003), L-aminopeptidase D-Ala-esterase (Bompard-Gilles *et al.*, 2000), penicillin acylase (Duggleby *et al.*, 1995), penicillin G acylase (McDonough *et al.*, 1999), and penicillin V acylase (Suresh *et al.*, 1999).

[1] http://www.neb.com/neb/inteins.html

4.3 Catalytic properties of ADC

One efficient biochemical way to decarboxylate an amino acid is to attach it to a molecule with an "electron sink" in order to facilitate bond cleavage of the α-carboxylate group. A large number of enzymes typically accomplishes this task by providing a pyridoxal 5-phosphate cofactor, invariably bound to the ϵ-amino group of a catalytic lysine residue (reviewed by John, 1995). The amino group of the amino acids then covalently binds to the cofactor through its aldehyde group. An alternative route, with respect to the cofactor that is used by pyruvoyl-dependent enzymes, is to covalently attach the substrate to a pyruvate-like moiety of the enzyme, the pyruvoyl group. A comparative computational investigation of the effect of the pyridine ring for amino acid decarboxylations and the pyruvoyl group came to the conclusion that the additional stabilization effect of the pyridine ring is only minor (Toney, 2001). Consequently, pyruvoyl-dependent enzymes appear to capture the major part of the catalytic power that pyridoxal 5-phosphate offers, are identical in stereospecificity, and appear to be similar in catalytic efficiency (Toney, 2001; van Poelje and Snell, 1990). HisDC is the only amino acid decarboxylase for which both types of cofactors have been observed: pyridoxal 5-phosphate dependent HisDC in Gram-negative bacteria (Tanase *et al.*, 1985) and animals (Watanabe *et al.*, 1979); pyruvoyl-dependent HisDC in Gram-positive bacteria, which show no relatedness in amino acid sequence. The general utility of both the pyruvoyl group and pyridoxal 5-phosphate stems from their ability to stabilize carbanions formed adjacent to the Schiff-base through delocalization into the extended π-bonded system. However, compared with pyridoxal 5-phosphate-dependent catalysis, pyruvoyl-dependent catalysis may be a more primitive mechanism that does not require exogeneous cofactors.

ADC uses a pyruvoyl-group as a cofactor to covalently bind L-aspartate. The nucleophilic attack of the amino-group of L-aspartate on to the carbonyl carbon of the pyruvoyl group results in formation of a Schiff-base (Figure 4.2) and creates a π-conjugated system between the bound sub-

strate and the adjacent amide bond. The created Schiff-base can easily be trapped by reduction with cyanoborohydride (Ramjee *et al.*, 1997), which was also shown for other pyruvoyl-dependent decarboxylases such as AdometDC (Markham *et al.*, 1982), HisDC (Recsei and Snell, 1970), and phosphatidylserine decarboxylase (Satre and Kennedy, 1978).

Upon decarboxylation of the α-carboxyl group, which dissociates as carbon dioxide, the negative charge on the α-carbon can be dispersed over the π-conjugated system, and even further into amino acid residues that H-bond to it, which could provide a "pull-effect" for the reaction. The presence of a "pull" and "push" effect appears to be required to facilitate decarboxylation (von Delft, 2000), because of the strength of the carbon-carbon bond. The decarboxylation results in formation a planar carbanionic intermediate. A "push" effect driving the negative charge from the carboxyl into the electron sink and towards the neutral carbon dioxide species could be provided by the hydrophobic environment at the active site. Based on the crystallographic analysis of enzyme-ligand complexes (see Section 4.4) von Delft (2000) proposed a detailed four-step catalytic mechanism for the decarboxylation reaction, which was, in part, supported by deuterium labeling experiments by Saldanha *et al.* (2001).

β-substituted D-amino acid analogues (i.e. with the opposite stereochemistry), such as β-fluoro/chloro-D-alanine, were demonstrated to be inhibitory for *E. coli* ADC (Smith, 1988), probably as a result of elimination of the β-substituted group and subsequent attack of a basic group of the enzyme on the vinyl group to cause irreversible inhibition (Smith, 1988; von Delft, 2000; Williamson and Brown, 1979). Of the tested compounds, mostly D-amino acids, i.e. with the opposite configuration of substrate, showed inhibition, although L-glutamate and L-serine also inhibited the enzyme. However, significant feedback inhibition by later pathway products, such as β-alanine, acetyl coenzyme A, coenzyme A, or pantothenate could not be observed (Williamson and Brown, 1979). All known pyruvoyl-dependent enzymes can be irreversibly inactivated by

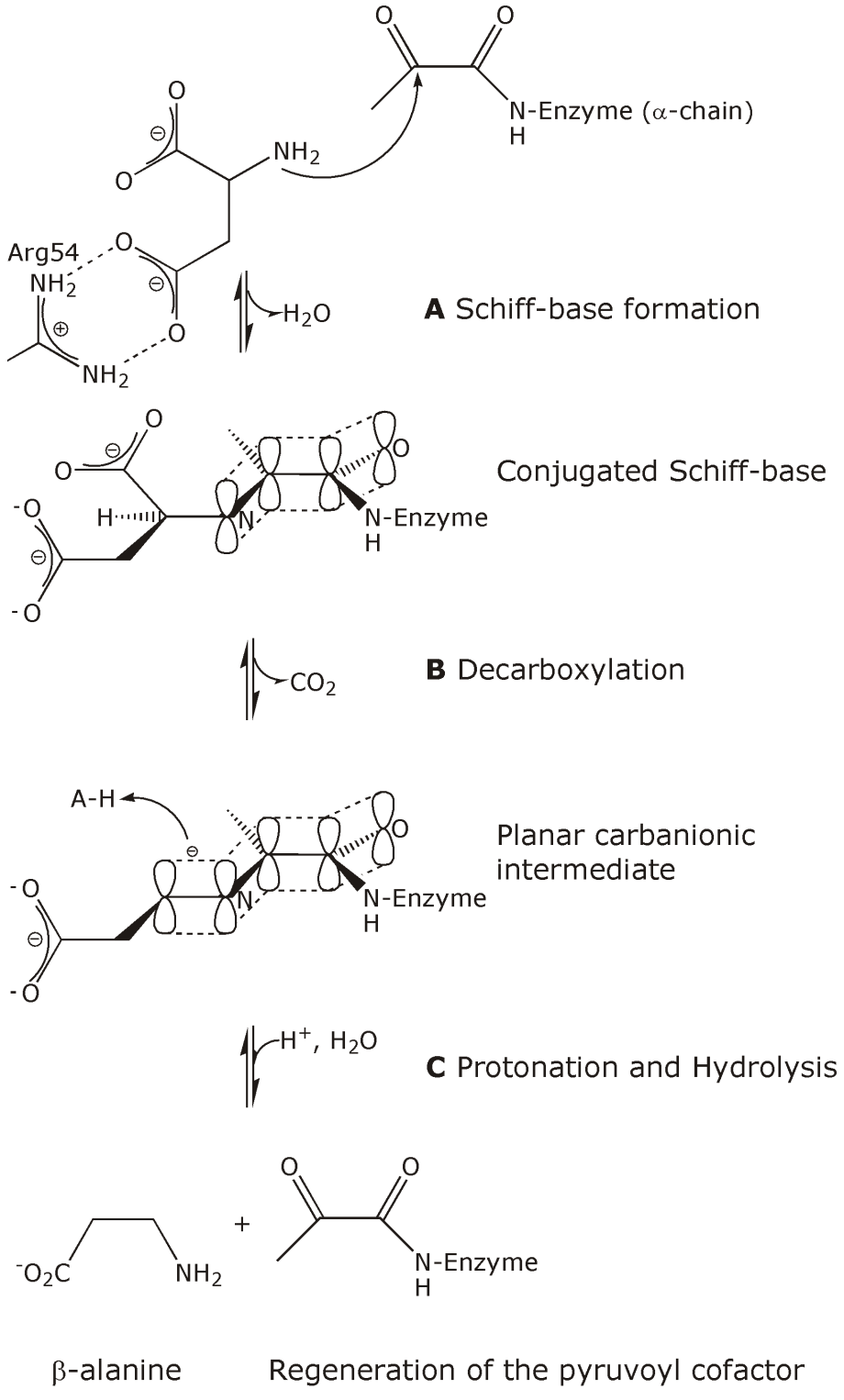

Figure 4.2 Schematic of the chemistry of the decarboxylation reaction catalyzed by ADC. (**A**) The first step involves formation of a Schiff-base or imine.(**B**) Decarboxylation of bound substrate. Atomic orbitals are drawn to highlight the π-conjugated system. (**C**) π-conjugation that stabilizes the negative charge after decarboxylation (**D**) In the final step, protonation and hydrolysis of the decarboxylated substrate results in formation of β-alanine and regeneration of the pyruvoyl cofactor.

a substrate mediated transamination in which they undergo an incorrect protonation during the enzymatic reaction, as was demonstrated for AdoMetDC (Diaz and Anton, 1991; Xiong *et al.*, 1999). The protonation occurs on the pyruvate instead of the α-carbon leading to conversion of the pyruvoyl group to an alanyl group. This can also explain why L-aspartate was observed to inhibit to some extent (Smith, 1988).

4.4 Structural features of ADC

The crystal structure of *E. coli* ADC (PDB-ID: 1AW8) was solved by multiple-isomorphous replacement (Albert *et al.*, 1998) to 2.2 Å maximum resolution, using a recombinant 126 amino acid protein with a molecular subunit weight of 13.8 kDa. The enzyme had been shown to be homotetrameric by electro-spray mass spectrometry (Ramjee *et al.*, 1997) and this was confirmed by the crystal structure (Figure 4.3 A, Figure 4.5). However, because of the crystallographic hexagonal symmetry, $P6_122$, only one dimer is located in the asymmetric unit and the tetramer is created by the crystallographic two-fold axis. The latter coincides with the internal pseudo-four-fold rotational symmetry of the tetramer. Apart from two small helices and one 3_{10} helix, the protein is completely β-stranded. Six β-strands form a β-barrel, which forms the core of the mushroom-shaped oligomer. The upper part of the protein (viewed in a side-orientation of the protein; Figure 4.5) adopts a rare tertiary structural arrangement, termed a "double-ψ" fold (Castillo *et al.*, 1999). This fold is formed by two interlocked motifs each of which comprise a loop and a sheet, that together resemble the Greek letter ψ (Figure 4.4).

E. coli ADC does not share significant sequence similarity with any other, non-ADC enzyme and, consequently, comparison based on sequence alone was unrewarding. However, the double-ψ fold (Figure 4.3) of ADC has been the subject of a detailed comparative structural analysis by Castillo *et al.* (1999). This suggested a common homodimeric ancestor

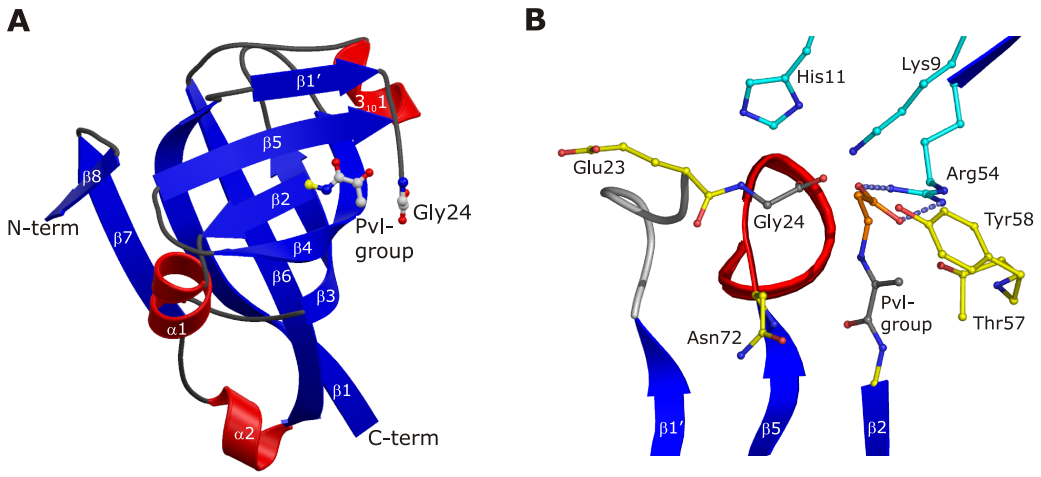

Figure 4.3 (**A**) Cartoon representation of one ADC subunit (Albert *et al.*, 1998). (**B**) View at the active site of ADC with the product β-alanine covalently bound to the pyruvoyl group (Pvl-group) (von Delft, unpublished results). Carbon atoms from an adjacent subunit are coloured in cyan. Figure **B** was prepared with PYMOL (DeLano, 2002).

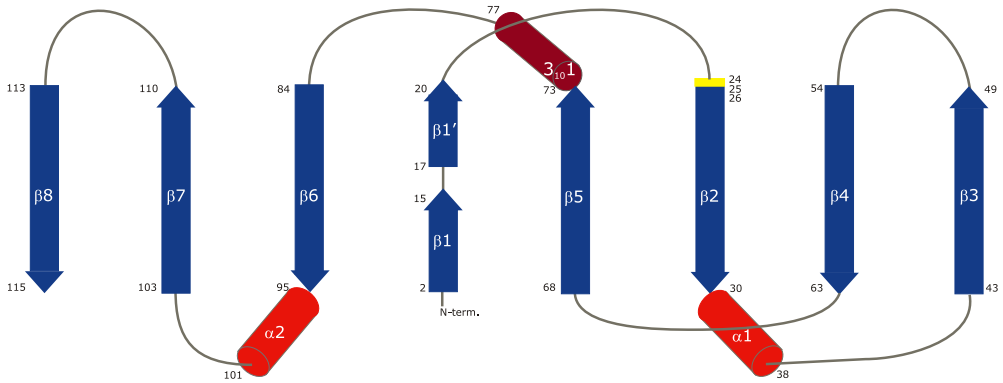

Figure 4.4 Schematic secondary structure topology representation of the ADC fold. Note the two crossovers that form the double-ψ arrangement. The location of the self-cleavage site is indicated with a yellow bar. Residues for the *E. coli* ADC amino acid sequence that border the respective secondary structural elements, are indicated with numbered labels. The scheme for this figure was adapted and modified from Castillo *et al.* (1999).

to ADC, barwin (Ludvigsen and Poulsen, 1992), dimethylsulfoxide reductase (Schindelin *et al.*, 1996), endonuclease V (Davies *et al.*, 1993), and formate dehydrogenase B (Boyington *et al.*, 1997). Although the enzymes do not exhibit significant sequence similarity, several common interactions including hydrophobic contacts and sidechain-mainchain H-bonds, that seem to be important for stabilizing the fold were identified. Residues at the ψ-turn (Ile28, Ala79) are hydrophobic and conserved among proteins of the fold, presumably for stabilising the β-barrel. Castillo *et al.* (1999) also identified a common functional role in these enzymes for the ψ-loop as a good place to position the catalytic or cofactor binding site, since in all these structures central active site residues are located in the ψ-loop between strands β1 and β2 (β1′ and β2 in ADC), indicating a common structural role for it (Figure 4.4).

Importantly, the ADC crystal structure revealed a fully processed cleavage site in one subunit, and a superposition of the ester intermediate and the fully processed cleavage site in the other subunit. Based on this observation the active site could be reliably identified (Figure 4.3). However, the interpretation of the electron density for the ester intermediate was not entirely straightforward but was eventually convincing because of a better data fit of the structural model with the modeled ester intermediate and more optimal sterochemistry (Albert *et al.*, 1998). The structure provided the first direct evidence for the existence of an ester intermediate in self-processing.

The self-cleavage site is located at the interface of two subunits. It was suggested that processing occurs after tetramer assembly and that it is a cyclical, co-operative process, more rapid for the first subunits of the proenzyme, but progressively slowing down as successive subunits cleave, because of subtle conformational change. The unexpected stability of the fourth ester may be explained by the fact that the residues likely to be involved in its further processing are then too far away for ester cleavage (Albert *et al.*, 1998).

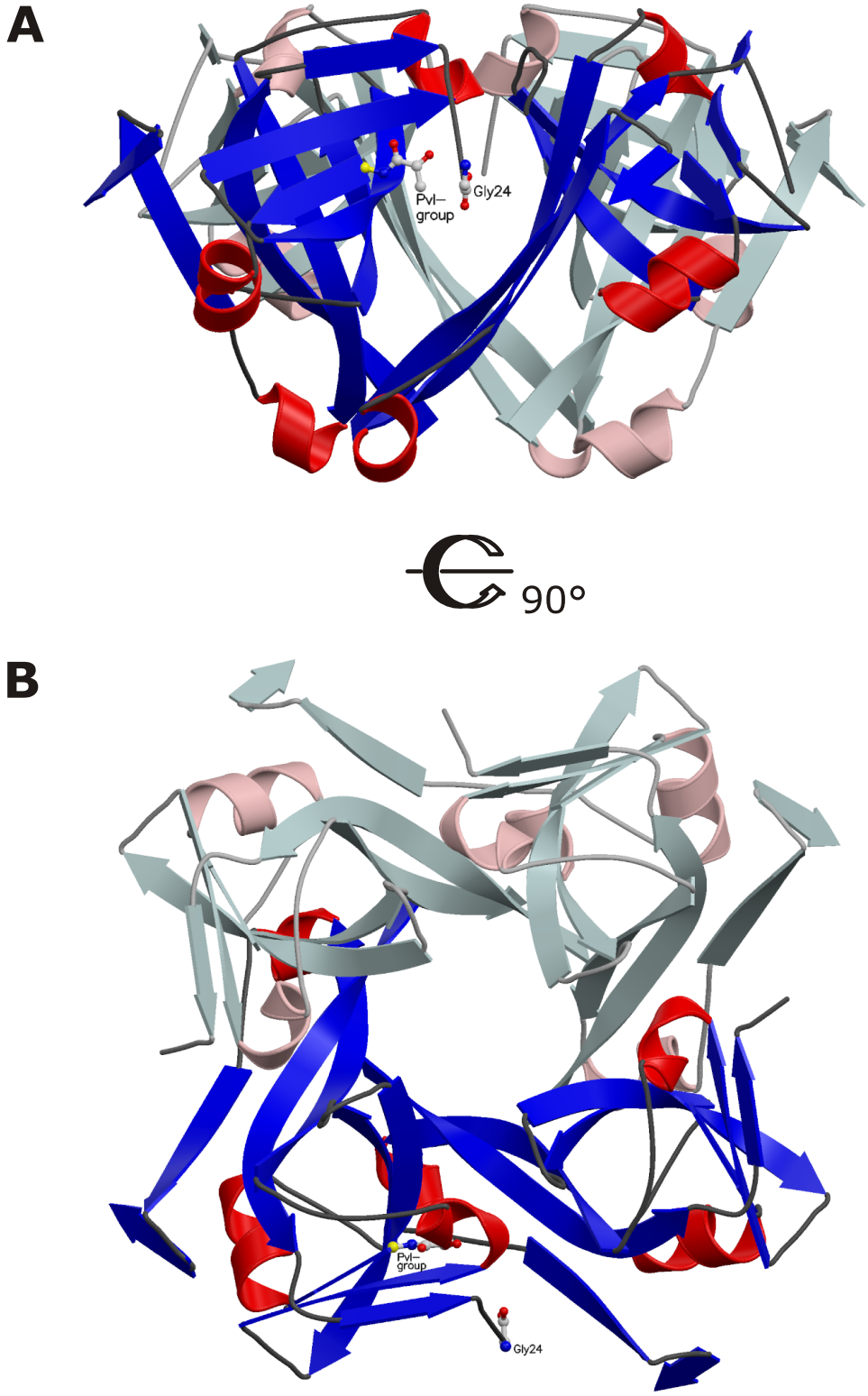

Figure 4.5 Cartoon representation of the ADC tetramer (Albert *et al.*, 1998). (**A**) View of the ADC tetramer with the internal four-fold symmetry axis in the plane of the paper. The pyruvoyl group (Pvl-group) and Gly24 of one subunit are shown in ball-and-stick representation. For this and all following figures of ADC the carbon atoms of the two self-cleavage residues 24 and 25 in ADC will be coloured in grey. (**B**) Rotated view of **A**, down the internal four-fold symmetry axis.

Based on the ADC crystal structure, it was also proposed that the nucleophilicity of the Ser25 may be enhanced by relaying the hydroxyl proton through a network involving a triad of conserved residues of Lys9, His11, from an adjacent subunit, and Tyr58 from the same subunit as the cleavage site (Figure 4.3). Increasing the nucleophilicity has been anticipated to be necessary for the self-processing reaction. Additionally, mechanical strain in the loop region, including residues His21, Tyr22, and Glu23 preceding the self-cleavage site (Figure 4.3), was suggested to play a role for ester formation (see Section 4.2).

In addition to the apo-enzyme crystal structure of ADC, several crystal structures of ADC in complex with various ligands have been determined by von Delft (2000). The enzyme was co-crystallized with substrate L-aspartate, product β-alanine (Figure 4.3), reduced product, and α-methyl aspartate as well as 3-amino-4-methyl-pentanoate. The crystallographic analysis enabled the authors to propose a catalytic mechanism for decarboxylation (see Section 4.3).

Crystallization and preliminary crystallographic analysis for *H. pylori* ADC have been described (Kwon *et al.*, 2002), but, to date, no other ADC structure has been reported.

4.5 Motives for studying ADC

Until recently, it has generally been assumed that the conversion of all inactive protein precursors to biologically active proteins is mediated by specific processing enzymes. Within the last two decades, self-catalyzed backbone rearrangements have been observed, described, and characterized in a number of physiologically highly important eukaryotic and prokaryotic enzymes. N→O and N→S acyl rearrangements represent common methods of activation and protein maturation of self-processing in diverse bi-

ological processes. Although first considered unrelated, it emerged in recent years that the molecular basis for this initial step in the mechanism for the four types of self-processing (Figure 4.1), is indeed closely related.

Although a number of processed structures of self-processing systems have been determined, structural information about unprocessed precursors of self-processing enzymes is relatively sparse. Because processing is generally a very rapid process in almost all self-processing systems, investigation of the unprocessed precursor enzyme is usually very difficult. Most investigators chose to circumvent this problem by mutating the nucleophile to a non- or less-nucleophilic residue (e.g. Ser→Ala mutants) and pursue structure solution for these precursor mutants. The nucleophile was then computationally modeled to deduce the native cleavage site conformation. Structure solution of a native precursor has so far proved impossible and no unprocessed cleavage site, from a self-processing enzyme that has not been mutated, has been reported. Consequently, it is not known to what extent the cleavage site structures of mutants that had their nucleophilic Cys, Ser, or Thr substituted resemble that of the native, unprocessed cleavage site. Therefore, the validity of the approach to computationally model the nucleophile in place of the mutated residue remains unproven. Dissecting the underlying molecular mechanism and the exciting (bio)chemistry in more detail will clearly require more structural data of self-processing enzymes in a pre-processed state.

In *E. coli* ADC, in contrast to other systems, self-processing is comparatively slow, which suggests that ADC is a particularly suitable model system for structural studies of the molecular mechanism of self-processing. Critical issues for the cleavage reaction in general and for pyruvoyl formation in ADC in particular are the steric relationship of the nucleophile to the carbonyl, the orientation of the leaving group (the amino nitrogen of the scissile peptide bond), identification of assisting catalytic residues in the protein, and perhaps most importantly, the energetic driving force be-

hind the isomerization event during self-processing. Addressing these issues should assist in coming towards a more comprehensive understanding of self-processing.

Insights into the molecular basis of the mechanism can potentially aid in producing drugs to prevent protein maturation and may be helpful in developing novel engineering tools. Intein splicing can already successfully generated multiple protein engineering applications, such as the IMPACT (intein-mediated purification with an affinity chitin-binding tag) (Chong *et al.*, 1997), EPL (expressed protein ligation) (Muir *et al.*, 1998), and IPL (intein-mediated protein ligation) (Evans *et al.*, 1999) techniques. Furthermore, since intein-containing precursors are inactive, chemical reagents that block splicing would have significant pharmacological value. Controllable inteins inserted into foreign contexts would provide a method for activating an enzyme on command, allowing biosynthesis of cytotoxic proteins or controllable gene knockouts. Vincent *et al.* (2003) substituted the N-terminal hedgehog domain with different proteins and by this means managed to target proteins to the cell membrane.

ADC presumably catalyzes the rate-limiting step in pantothenate biosynthesis and as such is a suitable target for the development of drugs. Preventing the protein maturation event in ADC, for instance by small molecule inhibitors, could render the protein inactive.

5 Scope of this thesis

Research on the four enzymes of pantothenate biosynthesis in Cambridge was initiated in the two laboratories of Prof. Chris Abell, in the Cambridge University Chemical Laboratory, and Dr Alison Smith, in the Plant Science Department, University of Cambridge. Consequently, the *pan* enzymes have all been cloned in their laboratories, and initial overexpression and purification of the enzymes has been developed there. In addition, much of the biochemical characterization has been carried out in their groups. The aim of the scientific investigation is to gain an in-depth understanding of the function and mechanism of the four enzymes in the pantothenate pathway.

The laboratory of Prof. Tom Blundell became involved in the research about seven years ago. Since then, crystal structures for KPHMT (von Delft *et al.*, 2003), KPR (Matak-Vinkovic *et al.*, 2001), ADC (Albert *et al.*, 1998), and PS (von Delft *et al.*, 2001), all from *E. coli*, have been determined in the colloboration with Prof. Chris Abell and Dr Alison Smith. With these structures at hand more detailed functional studies by crystallography of site-directed mutants, and complexes with ligands can now be pursued.

The aim of the research, described in this thesis, was to cocrystallize KPHMT with cofactor analogues, obtain crystals suitable for X-ray diffrac-

tion data collection, and solve the crystal structure. Furthermore, with the KPHMT crystal structure just solved, it was imperative to carry out a detailed comparative analysis of the structure, in order to possibly derive functional aspects from other enzymes.

For ADC, a series of site-directed mutants had been constructed to investigate the details of the self-processing mechanism. The aim was to develop a fast and reliable purification system, and to crystallize several of the ADC mutants. Furthermore, by solving their structure, and analysis of the latter, it was hoped to identify key residues involved in the self-processing reaction.

The following parts and chapters of this thesis are structured as follows:

In Part II the relevant materials, and biochemical, crystallographic, and computational methods, used in this study, will be described.

In Part III the results will be presented and discussed. This starts with Chapter 10, which will elaborate on the cocrystallization of KPHMT with two cofactor analogues; attempts to collect X-ray diffraction data; the encountered problems that were hampering successful data collection and structure solution; and a proposal for how to proceed for structure solution.

Chapter 11 discusses the detailed structural analysis of the KPHMT fold, which enabled the structure to be assigned to the phosphoenolpyruvate/pyruvate superfamily in the SCOP database; and interesting similarities in the active sites of enzymes in this superfamily.

Chapter 12 describes the establishment of a reliable purification system for ADC mutants, crystallization of several of them, and X-ray diffraction

data collection. Furthermore, the structure solution process (structure solution, refinement, and model building) of the first native precursor of a self-processing enzyme and three site-directed mutants of ADC, one of the latter at near atomic resolution, will be presented.

Chapter 13 encompasses the comparative analysis of seven ADC mutant structures with the native precursor ADC structure. It will be shown that conformational constraints apparently account for the loss of self-processing activity in certain mutants. A detailed possible reaction mechanism for the self-processing reaction will be suggested. Structural comparison carried out with other self-processing systems will show remarkable similarity, particularly with the enzyme glycosylasparaginase. Based on the investigation, several general features of the self-processing biochemistry appear to emerge.

In Part V the reader will find a description of the crystallization of morphine dehydrogenase in the presence of $NADP^+$. This research is however unrelated to the main focus of this study.

Part II

Materials and methods

6 Materials

6.1 Chemical compounds

All chemical compounds that were used for the biochemical procedures were purchased from Sigma-Aldrich or Fluka, unless stated otherwise in parentheses in the text. Sterile, de-ionized water was supplied by the MilliQ system (Millipore) in the laboratory.

E. coli growth media composition

All bacterial growth media were autoclaved (Astell Hearson, 2000 series) for one hour, and left to cool down to room-temperature, before usage.

2xYT medium (Difco Laboratories)	1.6% (w/v) bactotryptone, 1% (w/v) yeast extract, 86 mM NaCl.
LB medium (Difco Laboratories)	1% (w/v) bactotryptone,

	0.5% (w/v) yeast extract, 0.17 M NaCl.
SOC medium (Invitrogen)	0.5% (w/v) yeast extract, 2.0% (w/v) tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM $MgSO_4$, 20 mM glucose.
LB agarose plates	20 g ultra-pure agarose (GibcoBRL), dissolved in 1 l of MilliQ H_2O, supplemented with LB medium; after autocla- ving and cooling to below 60°C, 100 mg of ampicillin was added.

Components for SDS-PAGE

Buffer I	1.5 M Tris-HCl, pH 8.8.
Buffer II	0.5 M Tris-HCl, pH 6.8.
30% (w/v) acrylamide (Severn Biotech)	29.2% (w/v) acrylamide, 0.8% (w/v) bis-acrylamide.
Running buffer	25 mM Tris, 0.192 M glycine, 0.1% (w/v) SDS.

2x-sample buffer	0.35 M Tris-HC, 20% (v/v) glycerol, 4% (w/v) SDS, 2% (v/v) β-mercaptoethanol, 0.02% (w/v) bromphenol blue.
Coomassie blue stain solution	0.5% (w/v) Coomassie blue, 45% (v/v) methanol, 18% (v/v) acetic acid.
Destain solution	60% (v/v) MilliQ, 30% (v/v) methanol, 10% (v/v) acetic acid.
TEMED (BioRad)	
APS	

Table 6.1 Gel composition for two SDS-PAGE stacking gels with 1.5 mm spacers (BioRad) Upon mixture of these components 10 µl of TEMED was added to start the polymerization

Component	30% (w/v) acrylamide	Buffer II	MilliQ	10% (w/v) SDS	10% (w/v) APS
Volume	0.5 ml	1.0 ml	2.4 ml	40 µl	40 µl

Table 6.2 Gel composition for two 12% (w/v) acrylamide SDS-PAGE separating gel with 1.5 mm spacers (BioRad) Upon mixture of these components 10 µl of TEMED was added to start the polymerization

Component	30% (w/v) acrylamide	Buffer I	MilliQ	10% (w/v) SDS	10% (w/v) APS
Volume	4.0 ml	2.5 ml	3.3 ml	100 µl	80 µl

Components for tricine-PAGE

Buffer I:	3 M Tris, 0.3% (w/v) SDS, pH 8.45

5x Bottom electrode buffer	1 M Tris, pH 8.9
Top electrode buffer	0.1 M Tris, 0.1 M Tricine, 0.1% (w/v) SDS, pH 8.2
Fixative solution for gel development	25 mM Tris, 40 mM glycine, 50% (v/v) methanol, 5% (v/v) acetic acid
TEMED	
APS	

Table 6.3 Gel composition for one 8% (w/v) acrylamide tricine-PAGE stacking gel with 1.0 mm spacers (BioRad) Upon mixture of these components 3 µl of TEMED was added to start the polymerization

Component	30% (w/v) acrylamide	Buffer I	MilliQ	10% (w/v) APS
Volume	0.47 ml	0.73 ml	1.75 ml	36 µl

Table 6.4 Gel composition for one 10% (w/v) acrylamide tricine-PAGE spacer gel with 1.0 mm spacers (BioRad) Upon mixture of these components 2 µl of TEMED was added to start the polymerization

Component	30% (w/v) acrylamide	Buffer I	MilliQ	10% (w/v) APS
Volume	0.40 ml	0.66 ml	0.92 ml	10 µl

Table 6.5 Gel composition for one 16.5% (w/v) tricine-PAGE separating gel with 1.0 mm spacers (BioRad) Upon mixture of these components 2.6 µl of TEMED was added to start the polymerization

Component	30% (w/v) acrylamide	Buffer I	MilliQ	80% (w/v) Glycerol	10% (w/v) APS
Volume	2.72 ml	2.72 ml	1.59 ml	1.06 ml	40 µl

Tricine-PAGE is a modification of the original SDS-PAGE as developed by Laemmli (1970). The method was developed by Schagger and Von Jagow (1987) and allows separation of small molecular weight peptides, especially in the range of 2-20 kDa. Instead of a trailing glycine ion, tricine is used. Alternatively to self-prepared tricine-gels, 10-20% (w/v) acrylamide gradient tricine-gels (NovexInvitrogen) were used.

6.2 Chromatographic hardware

Fast performance liquid chromatography (FPLC) hardware (Pharmacia) was exclusively used for protein purification. For purification at room temperature and at 4°C the Äkta Explorer and the Äkta Prime system, respectively, were used.

7 Biochemical procedures

7.1 Transformation of plasmids into *E. coli* cells

Overexpression of recombinant proteins is today a routine-procedure in biochemistry and has the advantage of obtaining large quantities of the protein of interest, which are naturally often present only in relatively small quantities in the cell. Quantities of several milligrams of pure protein have, so far, been a prerequisite to successfully obtain protein crystals. For overexpression of many proteins, the protein production machinery of *E. coli* can be used. Because of the ease of use of this organism, it is often the first choice as host for overexpression. In order to insert a specific gene into bacteria, transformation of a plasmid through the cell membrane into *E. coli* cells is carried out. This procedure makes use of the fact that the cell wall and membrane of *E. coli* can be made permeable by treatment and partial destruction with chemicals, such as $RbCl_2$ and detergents, making the cells "competent" for transformation. During transformation, the sample is heated up in order to increase the mobility of the lipids in the membrane, thus allowing the plasmid to diffuse into the cell. The subsequent cold-shock procedure decreases the permeability of the membrane again, ensuring that the plasmid is kept inside the cells.

Plasmids containing the *panD* mutants, and the *panB* genes for the over-

expression of recombinant ADC mutants and for KPHMT respectively, were provided from the collaboration with the groups headed by Prof. Chris Abell and Dr Alison Smith. Competent cells were available from the respective laboratories and from members of the laboratory of Prof. Tom Blundell. In general, standard transformation protocols for inserting the plasmid into the expression host as described by Sambrook *et al.* (1989) were followed. The plasmids and *E.coli* expression strains that were used for the overexpression of ADC mutants and KPHMT are shown in Table 7.1.

Table 7.1 Plasmids and *E. coli* strains used for the transformation.

Protein	Gene	Plasmid	Antibiotic resistance	*E. coli* expression strain
ADC	*panD*	pRSETA	Ampicillin	BL21 C41
KPHMT	*panB*	pUC	Ampicillin	BL21 C41

The pUC plasmid (Vieira and Messing, 1991) contains a *lac* promoter upstream of both the gene of interest and the *lacI* gene, to which the *lac* repressor binds. Upon addition of IPTG (Melford Laboratories) the *lac* repressor is inactivated allowing the polymerase gene and the gene of interest to be expressed. The pRSETA plasmid (Invitrogen) is a small (2.9 kbases), high-copy number plasmid with a coding site for an N-terminal hexa-histidine tag and a thrombin protease cleavage site. Expression of the gene of interest is dependent on the T7 promoter (Hyvonen and Saraste, 1997), which relies on the fast and specific RNA polymerase of the T7 bacteriophage. Similarly to the pUC plamid, overexpression of the gene of interest is induced upon addition of IPTG.

The transformation protocol was as follows:

1. Transformation-competent *E. coli* cells, which were pre-treated with calcium and stored at $-80^{\circ}\mathrm{C}$ were thawed on ice.
2. 1 µl of plasmid ($\sim$20 ng/µl) was added to 50 µl of C41 BL21 *E. coli* cells.

3. The sample was incubated on ice for 20-30 min.
4. Subsequently, the mixture was heat-shocked for 1-2 min at 42°C on a water bath.
5. The mixture was kept on ice for another 2 min.
6. 1 ml of LB medium for ADC mutants or SOC medium for KPHMT medium, respectively was added.
7. The mixture was kept at 37°C in the incubator shaking for one hour.
8. The culture was centrifuged at 6500 r.p.m. for 5 min.
9. The cells were spread onto sterile, solid LB agarose plates, which were supplemented with ampicillin (Melford Laboratories) (final concentration: 100 µg/ml) with a sterilized, bent glass rod.
10. The plates were incubated at 37°C overnight and stored at 4°C in the coldroom or the fridge for subsequent picking of the colonies.

7.2 Protein overexpression in *E. coli*

Overexpression uses the cell machinery of the host to specifically produce a protein of interest in high quantities. Overexpression of the protein of interest in *E. coli* often relies on a specific RNA polymerase (predominantly the bacteriophage T7 polymerase), the gene for which has been inserted into the genome of the host cell. This RNA polymerase strongly interacts with a recognition site (T7 promoter) prior to the gene of interest on the plasmid, enabling production of large amounts of RNA transcripts and consequently overexpression of the protein. The plasmid then typically has a site for a gene that produces a repressor (such as the *lac*I repressor) that inhibits the transcription of the genomic RNA polymerase by binding to a specific DNA binding site (such as the lac operator site) prior to the

location of the RNA polymerase gene. Some plasmids, such as pRSETA used in this study, also have the same DNA binding site to which the *lac* repressor can bind located after the T7 promoter site.

Addition of a specific molecule (such as IPTG) that binds to the repressor, prevents the latter from binding to the operator site of the RNA polymerase allowing transcription of RNA polymerase. In this manner, overexpression can be controlled and selectively induced. Generally, overexpression is induced when the cells reach a late logarithmic phase, where a large amount of cells in conjunction with a high potential of them to produce the protein of interest exists. The RNA polymerase can then act on the promoter on the plasmid and transcribe the gene of interest. In order to prevent the cells from ejecting the plasmid, a gene for a protein (generally β-lactamase) is located and transcribed, independently of the gene of interest, on the plasmid. This enzyme can inactivate the antibiotic, that is added to the cell culture medium, such that cells containing the plasmid will survive.

In this study, after the plasmid was successfully transformed, overexpression of the gene to obtain sufficient amounts of protein was pursued. This was usually carried out as soon as possible after the transformation.

The overexpression protocol was as follows:

- A single colony of the transformed cells was picked from the agar plate with a sterilized pipette tip.
- The tip was then put into a sterile tube (Millipore) with 5 or 10 ml of autoclaved LB or 2xYT medium, to which ampicillin (usually stored as stock solutions at $-20^{\circ}\mathrm{C}$; alternatively it was prepared freshly) was added to reach a final concentration of 100 µg/ml.
- Thus obtained miniculture was then incubated in an orbital shaker, shaking at 200 r.p.m. at $37^{\circ}\mathrm{C}$ overnight, for $\sim$14 h.

- The next day the miniculture was added to 1 l in a 2 l Erlenmeyer flask with autoclaved LB medium, containing ampicillin (final concentration: 100 mg/ml).
- The flask with the culture was left shaking and the OD_{600} was spectrometrically controlled (Pharmacia, Novaspec II) in half hours steps.
- At an OD_{600} of 0.3-0.4, which usually took $\sim$ 3-4 h to reach, expression of the protein was induced by addition of IPTG (final concentration: 1 mM).
- Incubation at 37°C was maintained for a further 4 h (for ADC) and for up to 16 h (for KPHMT).
- The cells were centrifuged for 40 min at 4°C at 4300 r.p.m. with a Beckmann 1059G centrifuge.

The use of the pRSETA plasmid in conjunction with the BL21 C41 cells proved to be successful and up to 80 mg of protein per litre of culture could be routinely produced for the ADC mutants, as judged by the approximate absorbance at 280 nm of the lysate, using the theoretical extinction coefficient of the protein (see Subsection 7.3.1)

Likewise to ADC, overexpression of KPHMT from the pUC plasmid yielded amounts of up to 100 mg protein per litre of *E. coli* culture. However, following a suggestion by Dr Adrian Saldanha, induction with IPTG was kept for a longer period than for the overexpression of ADC. In general, it took a longer time for these cells to reach the equivalent optical density as for the cells overexpressing ADC. Once successfully overexpressed, a small amount of cells was flash frozen in liquid N_2 in an Eppendorf tube in the presence of glycerol at a final concentration of 20% (v/v). These stocks were then kept at -80°C. For further usage a pipette tip of cells was plated on an agarose plate, which produced colonies that were subsequently used for further overexpression.

7.3 Protein purification

Protein purification is a necessary step to obtain a pure and homogeneous protein sample. The term pure does not only refer to the lack of other contaminating macromolecules, but often also to the existence of a single molecular species of the macromolecule of interest, ideally identical in length, conformation, and charge. Because a crystal predominantly consist of an assembly of the same molecule interacting with each other, contamination with other proteins often prevents crystal formation or crystal growth. As a rule of thumb, a high purity of the protein can be indicated by a single band on SDS-PAGE, or, more so, by a monodisperse distribution of a single molecule species in dynamic light scattering experiments.

Liquid chromatography, such as affinity, hydrophobic interaction, ion-exchange, or size exclusion chromatography, is the most widely-used separation method for macromolecules. In general, it is based on the different interactions of the different macromolecular species, in the sample, with an immobilized solid phase (the resin), resulting in differing degrees of retardation, and therefore separation.

In this study, the following chromatographic techniques were used to purify protein:

- Affinity chromatography with Ni^{2+} and histidine-tagged protein

  This technique relies on the highly-specific interactions between the histidine residues and the Ni^{2+}.

- Anion-exchange chromatography

  The cationic group (CH_2-$N^+(CH_3)_3$) is immobilised on a cross-linked agarose hydrophilic bead similar to the one used for the gelfiltration (pre-packed Hiprep 16/10 SP XL, MonoQ HR 5/5 and Resource Q,

6ml columns). Anionic groups of the proteins will interact with the the cationic group of the resin, thus resulting in retardation.

- Hydroxyapatite chromatography

  Hydroxyapatite ($Ca_5(PO_4)_3OH)_2$ is a specific form of calcium phosphate that can be used for the separation of proteins (BioRad-gel hydroxyapatite, BioRad). Whereas the cationic groups of the protein can interact with the phosphate groups, the anionic groups interact with calcium sites of the resin.

- Size exclusion or gel-filtration chromatography

  A cross-linked matrix of agarose covalently attached to dextran (Sephadex resin) was used (pre-packed Superdex 75 10/30, Superdex 200 10/30, and Superdex 26/60 columns, Pharmacia). This matrix combines the gel filtration properties of cross-linked dextran with the high physical and chemical stability of highly cross-linked agarose. The matrix allows high resolution separation of macromolecules with relatively high flow-rates, at high pressure, and low, non-specific interaction of the matrix enables high recovery of protein.

7.3.1 Purification of ADC mutants

Purification of S25T and R54K ADC mutants overexpressed from pBluescript KS plasmid

After I started to work on the project, attempts were made to purify non-histidine tagged R54K and S25T ADC mutants. The purification protocol for these was similar to that for the wild-type ADC by Ramjee *et al.* (1997) and the whole procedure usually required seven working days. However, the purification was not successful to produce sufficient amounts of

pure protein for crystallization and was therefore not pursued further. It will however be mentioned briefly. The protein purified by this method yielded only very small crystals, unsuitable for X-ray diffraction studies. At the same time as I attempted to purify S25T and R54K ADC mutants by this method, Dr Brent Nabbs subcloned a number of ADC mutants from the pBluescript KS plasmid into pRSETA plasmid with a coding sequence for an N-terminal hexa-histidine tag.

Clones for the ADC mutants had been previously constructed by Dr Mike Witty, who also supplied them and introduced me to the purification protocol (Dr Michael Witty, personal communication). ADC mutants S25T and R54K were overexpressed from a pBluescript KS (Stratagene) plasmid, which was transformed into *E. coli* SJ16 cells according to a protocol similar to the one described in Section 7.1. SJ16 cells (Jackowski and Rock, 1981) are β-alanine auxotrophs, defective in ADC activity. The mutation in the *panD* gene encodes for an ADC version, that is identical to wild-type ADC with the difference of an additional eleven C-terminal residues.

A purification procedure similar to the purification of wild-type ADC, as described by Ramjee *et al.* (1997), was followed with small modifications. Harvested cells were resuspended in 10 mM Tris/HCl, 1 mM EDTA buffer pH 8.0. After disrupting the cell membranes with two passes through a French pressure cell press (SLM Aminco, SLM Instruments), at 20-40 000 psi, a 30-79% (w/v) $(NH_4)_2SO_4$ precipitation was carried out. This was potentially helpful to separate the protein from DNA or RNA that could influence the efficiency of the subsequent ion-exchange steps. The cell pellet was redissolved and dialyzed against 10 mM Tris/HCl, 1 mM EDTA, pH 8.0. Therafter, the protein sample was loaded on a Q-Sepharose column (Pharmacia), after which the protein was eluted with 10 mM Tris/HCl buffer, pH 6.8, with a 0.0-1.0 M KCl gradient. The eluted protein fractions were analyzed with tricine-PAGE and those containing ADC were concentrated to $\sim$10 ml volume using Centriprep 10 (Amicon) concentrators. Until this stage no problems were observed.

The next step involved purification with hydroxyapatite chromatography and at this stage large amounts of protein were lost. For the purification, hydroxyapatite (BioRad Laboratories; Bio-gel hydroxyapatite) packed into an Econo-column (BioRad) was used. The protein was attempted to be eluted with a 50-300 mM gradient of K_2HPO_4/KH_2PO_4 buffer pH 7.0, as recommended for acidic proteins (wildtype ADC has a theoretical pI of 5.75 as calculated with the Protparam online utility[1]). Dr Mike Witty had reported successful elution with this purification protocol for native ADC from hydroxyapatite resin before (personal communication). However, in the attempt to purify S25T and R54K ADC mutants the protein eluted between 0.2-0.3 M K_2HPO_4/KH_2PO_4 and no separation from contaminating proteins was achieved. Consequently, the eluted protein proved to be rather impure and the purity could not be improved by the following gel-filtration with Superdex 200 16/60 nor by anion-exchange chromatography on MonoQ resin (Pharmacia).

After this, Dr Mike Witty made another purification attempt, with a hydroxyapatite chromatography and changed the pH of the elution buffer for the hydroxyapatite step to pH 4.0. The column was equilibrated with 10 mM succinate buffer pH 4.0 and elution was then carried out with a 50-500 mM K_2HPO_4/KH_2PO_4 gradient. However, this attempt did not overcome the problem encountered before and also failed to yield sufficiently pure protein. Since by that time, histidine-tagged ADC mutants had been constructed, no further attempts to manipulate the purification protocol were pursued.

7.3.2 Purification of histidine-tagged ADC mutants

The centrifuged cell pellet obtained as described in Section 7.2 was left shaking in the incubator at 37°C overnight. This was done to allow for the

[1] http://ca.expasy.org/tools/protparam.html

fact that self-processing had previously been observed to be faster *in vivo* than *in vitro* (Dr Mike Witty, personal communication). On the following day, the purification protocol was started:

- Cell lysis

  Harvested cells were redissolved in ∼40 ml of lysis buffer (see Table 7.2) with one or two EDTA-free protease inhibitor tablets (Boehringer Mannheim) that were added immediately prior to the cell lysis.

Table 7.2 Component composition for the lysis buffer for the purification of ADC mutants. The pH was adjusted to 8.0 with NaOH.

Buffer component	NaH_2PO_4	NaCl	Imidazole
Concentration	50 mM	300 mM	10 mM

  Cells were then disrupted by sonication, on ice. For this, a 50 ml beaker with the lysis solution was kept in another 500 ml plastic beaker, filled with a water/ice mixture to prevent the solution from heating. A sonication protocol of 10 minutes total processing time, at medium intensity (5-7 on a 1-10 scale), with 3-5 seconds pulse exposure and 10-15 seconds recovery time was chosen using a sonicator (Misonix). Special care was taken that the sample did not heat up too much, by manually checking the temperature of the beaker with the lysis solution. Subsequently, cell debris was removed by centrifugation at 19 000 r.p.m. (33 902 g) for 45 mins at 4°C with a Sigma centrifuge and a 19776 rotor type. Apart from one ADC mutant, K9A, overexpressed protein from all ADC mutants could be predominantly found in the soluble fractions of the lysate, and purification could proceed immediately.

- Affinity chromatography on Ni^{2+}-NTA resin

  A 25 ml Econocolumn (BioRad) was packed with Ni^{2+}-NTA resin (Qiagen) and equilibrated with several column volumes of equilibration buffer (see Table 7.3).

Table 7.3 Component composition for the column equilibration buffer for the purification of ADC mutants The pH was adjusted to pH 8.0 with NaOH

Buffer component	NaH_2PO_4	NaCl	Imidazole
Concentration	50 mM	300 mM	20 mM

Before loading the protein sample on the column, the supernatant was filtered through a 0.22 µm syringe filter (Millipore). The filtered protein sample was loaded onto the affinity chromatography column, at $\sim 25^{\circ}C$ or room-temperature, and the column was washed with 5 column volumes of equilibration buffer to elute non-bound protein. Elution from the column was carried out by gravity-flow. This was followed by 5-8 column volumes of wash buffer I (see Table 7.4) and 5 column volumes of wash buffer II (see Table 7.5).

Table 7.4 Component composition for the wash buffer I for the purification of ADC mutants The pH was adjusted to pH 8.0 with NaOH

Buffer component	NaH_2PO_4	NaCl	Imidazole
Concentration	50 mM	300 mM	50 mM

Table 7.5 Component composition for the wash buffer II for purification of ADC mutants The pH was adjusted to pH 8.0 with NaOH

Buffer component	NaH_2PO_4	NaCl	Imidazole
Concentration	50 mM	300 mM	100 mM

The histidine-tagged protein of interest was eluted with elution buffer (Table 7.6), and collected in 4 ml fractions.

The NTA resin was generally re-used up to four times. The cleaning was carried out with 5 column volumes of 0.1 M NaOH, 5 column volumes of 0.1% (w/v) SDS, 5 column volumes of 50 mM EDTA, 10 column volumes of MilliQ H_2O and 5 column volumes of 70% ethanol. The column was then recharged with Ni^{2+} by adding

Table 7.6 Component composition for the elution buffer for the purification of ADC mutants The pH was adjusted to pH 8.0 with NaOH

Buffer component	NaH_2PO_4	NaCl	Imidazole
Concentration	50 mM	300 mM	250 mM

2 column volumes of 100 mM $NiSO_4$. The column was stored in 20% (v/v) ethanol at 4°C.

- Gel-filtration on Superdex 200 resin

  Protein fractions were analyzed by tricine-PAGE and those fractions containing the ADC enzyme were combined and concentrated to 1-3 ml with a Vivaspin 20ml concentrator (Vivascience) using a 5 kDa molecular weight cut-off membrane by centrifugation at 7500 r.p.m. at 4°C with a Sigma centrifuge and a 12150 rotor type. A Pharmacia Superdex 200-HR 30/2 (Pharmacia) size exclusion chromatography column was equilibrated with 2-4 column volumes of 25 mM HEPES (Melford Laboratories or Sigma) buffer, pH 7.4. The concentrated protein was then loaded on this column with a syringe (Vivasience), which was connected to the purification system at room temperature.

  The protein was eluted with 25 mM HEPES buffer pH 7.4, collecting 2 ml fractions. All ADC mutants eluted after 12-15 ml, with a flow rate of 0.5-1.2 ml/min (see Section 12.2). During elution the optical absorbance at 230, 260, and 280 nm was monitored. Protein samples were analyzed by tricine-PAGE and the pure fractions containing the desired protein combined and concentrated in a procedure similar to that described above.

  The protein was concentrated to a final concentration between 30 and 40 mg/ml as judged by its theoretical absorbance at 280 nm (ϵ_{280} = 15 340 M^{-1} cm^{-1} or 0.974 ml mg^{-1} cm^{-1}) with a Vivaspin 20 concentrator (Vivascience) with a 5 kDa molecular weight cut-off membrane. All absorption coefficients were calculated with the

Protparam utility[2] (Gill and von Hippel, 1989) from the Swissprot-Expasy internet server. The absorbance was measured using a spectrophotometer (BioSpec-1601E, Shimadzu). The protein solution was concentrated to 10-15 mg/ml and then stored at 4°C for further usage.

7.3.3 Purification of KPHMT

The purification protocol for KPHMT comprised the following steps:

- Cell lysis

  Centrifuged cells were resuspended in ∼30 ml of lysis buffer (see Table 7.7) in which 1-2 EDTA-free protease inhibitors tablets were dissolved.

Table 7.7 Component composition for the lysis buffer for the purification of KPHMT The pH was adjusted to pH 6.8 with NaOH. 10 protease inhibitor tablets per l of buffer were added

Buffer component	NaH_2PO_4	EDTA	DTT
Concentration	50 mM	1 mM	5 mM

  The cells were broken by passing them 2-4 times through an Emulsiflex C-5 homogenisor (Avestin). The apparatus was cooled with cold water during the procedure, which took ∼10-15 min. Subsequently, cell debris were removed by centrifugation at 19 000 r.p.m. (33 902 g) for 45 minutes at 4°C with a Sigma centrifuge and a 19776 rotor type. The overexpressed KPHMT protein could be predominantly found in the soluble fractions of the lysate.

- $(NH_4)_2SO_4$ precipitation

[2] http://ca.expasy.org/tools/protparam.html

Prior to the chromatographic separation techniques, a $(NH_4)_2SO_4$ precipitation was carried out. The supernatant was poured into a 250 ml Erlenmeyer beaker and the beaker put in a 500 ml plastic beaker filled with a water/ice mixture. While the supernatant was left stirring at 4°C , over a time-period of 20-25 min, $(NH_4)_2SO_4$ was slowly added to a final concentration of 15% (w/v) with a small metal spatula. The solution was kept stirring for 25 min and precipitated protein were removed by centrifugation at 19 000 r.p.m. (33 902 g) for 35 minutes at 4°C.

The step above was then repeated almost identically. An additional 25% $(NH_4)_2SO_4$ was added to the supernatant. The precipitated protein pellet was dissolved in ∼20-30 ml extraction buffer (see Table 7.8).

Table 7.8 Component composition for the extraction buffer for the purification of KPHMT The pH was adjusted to pH 6.8 with NaOH. 10 protease inhibitor tablets per l of buffer were added

Buffer component	NaH_2PO_4	EDTA	DTT	KCl
Concentration	50 mM	1 mM	5 mM	100 mM

In order to ensure smooth dissolving, the buffer was added to the pellet in the centrifuge tubes and the tubes were placed in the incubator, shaking at 200 r.p.m. and at a temperature below 10°C. Once the whole pellet had dissolved, the solution was poured into a dialysis tubing (Spectra/Por) with a molecular weight cut-off membrane of 13-18 kDa. The sample was then dialyzed overnight in 2 l of extraction buffer at 4°C.

- First anion-exchange chromatography step

  The dialyzed protein was loaded on an anion-exchange column (Hi-prep Q XL, Pharmacia) equilibrated with extraction buffer, at room temperature or at 4°C. The column was eluted at a flow-rate of 2.5 ml/min with extraction buffer and a KCl gradient from 0.2 to 1.0 M over 200 ml total elution volume, collecting 5 ml fractions. Two KCl gradient plateaus were kept, one at 0.4 M KCl for 40 ml and one at 0.5 M KCl for 25 ml. Protein fractions were analyzed with SDS-PAGE. KPHMT was eluted close the to 0.5 M KCl plateau (see Section 10.2). The combined fractions containing KPHMT, were poured into a dialysis tubing (Spectra/Por, Sprectrum) with a molecular weight cut-off membrane of 12-14 or 15 kDa and dialyzed overnight against 2 l of purification buffer I (see Table 7.9).

Table 7.9 Component composition for the purification buffer I for the purification of KPHMT The pH was adjusted to pH 6.8 with NaOH. 10 protease inhibitor tablets per l of buffer were added

Buffer component	NaH_2PO_4	EDTA	DTT	KCl
Concentration	50 mM	1 mM	5 mM	200 mM

- Second anion-exchange chromatography step

  The last purification step, described above, was repeated essentially identically, using a higher resolution column (ResourceQ 6 ml, Pharmacia), equilibrated with purification buffer I. After dialysis, fractions were concentrated to 2-5 ml with a Vivaspin 20 concentrator (Vivascience) with a 10 kDa molecular weight cut-off membrane by centrifugation at 7500 r.p.m. at 4°C with a Sigma centrifuge and a 12150 rotor type.

- Gelfiltration on Superdex 200 resin

  The concentrated protein sample was loaded onto a size exclusion column (Superdex 200 26/60), equilibrated with purification buffer II (see Table 7.10) at 4°C.

Table 7.10 Component composition for the purification buffer II for the purification of KPHMT The pH was adjusted to pH 6.8 with NaOH 10 protease inhibitor tablets per l of buffer were added

Buffer component	NaH_2PO_4	EDTA	DTT	KCl
Concentration	50 mM	1 mM	5 mM	1000 mM

A flow-rate of 1.3 ml/min was chosen and 4 ml fractions were collected. The protein eluted after ∼125 ml.

- Dialysis

  Protein samples containing pure KPHMT were combined and poured into the dialysis tubing with a molecular weight cut-off membrane of 12-14 kDa and dialyzed twice against 2 l of 25 mM HEPES, pH 7.0 at 4°C. After dialysis, the protein was concentrated to a final concentration of 40-45 mg/ml as calculated with its theoretical absorption coefficient at 280 nm (ϵ_{280} = 16 050 M^{-1} cm^{-1} or 0.568 ml mg^{-1} cm^{-1}) with a Vivaspin 20ml concentrator (Vivascience) with a 10 kDa molecular weight cut-off membrane. The protein was stored at this concentration at 4°C for further usage.

7.4 Assessment of the quality of the purified protein samples

7.4.1 Gel electrophoresis

The purity of the protein samples were assessed either SDS-PAGE by tricine-PAGE. PAGE predominantly separates proteins based on their molecular mass, after they have been denatured by SDS. For ADC tricine-PAGE with three zones, containing different acrylamide

concentrations (6.1) was chosen and for KPHMT 12% SDS-PAGE was carried out (see 6.1). The gels were prepared by using a gel pouring system (BioRad). Protein samples were mixed with an equal amount of 2×-sample buffer and heated for 10 min at 100°C to denature the protein. The samples were then loaded into the wells with a Hamilton syringe. Mark12 (Invitrogen), for ADC mutants, and Dalton 7 (Sigma), for KPHMT, were used as molecular weight markers. The gels were normally run at 75-100 mA and 100-150 V with a Powerpac 300 (BioRad) instrument. Both SDS and tricine gel were then stained with Coomassie blue stain solution for 10-15 min. The process of staining was accelerated by microwaving the solution with the gel for 1 min. tricine-gels were put into fixative solution first and then into Coomassie blue stain. After staining, the gels were put into the destain solution (6.1) for several hours. The gels were then dried between two cellophane sheets (BioRad gel drying apparatus).

7.4.2 Mass spectrometry

Mass-spectrometry is a accurate technique to acquire information of the purity of a protein sample, particularly to determine possible minor size differences of different protein species of the protein sample. What is measured is the mass/charge ratio. Proteins, in general can accommodate several different charges, leading to a series of peaks observed for each protein charge species, from which the mass can be deduced.

After purification, ADC protein samples were given to the Protein and Nucleic Acid Chemistry facility (Department of Biochemistry, University of Cambridge) for matrix-assisted-laser-desorption-ionization time-of-flight mass spectrometry analysis. These analytical steps were carried out by the personnel of the facility. For mass-spectrometry the protein was typically diluted to a concentration of ∼10 µm in a low salt buffer, such as 5 mM Hepes. Dialysis was carried out by repeated dilution and concen-

tration with Ultrafree (0.5 ml, Millipore) concentrators with a molecular weight membrane cut-off of 5 kDa, at 4°C.

KPHMT protein samples were analyzed by the mass spectrometry facility of the Chemical Laboratory (Cambridge University) with a nano electrospray ionization on an Abi Qstar spectrometer. The protein was diluted to a concentration of ∼5-10 pmol/µl, in 10 mM $NH_4CH_3CO_2$, pH 6.0.

7.4.3 Amino-terminal sequencing

After purification, KPHMT samples were submitted for N-terminal sequencing to the Protein and Nucleic Acid Chemistry facility to determine the N-terminal sequence of the protein.

7.4.4 Dynamic light scattering

Dynamic light scattering (see Carter and Sweet, 1997a, p.157) relies on the scattering of monochromatic light by aggregates or particles moving in solution. Because the diffusivity of the particles is a function of their size, measurement of diffusion coefficients can be transformed into hydrodynamic radii and a distribution of the sizes of the molecular species can be calculated.

Dynamic light scattering is an excellent technique to assess how monodisperse a protein sample is. Consequently, it is a very good indicator for the crystallizability of the protein. Dynamic light scattering with a PD2000DLS instrument (Precision Detectors) was attempted several times, but failed to give reproducable results. Even though the machine was repaired, this problem was not solved, and the technique could thus not be used, reliably.

8 Protein crystallography

Crystallography today is the most wide-spread and, probably, most direct technique to obtain atomic representations of macromolecules. Aspects of protein crystallography are covered in numerous books, such as by Blow (2002), Blundell and Johnson (1976), Carter and Sweet (1997a), and Carter and Sweet (1997b). More recently, a comprehensive coverage of macromolecular crystallography and structural biology was also published by Rossmann *et al.* (2001). An overview of the various stages of protein crystallography is given in Figure 8.1 and the individual steps will be discussed in more detail in the next sections.

8.1 Protein crystallization

Protein crystals are an absolute prerequisite to obtain discrete X-ray diffraction patterns of protein molecules extending to a reciprocal spacing, which corresponds to real space atomic resolution. Crystals present matter in a highly ordered and condensed form. In an ideal crystal, the protein molecules assemble in a regular way in three dimensions, so that every molecule is interacting identically with the neighbouring molecule. This crystalline array, with usually 10^{15} molecules (Blow, 2002), is necessary to obtain a sufficiently strong signal of the scattered X-rays. However,

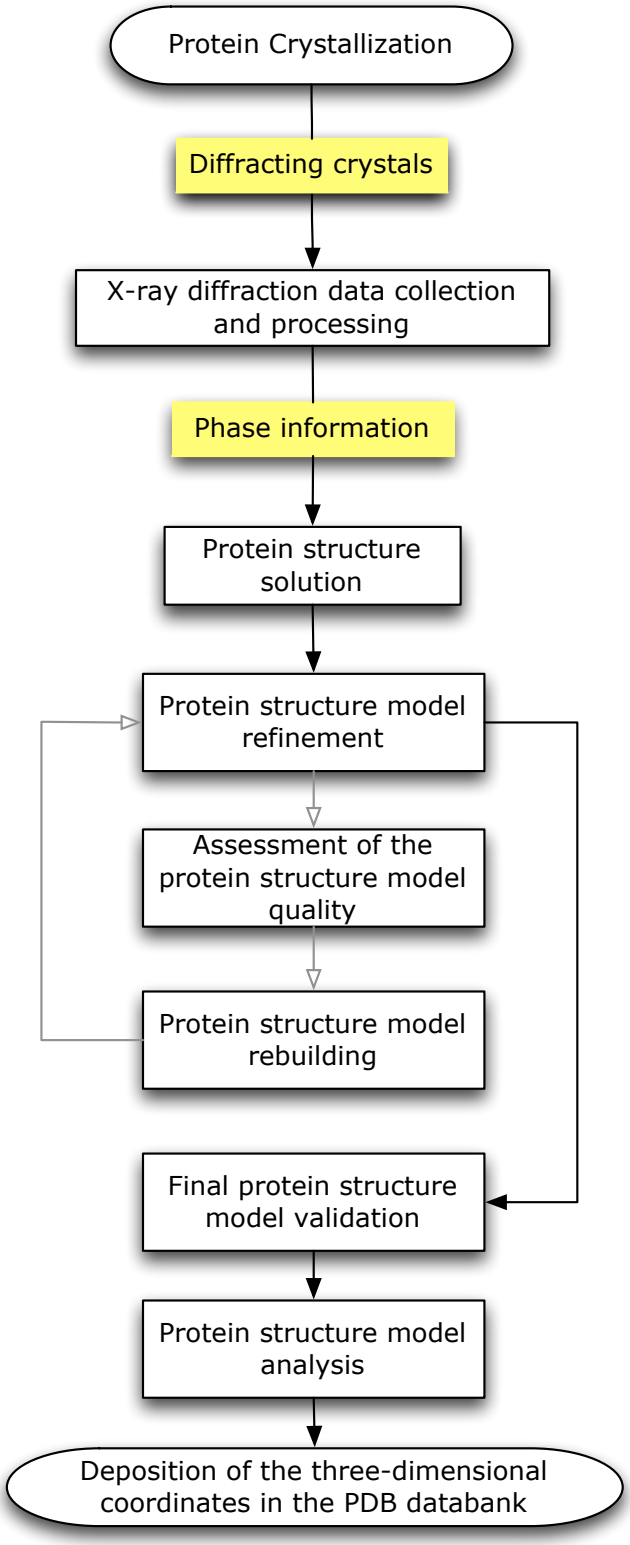

Figure 8.1 Schematic of the crystallographic process from protein crystallization to the final model for the crystal structure. Data information, whose acquisition is generally considered to be limiting for the process, is indicated in yellow colour.

protein crystals, in general, contain a relatively high amount of solvent (30-70% v/v). Also, the number of interactions, between the macromolecules in a crystal is small compared with the total number of atoms within each macromolecule. This often limit the possible achievable maximal resolution to which crystals of macromolecules, such as proteins, diffract.

In order to grow a protein crystal, the protein has to be brought into a supersaturated, thermodynamically unstable state, from which a crystalline phase may develop as it returns to thermodynamic equilibrium. Supersaturation can be achieved by slow evaporation of the solvent, such as by the vapour diffusion technique. Techniques, mechanisms, and practical aspects of protein crystallization are extensively covered by McPherson (1999).

In this study, protein crystallization trials were generally carried out at a constant temperature of 19°C or 4°C in temperature-controlled incubators (Sanyo). The hanging drop technique, with vapour diffusion, was used to screen for suitable conditions and also later to optimize crystal quality. For this technique, a reservoir was filled with 500 µl of an aqueous precipitant solution. The same solution was mixed with the protein solution, of a certain concentration, in different µl-proportions on a small, siliconised cover slip (Hampton Research). Normally, for screening, a 1µl:1µl ratio of precipitant solution to protein solution was chosen. After mixing, the cover slip was turned over to seal the reservoir, so that vapour diffusion could take place between the reservoir and the drop, leading to a volume variation of the drop until an equilibrium was reached between the reservoir solution and the drop. 24-well Linbro plates (Hampton Research, Molecular Dimensions) were used as reservoirs. Commercial screens with pre-prepared precipitant solutions (Hampton Research, Jena Bioscience, and Molecular Dimensions) were used for screening. Assessment of the crystallization trials and crystal growth were monitored with a microscope (Leica MZ12 and Olympus SZPT).

8.1.1 Choice of cocrystallization compounds for KPHMT

In solution, the natural cofactor of KPHMT, 5,10-meTHF, is unstable in the presence of air. A 1 mM solution reportedly degrades by 5% within 60 min (Schircks Laboratories)[1]. Therefore cofactor analogues were used for the cocrystallisation trials. (6*S*)-5-methyl-5,6,7,8-tetrahydrofolic acid ((6*S*)-5-meTHF) and (6*S*)-5-formyl-5,6,7,8-tetrahydrofolic acid ((6*S*)-5-foTHF) were identified as the most promising cofactor analogues. Both compounds share a high similarity with the actual substrate (see Figure 8.2) and carry a substitution at the N_5, so that the molecular structure could resemble the iminium form of 5,10-meTHF (see Section 3.2).

(6*S*)-5-meTHF calcium salt (Schircks Laboratories) is reported to be soluble in water up to a concentration of 10 mM (5g/l) at 25°C, whereas (6*S*)-5-foTHF calcium salt is soluble in water up to a concentration of 20 mM (10g/l) at 25°C. Both compounds are relatively stable when used within hours. However, 6(*S*)-5-meTHF calcium salt at 1 mM reportedly degrades by 4% within four hours (Schircks Laboratories), whereas the same concentration of (6*S*)-5-foTHF calcium salt degrades to that extent within fifteen days (Schircks Laboratories). Folic acid was also purchased, but proved very difficult to solubilize (reported solubility of 3 mM, or 1.6 mg in 1 l, at 25°C).

While for initial crystallization trials a racemic mixture of (6*R*)-5-foTHF and (6*S*)-5-foTHF was used, for the later trials, only the *S*-isomer was used which is known to be the biochemically relevant form for the enzyme (Powers and Snell, 1976). The solutions were always prepared freshly, immediately prior to the crystallization trials, either at 4°C or room temperature. After the solution had been made up, it was centrifuged at 9000 r.p.m. for 10 min to remove residual compound that did not solubilize.

[1] http://www.schircks.com/prices/sys/pteridines_frame21.htm

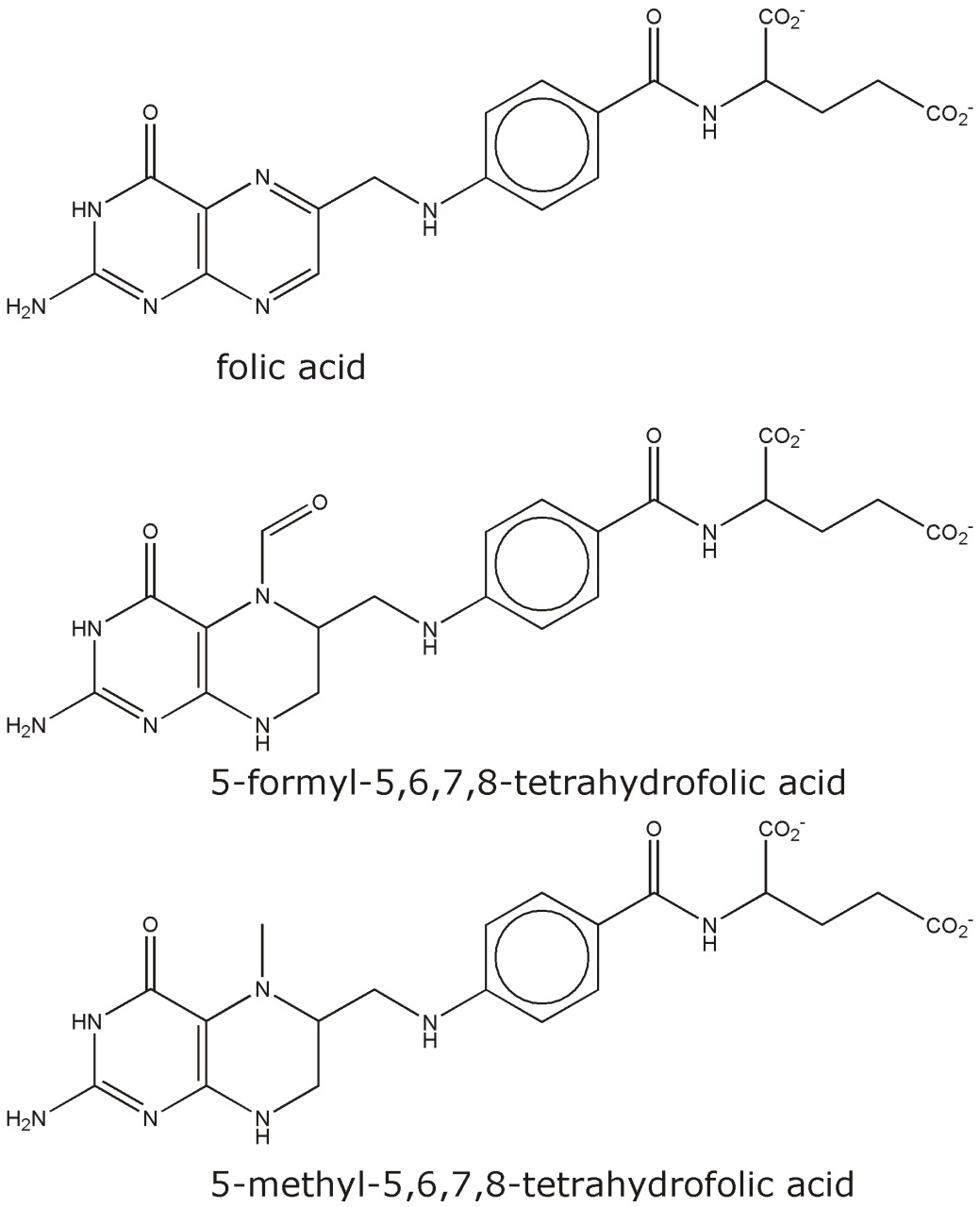

Figure 8.2 Illustration of the cofactor analogues of 5,10-meTHF used for cocrystallization with KPHMT in this study.

Dihydro-4,4-dimethyl-2,3-furandione (Aldrich) was used as the chemical compound for the product. The furan ring is known to be readily attacked by water to form ketopantoate in solution. A similar procedure to form ketopantoate was also used by Zheng and Blanchard (2003). Because the same compound could be cocrystallized and was shown to result in the product bound to the protein (von Delft *et al.*, 2003), it was also used for the crystallization trials pursued in this study. 3-methyl-2-oxobutanoic acid (α-kiva) sodium salt (ACROS Organics), was used for cocrystallization trials with the substrate. All compounds were kept at -20°C for storage.

8.2 Cryoprotection and crystal mounting

Cryocrystallography, reviewed by Garman and Schneider (1997), is used to minimize X-ray induced secondary radiation damage to the macromolecules in the crystal, by cooling down the crystal to a very low temperature. The absorbed energy of the X-rays produces heat and free radicals that damage the macromolecules in the crystal. Whereas the primary damage, caused by the incident X-rays, is independent of the temperature, the secondary damage caused by propagating radicals can be reduced by lowering the temperature. For X-ray diffraction data collection, this enables longer collection times on the same crystal. In this study, N_2 was used to flash-freeze the crystal specimens to a temperature of $-173°C$.

However, this can in general not be done without carrying out a procedure known as cryoprotection of the crystals, in which a low-molecular weight compound, such as glycerol or ethyleneglycol, is added to the crystal, or cocrystallized with the protein. Addition of such a compound (cryoprotectant) in sufficient concentration causes the aqueous solution surrounding the crystal to freeze as an amorphous glass, rather than in a crystalline form. Water in a crystalline form produces characteristic ice-

rings, which can impair X-ray diffraction data indexing (Section 8.3). The vitrified water will not produce X-ray diffraction patterns that could interfere with the X-ray diffraction of the crystallized protein. In addition, the use of cryo-loops (Teng, 1990) also obviates the need to mount the crystal in capillaries. As a consequence, during data collection, background X-ray scattering from the excess mother liquor and the capillary material is avoided.

In order to determine the concentration of the cryoprotectant, necessary to avoid ice-formation, a cryo-loop was dipped into different solutions of the mother-liquor, with varying concentrations of the cryoprotectant. The loop was then mounted on the in-house X-ray data collection facility into a stream of N_2 and an X-ray diffraction image (see Section 8.3) was collected. The lowest concentration of the cryoprotectant in the mother liquor solution, which did not result in significant ice-rings in the X-ray diffraction pattern was subsequently chosen for cryoprotection.

Cryocrystallography was used for all crystals in this study for mounting the crystals. For cryo-protection the coverslip was inverted, and the mother liquor containing the cryoprotectant was added to the drop with the crystal. Generally, the concentration of the cryoprotectant was increased stepwise, in order to minimize concentration gradients. The crystal was soaked in the cryoprotection solution for 1-10 min, during which the coverslip was covered with an inverted well, that had been removed from a 24 Linbro plate, to prevent evaporation. After the crystals had been cryo-protected in this way, they were collected in cryo-loops of appropriate size (CrystalCap system; Hampton Research) with cryo-tongs (Hampton research) under the microscope. The crystals were then quickly plunged into liquid N_2 and stored in crystal vials (Hampton research). These were kept on racks in larger storage dewars, which can maintain the temperature of below $-173^{\circ}\mathrm{C}$ for several months, until further usage. Alternatively, crystals were directly put in a flow of N_2 gas, produced by a cooling device (Oxford Cryosystems), on the in-house X-ray equipment for screening in the home laboratory.

For each protein a different cryoprotection protocol had to be devised and parameters, such as the time period of the crystal in cryoprotectant solutions, cryoprotectant concentration and speed of transfer in the liquid N_2 had to be optimized. Glycerol, ethyleneglycol, and sodium malonate were used as cryoprotectants in this study.

8.3 X-ray diffraction data collection and data processing

X-rays are electromagnetic waves (or photons) with a wavelength in the range of ∼0.05-0.35 nm. This coincides with the spacing between covalently attached atoms, polar interactions, and H-bonds, as existent in proteins. In general, X-rays can be generated in two physically different ways.

Firstly, they can be generated after an electron in an atom has been elevated from its ground state atomic orbital to a higher energy atomic orbital. On its transition back from the latter to a higher energy orbital than it has been initially in, the energy difference is emitted in the form of heat and an X-ray photon of a specific wavelength. The impulse for exciting an electron from its ground state can for instance come from another electron colliding with the atom. The rotating anode is an example where this effect is taken advantage of in order to produce X-rays. It consists of a filament, which constitutes the cathode, that emits electrons. The cathode has a potential difference of several kilovolts from an anode, which is usually made of Cu. The electrons that are released from the cathode hit the anode and X-rays are emitted (p.125 in Rossmann *et al.*, 2001). Because the heat development limits the intensity of the generated X-rays, the anode is cooled by water and rotated, to dissipate heat.

The second way to produce X-rays is to force electrons, whose speed

is close to the velocity of light, in an orbit by a magnetic field. Because the electrons are deviated from their linear path and undergo a path alteration, they are decelerated. During the deceleration they emit photons of a wider energy spectrum, some of which fall into the range of X-rays. This method is used at synchrotrons (see Lindley, 1999). Electromagnetic instruments such as bending magnets, undulators, and wigglers change the path of the electrons and produce the X-rays which can then be used at stations or beamlines around the storage ring. Synchrotron radiation has the advantage to be far more intense, of higher brilliance and collimation than radiation from a rotating anode. This often enables data collection of protein crystals to a higher resolution, better overall data quality, and in a shorter period of time.

Si, coated with materials such as Rb, is used for mirrors at both X-ray sources, rotating anodes and synchrotrons, to focus the beam. In addition, monochromators and slits are are used at the synchrotron to enable a monochromatic and collimated X-ray beam.

X-rays are predominantly scattered by the electrons in an atom. When an X-ray photon hits an electron, it can be absorbed and, if so, sets the electron into a vibration at the X-ray frequency. The excited electron can then emit an X-ray photon of the original energy and wavelength, in all directions. Coherent X-ray scattering from repetitive points, as in a crystal, can be understood in terms of Bragg's law (Equation 8.1).

$$2d \sin \theta = n\lambda \tag{8.1}$$

In the Bragg equation, d is the separation between two planes of atoms, θ denotes the angle of diffraction, n is an integer, defining the order, and λ is the wavelength of the X-ray radiation. This formula provides an explanation why X-ray diffraction maxima will only be observed at certain angles, at which the path difference between the waves, corresponding to the spacing between the atoms, equals an integer number of the wave-

length. A geometrical way to visualize the Bragg equation is the Ewald sphere, a sphere of radius $1/\lambda$ with the crystal at the centre (Figure 8.3). This schematic representation of diffraction is also useful to understand the relation of real-space and the imaginary, reciprocal space (see Rossmann *et al.*, 2001, p.58). The reciprocal lattice consists of points (with Miller indices hkl) that are the endpoints of reciprocal lattice vectors. The length of these vectors (1/d) is proportional to the inverse of the real-space spacing between two planes in the crystal, separated by distance d.

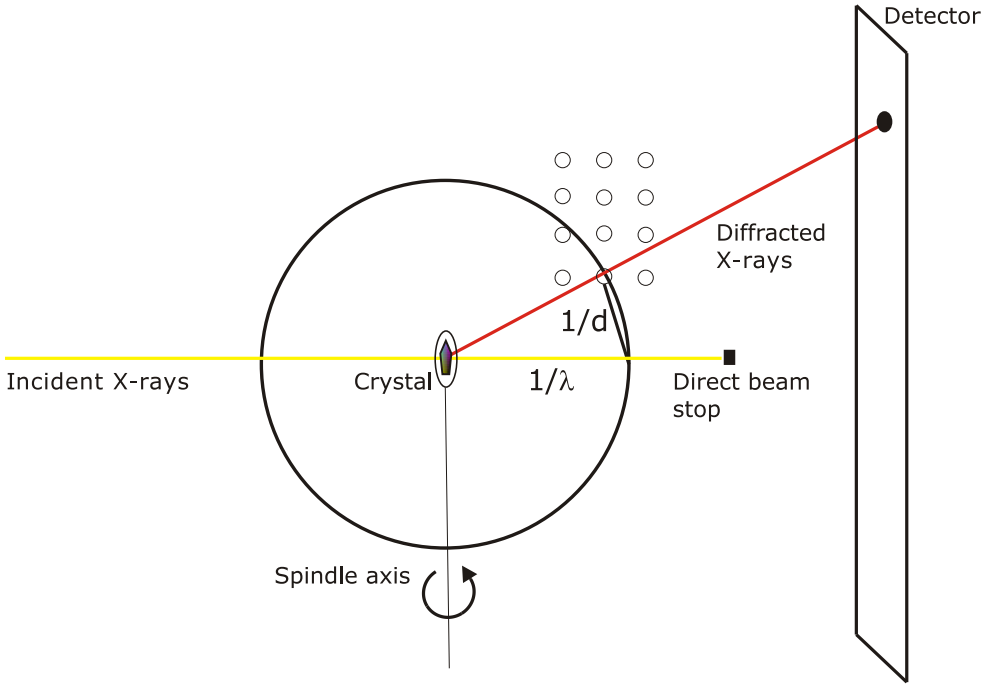

Figure 8.3 Schematic illustration of X-ray diffraction data collection in oscillation mode. The axis of rotation of the crystal, relative to the incident beam (yellow) is indicated. The circle, with radius $1/\lambda$, illustrates the Ewald sphere. A small set of reciprocal lattice points (small circles) are shown. Only reciprocal lattice points that are located on the surface of the Ewald sphere will give rise to a diffracted beam along the line joining the centre of the sphere to the reciprocal lattice-point on the surface of the sphere. Diffracted beams can then be recorded on the detector (black circle). Note that the relative size of the components of the illustration, in particular the Ewald sphere, the reciprocal lattice, and the detector, are not to scale.

X-rays, which are diffracted from the macromolecules in a crystal, are detected by X-ray detectors. These respond to the energy of the diffracted beam, or to the intensity, which is proportional to the number of photons per unit of time that are delivered. The X-ray diffraction detector measures the position of the diffracted X-rays (reflections), relative to the other diffracted X-rays, and its intensity. Two types of two-dimensional detectors were used in this study, image plates and charge-coupled devices.

Image plates store the diffraction image in a temporary form in which atoms are promoted to an excited state, by the X-ray photons, in which they remain trapped. The image plate can then be scanned by a laser beam that measures the intensity of the fluorescence. In general, image plates are storage phosphors made from BaFBr:Eu coated on a suitable backing. The energy of the trapped state can be released as blue fluorescence upon photostimulation by a red laser and the photostimulated emission can be recorded by a photomultiplier. After the diffraction pattern has been recorded the image plate is erased by exposing it to visible light. Many image plate instruments, such as the one used in this study, contain two plates, so that while one is being read out, another image can be collected in the meantime.

In contrast to image plates, charge-coupled devices (Phillips *et al.*, 2002) convert X-ray photons directly into electric current. Charge-coupled devices allow faster read out time (seconds), but are limited in size and in their dynamic range.

After the crystal has been mounted on the goniostat in a more or less arbitrary orientation (the "American method" (see Rossmann *et al.*, 2001, p.209)), X-ray diffraction data with a monochromatic X-ray source can be collected in several different ways (see Dauter, 1999; Evans, 1999, for possible data collection strategies). Among other considerations, optimal data collection strategy has to take into account the wavelength of the X-rays used for the data collection, the orientation of the crystal with respect to

the image plate and the X-ray beam, the position of the beamstop, the exposure time to the X-rays, the crystal-to-detector distance, the oscillation angle (rotation angle increment), the rotation start and the symmetry of the crystal lattice for the overall angle collected. These considerations affect later computational data calculations strongly.

During the X-ray diffraction data collection experiment, indexing of the X-ray diffraction images (see Subsection 8.3.3) is begun in order to monitor the X-ray data collection experiment. X-ray diffraction data processing of the whole dataset then becomes a short iterative process, in which the experimenter attempts to optimize the quality of the collected data, by manipulating parameters in the indexing software, such as spot radius, and/or by discarding images which are deemed of not high enough quality. The overall quality of the X-ray diffraction dataset can then be assessed by the following parameters:

- the low and high-resolution limits of the X-ray diffraction data;
- the mosaicity, a measure of long-range order of the unit cells within a crystal, which essentially defines the angular increment, by which a diffraction image can be rotated and still record the identical reflections as in the state before rotation;
- the completeness of the data as a percentage of the theoretical number of possible reflections;
- the intensities with its standard uncertainties $\langle \mathrm{I} \rangle / \sigma \langle \mathrm{I} \rangle$;
- the $\mathrm{R_{sym}}$ (Equation 8.2) a measure for the agreement of the data

$$R_{\mathrm{sym}} = \frac{\sum_h \sum_i ||I_{\mathrm{hi}}| - |I_{\mathrm{h}}||}{\sum_h \sum_i |I_{\mathrm{hi}}|} \tag{8.2}$$

  where, I_{hi} are symmetry-related intensities and I_{h} is the mean intensity of the reflection with unique index h;

- the redundancy or multiplicity of the dataset, which indicates how often a reflection has been measured on average.

8.3.1 X-ray diffraction data collection facilities

The in-house facility was generally used to screen crystals for diffraction. In the laboratory, in which the research for this study was carried out, X-rays were generated with a Cu-rotating anode (Rigaku MSC), at 50 mA and 100 kV, with a wavelength of 1.541 Å of the Cu Kα radiation. The radiation was collimated using Osmic confocal mirrors. For recording the X-ray data a RAXIS IV image plate (Rigaku MSC) was used.

Synchrotron radiation at Daresbury stations 9.6, 14.1, and 14.2 was then used for X-ray diffraction data collection. The wavelength used at the three stations was 0.870, 1.488, and 0.978 Å, respectively. A Quantum 4R ADSC (ADSC) or MAR 150 mm (MAR) digital charge-coupled device were used for recording of the X-ray diffraction data. Collimation of the X-rays at Daresbury synchrotron is carried out using slits and Rh-coated Si-mirrors.

8.3.2 X-ray diffraction data collection

All X-ray diffraction data were collected using the oscillation method (Arndt and Wonacott, 1977; Otwinowski, 1991), where the crystal is rotated about an axis perpendicular to the incident X-ray beam and a static two-dimensional X-ray detector. For the X-ray diffraction data collection in this study, the plane of the X-ray detector was kept perpendicular to the incident beam. Software packages used to control oscillation angle, exposure time to X-rays, and the total data collection time, included CRYSTALCLEAR (Rigaku MSC) for the in-house facility and PXGEN++ (Mar) and ADX (ADSC) at Daresbury synchrotron.

8.3.3 X-ray diffraction image processing

All computational calculations, described in the following, were carried out on Unix software-operated Silicon graphics Origin 2000 server, a Silicon graphics O_2 machine, or a Linux-based personal computer with Intel Celeron processor.

Diffraction data were indexed and processed with DENZO V1.97.2 (Otwinowski, 1991). The program is used to produce accurate unit cell dimensions, and to assign both Miller indices and a value for the intensity for all the recorded X-ray reflections of a dataset. The program carries out a sequence of calculations and steps, as follows:

Peak search The first step constitutes an initial peak search on a single image, by searching for intensity maxima on the X-ray diffraction image. Each spot must meet certain criteria to be retained as a diffraction spot, such as the signal-to-noise ratio and the size of the reflection.

Auto-indexing Auto-indexing is a required procedure if the crystals are mounted in an arbitrary orientation, as was the case in this study. In order to index and integrate the X-ray diffraction data it is necessary to know the unit cell dimensions as well as the orientation of the unit cell axes, relative to the X-ray beam, and the X-ray detector. To auto-index an image, the only information the user has to supply is the crystal-to-detector distance, the wavelength, the direct X-ray beam position and the image rotation parameter that relates the raster grid orientation to the camera axes. The auto-indexing procedure for DENZO has not been described in detail. However, the algorithm may be assumed to have significant similarities to the one used in the indexing software of DPS and MOSFLM as described by Steller *et al.* (1997).

In these programs, a set of observed reflection positions, determined by the peak searching procedure, is used to compute the corresponding reciprocal-lattice vectors, assuming a stationary crystal. These vectors are then projected onto a chosen orientation and the projections are subjected to a one-dimensional fast Fourier transform. The program explores all possible directions, separated by a small increment, and analyzes the frequency distribution of the projected reciprocal-lattice vectors for each of them. Exploration of such directions establishes the best potential basis vectors of the real cell. The program then determines a reduced cell from the obtained basis vectors, and calculates the best cells for all fourteen Bravais lattices, including a distortion index. The experimenter then has to choose a lattice, which in general will be the highest-symmetry lattice with the lowest distortion index. However, the cell reduction routine can not determine the lattice symmetry.

Diffraction geometry refinement After the unit cell parameters for the first image have been determined, it can be used to predict the position of reflections on other oscillation images and parameters can be refined, so that predicted spots fit the measured reflection positions better. Several parameters, such as the crystal and detector orientation, the unit cell dimensions, the position of the direct beam, and the crystal to detector distance, require refinement for each processed image. For this, a least-squares method is used, which minimizes the deviation of the reflection centroids from their predicted positions. An accurate prediction of the positions of the reflections is crucial to achieve a precise integration of the intensities in the next step.

Integration of X-ray diffraction data intensity Data integration refers to the process of obtaining estimates of diffracted intensities of the individual reflections and their standard deviations. In DENZO the intensity of each reflection is estimated using a method known as profile fitting. In

order to calculate the diffraction intensity, the detector background has to be estimated and then subtracted from the reflection profile. Profile fitting assumes that the actual spot shape or spot profile is known, and the intensity is derived by finding the scale factor that, when applied to the known profile, gives the best fit to the observed reflection profile. For this, an averaged profile is first predicted based on the shape of the other reflections within a chosen radius. The predicted profile is an average of profiles shifted by the predicted separation between the spots. If the predicted reflections are wrong, the profile will be broadened. In the second step, the information from the actual profiles is combined with the predicted one, and the intensity is given by the scale that has to applied to the profile to best fit the reflection. The background is calculated using an average detector signal in the immediate neighbourhood of a specific reflection. The experimenter can influence the integration process by changing spot radius and the size of the background.

Scaling and post-refinement Once the diffraction spots of all images have been assigned an index, the integrated and indexed data have to be placed on a common scale and multiple observations merged together. For data scaling in this study, the intensities were scaled with SCALEPACK V1.97.2 (Otwinowski, 1991). Scale factors are determined by comparison of symmetry-related reflections, which should be equal in intensity. The scale-factors are then adjusted to obtain the best internal consistency of intensities, and applied to all recorded reflections. During the process redundant measurements and symmetry-related measurements are merged and outliers, reflections which do not meet certain criteria, are rejected.

In addition to the scaling procedure, the program refines crystal parameters (unit cell and orientation) for all images. This process of refining crystal parameters, using the combined reflection intensity measurements, is known as global or post refinement. It allows for separate refinement of the orientation of each image, but with the same unit-cell value used for

the whole dataset. The post refinement is also more precise than the processing of a single image in the determination of crystal mosaicity and the orientation of each image. At the end of the scaling and post-refinement procedure the experimenter ends up with a file containing the precise unit cell dimensions and scaled intensities. In order to convert the intensities into structure factor amplitudes and to generate the intensity distribution of the reflections (Wilson plot), the program TRUNCATE (Collaborative Computational Project Number 4, 1994) was used. In this program the intensities are converted into amplitudes and the amplitudes are put on an approximate scale using the scale factor from the Wilson plot.

8.4 Molecular replacement

Because no X-ray lenses exist to recombine the diffracted X-ray, to directly form a visual representation of the diffracting model, indirect ways have to be pursued to reconstruct a representation of it. Therefore, calculation of a three-dimensional representation of the protein structure, always requires a second piece of information, other than the Miller indices and the intensities of the reflections, and that is the phase angle associated with each reflection. Considering X-rays as waves, the phase angle refers to the relative times the waves, which create the different reflections, arrive at the detector. However, the direct phase information of the reflections is lost, because it cannot be recorded on the two-dimensional X-ray diffraction detector. Several indirect ways have been deviced in order to acquire the phase information.

Two proteins that are structurally related or identical will have some similarities in the structure factors, even if they are located in different Bravais lattice types, space groups or unit cells. If a protein model is already known to be structurally similar to the protein, for which X-ray diffraction data has been collected, this model can be used to obtain phase angle

estimates for the unknown protein structure. The technique is known as molecular replacement (see Rossmann *et al.*, 2001, p.263-278) and essentially places the related model (thereafter referred to as the probe) in the unit cell in place of the molecule, one is searching for (thereafter referred to as the target). Then the phases from the known probe are used as first estimates for the unknown protein structure.

Since the structures of both ADC and KPHMT were already available, molecular replacement could be pursued for structure solution in this study. In order to acquire information about the non-crystallographic symmetry and to calculate internal Patterson correlations, the program POLARRFN (Collaborative Computational Project Number 4, 1994), which carries out a fast rotation function search (see below) in polar coordinates, was used. The corresponding Patterson maps were displayed with XPLOT84DRIVER (Collaborative Computational Project Number 4, 1994). For molecular replacement, AMORE (Navaza, 2001) was used as the program of choice to determine the molecular replacement solutions. Two successive steps are carried out.

Rotation function search First, the program has to calculate a Patterson vector map (see Rossmann *et al.*, 2001, p.61) of the target, based on the observed structure factor amplitudes, and of the probe, in a P1 cell. It then tries to determine the angular orientation of the target to the probe by rotating the probe around an origin point. The rotational position of the Patterson map of the probe is changed in increments and correlation with the Patterson map of the target is evaluated using a rotation function. Overlap between the two Patterson maps will result in maxima or peaks for the rotation function, which are then chosen by the experimenter for the subsequent translation function search.

Translation function search If the crystal has symmetry other than the lattice symmetry, a translational search has to be carried out, to determine

the position of the crystal origin relative to the position of the rotation axis relating symmetry-related molecules. The translation function evaluates the overlap between the Patterson maps of the probe and the target. For this the inter-molecular inter-atomic cross-vectors, of symmetry-related molecules, far from the origin of the Patterson map are important. Similarly, to the rotational search, the positional parameters of the probe in the unit cell are varied on a grid. The correct orientation and position of a search model in the crystal unit cell should lead to the maximal overlap of both the self- and cross vectors throughout the entire unit cell. In addition to the peak height of the maxima, a correlation coefficient and the crystallographic R-factor (see below) are calculated to evaluate the molecular replacement solutions. The experimenter then, in general, chooses the highest translation function peak(s) with the lowest R-factor and the highest correlation coefficient(s).

In order to assess the validity of the molecular replacement solutions, the orientation and position in the unit cell has to be examined, which was carried out with MOLPACK (Wang *et al.*, 1991). There should be no clashes between the molecules in the unit cell. After the probe has been positioned, phases can be calculated from the probe, which, combined with the observed structure factor amplitudes of the target, produce an initial model for the target.

8.5 Protein structure model refinement

After an initial atomic model has been derived, the agreement between the model and the experimental data has to be maximized. This is carried out in an iterative process, alternating between structure refinement and manual protein model rebuilding (Figure 8.1, Section 8.6). In macromolecular refinement one searches for the conformation of the macromolecule that fits the observed diffraction data best, while maintaining reasonable

stereo-chemistry. Therefore, in general, the properties that are modified during the process of refinement are the three-dimensional coordinates of the atomic positions and the temperature-factor (known as the anisotropic U or the isotropic B-factor) that characterizes the displacement from the atom's mean position. The latter value is often used as an indicator of the local disorder.

Refinement is a particularly important process in structure solution of proteins, because the number of observations is smaller than the number of variables that can be refined (underdetermined system), unless very high-resolution data are available. This does not allow refinement of the atomic positions, independently from each other, and in turn requires the input of additional, often chemical, knowledge to simplify the approach. Addition of this information, such as restriction on bond angles, bond length and peptide geometry, can then be used to restrain or even constrain certain variables. Consequently, this increases the observations-to-variables ratio, making refinement of macromolecules feasible. In addition to modelling the atomic positions of the atoms, the diffraction will not be modelled accurately without accounting for the bulk solvent, because of the large solvent volume. Bulk solvent, as represented by low-resolution reflections, is generally modelled as a continuum of electron density with a high atomic displacement parameter (see Rossmann *et al.*, 2001, p.372 and 400).

Aspects of reciprocal space refinement and several program-approaches are described by Rossmann *et al.* (2001)(p.369-414), and have been reviewed by Brunger *et al.* (1998). As for any optimization problem, the aim in refinement is to find a global minimum of a defined target function. Both reciprocal space refinement-programs, used in this study, apply a mathematical approach known as cross-validated maximum-likelihood refinement (cross-validation will be discussed below), which has been described by Murshudov *et al.* (1997) and Pannu and Read (1996). The basis for the maximum-likelihood approach is that the best model will be

the one most consistent with the data. The consistency is measured statistically, by the probability that the observations would be made, given the current model, which is known as likelihood. The probability of the model given the data is then proportional to the likelihood multiplied by the probability of the model, known without the data (Bayes' theorem). The probability of the model, known without the data, is reflected by the prior knowledge, such as the stereochemical information about bond lengths and bond angles. These parameters on bond angles and length are in general based on the ones described by Engh and Huber (1991), which were derived from small molecule crystallography. Changing the model parameters so that the values of the observations become more probable, increases the likelihood, which indicates a better model. The program attempts to find the best model parameters by effectively maximizing the likelihood (in fact it is minimizing the negative logarithmic likelihood).

Different mathematical optimization methods exist to optimize the maximum-likelihood target function. The programs used for the classical energy minimization in this study apply a combination of the steepest descent and conjugate-gradient approach (first derivative of the target function) to find a minimum (Murshudov *et al.*, 1999). This allows prediction of which directional change will reduce the value of the target function.

An alternative approach to this method is optimization by simulated annealing, as described by Adams *et al.* (1997) and Brunger in Rossmann *et al.* (2001). Annealing is a physical process of heating up a solid phase until its particles randomly arrange themselves in a viscous liquid phase. This phase is then cooled slowly so that particles arrange themselves in the lowest energy state. By setting the target energy to be equivalent to the potential energy of the system, the process of annealing can be simulated. The objective function is then essentially sampled at a new point in parameter space. If the value of the objective function at the new point is less than that at the current point, the new point becomes the current point. If the value is greater at the new point, the Boltzmann probability

of the difference in the energy function is compared to a random number. If it is less than the random number the new point is accepted as the current point. This process continues until a sufficiently deep minimum is found that the sampling process never leaves that region of parameter space. At this point the temperature for the Boltzmann term is reduced, which lowers the probability that the current point will move out of the region. This produces a finer search of the local region. The cooling process is continued until the solution has been restricted to a sufficiently small region.

In contrast to the optimization methods mentioned above, simulated annealing allows motion against the gradient. The temperature of the system can be altered, where the temperature has not its normal physical meaning, but determines the probability of overcoming barriers or maxima of the target function. The higher the temperature, the more likely it is to cross these barriers. The optimal temperature, in general, has to be found by trial and error. The simulated-annealing procedure requires a Boltzmann distribution to be calculated at a given temperature and an annealing schedule, which consists of a sequence of temperatures at which the Boltzmann distribution is calculated. Simulated annealing procedures can differ in the way a transition or move is generated from one set of parameters to another. In the program used for this study, molecular dynamics simulation, which limits the search to physicochemically reasonable moves is applied. Molecular dynamics, essentially, calculates a solution for Newton's second law. The initial velocities for this calculation are chosen randomly and each different starting velocity will result in slightly different models. The best model, among the different ones generated by this multi-start approach, can then be chosen by the experimenter. Simulated annealing requires that the temperature is controlled during the molecular dynamics simulation, which is achieved by coupling the equations to a heat bath through a friction term, known as temperature coupling (see Rossmann *et al.*, 2001, p.378).

For the annealing schedule, in general, a linear slow-cooling protocol is chosen which ensures that once a global minimum for a model is found it will not be moved out of it. Molecular dynamics will result in small changes of the model, some of which lower the energy term of the target, which increases the kinetic energy (i.e. temperature) of the system. The temperature-coupling removes the excess of energy, and makes the reverse motion and escape from a global minimum more unlikely. The program also allows constraints on bond lengths and bond angles to be used, such that, essentially, only the torsion angles are refined as variable parameters (Rice and Brunger, 1994), in contrast to cartesian molecular dynamics (flexible bond lengths and angles).

Simulated annealing as an optimization technique is particularly well-suited to overcome the multiple minima problem, because it can cross barriers between minima. Therefore it can explore a greater volume of the parameter space to find better models and has a large radius of convergence compared with gradient minimization techniques.

The process of refinement, or the end result of each refinement cycle run, has to be monitored, in order to find out whether it improves the protein model. The main criterion used to measure the agreement is the crystallographic R-factor. This residual measures the discrepancy between the calculated and the observed data (Equation 8.3).

$$R_{\text{cryst}} = \frac{\sum_h ||F_{\text{o}}| - |F_{\text{c}}||}{\sum_h |F_{\text{o}}|} \tag{8.3}$$

where $|F_{\text{o}}|$ refers to the observed structure factor amplitudes of index h and $|F_{\text{c}}|$ to the calculated structure factor amplitudes of index h.

However, it was shown that the R-factor can be made very low by increasing the number of parameters used to describe the model and that incorrect models can be refined to acceptable R-factors (Branden and Jones, 1990). In order to overcome the model bias of the R-factor, Brunger (1992)

introduced the use of a set of randomly selected reflections (generally 5-10% of all reflections) that are put aside and are not used during the refinement. When a separate R-factor is calculated for this set of data, it can provide an objective cross-validation of the R-factor. This R-factor is known as R_{free}.

Because every reflection contains information about the entire structure, changes made to the model that do not improve the fit of the model to the X-ray diffraction data, should not improve the fit of the test-set data to the model. Consequently, the R_{free} should remain constant or increase. In contrast to the conventional R_{cryst}, the R_{free} is highly correlated with the accuracy of the model (Brunger, 1993; Kleywegt and Brunger, 1996; Kleywegt and Jones, 1995). Thus, monitoring of the R_{free} can avoid over-interpretation of the data. If the model is correct the test set reflections should have an R-factor close to that of the reflections used in the refinement. The process of refinement is then normally considered to have converged when the changes in R_{free} are no longer significant.

In this study, the programs CNS V1.1 (Brunger *et al.*, 1998) and REFMAC V5.1.24 (Murshudov *et al.*, 1997) were used to refine the protein models. CNS V1.1 was used for the initial refinement, which consisted of cycles of rigid-body refinement, to improve the overall position of the model, followed by molecular dynamics simulated-annealing. This was particularly important to reduce the model bias, which is often significant in models derived by molecular replacement. The settings for the simulated-annealing procedure involved torsion-angle molecular dynamics and a slow-cooling protocol. In general, 5-10 starting velocities were chosen with a starting temperature of 1727°C. Other than that, the default parameters, from the provided script-template were used. Subsequently, REFMAC was used for the conventional minimization refinement. The diagonal weighting term, representing the weight put on the restraints, was put lower for the earlier stages of refinement to restrain the geometry. While progressing in the refinement cycles this was then increased to relax the re-

straints. A 5% test set of randomly-chosen reflections, using the program UNIQUE (Collaborative Computational Project Number 4, 1994) was used for calculation of the R_{free} throughout the whole refinement process.

At high resolution it is sometimes possible to model the anisotropic displacement of the atoms either in a full anisotropic refinement or by a single tensor that describes translation, libration, and screw (TLS) for the correlated motion of atoms in rigid groups. The latter may be an extremely accurate way to model the behaviour of the molecules (Rossmann *et al.*, 2001, p.372), possibly even at medium resolution. TLS refinement has been described for small molecule crystallography (Schomaker and Trueblood, 1998) and for macromolecular refinement (Howlin *et al.*, 1989). TLS refinement has more recently also been implemented into REFMAC. TLS refinement (Winn *et al.*, 2001) was applied in the last refinement cycles to improve the data fit. The TLS groups were defined according to secondary structural motifs, such as α-helices, β-strands, and loops. The number of TLS groups was then varied, while monitoring the R_{free} and R_{cryst}. The TLS parameters were evaluated with the program TLSANL (Howlin *et al.*, 1993). For the TLS refinement, the initial B-factor was fixed to the mean B-factor as derived from the Wilson plot.

In addition, electron density modifications, such as solvent flattening and non-crystallographic symmetry averaging with the program DM (Cowtan and Zhang, 1999) were occasionally used in this study. Electron density modification essentially adds real-space restraints or constraints that potentially can improve the phase information and consequently the electron density maps (see Carter and Sweet, 1997b, p.55). Solvent flattening exploits the fact that the solvent region is flat at medium resolution owing to the disorder of the solvent. Non-crystallographic symmetry averaging (see Carter and Sweet, 1997b, p.18-53) essentially increases the signal-to-noise ratio by averaging equivalent parts in molecules, related by non-crystallographic symmetry.

In order to add water molecules to the model, the automatic water-

adding procedure from ARP/WARP (Perrakis *et al.*, 1999) was used occasionally. The overall refinement was continued until convergence of the R_{free} and until the model was deemed correct.

8.6 Protein structure model building

Alternating with refinement cycles, rebuilding of the model has to be carried out. The electron density maps are derived by calculating the Fourier transform of the structure factor amplitudes. During each refinement cycle, adjustments in the position of the atoms in the protein model are made, so that they fit the electron density map better. Furthermore, water molecules and other hetero-groups such as ions and ligands are added during the rebuilding process. Errors in models can generate peaks in the Fourier maps that can be interpreted as solvent peaks. Therefore, adding waters too early in the process may lead to model errors and should be carried out at an intermediary stage of rebuilding and refinement. As a rule of thumb one water molecule can be placed per amino acid residue in a structural model with structure factors to 2 Å resolution.

The adjustments are facilitated by graphic programs, that allow movement of the atoms on the screen. Changes of the atomic coordinates will result in a new model, the phases of which, after refinement, will be combined with the observed structure factor amplitudes to calculate a new electron density map, which ideally will be improved and clearer to interpret.

Practical advice for good rebuilding as well as refinement procedures has been described by Kleywegt in Carter and Sweet (1997b). In order to rebuild and interpret the models in this study, σ_A-weighted (Read, 1986) difference electron density maps of the $2mF_o$-DF_c and mF_o-DF_c type were used. F_o-F_c electron density maps have ”positive” and ”negative” peaks,

for parts of the protein model that have not been adequately represented and parts that are not supported by the observed data, respectively. Manual rebuilding cycles were performed in the programs O (Jones *et al.*, 1991) and XTALVIEW (McRee, 1999). The parameters for bond angles and lengths for ligand molecules and inorganic small molecule compounds, which were modelled into the electron density maps, were retrieved from the Uppsala HIC-UP server[2] (Kleywegt and Jones, 1998).

8.7 Assessment of protein structure model quality

At the end of the refinement and rebuilding procedure, as well as at intermediate model refinement and rebuilding stages, the protein structure model should be examined for (in)correctness. In general, the accuracy of a protein structure model and the level of detail it contains is a function of the quality and quantity of the experimental data on which it is based, and the care and protocols used during data processing and refinement. The objective of validation of the data and the model is to give a realistic estimate of the true accuracy of the model as well as an indicator of the precision with which the model explains the data.

Owing to the limited observation-to-parameter ratio and poor phase information that are often encountered in macromolecular crystallography, constructing and refining the protein structure model is not an exact scientific process with a unique solution. Therefore, it is particularly important to assess the confidence level the experimenter, as well as the users of the model coordinates, should assign to the overall and individual parameters of the final model. Moreover, the evaluation process can help detecting problematic regions in a model structure, or identifying unusual atoms

[2] http://xray.bmc.uu.se/hicup/

or residues, which may have a functional role. Protein structure model assessment and its central role in crystallography has been reviewed by Dodson *et al.* (1998), Dodson (1998), and Kleywegt (2000).

The major global quality indicator for validating protein crystal structures is the R_{free}. In general, better and higher resolution X-ray data will lead to a lower R_{free} and better R_{free}/R_{cryst} ratio (Kleywegt and Jones, 2002; Tickle *et al.*, 2000). Furthermore, it can be used to estimate the coordinate error (Kleywegt and Brunger, 1996), a second important indicator. Coordinate-error estimates from cross-validated σ_A (Read, 1986) give a reliable impression of the positional error of the model.

Global stereochemical statistics such as overall bond angles, bond length and dihedral angles variability are often used as global statistical indicators. The r.m.s.d. from the parameters, retrieved from small molecule crystal structures by Engh and Huber (1991) are monitored. R.m.s.ds. for specific residues can then also be used as a local indicators of model quality. A particularly powerful structure validation tool is the distribution of the ϕ/ψ torsion-angles in the Ramachandran plot (Ramakrishnan and Ramachandran, 1965). In contrast to the bond lengths and bond angles which are essentially constrained during refinement by the same values with which they are compared at the evaluation stage, the ϕ/ψ are in general not constrained during the refinement.

The Ramachandran plot is one of the most useful stereochemical statistics. It essentially provides a map of angular dimensions, in which the ϕ/ψ angles of all the residues of the protein model are placed. The side-chains of the amino-acids clash at certain angles and such conformations of amino-acids are energetically unfavourable. In general, this allows only a range of combinations of ϕ/ψ. A global indicator of a model is the percentage of the residues in the allowed, generously allowed or forbidden region of the Ramachandran plot. Most often, a protein model with a good quality will only have a small number of outliers, which are residues

in the forbidden part of the Ramachandran plot. The outliers in the Ramachandran plot can then be examined by the experimenter and as such the Ramachandran plot is also an indicator for the local quality.

The average B-factor of a protein structure model can also be a good indicator for its quality, but care should be taken when TLS refinement was used. In this case the individual B-factors may only represent a fraction of the real B-factor, unless TLSANL was applied to account for the contribution of the TLS group.

Local analysis tools for the protein structure model include H-bonding analysis, such as for the correct orientation of His, Asn, and Gln side chains, which can often not be inferred from electron density alone. Analysis of the H-bonding networks (Hooft *et al.*, 1996b) can help here. Furthermore, evaluation of the side-chain planarity, chiral volumes, that can identify potential clashes, abnormally short interatomic distances, individual B-factors, and unusual rotamers. In general, the values for these parameters are compared to statistics derived from previously solved macromolecular and/or small molecule crystal structures.

All these assessment procedures can help to detect outliers, residues with properties, that are unusual compared with the most commonly known values. The difference of a true outlier, supported by the data, or an error, has to determined by inspection of the local electron density map by the experimenter. Outliers, such as residues with unusual mainchain ϕ/ψ torsion-angle combinations, that do not have unequivocally clear electron density are often errors.

After this validation has been carried out, it is good practice to submit the three-dimensional coordinates together with the observed structure factor amplitudes to the Protein Data Bank[3] (Berman *et al.*, 2000), an online archive for experimentally determined protein structure models.

[3] http://www.rcsb.org/pdb/

From this open database coordinates and observed structure factors can be easily retrieved in a standard form by everyone.

Validation was used throughout the model building and refinement process. The quality of the protein models was evaluated with RAMPAGE[4] (Lovell *et al.*, 2003). The program calculates ϕ/ψ angles and evaluates them with respect to their position in the Ramachandran plot. Furthermore, validation was pursued with WHATCHECK (Hooft *et al.*, 1996b), which also assesses the correct nomenclature of the atom assignment, and PROCHECK (Laskowski *et al.*, 1993). All models, together with the observed structure factors, were deposited in the PDB through the online AUTODEP tool [5].

[4] http://www-cryst.bioc.cam.ac.uk/rampage
[5] http://deposit.pdb.org/adit/

9 Bioinformatical methods

A range of bioinformatical methods was used in this study to analyze the protein structure models and examine similarities between them.

In order carry out comparative analysis of protein structure models, superposition of the various models is often useful. Structural superpositions of atomic coordinates and calculations of r.m.s.ds. between similar protein models (such as different ADC mutants) were carried out with LSQMAN (Kleywegt and Jones, 1994) and LSQKAB (Kabsch, 1976). Both programs calculate a translation and rotation matrix for the superposition, which can then be applied to one of the molecules to bring it into an equivalent global position to the other. To apply the operator, XPAND (see Rossmann *et al.*, 2001, p.353) was used. In order to calculate geometric residue parameters, such as torsion angles, DANG from the KINEMAGE package by Prof. Jane Richardson as described in Rossmann *et al.* (see 2001, p.729-731) was used.

The objective of amino acid sequence-to-sequence, amino acid sequence-to-structure, and structure-to-structure searches in this study was to identify homologues, which in turn could help to establish functional and structural relationships. In order to generate amino acid sequence alignments BLAST and PSI-BLAST, basic local alignment tools (Altschul *et al.*, 1990), were used to automatically retrieve sequences.

These tools essentially calculate a matrix of similarity scores, in which identities have positive values and mismatches have negative values, for all possible pairs of residues. The similarity score for a contiguous segment is then the sum of the similarity scores for each pair of aligned residues. A maximal segment pair is defined in the algorithm, which is the highest scoring pair of identical length segments chosen from two sequences. A segment pair is locally maximal if its score cannot be improved by extending or shortening both segments. As a measure of local similarity, BLAST calculates the score of each maximal segment pair and then seeks all locally maximal segment pairs with scores above some cutoff. The program BLAST first identifies similar segments between the query sequence and a database sequence, then evaluates the statistical significance of any matches that were found, and finally reports only those matches that satisfy a user-selectable threshold of significance. Amino acid sequences were retrieved from the Swiss-Prot protein sequence database[1] (Bairoch and Boeckmann, 1991).

Subsequently, the retrieved protein sequences were aligned with CLUSTALW (Thompson *et al.*, 1994) to generate multiple-sequence alignment, and alignments were displayed in the CLUSTALX interface (Thompson *et al.*, 1997). CLUSTALW assigns individual weights to each sequence, and amino acid substitution matrices are varied at different alignment stages according to the divergence of the sequences to be aligned. Residue-specific gap penalties encourage new gaps in potential loop regions rather than regular secondary structure.

In order to find homologues, a more sensitive method to sequence-to-sequence alignments constitute sequence-structure alignments. Amino-acid sequence-to-structure searches and alignments were carried out by using the fold recognition server 3D-PSSM (Kelley *et al.*, 2000) and the homology recognition server FUGUE (Shi *et al.*, 2001). Similar to CLUSTALW, FUGUE initially generates a multiple-sequence alignment with the query

[1] http://ca.expasy.org/sprot/

sequence. This sequence profile is then used to search structural profiles. The structural profiles are derived from homologous structural alignments in the HOMSTRAD database (Mizuguchi *et al.*, 1998b), using environment-specific substitution tables and structure-dependent gap penalties.

When structural information is available for the search model, recognition of distant homology between two proteins, is more probable than with sequence-to-structure searches alone. The DALI software (Holm and Sander, 1993), which carries out automatic structure-to-structure comparisons to all protein in the PDB. The coordinates of each protein are used to calculate residue-residue Cα distance matrices. Amino acid sequence alignments taking into account structural information (structural alignments) were generated with COMPARER (Sali and Blundell, 1990), a structure-based alignment program. For visualization, the structural alignments were formatted in JOY (Mizuguchi *et al.*, 1998a). Essentially the program annotates the one-dimensional information contained in sequence alignments with a code for several aspects of three-dimensional information, such as H-bonding, solvent accessibility, and secondary structure of residues. Structural superpositions, as generated by COMPARER, for the visual inspection of the pairwise superposed three-dimensional coordinates were produced with MNYFIT (Sutcliffe *et al.*, 1987).

Catalytically-important residues were identified from their conservation in multiple sequence alignments and their relevance judged from visual inspection of the structures. Once homologues have been identified, their classification in the SCOP database (Murzin *et al.*, 1995), a database that classifies homologous protein structures into families and superfamilies, was investigated. This was done to potentially determine both more distant relationships with proteins classified in the same (super)family and evolutionary aspects.

The DSSP algorithm (Kabsch and Sander, 1983), which assigns sec-

ondary structure based on pattern recognition of H-bonds and geometrical features, was used to assign secondary structures to produce figures in PYMOL (DeLano, 2002). MOLSCRIPT (Kraulis, 1991), and RASTER3D (Merritt and Bacon, 1997) for rendering, were also used for illustration of the protein structure models. Schematic topology diagrams of the protein secondary structure were drawn following the representation developed by Prof. Jane Richardson (see Wyckoff *et al.*, 1985, p.351).

Part III

Results and discussion

10 Crystallization of KPHMT with cofactor analogues

10.1 Introduction

In order to understand the catalytic mechanism of KPHMT, a structure model of the protein in complex with a cofactor analogue would certainly be helpful. Identifying the binding residues for the cofactor, will also be important for the development of THF-based enzyme inhibitors.

Several crystal structures of enzymes in complex with THF and/or THF analogues have been solved. Among them are dihydrofolate reductase in complex with folic acid (Sawaya and Kraut, 1997), dimethylglycine oxidase in complex with 5-fo-THF (Leys *et al.*, 2003), nitric oxide synthase in complex with THF (Pant *et al.*, 2002), ribonucleotide transformylase with 5-deazafolic acid (Almassy *et al.*, 1992), serine hydroxymethyltransferase with 5-fo-THF (Trivedi *et al.*, 2002), formiminotransferase-cyclodeaminase with 5-fo-THF (Kohls *et al.*, 2000), and thymidylate synthase in complex with 5-hydroxymethylene-6-hydrofolic acid (Perry *et al.*, 1993). These crystal structures proved that such THF cofactor analogues are, in principle, amenable to cocrystallization with enzymes.

Structural models for binding of 5,10-meTHF to KPHMT based on the structure of THF-derivatives in serine hydroxymethyltransferase have been constructed (Louise Birch, personal communication). However, the reliability of such computationally-derived models is often limited and can hardly substitute experimental results. Therefore, an experimental model of the binding was clearly desirable.

10.2 Overexpression and purification of KPHMT

The purification protocol used in this study was an adaptation of the protocol described by von Delft (2000), which in turn was based on the purification protocol developed by Jones *et al.* (1993) and modified by Drs Michael Witty and Adrian Saldanha.

KPHMT could be overexpressed in large quantities, following the protocol described in Section 7.2, as judged by the final amount of purified protein of up to 80 mg of pure protein expressed per litre of *E. coli* culture. Despite the high level of overexpression, the protein was predominantly found in the soluble fraction of the crude lysate of the *E. coli* cells.

The $(NH_4)_2SO_4$ precipitation (see Subsection 7.3.3) was carried out to potentially separate the protein from polynucleotides (DNA and RNA) which could have affected the efficiency of the subsequent anion-exchange steps. However, given the amount of overexpressed protein, this step presumably could be omitted and one could possibly start the purification with the first anion-exchange step. Furthermore, the first $(NH_4)_2SO_4$ precipitation step did apparently hardly result in significant precipitation and may consequently be a redundant separation step.

In order to minimize the risk of proteolytic cleavage of KPHMT during the purification procedure, the buffers for the two anion-exchange steps

contained protease-inhibitors. For the same reason, the purification steps for KPHMT in this study were normally carried out at 4°C. However, depending on the availability of the purification facilities, the two anion-exchange purification steps, described in Subsection 7.3.3, were also carried out at 19°C, without apparent effects on the quality of the expressed protein sample. The previous purification procedure had involved a heat step of 80°C for 5 min (Jones *et al.*, 1993), indicating that the protein is relatively heat-resistant and stable. Therefore, the anion-exchange steps can also be carried out at 19°C or room temperature.

In terms of separation, the first anion-exchange step proved quite efficient (see Figure 10.1). The second anion-exchange step was repeated identically to the first one but with a higher resolution column. The gel-filtration step was necessary to remove further contaminants and to ensure homogeneity and monodispersity of the sample. On one occasion, the last two purification steps were exchanged in their time-order, but it did not appear to have a major effect on the purity of the protein sample. The elution profile of the gel-filtration step indicated the possibility of two peaks that were not well separated (Figure 10.1). Therefore, fractions 32, 33, 34, and 35 and fractions 36, 37, and 38, respectively were combined. However, the two fractions did not show differences on 12% SDS-PAGE, and the purified protein migrated as a single band (Figure 10.1), indicating a pure sample. Mass-spectrometry analysis confirmed the high purity of the protein sample, and N-terminal sequencing confirmed the correctness of the first 10 N-terminal residues (data not shown).

The protein has before been observed to be stable on storage at 4°C for several months without apparently influencing crystallizability (Dr Adrian Saldanha, personal communication). Therefore, in this study the protein was also stored at 4°C, after being concentrated to ∼40-45 mg/ml.

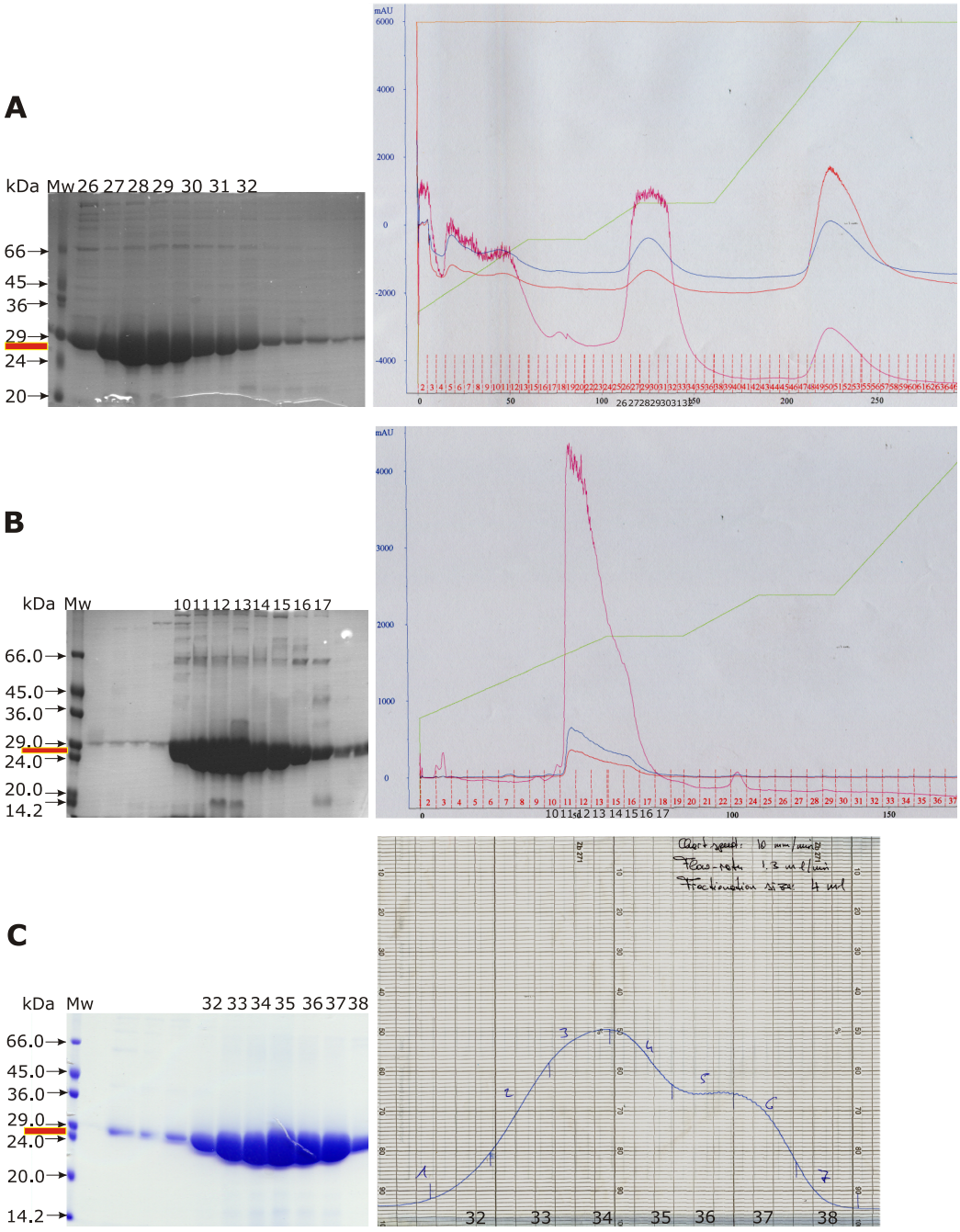

Figure 10.1 (**A**) 12% SDS-PAGE of the fractions containing KPHMT from the first anion-exchange step. Molecular weight markers are shown on the left (Mw) and the molecular weight is indicated in kDa. The relative position of KPHMT is indicated with a red bar. On the right; the corresponding purification graph with the x-axis showing the elution volume in ml, whereas the y-axis shows the optical absorbance for 280 nm, blue. The absorbance is not to realistic scale, because it was calibrated with the flow-through, which, for the first step, is very high. Also monitored is the optical absorbance at 230 nm: red and at 260 nm: maroon; the KCL concentration is shown in green. Fractions that were pooled together for the next step are indicated as black numbers and shown on the SDS-PAGE. (**B**) 12% SDS-PAGE of the fractions containg KPHMT and graph of the second anionexchange step. Fractions that were pooled together are shown as black numbers. (**C**) 12% SDS-PAGE and graph of the gelfiltration step. Fractions 32, 33, 34, and 35, and, separately, fractions 36, 37, and 38 were combined respectively. The optical absorbance at 280 nm was monitored

10.3 Crystallization conditions for KPHMT with cofactor analogues

10.3.1 Background

From initial crystallization trials on KPHMT, carried out by Drs Venugopal Dhanaraj and Galib Khan, three conditions could be identified, which produced crystals. However, only one of them resulted in stable crystals with reasonably strong X-ray diffraction patterns, and this condition was further optimized by Dr Frank von Delft (von Delft, 2000). The overall crystallization protocol was rather complex, with difficulties in reproducing crystals with a single precipitant condition and temperature.

The final crystallization conditions Dr Frank von Delft developed, included PEG8000, NaCl, $NaCH_3CO_2$ with sodium citrate, pH 6.8 as a buffer (von Delft, 2000). A series of six drops, varying in the concentration of PEG8000 (9-11% (w/w)), and NaCl (in reverse order than PEG8000; 200-100 mM) was set up within a few minutes. Crystals were grown at 4°C with an intermediate temperature equilibration step at 19°C in a polystyrol box and further crystal growth at 4°C. This was considered to be necessary in order to obtain a sufficient nucleation rate, while still gaining a reasonable size of the crystals.

The crystallization conditions resulted in two crystals forms of apparently similar morphology, clusters of rhombic crystals with dimensions of 100×100×100 μm^3 or larger, plate-like rhombic crystals with dimensions of 500×300×100 μm^3 (von Delft, 2000). X-ray diffraction patterns collected from these crystals showed that both crystals were monoclinic in space group $P2_1$, with unit cell dimensions a=86 Å, b=156 Å, c=208 Å, β=98° (subsequently referred to as "large unit cell" crystal form). These crystals would diffract to a minimum Bragg spacing of 2.8 Å at a syn-

chrotron X-ray source. However, on one X-ray diffraction data collection visit to the APS synchrotron in Chicago, unusually regular tear drop-shaped crystals were found in a crystallization drop, prepared five month before crystal mounting. X-ray diffraction data of one of these crystals, could be collected up to a resolution of 1.8 Å. Subsequent data indexing showed that the crystals were in the same space group, but in contrast to the large unit cell had unit cell dimensions of a=86, b=157, c=100 Å and β=98° (subsequently referred to as "small unit cell" crystal form). These small unit cell crystals, however, could not be reproduced by von Delft (2000). Whereas the large unit cell contained two KPHMT decamers, the small unit cells contained only one decamer. Both crystal forms apparently grew in the presence of α-kiva and ketopantoate, although no KPHMT crystal structure with α-kiva was eventually solved.

After Dr Frank von Delft had left, Dr Tsuyoshi Inoue took over the project temporarily. He encountered difficulties in reproducing the previous crystallization conditions and consequently started to screen for new crystallization conditions. When I took over the project I started to explore conditions where Dr Tsuyoshi Inoue had left the project. The protein used for these initial crystallization trials was purified by Dr Adrian Saldanha.

10.3.2 Crystallization trials

Initial screening for crystallization conditions was carried out with PEG4000 because with lower molecular weight PEGs the response time for crystallization is often shorter than with higher molecular weight PEGs. In order to gain a faster nucleation rate, the temperature used for these trials was 19°C. A solution of 100 mM NaCl, 100 mM $NaCH_3CO_2$, 10 mM $Mg(CH_3CO_2)_2$, 50 mM Bis-tris buffer pH 6.0 with 6-13% (w/v) PEG4000 was initially used as precipitant solution. The final protein concentration in the drop was varied between 23 and 38 mg/ml for different crystallization trials.

This condition was also pursued for cocrystallization with a cofactor-analogue. A saturated solution (10 mM) of (6*R*,*S*)-5-foTHF calcium salt was prepared freshly immediately before setting up the crystallization drops. This solution was then diluted with the protein solution to a final (6*R*,*S*)-5-foTHF concentration of ∼4-8 mM, which was well above the protein concentration of ∼1-1.34 mM. The protein—cofactor analogue solution was incubated on ice for 30 min, to allow diffusion of the cofactor analogue into the protein and for binding, prior to setting up crystallization trials. This early screening surprisingly yielded rather big crystals of up to 900 µm (Figure 10.2 A) in size in one dimension, with a precipitant composition of 100 mM NaCl, 100 mM $NaCH_3CO_2$, 10 mM $Mg(CH_3CO_2)_2$, 50 mM Bis-tris buffer pH 6.0, 7.5% (w/v) PEG4000. However, these crystals were very unstable in the X-ray beam and proved difficult to cryoprotect with either ethyleneglycol or glycerol. The X-ray diffraction pattern collected on the in-house X-ray data collection facility was very weak (data not shown) and the crystals showed strong decay within a few minutes. These crystals could not be reproduced, and the unusual growth may have resulted from temperature instability due to a malfunctioning temperature control at that time in the crystallization room.

Because of the observed instability of the crystals, subsequent crystallization trials were carried out at 4°C in temperature-controlled incubators. The initial protein concentration was maintained between 23 and 30 mg/ml, but the pH and buffer component of the precipitant solutions was varied. The tested buffers and the corresponding pH-values for the precipitant solutions were 50 mM sodium citrate pH 6.0, 50 mM Bis-tris pH 6.5, 50 mM imidazole pH 7.0, 100 mM Hepes pH 7.4, and 50 mM Tris/HCl pH 7.5 and 8.0, while keeping the other components that were mentioned above relatively constant. Crystallization screening only around the neutral pH range was pursued because all cofactor analogues are more stable at neutral pH. Crystals would usually grow relatively quickly within 2-8 days.

All these trials were carried out with product ketopantoate and, in separate drops, with α-kiva as well. The protein was mixed with a ketopantoate or α-kiva solution to obtain a final concentration of 30 mM of the product or substrate, and the solutions were incubated on ice for 30 min, prior to the crystallization trials. Apart from the last two conditions with Tris/HCl pH 7.5 and 8.0, all conditions with the different buffers produced crystals. Nucleation in the drop started quickly, as monitored under the microscope and, therefore, care was taken that the trays were moved to $4^{\circ}\mathrm{C}$ within 5 min after the first drop of protein was mixed with precipitant solution. This meant setting up only six wells out of a possible twenty four wells from the Linbro plates.

The precipitant conditions with 50 mM citric acid pH 6.0 and 100 mM Hepes pH 7.4 were found to be the most promising. A large amount of small (maximal dimension 100 µm) rhombic crystals could be grown, similar to the smaller of the two morphologically different crystal forms Dr Frank von Delft had obtained. However, these conditions yielded many small and often cross-linked crystals and thus required optimization to enable screening for X-ray diffraction. First, a range of additives, such as dioxane, ethanol, 2-methyl-2,4-pentanediol; and PEG400 in concentrations up to 5% (v/v) added to the precipitant solution, were tested for their effect on the crystallization. However, none of these additives improved the crystals. Almost all had a rather contrary effect, resulting in smaller crystals or none at all.

For this reason, attempts to improve the crystal quality and size, focused on refining the condition by changing protein and the PEG4000 concentration. A concentration of 6-8% (w/v) of PEG4000 with a final protein concentration of 28-30 mg/ml was found to be optimal for crystallising KPHMT in the presence of ketopantoate. Crystals with 50 mM citric acid as a buffer could be grown to a size of $\sim 100 \times 100 \times 50$ µm^3. The crystals were cryoprotected with PEG400 in the following way: mother liquor solutions containing PEG400 at concentrations of 5, 10, 15 and 20% (v/v) were prepared; 0.4 µl of the 5% solution was then added to the drop with

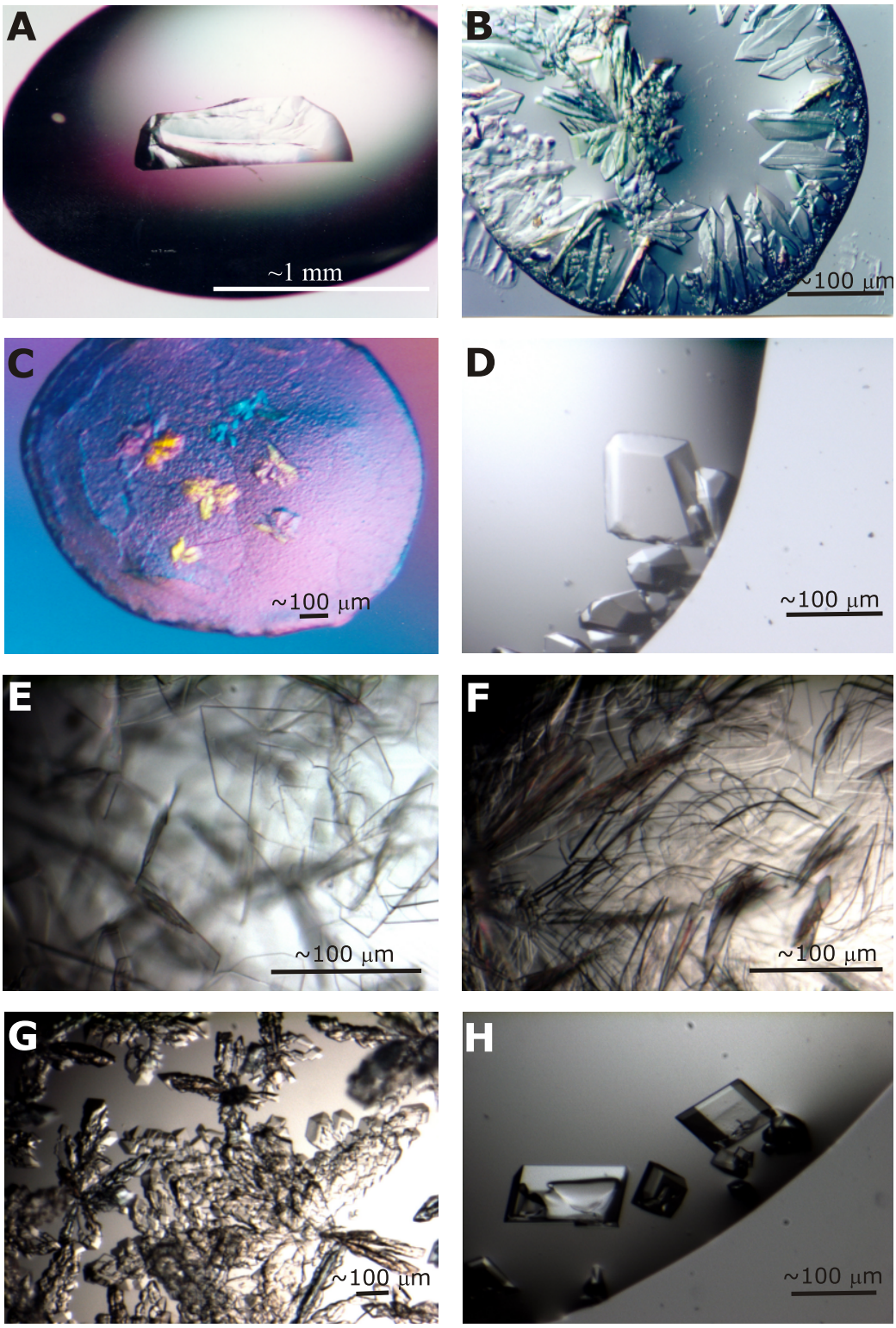

Figure 10.2 Morphology of the crystal forms obtained for KPHMT. (**A**) Crystals of KPHMT incubated with (6*R*,*S*)-5-foTHF. (**B**) Small unit cell crystals of KPHMT incubated with ketopantoate grown with the crystallization condition of Table 10.2. (**C**) Crystals of KPHMT incubated with α-kiva and (6*R*,*S*)-5-meTHF. (**D**) Crystals of KPHMT incubated with α-kiva (**E**) KPHMT crystals grown in the presence of α-kiva and (6*S*)-5-foTHF, with Bis-tris buffer pH 6.0. (**F**) KPHMT crystals grown in the presence of α-kiva, (6*S*)-5-meTHF, with Bis-tris buffer, pH 6.0. (**G**) KPHMT crystals grown in the presence of α-kiva, (6S)-5-meTHF with Bis-tris buffer, pH 6.5. (**H**) KPHMT with α-kiva, (6*S*)-5-meTHF, Bis-tris, pH 6.5.

the crystals on a coverslip and left for 5 min. The inverted coverslip was covered with a well from a crystallization plate to reduce evaporation. Another 0.5-1 µl of 5% solution was added to the drop and left for another 5 min. Then the crystal was successively transferred in 5 min intervals, first into the 5% solution and then stepwise to the final 20% solution. The crystals were then directly mounted at the in-house X-ray machine in a flow of N_2 and screened for X-ray diffraction.

Table 10.1 Crystallization condition A for KPHMT that was incubated with ketopantoate obtained with sodium citrate as a buffer

Component	PEG4000	NaCl	$NaCH_3CO_2$	$Mg(CH_3CO_2)_2$	Sodium citrate pH 6.0[a]	Protein
Concentration	6% (w/v)	100 mM	100 mM	10 mM	50 mM	24-30 mg/ml

[a] The pH was adjusted with NaOH

X-ray diffraction reflections were extended out to 2.8 Å resolution (see Figure 10.3 A). Indexing showed that the crystal was of monoclinic space-group with unit cell dimensions of 101×164×172 $Å^3$ and β=96.4°. Although the c-axis was 30 Å shorter than that of the large unit cell previously obtained by Dr Frank von Delft, there was no indication that these crystals could be crystals with small unit cell type. However, it was the crystals with the small unit cell that were expected to give higher resolution.

While further optimizing the crystallization conditions, regularly-shaped crystals of KPHMT (Figure 10.2 B) incubated with ketopantoate were also obtained with Hepes (Table 10.1). These crystals were cryoprotected similar to the crystals obtained with sodium citrate. While initial diffraction indicated a high mosaicity (Figure 10.3 B), slower increase of the cryoprotection solutions resulted in X-ray diffraction patterns extending to a maximum resolution of 4.0 Å using the in-house X-ray data collection facility. However, strong lower-resolution reflections could be observed (Figure 10.3 B,C,D). Crystals were monoclinic with unit cell dimensions of 86×159×109 $Å^3$ and β=159°.

Table 10.2 Crystallization condition B for KPHMT that was incubated with ketopantoate yielding crystals with small unit cell dimensions

Component	PEG4000	NaCl	$NaCH_3CO_2$	$Mg(CH_3CO_2)_2$	Hepes pH 7.4[a]	Protein
Concentration	7% (w/v)	100 mM	100 mM	10 mM	100 mM	24-30 mg/ml

[a] The pH was adjusted with NaOH

Importantly, these unit cell dimensions were close to the dimensions of the crystals with small unit cell dimensions from Dr Frank von Delft. This was somewhat promising in that a condition (Table 10.2) was found that appeared to produce KPHMT crystals in the same space group as before with small unit cell dimensions. However, further improvement of the crystallization conditions was necessary to increase the maximum resolution (Figure 10.3).

Once reliable crystallization conditions for the small unit cell crystals had been found, they were then pursued for the potential to produce crystals with cofactor analogues. Attempts were then made to soak these crystals with (6*R*,*S*)-5-foTHF. The cofactor analogue was either added in solid form with a spatula or in mother liquor solution. However, in both cases the KPHMT crystals dissolved, and consequently soaking this cofactor analogue into the crystals was not successful. Therefore the crystallization conditions, using either Hepes pH 7.4 or sodium citrate pH 6.0 as buffers, and α-kiva, were then pursued further including KPHMT incubated with (6*R*,*S*)-5-foTHF in 4-8 molar excess. No crystals could be obtained with sodium citrate. In contrast, using the crystallization conditions from Table 10.2 to cocrystallize with the cofactor resulted in small crystals (maxium dimension 60 µm) that took 5-9 days to grow (Figure 10.2 C). Regularly, precipitate would form upon mixture of the protein-cofactor solutions and the precipitant solution and after several days, crystals would form within this precipitate. However, these crystals were too small to mount and to screen them for X-ray diffraction on the in-house X-ray diffraction facility. KPHMT incubated with cofactor analogue and ke-

topantoate did not result in crystals and the solution only contained precipitate.

Subsequently, new protein samples were purified by the author and further crystallization trials could be pursued, jointly with Dr Loïc Bertrand who temporarily joined the project. The crystallization trials with the new protein sample started with the conditions described in Table 10.2 with 6-12% (w/v) PEG3350 instead of PEG4000, at 4°C. However, instead of a racemic mixture of the cofactor analogue, the stereochemically *S*-pure form, known to be used by the enzyme was used (Subsection 8.1.1). In addition to (6*S*)-5-foTHF two other analogues, (6*S*)-5-meTHF calcium salt and folic acid, were used.

α-kiva was added to the protein solution in five- and ten-fold excess to the protein concentration. Crystals of KPHMT incubated with α-kiva could readily be obtained (Figure 10.2 D) with the condition described in Table 10.2. This indicated that the protein from this sample crystallized under the same conditions as before. With respect to the concentration of α-kiva, no marked effect on the crystallization of KPHMT (without cofactor analogues), could be observed between using a five- or ten-fold excess.

Crystallization trials were then pursued further with KPHMT incubated with both α-kiva and cofactor analogues. The cofactors were only added in slight excess at a concentration of 1-2 mM, while initially using a protein concentration between 28-31 mg/ml (0.9-1.1 mM) and an α-kiva concentration of 5 or 10 mM. Crystals could be obtained with and without α-kiva for all three cofactor analogues. The optimal PEG3350 concentration was in the range of 6-8% (w/v). The crystallization drops in all cases were covered by a high number of thin plate-like crystals. The maximum dimension of a single crystal was approximately 80 µm. However, crystals would frequently grow attached to each other.

Although crystals with KPHMT and folic acid, both with and without

α-kiva, were obtained, cocrystallization with this cofactor analogue was then not pursued further because the other two analogues resulted in more regularly-shaped and bigger crystals. Additionally, the solubility of folic acid was rather low (see Subsection 8.1.1) and it was difficult to solubilize it.

Because of the high number of nucleation sites, the protein concentration was lowered to 25 mg/ml for subsequent crystallization trials. Reducing the protein concentration below 20 mg/ml did not produce crystals. In addition, the pH was lowered using 25 mM Bis-tris buffer pH 6.0, 6.5, and 7.0. Crystals of KPHMT incubated with α-kiva and either (6*S*)-5-meTHF or (6*S*)-5-foTHF grew more regular and bigger with Bis-tris buffer pH 6-6.5 in the precipitant solution (see Figure 10.2 E, F, G). Similar crystals could also be produced without α-kiva and with either of the two cofactor analogues. These crystals were screened for X-ray diffraction. Cryoprotection was carried out with glycerol similar to the protocol described above for PEG400. X-ray diffraction screening and data indexing (Figure 10.4) indicated that the crystals all were monoclinic and had small unit cell dimensions (see Section 10.4).

However, quite frequently, crystals would dissolve upon moving the crystallization plate to 19°C from the 4°C incubator. This phenomenon could not be associated with a specific crystallization condition. It was assumed that a higher molecular weight PEG could circumvent that problem. Therefore, in order to increase the size of the crystals, PEG3350 was substituted for PEG6000, using it at a concentration of 6-10% (w/v).

Optimization of the crystallization conditions with PEG6000 by lowering the protein concentration to 20-23 mg/ml finally resulted in a condition that yielded regularly shaped crystals with a clear monoclinic morphology (Figure 10.2 H). The largest crystal in this drop had dimensions of approximately 150×80×100 µm^3. The protein concentration used for the crystallization trials was 23 mg/ml and α-kiva was added in a two-fold

excess. The composition of the precipitant solution is shown in Table 10.3. However, X-ray diffraction data from these crystals could not be collected, because the crystals did not cope well with the cryo-protection protocol with glycerol, developed before (see above).

Table 10.3 Crystallization condition C for KPHMT that was incubated with α-kiva and (6*S*)-5-meTHF

Component	PEG6000	NaCl	$NaCH_3CO_2$	$Mg(CH_3CO_2)_2$	Bis-tris pH 6.0[a]	Protein
Concentration	8% (w/v)	100 mM	100 mM	10 mM	25 mM	20-23 mg/ml

[a] The pH was adjusted with HCl.

10.4 X-ray diffraction screening and processing

In this study, cryo-cooling was systematically applied to screen crystal specimens for X-ray diffraction. The crystals were mounted on the in-house X-ray diffraction facility, as described in Subsection 8.3.1 and Subsection 8.3.2. Screening for X-ray diffraction was carried out with a 15-30 min exposure time, with a 1°C oscillation angle. Always, two orientations of the crystals rotated by 90°C were examined.

Crystals of KPHMT, incubated with α-kiva and either (6S)-5-meTHF or (6S)-5-foTHF, diffracted only weakly on the in-house X-ray diffraction facility. Because of the weak X-ray diffraction of the crystals, manual indexing —as opposed to auto-indexing— had to be carried out. The diffraction spots were manually selected in DENZO and then used to calculate unit cell dimensions. All X-ray diffraction patterns that could be screened indicated a monoclinic spacegroup and very similar unit cell dimensions of between 86-87.5 Å for a, 153-158.2 Å for b, 118-119.5 Å for c, and 98-108° for β.

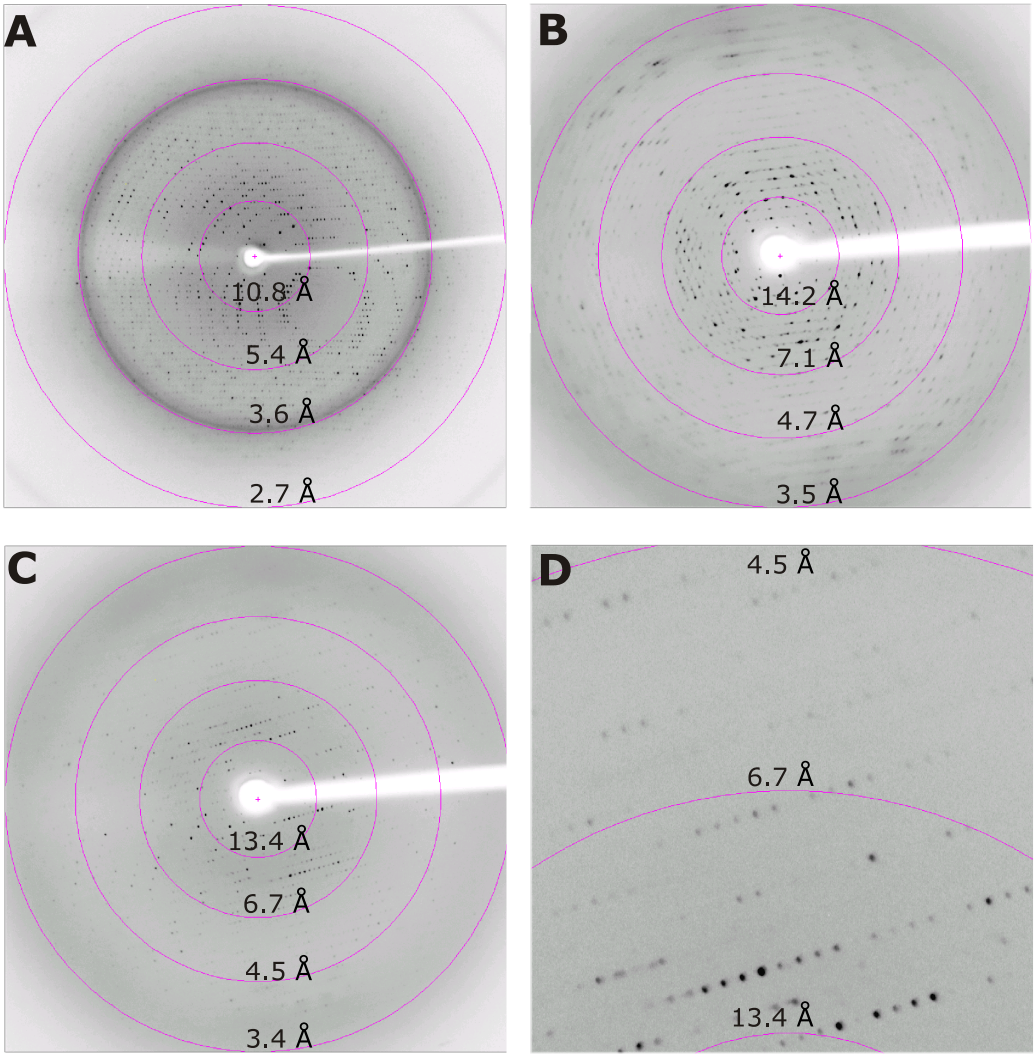

Figure 10.3 X-ray diffraction images of crystals of KPHMT grown in the presence of ketopantoate. The pink circles in this and the following X-ray diffraction images indicate resolution limits. The corresponding real-space resolution limits are indicated. (**A**) Diffraction obtained from crystals of KPHMT and ketopantoate with sodium citrate as the buffer component of the precipitant solution. (**B**) Initial diffraction pattern with high mosaicity obtained from crystals of KPHMT and ketopantoate with the precipitant condition from Table 10.2. (**C**) Diffraction obtained from optimized crystals of KPHMT and ketopantoate with Hepes as the buffer component with the precipitant from Table 10.2. Indexing of this image indicated small unit cell dimensions. (**D**) The same diffraction pattern as in **D** but in close-up view.

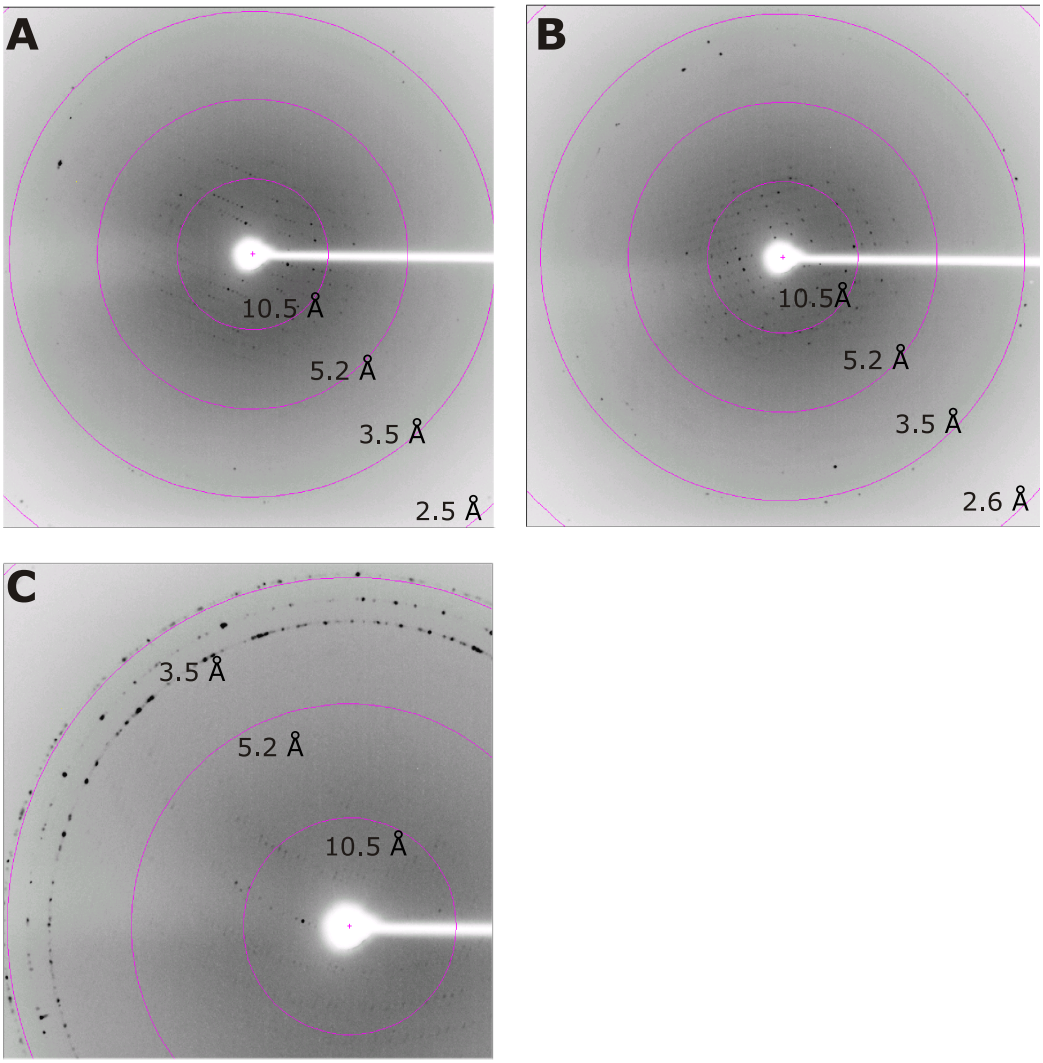

Figure 10.4 X-ray diffraction images of crystals of KPHMT, incubated with α-kiva and co-factor analogues. The crystallization conditions for these crystals were similar to the one described in Table 10.3, but with PEG3350. (**A**) X-ray diffraction obtained from crystals of KPHMT grown on the presence of α-kiva and (6*S*)-5-meTHF. (**B**) X-ray diffraction obtained from crystals of KPHMT grown in the presence α-kiva with (6*S*)-5-foTHF. (**C**) Diffraction obtained from crystals of KPHMT incubated with (6*S*)-5-foTHF, without α-kiva.

10.5 Discussion

Obtaining crystallization conditions that result in well-diffracting crystals of KPHMT, incubated with cofactor analogues, certainly proved more difficult than could be expected from the rather positive results at the beginning. A significant area of crystallization space around crystallization conditions that were found to be optimal for either KPHMT alone or for KPHMT, incubated with product ketopantoate or substrate α-kiva, was explored. A set of crystallization conditions was found that appear to be promising. However, the problem was one to optimize crystallization conditions, not to obtain crystals.

KPHMT crystals, incubated with two cofactor analogues and/or α-kiva can now be produced. The corresponding X-ray diffraction of these crystals extended to $\sim$4.0 Å on the in-house X-ray generator. Bringing such kind of crystals to a synchrotron source could presumably improve the diffraction limit. However, in order to collect a full dataset, crystals of larger three-dimensional proportions, such as the last crystal that were produced in this study, will be necessary. The author presumes that with the the crystallization condition of Table 10.3 large enough KPHMT crystals can be reproduced. Although all cofactor analogues are relatively unstable (see Subsection 8.1.1), once the cofactor is bound it could be assumed to be stabilized by protein—cofactor interactions. Whether the cofactor analogues actually bound to the protein will only be revealed by difference Fourier maps.

Of the three tested cofactor analogues, folic acid was not used further because of the low solubility and because binding may be rather weak in general, compared to both 5-foTHF and 5-meTHF. Enzyme inhibition experiments by Saldanha (2002) indicated that the substitution at the N_5 appears to be important for efficient binding to the enzyme. However, no crystals of KPHMT could be obtained in the presence of both ketopantoate and cofactor analogues. Presumably, the product could sterically hinder the binding of cofactor analogues.

11 Comparative analysis of the KPHMT crystal structure

11.1 Introduction

The features of the *E. coli* KPHMT crystal structure are described in the introduction (see Section 3.4) and show that the tertiary structure adopts a classical $(\beta\alpha)_8$ or TIM-barrel fold.

The $(\beta\alpha)_8$-barrel fold motif is the most frequently occurring enzyme folding pattern found in the PDB, and it is anticipated that as many as 10% of the structures of all enzymes adopt this fold (Gerlt, 2000). Similarly, transcriptome analysis of yeast genome expression studies showed the $(\beta\alpha)_8$-barrel to be the most common transcriptome fold (Jansen and Gerstein, 2000). Enzymes with this fold, so far, include examples from five of the six primary types of enzymes as classified by the Enzyme Commission and are performing a particularly wide range of often unrelated functions (Copley and Bork, 2000; Hegyi and Gerstein, 1999; Nagano *et al.*, 1999; Pujadas and Palau, 1999). Most of them share no significant sequence similarity, although the overall three dimensional structure of their $(\beta\alpha)_8$-barrel domains are surprisingly similar (Branden, 1991). Approximately

one-third of the known $(\beta\alpha)_8$-barrels have their polypeptide chains arranged in a single barrel domain and in most of them the homo-dimer seems to be the functional oligomeric unit.

Notwithstanding their functional versatility, in all $(\beta\alpha)_8$-barrel enzymes found so far, the active site is located at the C-termini of the β-strands (Wierenga, 2001). In general, the geometry of the active site itself is predominantly shaped both by residues at the C-termini of the β-strands and the residues of the loops connecting the β-strands with the consecutive α-helices. These $\beta\alpha$-loops are of variable length and residue composition, accounting for the different functions, and it was shown that small sequence changes in these loops can change function (Jurgens *et al.*, 2000). Consequently, this fold constitutes a striking example for the separation of catalytically and structurally relevant residues.

The abundance and stability of the fold has led to much interest and speculation as to whether it evolved independently or whether it diverged from a single or more limited number of progenitors. Farber and Petsko (1990) have argued for a common ancestry of a limited number of $(\beta\alpha)_8$ barrels on the basis of similarity in the positions of the active sites as well as other structural features. Structural evidence that a $(\beta\alpha)$8-barrel fold has evolved by gene duplication and fusion from an ancestral half-barrel has recently been brought forward for N-[5-phospho-ribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (ProFAR isomerase or HisA) and imidazole glycerol phosphate synthase (ImGPS or HisF), two enzymes involved in histidine biosynthesis (Hocker *et al.*, 2001; Lang *et al.*, 2000). The sequence of both proteins shows a clear internal duplication, along with a detectable repeat related by a two-fold axis in the structure. Copley and Bork (2000) provide evidence that 12 out of the 23 TIM-barrel superfamilies listed in the SCOP database at the time of the analysis share a common origin and that the common ancestor was most reminiscent of HisA. The authors also support a model of divergent evolution for most of the $(\beta\alpha)_8$-barrel enzyme families involved in pathways of

the central metabolism. However, it is not clear if a common evolutionary origin can be assigned to all $(\beta\alpha)_8$-barrel families. Common structure is often taken as an indicator of common ancestry the simplicity of the core fold. However, the absence of sequence similarity and the vast range of functions also implies convergent evolution, as might be the case with the $(\beta\alpha)_8$ glycosidases (Nagano *et al.*, 2001).

As a consequence of the above mentioned, comparative analysis of structures with a $(\beta\alpha)_8$-fold requires careful attention to details in the structures, as the overall fold can be very similar, even though there is no significant sequence identity among the different $(\beta\alpha)_8$ enzymes.

One approach to gain functional insights into the structure of a protein is to compare its active site with homologous proteins and proteins known to have similarities in the catalytic chemistry. Sequence searches using PSI-BLAST with the KPHMT amino acid sequence failed to detect any enzymes similar or homologous to KPHMT. However, comparative analysis of three-dimensional structures is a powerful tool in identifying evolutionary relationships and in gaining insight into the mechanism. This approach was, therefore, used to identify homologues of KPHMT and to compare their active sites. Major parts of the analysis have also been described elsewhere (Schmitzberger *et al.*, 2003b; von Delft *et al.*, 2003).

11.2 Analytical procedures

KPHMT amino acid sequence and structure searches were carried out as described in Chapter 9, by using the homology recognition server FUGUE and the fold recognition server 3D-PSSM . The structure of KPHMT was compared to those of all proteins in the PDB with the DALI software. Potential homologues with high Z-scores and low r.m.s.d. were then aligned with COMPARER and the structural alignments were formatted in JOY.

Conservation of the fold and global structural features was assessed by inspection of the superposed coordinates of the aligned structures. Catalytically important residues were identified from their conservation in multiple sequence alignments and by visual inspection of their position in the structures. In addition to the enzymes identified with DALI, FUGUE and 3D-PSSM, enzymes which were classified in the same Enzyme Commission group and for which structures were available were compared to KPHMT by using the methodology described above.

The SCOP database superfamily classifications both of proteins considered homologues and of enzymes classified in the same Enzyme Commission group were investigated. All the other members of the superfamily in which homologues of KPHMT were identified were subjected to the comparative structural analysis described above. In that way, more distant homologues could potentially be identified and relationships within the superfamily could be revealed. Particular attention was then paid to the conservation of the active-site residues common to homologues and positions in the structures of those having equivalent functions.

11.3 Structural homologues

A search of protein folds with KPHMT expectedly found similarities to numerous proteins with a $(\beta\alpha)_8$-barrel fold motif. Closest structural similarity could be observed with phosphoenolpyruvate mutase (PEPM; EC 5.4.2.9; PDB-ID: 1PYM) and isocitrate lyase (ICL; EC 5.4.2.9; PDB-ID: 1IGW). Both of these enzymes could also be detected as homologues by a KPHMT sequence-to-structure homology search using FUGUE, with PEPM scoring particularly high, as it also did in a search with the 3D-PSSM server. In spite of lack of significant overall sequence similarity between KPHMT and either of the two enzymes, structure-based sequence alignments with COMPARER revealed remarkable similarities both in the locations and properties of important active-site residues.

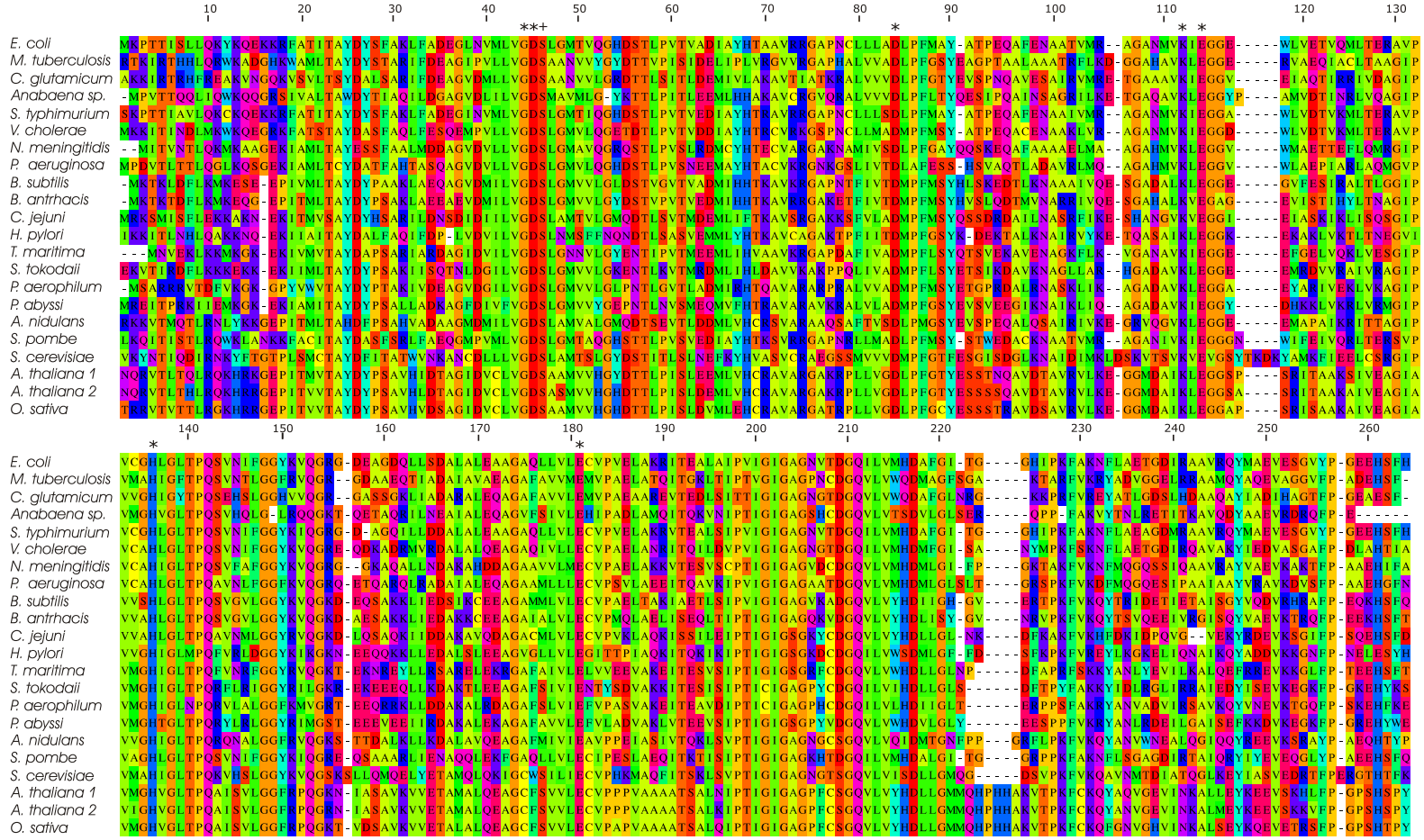

Figure 11.1 Representative amino acid sequence alignment of KPHMT orthologs, including representatives of bacteria, archaea and eukaryotes. The alignment is numbered according to the full-length *E. coli* sequence, without counting the gaps. Several amino acid sequences have been truncated at the N- and C-termini to simplify the alignment. Invariant residues are indicated with a star and conserved residues with a +.

PEPM from *Mytilus edulis* is a homotetrameric (dimer of dimers) enzyme that catalyzes the conversion of PEP to phosphonopyruvate in the biosynthesis of phosphonates, involving cleavage of an O-P bond and formation of an C-P bond (Huang *et al.*, 1999; Kim and Dunaway-Mariano, 1996) (see Figure 11.2). The overall globular architecture of the $(\beta\alpha)_8$-barrel is very similar to that in KPHMT. Both enzymes have an additional N-terminal helix of the same length, prior to the $(\beta\alpha)_8$-barrel domain that lies across the N-terminus of the barrel. The major difference in the macromolecular assembly lies in the helix swapping of the eighth C-terminal helix in PEPM, which, together with two ensuing smaller helices, protrudes from one subunit to pack against a neighboring subunit, a phenomenon that is not observed in *E. coli* KPHMT.

Conservation in the overall sequence between KPHMT (Figure 11.1) and PEPM (Figure 11.4) is most obvious in helix $\alpha 7$ and in the loop connecting helices $\alpha 7$ and $\beta 6$ (see Figure 3.5). The reason may be solely structural, because this part of the $(\beta\alpha)_8$-barrel has no direct catalytic role and is not involved in dimer formation either. Further significant sequence similarity can be observed in strand $\beta 3$, which is important for the orientation of Asp84. The insertion in the loop region between strands $\beta 4$ and $\alpha 5$ in PEPM is thought to be involved in conformational change upon substrate binding and blocks the C-terminal region of the active site in the crystal structure. However, a similar functional region in KPHMT could not be identified.

The most striking resemblance exists in the coordination sphere of the active-site Mg^{2+} ion (see Figure 3.4) and the coordination of the ligands in the two enzymes. Ser46 at the end of strand $\beta 2$, Asp84 in strand $\beta 3$, and Glu114 at the C-terminus of $\beta 4$, all of which are completely conserved in KPHMT (see Figure 11.1), have their equivalents in the invariant residue Ser46 at the end of $\beta 2$, Asp85 in $\beta 3$, and Glu114 at the C-terminus of $\beta 4$ in PEPM (Figure 11.4). Superposition and inspection of the active site shows that these residues are in fact in the same absolute position and coordi-

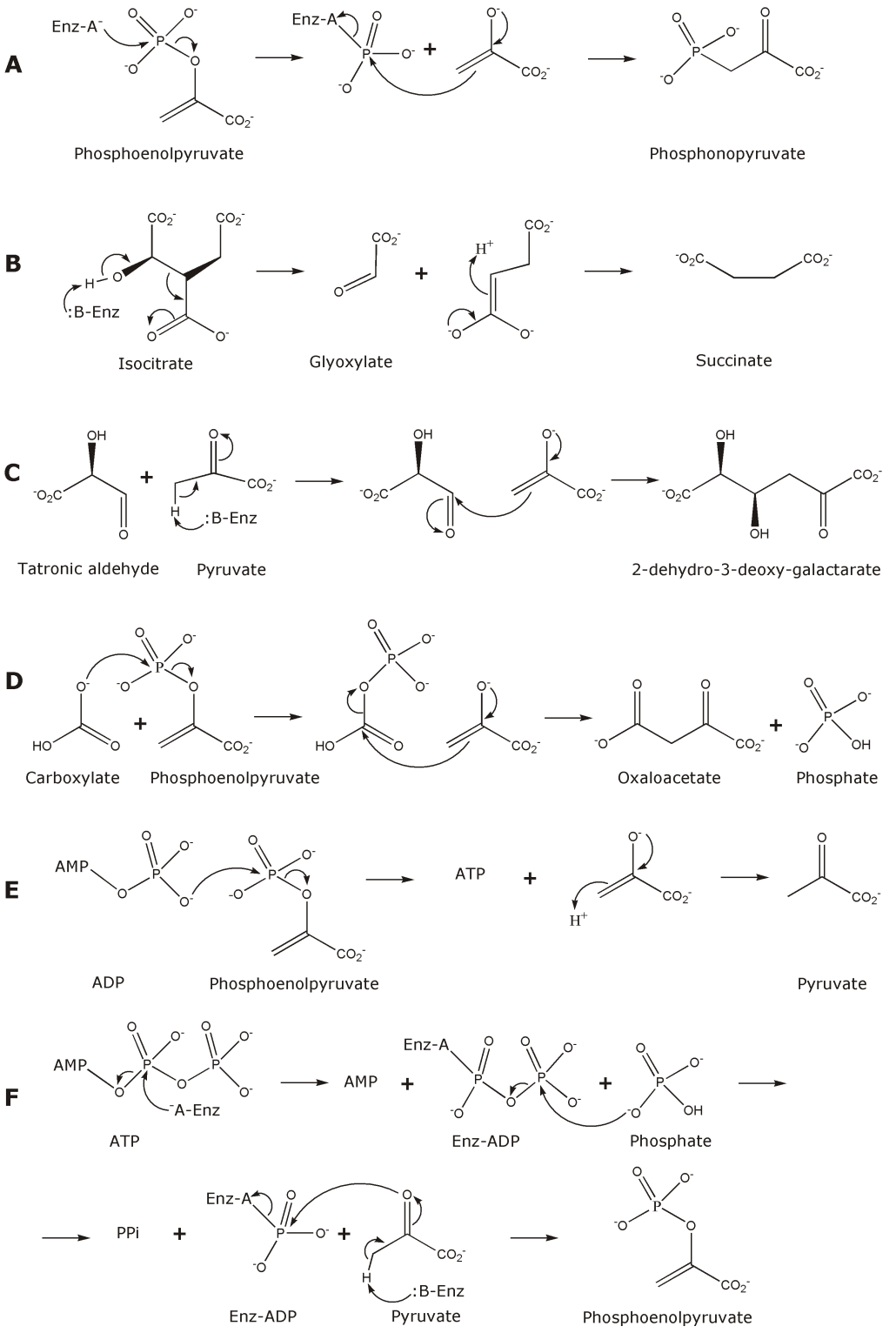

Figure 11.2 Schematic illustration of the biochemical reactions catalyzed by the enzymes in the PEP/pyruvate superfamily. (**A**) Phosphoenolpyruvate mutase. (**B**) Isocitrate lyase. (**C**) 2-dehydro-3-deoxy-galactarate aldolase. (**D**) Phosphoenolpyruvate carboxylase. (**E**) Pyruvate kinase. (**F**) Pyruvate phosphate dikinase.

nate the ligand and Mg^{2+} in exactly the same way (Figure 11.5). Asp85 coordinates Mg^{2+} directly in an equatorial position but is also H-bonded to a water residue and Glu114 H-bonds to this water-residue which also coordinates to Mg^{2+} equatorially. Asp45 in KPHMT has been replaced by a water molecule, which is held in place by H-bonding to the completely conserved residues Asp58 and Asp87 in PEPM. Asp58 also H-bonds to the Mg^{2+}-coordinated water in the same way as Asp45 does in KPHMT. Another closely similar feature is the H-bond of the carboxyl-group of oxalate to the invariant Ser46 in PEPM, in a related way as the carboxyl-group of ketopantoate H-bonds to the completely conserved Ser46 in KPHMT. The carboxyl groups of the oxalate residue in PEPM coordinate axial and equatorial positions of Mg^{2+} as do the carboxyl and keto-group of ketopantoate. However, in general, a Mg^{2+} ion that uses three water molecules as a bridge to the protein, as in PEPM, is rather unusual.

Table 11.1 Statistical data on the comparison of some indicators of similarity of the enzymes currently assigned to the PEP/pyruvate superfamily in SCOP

Enzyme[a]	PID (%)	DALI Z-score	No. of aligned residues	R.m.s.d. (Å)	FUGUE Z-score
PEPM	13.5	17.1	198	3.2	14.47
ICL	16.5	14.6	201	3.8	3.35
DDGA	11.4	12.3	173	3.0	N/A
PK	16.1	11.8	179	3.5	N/A
PPDK	10.6	9.9	178	3.7	N/A
PEPC	12.9	6.2	183	3.8	N/A

[a] All comparisons in the table are with respect to KPHMT. Sequence identities were calculated from structure-based COMPARER (Sali and Blundell, 1990) alignments. Abbreviations: PID: percentage amino acid sequence identity. For the comparison with ICL, the *E. coli* structure was used, and for the multidomain proteins PEPC, PPDK, and *E. coli* PK, only their $(\beta\alpha)_8$ domains were used.

PEPM shows closest structural and sequence-similarity to ICL and *vice versa* (see Figure 11.4). *E. coli* ICL and PEPM superimpose with an r.m.s.d. of 1.51Å for 221 structurally-aligned residues and the sequence identity between the two enzymes is 20.5%. Furthermore, the homotetrameric ICL

is structurally the second most similar enzyme to KPHMT. ICL catalyzes the reversible conversion of isocitrate into glyoxylate and succinate, the first step in the glyoxylate bypass pathway. The crystal structure of the *E. coli* enzyme in complex with Mg^{2+} and pyruvate (Britton *et al.*, 2000), the *A. nidulans* enzyme in complex with Mn^{2+} and glyoxylate (PDB-ID: 1DQU) (Britton *et al.*, 2001) and the *M. tuberculosis* enzyme complex with Mg^{2+} and with two inhibitors (PDB-ID: 1F8I) (Sharma *et al.*, 2000) have been solved. For the comparison with KPHMT in this study, the *E. coli* ICL structure was used.

The conserved amino acid sequence between ICL and KPHMT overlaps largely with the residues that are conserved between KPHMT and PEPM. In *E. coli* ICL there are three α-helices before the first β-strand, of which α3 lies across the N-terminal part of the barrel. As in PEPM, the amino acids chain folds in a series of α-helices at the C-terminus, and similarly, the α12 and α13 helices are involved in helix swapping. The conserved residues 194 to 199 in the region between β4 and α5 in *E. coli* ICL are disordered in three out of four subunits. Comparison with PEPM shows that this region is spatially euqivalent to the one expected to undergo conformational change in PEPM between strands β4 and β5 (see above).

The active site of ICL shows almost identical features to those in PEPM (Figure 11.6), and the key active site residues in PEPM are also invariant in ICL. Similarly, the mode of ligand binding is essentially the same. The conserved Asp157 after the β3 in *E. coli* ICL is equivalent to Asp84 in KPHMT and Asp85 in PEPM, as is the conserved Glu186 after β4 , which is equivalent to Glu114 in both KPHMT and PEPM (Figure 11.4). Similarly, the conserved Ser81 has its equivalent in both KPHMT and PEPM, with Ser46.

The only difference in the coordination sphere of Mg^{2+} between PEPM and *E. coli* ICL seems to be the conservative substitution of Asp87 to its equivalent Glu159 in the *E. coli* ICL structure, although in *A. nidulans* and

M. tuberculosis ICLs, this residue is also an aspartate (Asp170 and Asp153, respectively).

11.4 Mechanistic comparison of KPHMT with ICL and PEPM

Comparison of the mechanism of both ICL and PEPM with KPHMT is not straightforward. The reaction mechanism in PEPM is not very well understood, but it has been proposed that it involves a covalent phosphoenzyme intermediate of the phosphoryl group with Asp58. Huang *et al.* (1999) proposed that during formation of this phosphoenzyme intermediate, Mg^{2+} shifts toward the position of the water to the left of the Mg^{2+} ion (Figure 11.6), interacting directly with the carboxylate groups of Asp85, Asp87, and Glu114, a process for which no equivalent in KPHMT has been found so far. What appears to be much the same in KPHMT and PEPM is the need to stabilize an enolate intermediate. This is derived from pyruvate in PEPM (Kim and Dunaway-Mariano, 1996) and from α-kiva in KPHMT.

In ICL the initial step, as in KPHMT, is a proton abstraction (Figure 11.2). As in KPHMT and PEPM, the reaction intermediate for the reverse reaction in ICL involves an enolate, which then undergoes a Claisen condensation. For *E. coli* ICL, it has been suggested that Cys195, which is located in the loop region between strands β4 and β5 (Figure 11.4), is the base for the proton abstraction. In *M. tuberculosis* ICL, a large conformational movement of this cysteine-containing loop together with the ordering of the final helix at the C-terminus could be observed. However, neither ICL nor PEPM appear to use any cofactor other than Mg^{2+} in the reaction.

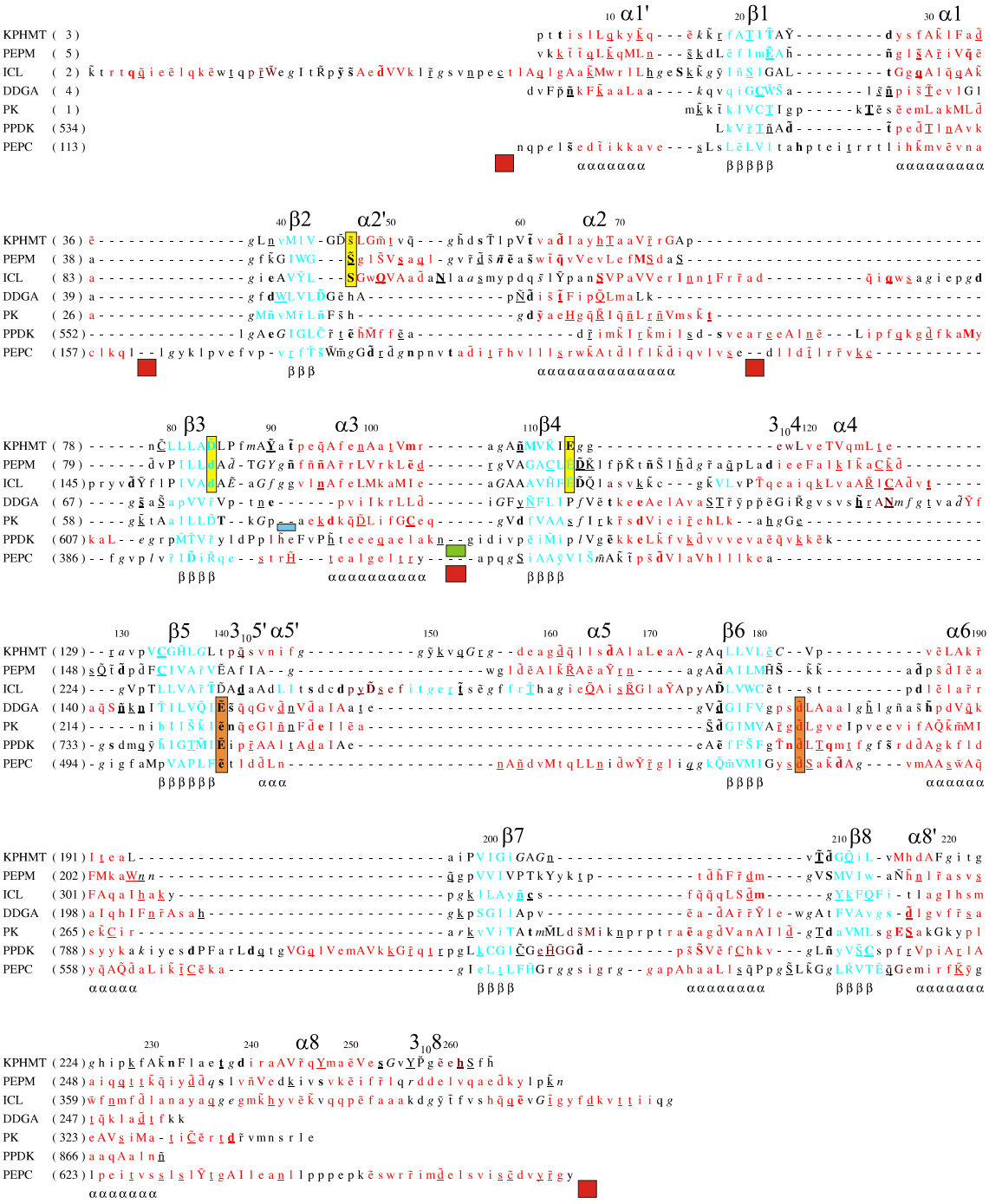

Figure 11.4 Structure-annotated amino acid sequence alignment of KPHMT, in JOY-format (Mizuguchi *et al.*, 1998a), with all the enzymes currently classified in the PEP/pyruvate superfamily. The full-length sequences of KPHMT, PEPM, *E. coli* ICL, and DDGA are shown. From the three-domain *E. coli* PK, only the $(\beta\alpha)_8$ domain comprising residues 1–70 and 171–345 was used (residues 70–171 have been removed; the position of this segment is indicated by a blue box). Likewise, only the $(\beta\alpha)_8$ domain of PPDK, including residues 534–874, was used. In order to simplify and compress the alignment, residues 640–696 were removed and the deletion is indicated by a green box. For the initial alignments, the full-length PEPC protein was used. The following, mainly α-helical, insertions were removed: residues 1–112 at the N-terminus, 161–232 between strands β1 and β2, 289–375 between β2 and β3, 412–464 between strands β3 and β4, and 671–883 at the C-terminus. All deletions are indicated by a red box, below the alignment.(**legend continues**)

Figure 11.4 continued The yellow and brown boxes around the residues highlight the conserved catalytic residues among KPHMT, ICL, and PEPM and among DDGA, PEPC, PPDK, and PK, respectively. The alignment is numbered according to the KPHMT amino acid sequence, with every 10th residue number shown. Secondary structural elements are also denoted according to those of KPHMT. The key to the JOY alignment is as follows:

α-helix	red x
β-strand	blue x
3_{10} helix	maroon x
cis-peptide	breve x̆
solvent accessible	lower case x
solvent inaccessible	UPPER CASE X
hydrogen bond to main-chain amide	**bold x**
hydrogen bond to main-chain carbonyl	underline x
hydrogen bond to other side-chain	tilde x̃
disulfide bond	cedilla ç
positive ϕ torsion angle	*italic x*

11.5 Comparison with members of the PEP/pyruvate superfamily

At the time of the analysis, the SCOP database distinguished 25 superfamilies of $(\beta\alpha)_8$-barrel proteins in which the members of each superfamily have a probable common evolutionary origin. Both ICL and PEPM are assigned to the PEP/pyruvate superfamily. Other members of this superfamily include *E. coli* 2-dehydro-3-deoxy-galactarate aldolase (DDGA; EC 4.1.2.20; PDB-ID: 1DXF), *E. coli* PEP carboxylase (PEPC; EC 4.1.1.31; PDB-ID: 1QB4), *E. coli* pyruvate kinase (PK; EC 5.4.2.9; PDB-ID: 1PKY) and pyruvate phosphate dikinase (PPDK; EC 2.7.9.9; PDB-ID: 1KBL) from *C. symbiosum*.

Like KPHMT, the homohexameric (trimer of dimers) DDGA is a class 2 aldolase with a single-domain $(\beta\alpha)_8$- fold (Izard and Blackwell, 2000).

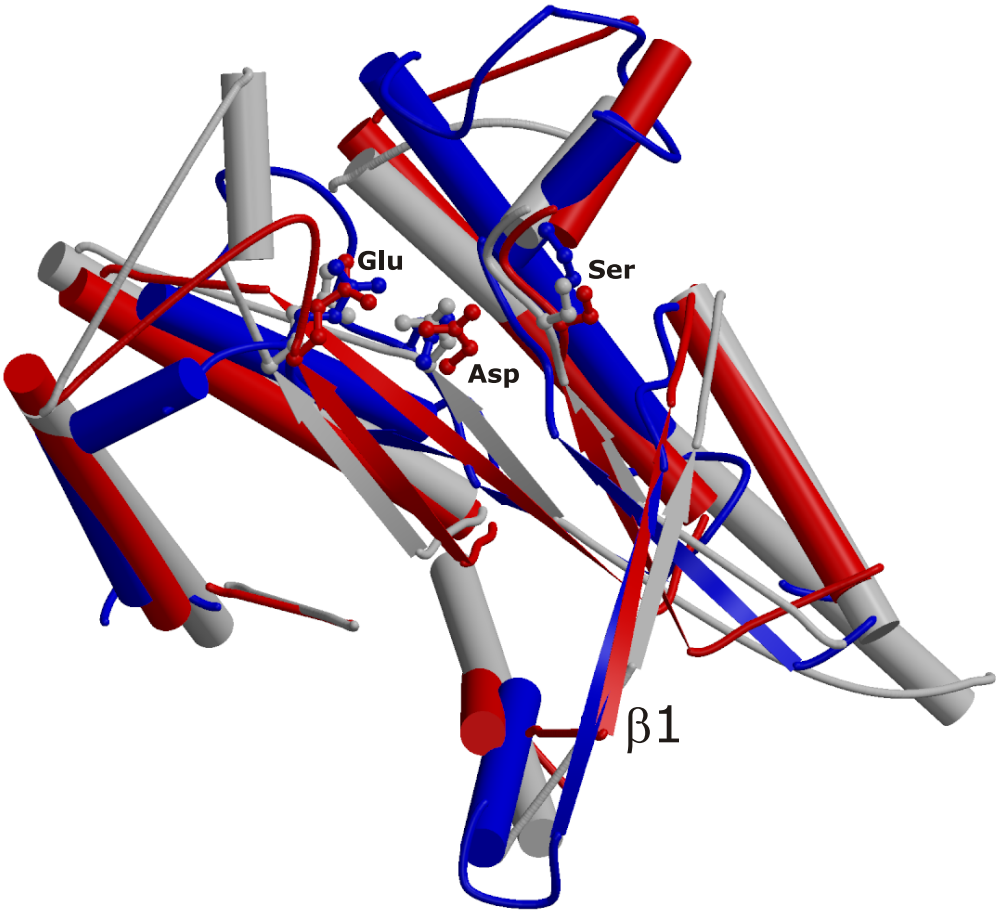

Figure 11.5 Cartoon illustration of the three-dimensional superposition of the coordinates of KPHMT (blue), *E. coli* ICL (grey), and PEPM (red). Shown are the secondary structural elements around the first four β-strands comprising residues 1–130 in KPHMT, 43–224 in ICL (for clarity the first 42 residues are omitted), and 1–150 in PEPM. Also shown are the catalytically important residues Ser46, Asp84, and Glu114 in KPHMT, Ser46, Asp85 and Glu114 in PEPM and Ser91, Asp157, and Glu186 in *E. coli* ICL in ball-and-stick representation.

The enzyme catalyzes the reversible aldol cleavage of 2-dehydro-3-deoxy-galactarate into pyruvate and tartronic semialdehyde (Figure 11.2) in the catabolic pathway for the utilization of D-glucarate/galactarate. The enzyme has a low substrate specificity, condensing a wide range of aldehydes with pyruvate. A DALI search with KPHMT listed DDGA only as the 31st hit in the list. However, PEPM is structurally the second most simi-

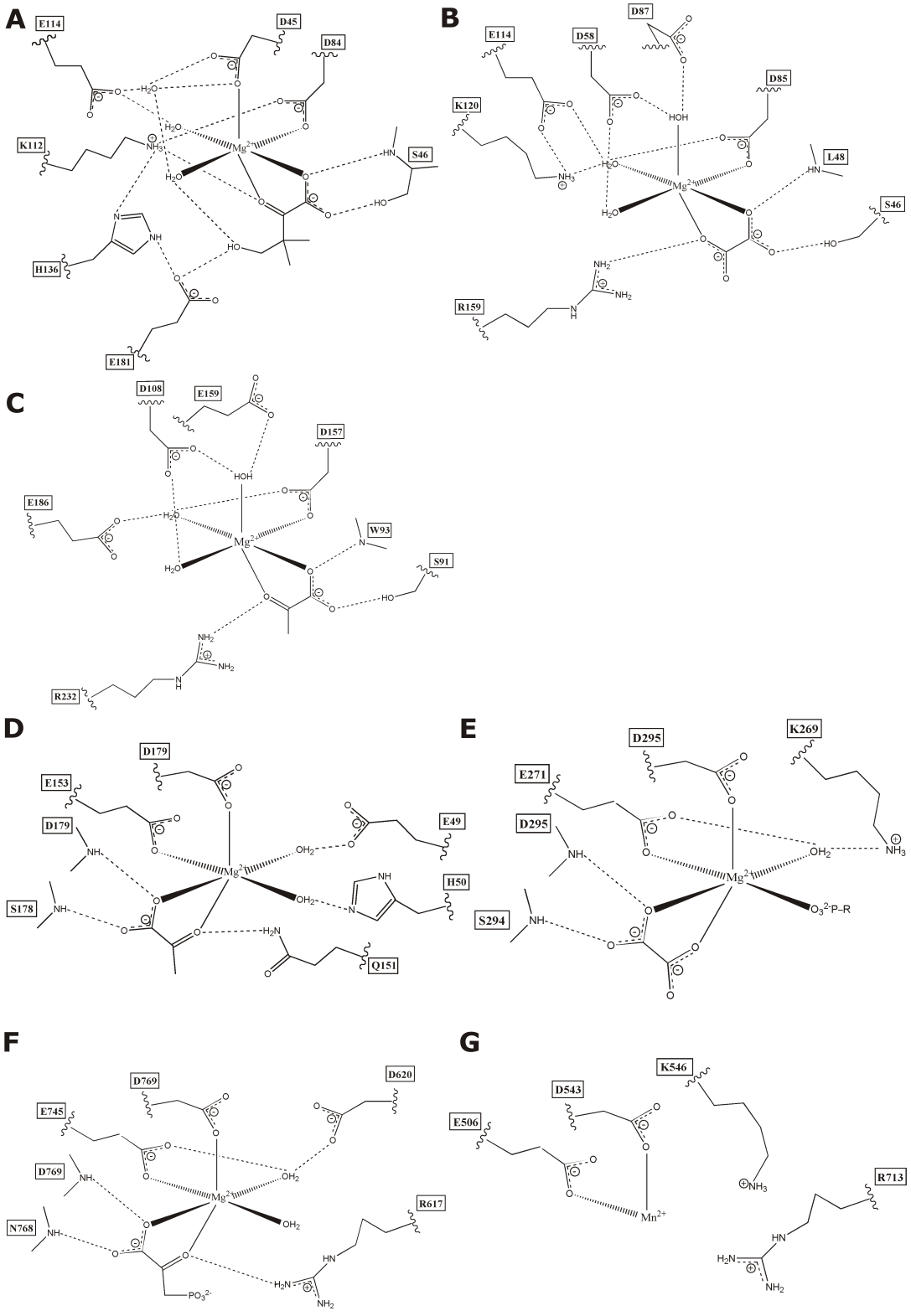

Figure 11.6 Schematic active site representations of KPHMT and the current members of the PEP/pyruvate superfamily. All are viewed with the vertical coordination of the ion in line with the vertical axis of the $(\beta\alpha)_8$ barrel. (**A**) KPHMT; (**B**) PEPM; (**C**) ICL; (**D**) DDGA; (**E**) rabbit muscle PK (PDB-ID: 1A49); (**F**) PPDK; (**G**) PEPC. The crystal structure of *E. coli* PEPC has been solved in complex with Mn^{2+}, but no ligand-bound structure of PEPC is available yet.

lar enzyme to DDGA within the PEP/pyruvate superfamily. The two enzymes superimpose with an r.m.s.d. of 3.1 Å for 174 structurally-aligned residues, whereas sequence identity between the two enzymes is not significant (12.1%). DDGA shows some features common with KPHMT, such as the N-terminal α-helix. Analogous to findings for ICL and PEPM, helix swapping of the eighth α-helix is observed. The active site Mg^{2+} is directly complexed by Glu153 in the loop after strand β5 and by Asp179 in the α-helix after β6 (Figure 11.6). The cocrystallized pyruvate occupies two Mg^{2+} binding sites and H-bonds to Ser178. Two water molecules complete the coordination sphere, one of which H-bonds to Glu49, which aligns with Asp45 of KPHMT (Figure 11.4). Consequently, the overall Mg^{2+} and ligand coordination bears some resemblance to that of KPHMT, but with inverted symmetry. The absolute position of the metal ion has moved from between strands β3 and β4 to a position between strands β5 and β6. However, not only the absolute position but also the position of the aspartate and glutamate residues relative to each other are different in DDGA. The possibility of a circular permutation was investigated. Nonetheless, no evidence for such an event could be found. A circular permutation in which β-strands are joined and the barrel is opened at another place would not explain why the relative positions of Asp and Glu have been exchanged.

The first step in the reaction catalyzed by DDGA involves enolization of pyruvate (Figure 11.2) in an aldol reaction chemically very similar to the one in KPHMT. However, no base close enough for the proton abstraction could be identified in the active site of the enzyme. Instead, a phosphate, which was observed in the active site in the ligand-free structure, is thought to accept the proton of the pyruvate's methyl group (Izard and Blackwell, 2000).

PEPC, *E. coli* PK, and PPDK are not very similar, structurally, to KPHMT. However, investigation of the structural similarity relationship between the first three and DDGA shows high structural similarity between them, with closest structural similarity between DDGA and PK (Figure 11.4). PK

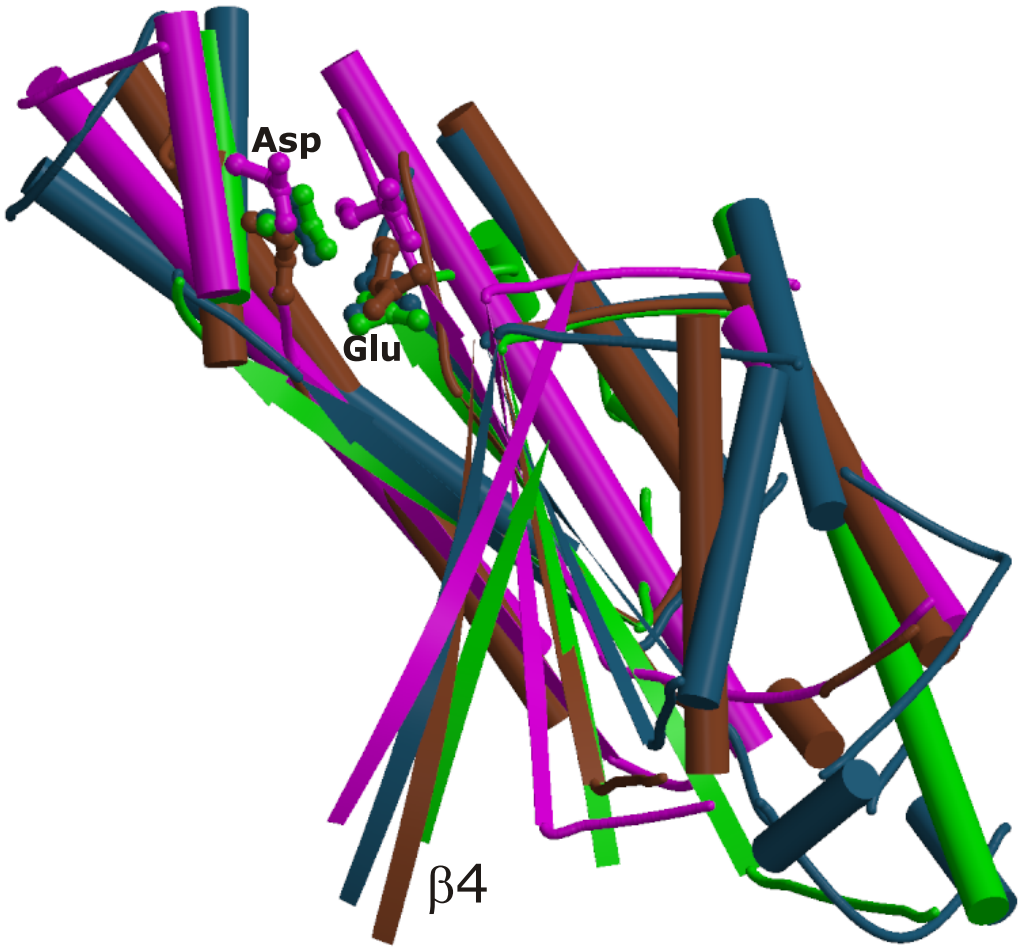

Figure 11.7 Cartoon illustration of a the superposition of the three-dimensional coordinates of DDGA (steel blue), PEPC (magenta), PPDK (green) and *E. coli* PK (brown). Shown are the secondary structural elements around β-strands 3–6 in the range of residues 70–208 in DDGA, 391–571 in PEPC, 613–775 in PPDK, and 61–271 in PK (residues 398–412 in PEPC, and residues 618–700 in PPDK were omitted for clarity). The catalytically important residues Glu153 and Asp179 in DDGA, Glu745 and Asp769 in PPDK, Glu506 and Asp543 in PEPC, and Glu223 and Asp246 in PK are shown in ball-and-stick representation.

in turn is structurally the most similar enzyme to both PEPC and PPDK. The homotetrameric PK, a three domain protein with a $(\beta\alpha)_8$-barrel core domain, catalyzes the conversion of PEP to pyruvate coupled to the synthesis of ATP (Larsen *et al.*, 1998; Mattevi *et al.*, 1995) (Figure 11.2). The ho-

motetrameric PEPC catalyzes the irreversible carboxylation of PEP to form oxalacetate and inorganic phosphate (Kai *et al.*, 1999; Matsumura *et al.*, 1999), whereas PPDK catalyzes the the interconversion of ATP, phosphate, and pyruvate into AMP, pyrophosphate, and PEP, respectively (Herzberg *et al.*, 1996, 2002) (Figure 11.2).

Structure-based sequence alignment of the $(\beta\alpha)_8$ domain and superpositions of the active site show that the Mg^{2+}-coordinating residues in PEPC, PK, and PPDK are in identical positions (Figure 11.7) and align with those in DDGA (Figure 11.4). The mode of binding of the α-ketoacid moiety of the co-crystallized ligands to Mg^{2+} is essentially the same in DDGA, PK and PPDK (Figure 11.6). PEPC, PPDK, and PK, together with PEPM, also share the ligand PEP as their substrate or product. Whereas PK uses one ADP molecule and PPDK uses one ATP molecule in the course of the reaction, PPDC does not have any second substrate.

11.6 Comparison with class 2 aldolases and THF-binding enzymes

Apart from 2-dehydro-2-deoxy-galactarate aldolase, structures of four class 2 aldolases, all from *E. coli*, are available. These include fructose-1,6-bisphosphate aldolase (Hall *et al.*, 1999; Plater *et al.*, 1999) and tagatose-1,6-bisphosphate aldolase (Hall *et al.*, 2002), both adopting a $(\beta\alpha)_8$-fold. Moreover, there are L-fuculose-1-phosphate aldolase (Dreyer and Schulz, 1996) and L-ribulose-phosphate 4-epimerase (Luo *et al.*, 2001). As seen in the SCOP database, the latter two enzymes do not adopt the $(\beta\alpha)_8$-fold and belong to the class 2 family in the class 2 superfamily within the α and β (α/β)-fold.

The dimeric fructose-1,6-bisphosphate aldolase and the tetrameric

tagatose-1,6-bisphosphate aldolase are assigned to the class 2 aldolase family in the aldolase superfamily in SCOP. In each of these structures, the catalytic Zn^{2+} metal ion has an environment different from that of the Mg^{2+} coordination site found in KPHMT. However, although the catalysis of all of these class 2 aldolases involves essentially the same chemical mechanism as in KPHMT, comparison of their active sites, overall three-dimensional structures, and functions with the respective properties of KPHMT did not show any significant analogies. None of the class 2 aldolases uses a second substrate that is even remotely similar to 5,10-meTHF. As the class 2 aldolases appear to be spread out across folds and superfamilies, membership in this functional group is not a particularly helpful criterion in understanding structure and function in KPHMT.

The identity of the second substrate or electrophilic donor was also not a basis for classification, as folate binding is not a trait of any existing $(\beta\alpha)_8$ superfamily.

11.7 Conclusions

Comparison of the structure of KPHMT with the structures of PEPM and ICL has shown considerable similarity in the overall folds and, more importantly, convincing similarity in their active sites. The two Mg^{2+}-coordinating carboxylate-containing residues, together with a serine residue that interacts with the ligand, are completely conserved throughout the three enzymes. Additionally, an enolate is an intermediate in each of the reactions catalyzed by the these enzymes. On these grounds, it is proposed that KPHMT should be assigned to the PEP/pyruvate superfamily and should form a family of its own within this superfamily.

However, DDGA, PEPC, PPDK and PK, which are all in the PEP/pyruvate superfamily, show a different active site architecture, which

appears as a distorted mirror image to the one found in KPHMT. It seems difficult to imagine that a circular permutation can explain such a change, and the evolutionary step behind such an alteration could not determined. Based on the latter observation, it is conceivable that there should be a subdivision within the PEP/pyruvate superfamily into two groups; one comprising KPHMT, PEPM and ICL and the other comprising DDGA, PEPC, PPDK and PK. DDGA appears to be a link between the two groups, as it shares significant structural similarity to members of both groups.

Nonetheless, at this stage not much can be infered from this for the mechanism in KPHMT. KPHMT being a member of the PEP/pyruvate superfamily is yet another example of the catalytic proficiency of the enzymes of this particular superfamily and of the functional plasticity of the $(\beta\alpha)_8$-fold in general. Although functionally distinct and mechanistically diverse, members of the PEP/pyruvate superfamily seem to exploit the same chemical strategy, namely, the use of a catalytic Mg^{2+} ion in a very specific way, namely as an electron sink, and to position and orient the substrate, which always resembles an α-ketoacid moiety. However, apart from KPHMT, only PK and PPDK use a second substrate in the reaction.

The enzymes in the PEP/pyruvate superfamily appear to have evolved divergently, while conserving modes of binding to substrates and producing similar reaction intermediates but not maintaining a similarity in their mechanisms, such as proton abstraction. Also, selection for a specific cofactor does not seem to have played any role. The analysis thus appears to stand in disagreement to recent studies by Teichmann *et al.* (2001b) that supported the theory that chemistry and cofactor binding are more likely to be conserved during evolution than substrate binding. However, it is not possible to determine whether any enzyme may have evolved from another one within the PEP/pyruvate superfamily. There may have been a common ancestor which no longer exists. Given that pantothenate is likely to have existed in the prebiotic soup and that the pantothenate pathway is part of primary metabolism, it seems reasonable to argue that KPHMT must be one of the oldest members of this superfamily.

Added note The analysis has now also received independent support by the recent SCOP classification of KPHMT into this superfamily and into a family of its own within this superfamily.

11.8 Implications of the analysis for the evolution of the pathway

The sequence and structure analysis of KPHMT, described above, completes this kind of analysis of the enzymes in the pantothenate pathway. Similar approaches for ADC by Castillo *et al.* (1999), for KPR by Matak-Vinkovic *et al.* (2001), and for PS by von Delft *et al.* (2001) have identified enzymes related to them.

Interestingly, none of the four enzymes in the pantothenate pathway have significant similarities in their active sites or tertiary structure to each other. ADC has a double-ψ fold, while KPR is a two domain protein with an N-terminal NAD(P) binding Rossmann-fold domain, classified in the NAD(P) binding Rossmann-fold domain in SCOP, and an entirely helical C-terminal domain. Although the N-terminal domain of the two domain protein PS, which is classified in the nucleotidylyl transferase superfamily in SCOP, also adopts a (β/α) mononucleotide binding or Rossmann fold, its architecture is very distinct. Consequently, the enzymes in the pantothenate pathway can not have evolved according to the Horowitz theory of retrograde evolution, but appear to have been recruited in a manner that would be consistent with the patchwork theory (see Section 1.4). A respective analysis of the four *E. coli* enzymes of the pantothenate pathway has been described by Lobley *et al.* (2003).

In addition, none of the enzymes in the whole pantothenate and Coenzyme A pathway (see references in Section 2.2) adopts a $(\beta\alpha)_8$- or TIM-

barrel fold, which supports the assumption that KPHMT may have been recruited far away from the immediate pathway neighborhood, across from an already existing pathway.

With the KPHMT structure in hand, FUGUE searches with the amino acid annotated sequences of fully sequenced *A. thaliana* [1] (The Arabidopsis genome initiative, 2000) and *S. cerevisiae*[2] (The Yeast genome directory, 1997) genomes were carried out, jointly with Dr Harald Ottenhof and Hyun Gweon. Essentially, the KPHMT structure was aligned with each annotated protein sequence in the database, and the proteins, whose sequences produced the best alignments, were examined further. The observation of two KPHMT sequences in *A. thaliana* and one in *S. cerevisiae* confirmed the previous observation, based on PSI-BLAST searches, of the existence of KPHMT in these two organisms. The same methodology was also applied to the ADC structure. However, no convincing alignments were found for a potential ADC sequence in these organisms. The latter results provide further support for the absence of ADC in lower eukaryotic organisms and plants. From some of these organisms, alternative ways to supply β-alanine are known (see Subsection 2.1.1).

[1] http://arabidopsis.org/
[2] http://mips.gsf.de/genre/proj/yeast/index.jsp

12 Crystal structure solution of ADC enzymes

12.1 Background

Before I started research on the project, Dr Michael Witty had purified a fraction of unprocessed ADC (pro-ADC) by adding an additional purification step with MonoQ anion-exchange chromatography to the purification scheme as described by Ramjee *et al.* (1997). Drs Diana Matak-Vinkovic and Mladen Vinkovic had obtained crystals from this purified sample and collected X-ray diffraction data. In addition, they crystallized an ADC mutant where Ser25 had been mutated to Ala (S25A), collected X-ray diffraction data and solved the structure. I started where they left the project, refining the pro-ADC and S25A crystal structures and crystallizing other site-directed mutants of ADC. Dr Michael Witty had constructed a set of such site-directed mutants of ADC, had cloned them into pBluescript vector, and could overexpress them in SJ16 *E. coli* cells.

Four of these mutants have a substitution at the cleavage site: Gly24→Ser (G24S), Ser25→Thr (S25T) and Ser25→Cys (S25C). Two are insertion mutants, one with an alanine inserted between Glu23 and Gly24

(A24a-G24b) and one with an alanine inserted between Ser25 and Cys26 (S25a-A25b), in order to analyze the effect of the ψ-loop size. One, a His11→Ala (H11A) mutation, was made to investigate the possible role of this residue from an adjacent subunit in Hα proton abstraction from the ester intermediate. Furthermore, the invariant Arg54 was mutated to a lysine (R54K), the invariant Tyr58 to a phenylalanine (Y58F), and the conserved Lys9 to glutamine (K9Q).

Dr Brent Nabbs, subsequently subcloned all of these mutants into pRSETA vector, that could be used with T7 RNA polymerase-based *E. coli* expression systems, such as BL21 C41 cells. After a suitable overexpression and purification protocol had been devised by Michael Webb and myself, crystallization trials with all those site-directed mutants were pursued.

The crystallization of A24a-G24b and S25C, as well as H11A and S25a-A25b was then continued by Mairi Kilkenny and Carina Lobley, respectively who solved the crystal structures of these mutants. The comparative structural investigation described further in Chapter 13, will also include the analysis of those structures. The major part of the results has also been described by Schmitzberger *et al.* (2003a).

12.2 Overexpression and purification of ADC mutants

Native ADC, on initial purification, comprises a mixture of pro-enzyme, or π-chain, and processed α- and β-chains (Ramjee *et al.*, 1997), because the enzyme self-processes over time. Further chromatography of this purified protein by Dr Michael Witty on MonoQ anion-exchange resin yielded three peaks (Figure 12.1, A), the second and third of which again com-

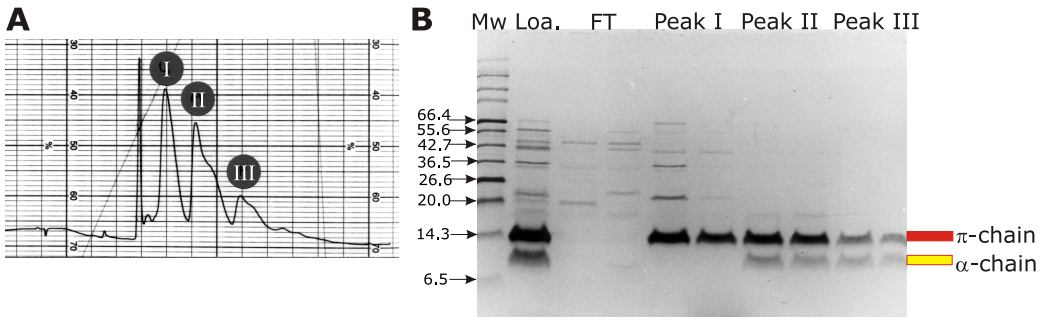

Figure 12.1 (**A**) Chromatography graph of the three separated peaks (I, II and III), representing native ADC in different self-processing states. The optical absorbance at 280 nm was monitored (y-axis), and the x-axis shows the elution volume. (**B**) Corresponding SDS-PAGE of the three chromatography peaks ADC in different self-processing states. The relative positions of the α- and π-chain in this gel are indicated with a yellow and red bar, respectively. Abbreviations: Loa.: loaded native ADC sample; FT: flow-through of the MonoQ anion-exchange chromotography; Peak I: pro-ADC; Peak II: mixture of π-chain and α-chain; Peak III: mixture of π-chain and α-chain. In the last two, the β-chain is present, but cannot be detected on SDS-PAGE. Both figures are courtesy of Dr Michael Witty.

prised a mixture of π-, α-, and β-chains, as observed on SDS-PAGE (Figure 12.1, B). In contrast, the first peak (peak I) was essentially unprocessed ADC, although Western blotting revealed a trace amount of α-chain (Schmitzberger *et al.*, 2003a). Re-chromatography of peak I on MonoQ resin again resulted in the three peaks, showing that the unprocessed enzyme from peak I, over time, converts into the processed enzyme of peaks II and III (Dr Michael Witty, personal communication). This fractionation of ADC was possible, because ADC processed significantly more slowly (Chopra *et al.*, 2002; Ramjee *et al.*, 1997) than any other self-processing systems, described in the literature to date. However, as observed previously for recombinant ADC, after storage at 4°C, the peak I fraction processed further, demonstrating that it was active and not merely a denatured fraction. This fraction (pro-ADC) therefore provided the opportunity to determine the structure of unprocessed ADC, which was nevertheless competent for processing.

When I started on the project, I initially attempted to purify R54K and

S25T from pBluescript vector, as described in Subsection 7.3.1. This purification protocol has been adapted from the protocol for native ADC (Ramjee *et al.*, 1997) and modified by Dr Michael Witty. However, the purification protocol proved difficult to apply for the ADC mutants, was very time- and labour-expensive and did not yield sufficient amounts of pure protein.

As a consequence, for the site-directed histidine-tagged ADC mutants encoded on pRSETA plasmid, a reliable overexpression and purification protocol had to be developed , to allow "high-throughput" preparation of pure protein for crystallization. A standard overexpression protocol as described by Sambrook *et al.* (1989) was found to be sufficient (see Section 7.2), and enabled a high-level of overexpression (see Section 7.2). Jointly with Michael Webb an efficient two-step purification strategy, as described in Subsection 7.3.2, was devised. The purification approach was much faster, more efficient, and less error-prone than the previous purification of non-histidine tagged mutants (Subsection 7.3.2). This purification method allowed purification of up to 80 mg of protein per litre of *E. coli* culture. After analysis of the produced protein samples by mass-spectrometry (data not shown), crystallization could be pursued within four days of the start of the overexpression. Attempts to cleave off the histidine-tag with a thrombin protease resulted in proteolysis of the protein (Michael Webb, personal communication), indicating that the protease cleaved the protein at an internal cleavage site in addition to the thrombin cleavage site at the end of the histidine-tag, and thus was not suitable in our case.

The affinity purification on Ni^{2+}-NTA resin proved to be a very powerful purification step and on its own resulted in a protein sample of high purity (Figure 12.2). Nonetheless, to improve homogeneity and to remove contaminants, purification with gel-filtration was carried out. Superdex 75 resin gave noticeably worse results than Superdex 200, with a trailing peak. However, purification on Superdex 200 resin yielded a homogeneous protein sample (Figure 12.2, E).

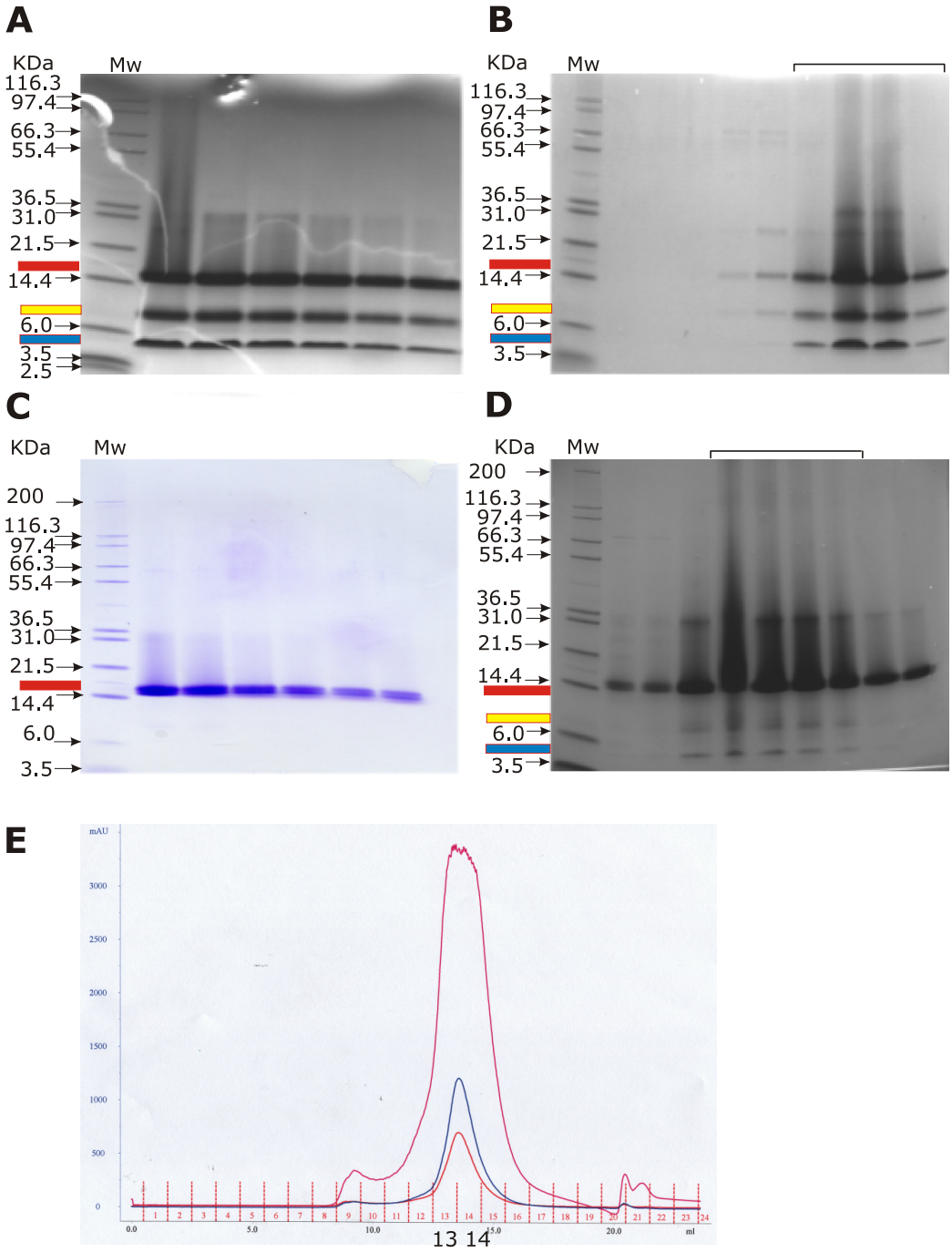

Figure 12.2 (**A**) Representative 10-20% gradient tricine-PAGE for ADC mutants that self-process after purification on Ni-NTA resin and (**B**) after gel-filtration. The relative position of π-, α-, and β-chain in this gel are indicated with a red, yellow, and blue bar, respectively. Fractions, combined for crystallization, are indicated with a black bar above the gel. The faints band above the π-chain represents undissociated ADC multimers. (**C**) Representative 10-20% gradient tricine-PAGE for ADC mutants that do not process, after purification on Ni-NTA resin and (**D**) after gelfiltration. Faint bands coorresponding to traces of β- and α-chains. (**E**) Representative gelfiltration purification graph with the x-axis showing the elution volume in ml, whereas the y-axis shows the optical absorbance for 280nm: blue; for 230nm: red; for 260nm: maroon.

The following ADC mutants were overexpressed and purified successfully: A24a-G24b, G24S, H11A, S25a-A25b, S25C, S25T, R54K, and Y58F. All overexpressed ADC mutant protein samples were predominantly found in the soluble fractions of the crude lysis. In contrast, K9Q, was largely insoluble. Whereas A24a-G24b, G24S, S25a-A25b, and S25T did not show any self-processing activity within the time span of the purification, as judged by tricine-PAGE and mass-spectrometry analysis. H11A, R54K, S25C, and Y58F were capable of forming significant amounts of α- and β-chains in solution, within the time-span of the purification, as judged by the occurrence of π-, α-, and β-chains on the tricine-PAGE and by mass-spectrometry analysis. Representative tricine-gels for the enzymes that do self-process and those that do not self-process are shown in Figure 12.2 C,D and A,B, respectively. However, very faint bands, corresponding to trace amounts of α- and β-chains were detected for all mutants in solution, except S25A-A25b and S25A, on prolonged storage and incubation (Michael Webb, personal communication). There was no apparent difference between ADC mutants that self-processed and those that did not with respect to the migration on gel-filtration (Figure 12.2 E). A small peak before the main peak was characteristic of the purification on Superdex 200 resin (see Figure 12.2). Analysis of tricine-PAGE indicated that similar to the main peak, this peak contained ADC.

Previous experimenters (Dr von Delft, personal communication) reported that the protein is very stable and can be kept at $4^{\circ}\mathrm{C}$ for up to a year and could then still be crystallized. Therefore a storage temperature of $4^{\circ}\mathrm{C}$ was chosen for the protein.

12.3 Crystallization of ADC mutants

For crystallization screening, the protein solution was diluted to protein concentration of 5-9 mg/ml. All crystallization trials were carried out at

19°C with vapour diffusion technique as described in Section 8.1 with a 1:1 ratio of precipitant to protein solution. The crystallization drops had a final volume of 1-3 µl.

Initial crystallization screening for R54K and S25T, without a histidine-tag, resulted in a large number of small crystals, with maximal dimension of 40 µm under several conditions. The most promising conditions appeared to be a precipitant solution of 8-9% (w/v) PEG2000 MME, 10 mM $NaCH_3CO_2$, pH 5.0. However, any attempts to increase the size of the crystals by micro-seeding or macro-seeding were unsuccessful. Seeds were also taken from older S25A crystals, but had no apparent positive effect on the crystal size. Similarly, adding compounds such as dioxane, ethanol, and dimethylsulfoxide did also not lead to any improvement. Most probably, this was a result of the problems with the purity of the protein from the initial purification protocol (Subsection 7.3.1). X-ray screening in-house was attempted, but failed to produce any significant diffraction pattern.

Therefore, after histidine-tagged mutants had been overexpressed and purified, crystallization trials were pursued with those. As soon as possible after the purification, in general within 1-3 days, the ADC mutants were crystallized. Initial screening with $(NH_4)_2SO_4$ as the main precipitant was successful and produced crystals for G24S, S25C, S25T, S25a-A25b, and Y58F.

S25T crystallized in two crystal forms. Initial crystallization trials produced hexagonal prismatic crystals with pyramidal ends at the top and bottom that would form both in 1.2-1.4 M $(NH_4)_2SO_4$, 0.1 M MES pH 6.0 as well as in 1.2 M unbuffered $(NH_4)_2SO_4$ (crystal form A) (see Figure 12.3 A). A different crystal form of bi-pyrimidal hexagonal morphology would grow in 2.4 M sodium malonate, NaOH, pH 4.0 (crystal form B) (see Figure 12.3 B). The latter form had been observed before for native ADC by von Delft (2000). Two crystal forms were also obtained for G24S, although

both would grow in the same condition. Crystals of a similar morphology as crystal form B of S25T would grow in 1.4-1.6 M $(NH_4)_2SO_4$, 0.1 M citric acid pH 4.0. In the same condition, prisms similar to crystal form A of S25T would also grow (G24S crystal form A) (see Figure 12.3 C).

The pH of both crystallization conditions for G24S and S25T was also close to the pH 4.6 at which processed ADC was crystallized (Albert *et al.*, 1998). Crystals would usually grow within seven days to a size of up to $0.6\times0.6\times1$ mm^3 for S25T crystal form B and to about half that size for S25T crystal form A and both crystal forms of G24S. Y58F crystals of the same morphology as crystal form B of S25T grew to dimensions of about $100\times50\times80$ μm^3 in 1.4 M $(NH_4)_2SO_4$, 0.1 M citric acid pH 4.0. However, these crystals were rather unstable and never diffracted to more than 2.9 Å at a synchroton source. A full dataset could not be collected on them.

For cryo-protection the crystals were successively incubated for 3-5 min in mother liquor solutions with 1, 5, 10, 15, 20, and 25% (v/v) glyerol and after being mounted in cryo-loops, they were stored in liquid N_2. Special care had to be taken to follow a step-wise protocol to increase the concentration of glycerol in order to keep crystal mosaicity low and to ensure diffraction to high resolution. S25T crystals of form B were cryoprotected in a similar way by gradually increasing the sodium malonate concentration to 3.2 M, since sodium malonate is a suitable cryoprotectant itself (McPherson, 2001). 3.2 M sodium malonate, NaOH, pH 4.0 was added to the drop with the crystal in small amounts and, eventually, the crystal was transferred to a drop with the final concentration (i.e. sodium malonate at 3.2 M).

Crystals for pro-ADC had previously been obtained by Drs Diana Matak-Vinkovic and Mladen Vinkovic in 1.6 M $(NH_4)_2SO_4$ pH 5.5, and for S25A in 1.5 M $(NH_4)_2SO_4$, 0.15 M Tris/HCl pH 8.0. A24a-G24b and S25C, as well as H11A and S25a-A25b were all crystallized in 1.4-1.6 M $(NH_4)_2SO_4$ pH 4.0, by Mairi Kilkenny and Carina Lobley, respectively.

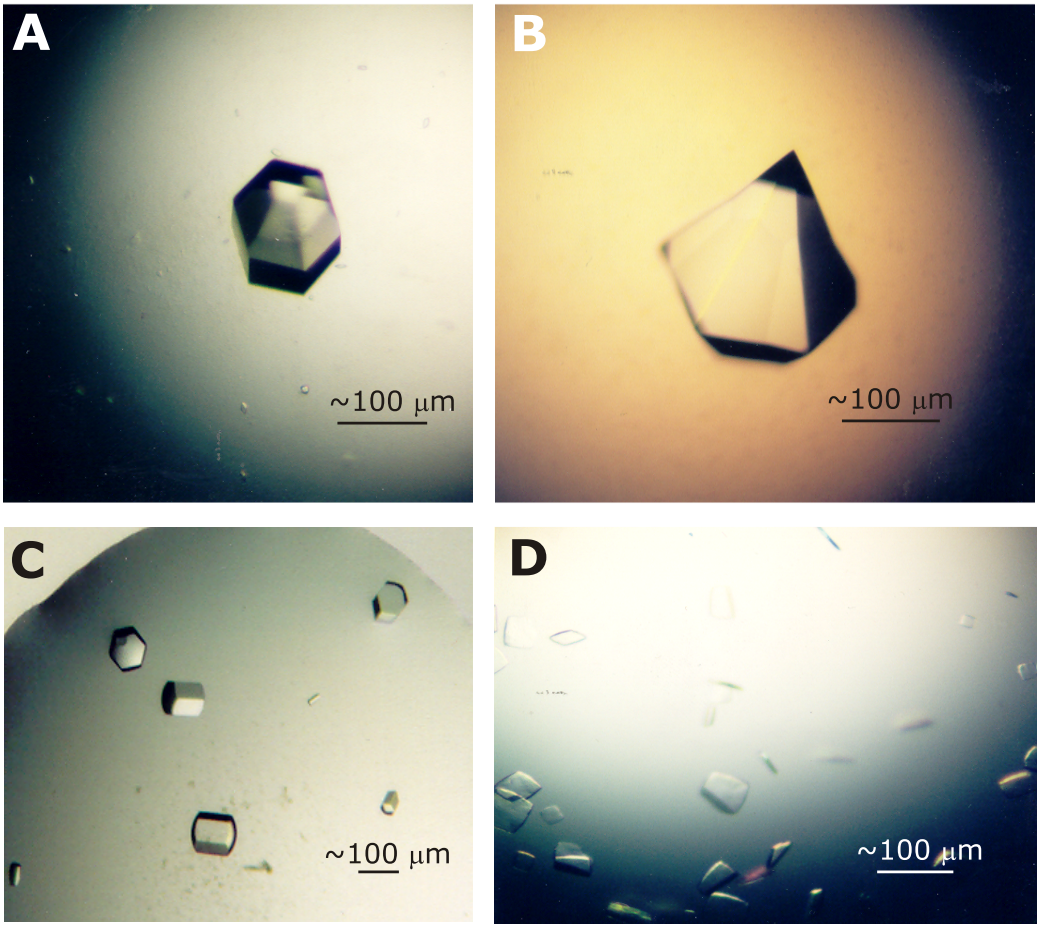

Figure 12.3 Crystal morphology of the crystallized ADC mutants. (**A**) S25T crystal form A (**B**) S25T crystal form B (**C**) G24S crystal form A (**D**) S25C crystals.

Apart from S25T, which was cryo-protected with sodium malonate, all mutant crystals as well as pro-ADC crystals were cryo-protected with the same compound, namely glycerol.

12.4 X-ray diffraction data collection of ADC mutant crystals

The initially obtained crystal form B of S25T diffracted up to 3.2 Å at the in-house X-ray data collection facility and 2.6 Å minimum Bragg spacing at Daresbury SRS. However, the X-ray diffraction pattern was highly mosaic and the X-ray data quality was limited by the low resolution. Scaling of the data indicated that the Bravais lattice was still hexagonal, however unit cell dimensions (see Table 12.1) indicated a different unit cell and a different space group.

Crystal form B of S25T ADC diffracted up to 1.4 Å maximum resolution at the SRS, however at the first synchroton journey X-ray diffraction data to a resolution of 1.56 Å was collected, because of time-limitations. Subsequently, on a second data collection with more time to carefully set up data collection, X-ray diffraction data to 1.26 Å were collected.

In the latter case, X-ray diffraction data were collected with X-rays of a wavelength of 0.978 Å in two steps. To record the high-resolution reflections a data collection with a short detector-to-crystal distance (95 mm) with long X-ray exposure time (16 sec), for good signal-to-noise ratio was chosen. Additionally, the data collection setup was adjusted for a short beamstop to crystal distance, to minimize background scattering of the X-rays. Subsequently, the low resolution data were collected, starting from the same angle as before, with a long detector-to-crystal distance (180 mm) and short X-ray exposure time (5 sec) with the beam stop far away from the crystal to collect low-resolution reflections. For both data collection steps, a total of 100° with 0.5° oscillation angle increments was collected, to achieve high data redundancy.

One problem which arose from the rather large unit cell dimension in one orientation (see Section 12.5), was that reflections started to overlap

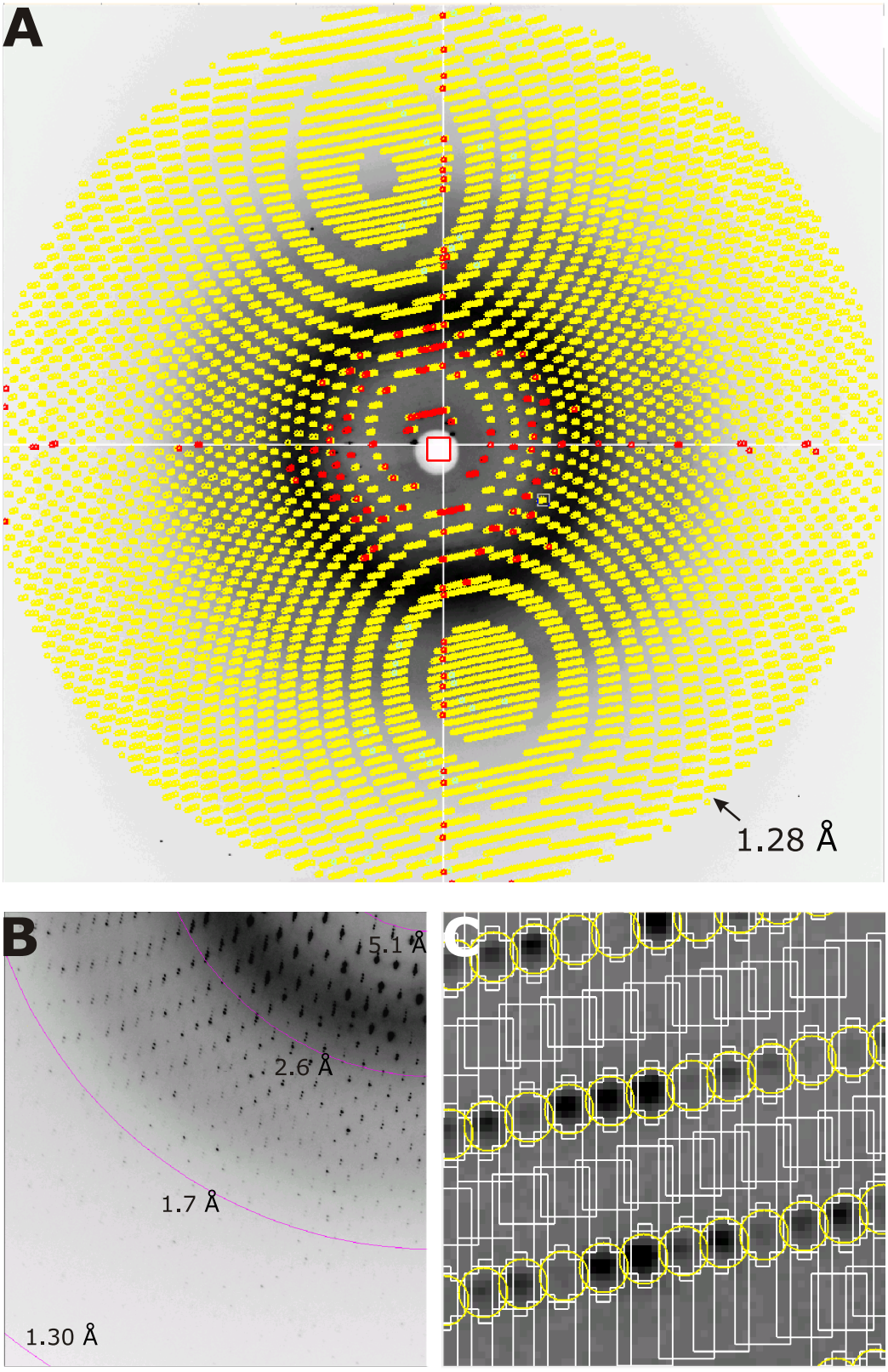

Figure 12.4 (**A**) An auto-indexed X-ray diffraction pattern of S25T. Rejected reflections are shown as red circles and partially recorded reflections are indicted by yellow circles. (**B**) Illustration of the peak separation problem at short detector-to-crystal distance. Particularly, at low resolution the peaks are poorly separated. (**C**) Illustration of how the background box parameter (white rectangular box) was adjusted to enable indexing of the dataset.

Table 12.1 Properties of the different S25T crystal forms

Space group	Unit cell dimensions a (Å)	c (Å)	Molecules per asymmetric unit[a]	Maximum resolution (Å)	Mosaicity (°)	Data collection facility
$P6_3$	102.4	180.7	4	3.20	1.40	In-house
$P6_3$	102.5	180.7	4	2.60	1.20	SRS-14.2
$P6_122$	70.1	215.4	1	1.56	0.25	SRS-14.2
$P6_122$	70.5	215.3	1	1.26	0.45	SRS-14.2

[a] One molecule comprises two ADC subunits, the search model for the molecular replacement

(see Figure 12.4 B). This posed a serious problem for data processing, as overlapping spots cannot be distinguished and are consequently rejected by the processing software. One way to circumvent this, would have been to orient the c-axis with the incident beam, at the risk of a lower completeness of the dataset. For this study, care was taken to find a good initial orientation of the crystal, in which the longest unit cell axis, would be oriented such as to allow high resolution data collection, but not be oriented coincident with the X-ray beam. Secondly, in order to allow indexing and scaling of the data, the background box in the input script for DENZO had to be made narrower, as shown in Figure 12.4 C. This then even enabled indexing of the low-resolution reflections (Figure 12.4 A). However, when scaling the two datasets together, most of the low resolution reflections from the high-resolution data collection were discarded, because they were of lower quality than those from the low-resolution pass. Both datasets combined resulted in almost 100% X-ray diffraction data completeness with high data redundancy. For the X-ray diffraction dataset, used for the refinement, reflections up to a nominal resolution of 1.29 Å were included for scaling.

X-ray data collection of G24S crystals was relatively straightforward. A total of 100° of X-ray diffraction data was collected, with a 0.5° oscillation increment, 10 sec X-ray exposure time, and with X-rays of 1.488 Å wavelength. The crystal form similar in morphology to crystal form B of S25T

diffracted up to 2.0 Å whereas the other crystal form only diffracted to a maximum resolution of 2.6 Å. X-ray diffraction data were collected from both forms, but only the higher resolution data was eventually used, because of the higher data quality. Presumably because of the long wavelength the crystals diffracted to ~2.0 Å, but the X-ray diffraction data was of good quality. Complete data statistics for the X-ray diffraction data collection are shown in Table 12.3.

12.5 Structure solution of ADC mutants

In order to solve the structure of crystal form A of S25T, molecular replacement, as described in Section 8.4, with the previously solved ADC structure (Albert *et al.*, 1998) was carried out. Residues 20-26, comprising the ψ-loop region, were removed in order to reduce the bias from the molecular replacement probe for the subsequent refinement. Four clear rotation and translation solutions for the dimer search probe could be unequivocally determined, locating two tetramers in the asymmetric unit (Table 12.2).

Table 12.2 Rotation and translation function solutions for the Patterson search for S25T ADC[a]

α (°)	β (°)	γ (°)	Tx_{fract}	Ty_{fract}	Tz_{fract}	R-factor	Correlation coefficient
24.00	17.34	23.98	0.4214	0.9357	0.0000	47.2	32.8
10.64	109.91	122.43	0.6901	0.6704	0.9048	44.9	40.6
44.81	93.23	233.65	0.0354	0.4010	0.2398	42.8	46.9
34.36	70.54	302.20	0.7309	0.9541	0.7644	40.7	53.7

[a] The translation functions are listed in the sequence in which they were found. All solutions were clearly distinguishable from the highest background peak

Refinement of the protein structure model was continued until an R_{cryst} of 30 and an R_{free} of 36, at 2.6 Å resolution. Placement of water molecules and further refinement proved difficult because of the low resolution. The

electron density maps of this structure were difficult to interpret, but indicated an unprocessed cleavage site. Higher resolution data was clearly necessary to obtain a clear picture. A better model could eventually be calculated with the X-ray diffraction data, collected from crystal form B at 1.56 Å and 1.29 Å resolution, respectively (Section 12.3).

The indexed, integrated, and scaled X-ray diffraction data from crystal form B as well as from the data of G24S indicated that the protein crystallized in an isomorphous form in the same space group as the previously solved ADC structure. The unit cell dimensions were within 3% of the values of the parameters of the pro-ADC crystals. Consequently, the structure could be solved by directly using the native ADC structure in the refinement with CNS, and to calculate difference Fourier maps. This confirmed that the space group is $P6_122$ with two subunits (subunit A and B) in the asymmetric unit, which were related by the crystallographic two-fold axis to form the tetramer. Similarly to the molecular replacement approach above, residues 20-26, comprising the ψ-loop region, were removed prior to using the structure as a model. The structure of S25T with the 1.29 Å resolution data could not be directly solved with the processed ADC structure as a search model. This may have been due to the fact that the sum of the coordinate differences were too large for convergent refinement with simulated annealing. Therefore, the already partially-refined model derived from the 1.56 Å dataset was used as the initial model, and with this, refinement started to converge.

The crystal structure of S25A was solved in tetragonal space group I422 by Drs Diana-Matak Vinkovic and Mladen Vinkovic. In this case, the crystallographic four-fold axis coincides with the internal four-fold axis of the tetramer. Pro-ADC, as well as the other ADC mutant structures A24a-G24b, H11A, S25a-A25b, and S25C all crystallized in space group $P6_122$.

12.6 Structure refinement and model building of ADC enzymes

Refinement and model building of the four structures, G24S, pro-ADC, S25A, and S25T was carried out as described in Section 8.5 and Section 8.6. Model building proved very difficult in the ψ-loop region preceding the cleavage site. Electron density for the side-chains of residue His21, Tyr22, and Glu23 only appeared at very low contour level, in general below 1.0σ. Even when contoured at 0.6σ the electron density maps were less than clear, indicating strong disorder of these residues in the crystal.

In order to improve the electron density maps several approaches were pursued. Non-crystallographic symmetry averaging of the two subunits in the asymmetric unit in conjunction with solvent flattening (Section 8.5) was applied, but failed to make the electron density maps significantly clearer. In the last cycles of the refinement, TLS groups were defined according to secondary structure elements, as calculated by DSSP and JOY. This improved electron density maps in general but hardly in the disordered region. However, it led to a better data fit of the model, as judged by the decrease in the R_{free} value. Because of the lack of electron density, an alanine instead of the correct amino acid, known from the sequence, was modeled when the electron density was particularly poor. Overall refinement was continued until convergence of the R_{free}.

Characteristic for the histidine-tagged mutants, in subunit A the last 3-5 residues from the histidine-tag, prior to the N-terminal residue in the ADC sequence, could be observed and modeled in the electron density maps. In the previously solved ADC structure (Albert *et al.*, 1998), residues Arg116–Ala126 at the C-terminus of the α-chain were not modelled. However, in subunit A of G24S the mainchain atoms, and, partly, the side-chains, of eight additional residues (Arg116, Thr117, Ala118, Lys119, Ala120, Ile121, Pro122, Val123) at the C-terminus of the protein could be modelled. The

Table 12.3 X-ray diffraction data collection-, refinement-, and stereochemistry- statistics for the final models of the refined crystals structures of pro-ADC, G24S, S25A, and S25T

		Pro-ADC	G24S	S25A	S25T
Data collection					
Space group		$P6_122$	$P6_122$	I422	$P6_122$
Unit cell parameters (Å)	a=b	71.0	71.4	73.1	70.5
	c	216.4	219.1	111.1	215.3
Resolution range (Å)	overall	62-1.95	25-2.0	25-1.95	60-1.29
	highest res. shell[a]	2.01-1.95	2.05-2.00	2.02-1.95	1.30-1.29
No. of unique reflections		24 605	23 414	11 321	81 391
Multiplicity		17.9	30.9	9.3	52.5
R_{sym}[b]	overall	5.9	7.8	4.0	9.7
	highest res. shell[a]	29.0	34.5	32.8	77.0
Average $\langle I \rangle / \sigma \langle I \rangle$	overall	24.5	33.8	36.6	25.7
	highest res. shell[a]	3.2	5.0	4.2	4.4
% completeness	overall	96.4	98.6	95.2	99.0
	highest res. shell[a]	89.3	90.5	92.7	97.9
Estimated mosaicity (°)[c]		0.35	0.27	0.67	0.45
Wilson B ($Å^2$)		24.8	24.9	26.7	15.3
Refinement					
R_{cryst}[d]	overall	15.9	16.5	15.8	15.3
	highest res. shell[a]	19.6	18.7	17.3	23.7
R_{free}	overall	19.5	18.5	20.5	16.6
	highest res. shell[a]	23.8	23.0	20.6	25.1
No. of reflections	working set	22 229	21 839	10 244	76 434
	test set	1183	1180	518	4045
No. of non-H atoms	protein	1793	1876	902	1865
	water	274	237	141	387
	ligand	5	10	0	22
Model quality					
Estimated coordinate error (Å)[e]		0.12	0.13	0.13	0.04
R.m.s.d. bonds (Å)[f]		0.014	0.016	0.017	0.011
R.m.s.d. angles (°)[f]		1.60	1.57	1.56	1.45
No. of residues[g] in	most favoured region	217	225	110	219
	generously allowed region	5	6	5	5
	disallowed region	1	2	0	0

[a] highest resolution shell

[b] $R_{\text{sym}} = \frac{\sum_h \sum_i ||I_{\text{hi}}| - |I_{\text{h}}||}{\sum_h \sum_i |I_{\text{hi}}|}$ (see Equation 8.2)

[c] Calculated with SCALEPACK (Otwinowski, 1991).

[d] $R_{\text{cryst}} = \frac{\sum_h ||F_{\text{o}}| - |F_{\text{c}}||}{\sum_h |F_{\text{o}}|}$ (see Equation 8.3)

[e] Estimated coordinate error (see Section 8.7) of the structure based on the R_{free}, calculated with REFMAC (Murshudov *et al.*, 1997).

[f] R.m.s.d. calculated with WHATCHECK (Hooft *et al.*, 1996a).

[g] Distribution of the number of residues in the Ramachandran plot. ϕ/ψ distributions were calculated with RAMPAGE (Lovell *et al.*, 2003).

final protein model structures were assessed as described in Section 8.7. Crystallographic statistics and Ramachandran diagrams of the main-chain torsion angle distribution for the final models are shown in Table 12.3 and Figure 12.5, respectively. Final $2F_o$-F_c electron density maps of the cleavage site and its immediate environment will be illustrated in Section 13.2.

The three-dimensional coordinates as well as the observed structure factors have been deposited in the PDB-databank under the ID codes 1PPY (pro-ADC), 1PQF (G24S), 1PQE (S25A), and 1PQH (S25T).

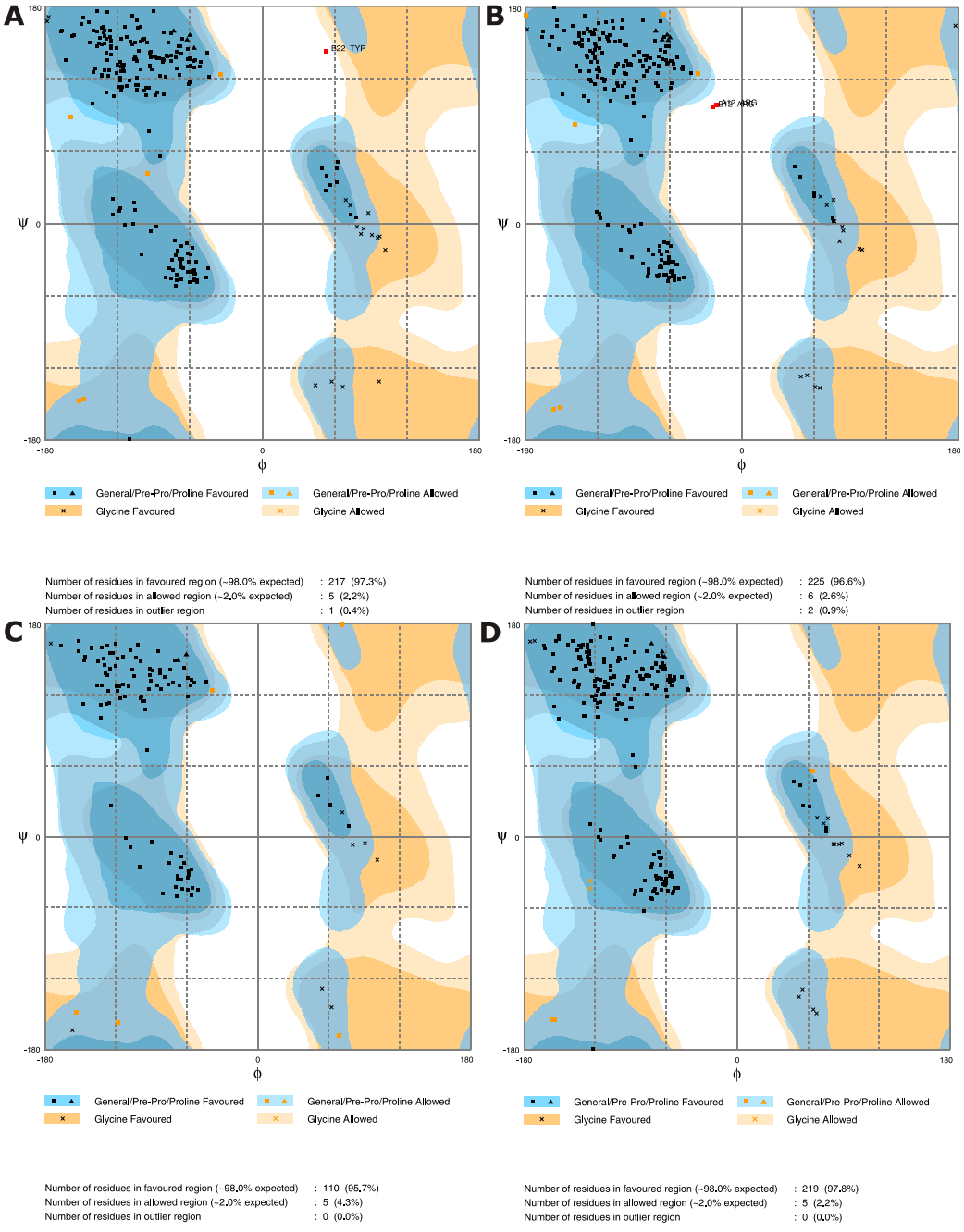

Figure 12.5 Ramachandran diagram of the ϕ/ψ angle distribution. (**A**) pro-ADC. (**B**) G24S. (**C**) S25A. (**D**) S25T. Both axis are in units of °. All diagrams were calculated with RAMPAGE (Lovell *et al.*, 2003).

13 Structural analysis of ADC enzyme crystal structures

13.1 Introduction

Protein self-processing is an important intramolecular post-translational mechanism to activate enzymes, as described in Section 4.2. General and specific features of the self-processing biochemistry have been illustrated in Figure 4.1. A more detailed scheme of the self-processing mechanism in ADC, as it is currently understood, is shown in Figure 13.1. Although no direct evidence has yet been provided for the oxyoxazolidine intermediate, chemical knowledge and experiments in model systems imply the existence of it (Iwai and Ando, 1967).

ADC seemed to be an adequate, and representative model system to study the mechanism of self-processing in more detail by high-resolution crystal structure analysis, because structural information for both the processed cleavage site and the ester intermediate of the enzyme have already been available (Albert *et al.*, 1998). Although the analysis of these structures suggested several residues important for the self-processing catalysis, it could not determine amino acids that could serve as catalysts des-

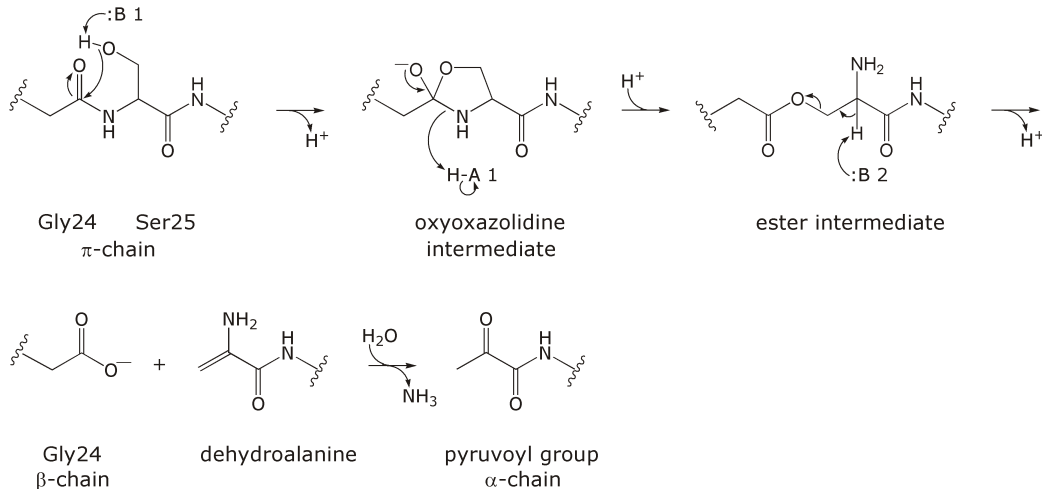

Figure 13.1 Schematic representation of the self-processing reaction, as it is currently understood, applied to ADC. Base 1, acid 1, and base 2 are designated as :B1, H-A 1, and :B2 respectively.

ignated base 1, acid 1, and base 2 in the mechanistic scheme Figure 13.1. Consequently, one of the main objectives of this study was to attempt to identify such functional residues. Another purpose of this study was to investigate the roles of the evolutionarily conserved cleavage site residues, Gly24 and Ser25, in self-processing by crystallographic studies of site-directed mutants.

This study represents the first comprehensive structural investigation of a cleavage site of a self-processing enzyme. Importantly, the first native, cleavage site of a self-processing enzyme that has not been mutated will be described, along with a detailed structural comparison with the solved ADC mutant structures, mentioned in Section 12.6.

13.2 Structural features of pro-ADC and ADC mutants

Pro-ADC The global model of the crystal structure of pro-ADC is basically identical to the processed ADC structure solved by Albert *et al.* (1998), apart from the significant differences at and around the cleavage site. This also applies to the structure models of all the ADC mutants discussed below (see Table 13.1). Subunits A and B of processed ADC and that of pro-ADC superpose with a main-chain r.m.s.d. of 0.22 Å and 0.19 Å, respectively, for residues 1–16, 26–71, and 77–115. Care had to be taken to generate accurate and precise superpositions to obtain realistic estimates of the global and local differences. Therefore, residues 16–26 around the ψ-loop region as well as residues 72–77 around the 3_{10} helical region were omitted for producing superposed coordinates. This superposition procedure was also followed for comparison of the ADC mutant structures described below and in Table 13.1. Obviously, this procedure also had to take into account the 3-5 residues that were modelled at the N-terminus of the histidine-tagged mutants. Once a suitable global structural superposition had been acquired in this way, r.m.s.ds for individual residues could be calculated and compared. The two subunits of pro-ADC in the asymmetric unit superpose with a Cα atom r.m.s.d. of 0.29 Å for superposed coordinates of residues 1–16 and 26–71 and 77–115, indicating that they are virtually identical in this region.

Pro-ADC is able to process (Section 12.2), but the model for the crystal structure showed an unprocessed cleavage site. $2F_o$-F_c Fourier maps showed clear electron density, above 1σ contouring level, for the peptide bond between Gly24 and Ser25 (see Figure 13.2 A) from the start of the refinement. Fitting an ester at partial occupancy, as has previously been carried out with the processed ADC structure (Albert *et al.*, 1998), in a position similar to the ester in apo-ADC was investigated. However, a better fit of the experimental data by doing so could not be obtained, and thus the

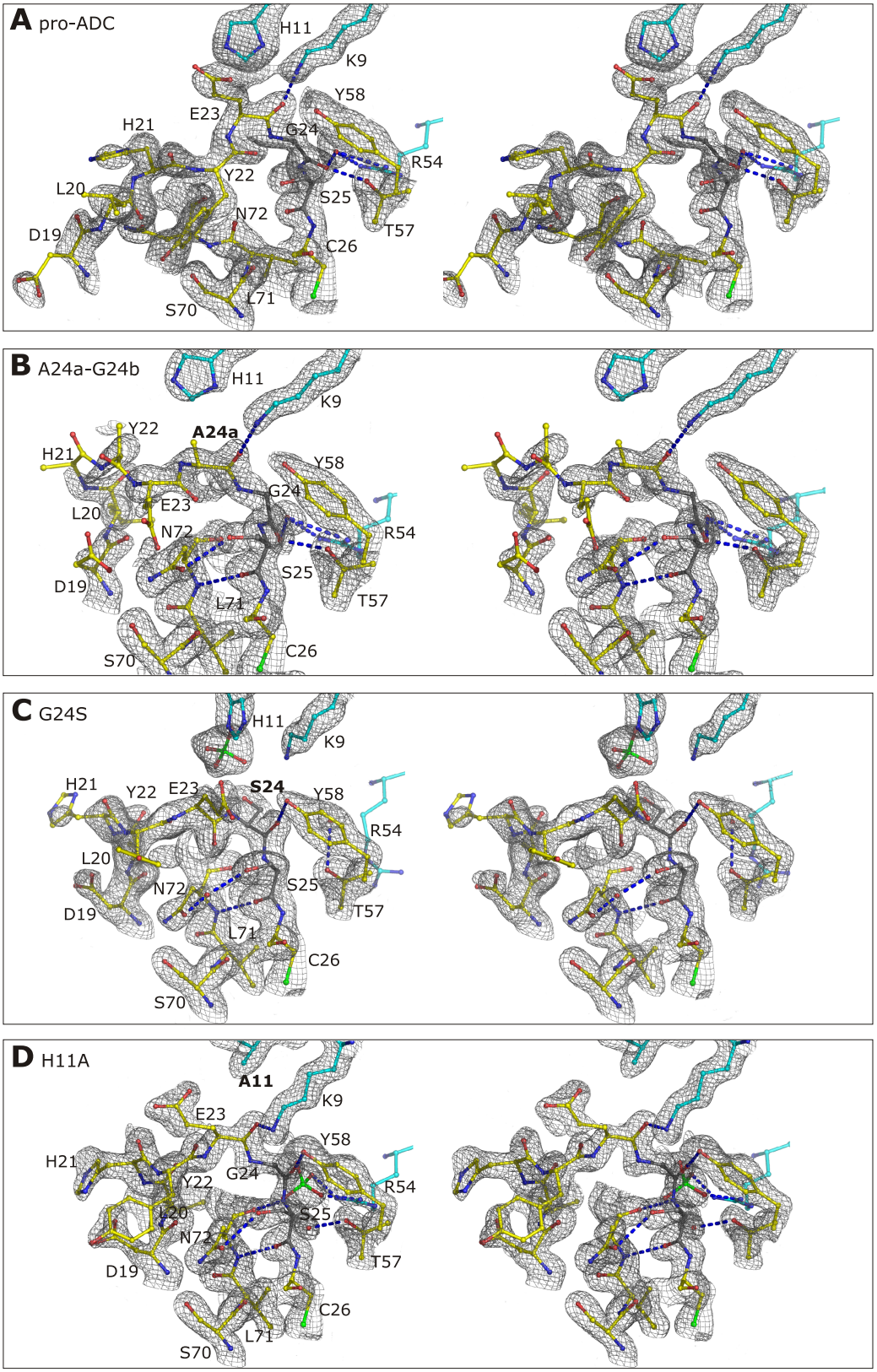

Figure 13.2 Stereo-illustration of the $2F_o$-F_c electron density maps (grey colour), of the self-cleavage site and its environment, contoured at 1.2 σ. All figures are from subunit A in a similar orientation and show Asp19–Cys26, Ser70–Asn72, Thr57, Tyr58 and Lys9, His11, Arg54 from an adjacent subunit. The mutations are shown in bold letters. (**A**) Pro-ADC. (**B**) A24a-G24b. (**C**) G24S (**D**) H11A. See Figure 3.4, Figure 4.5, and Figure 4.3 for colour scheme. All Figures prepared with PYMOL (DeLano, 2002).

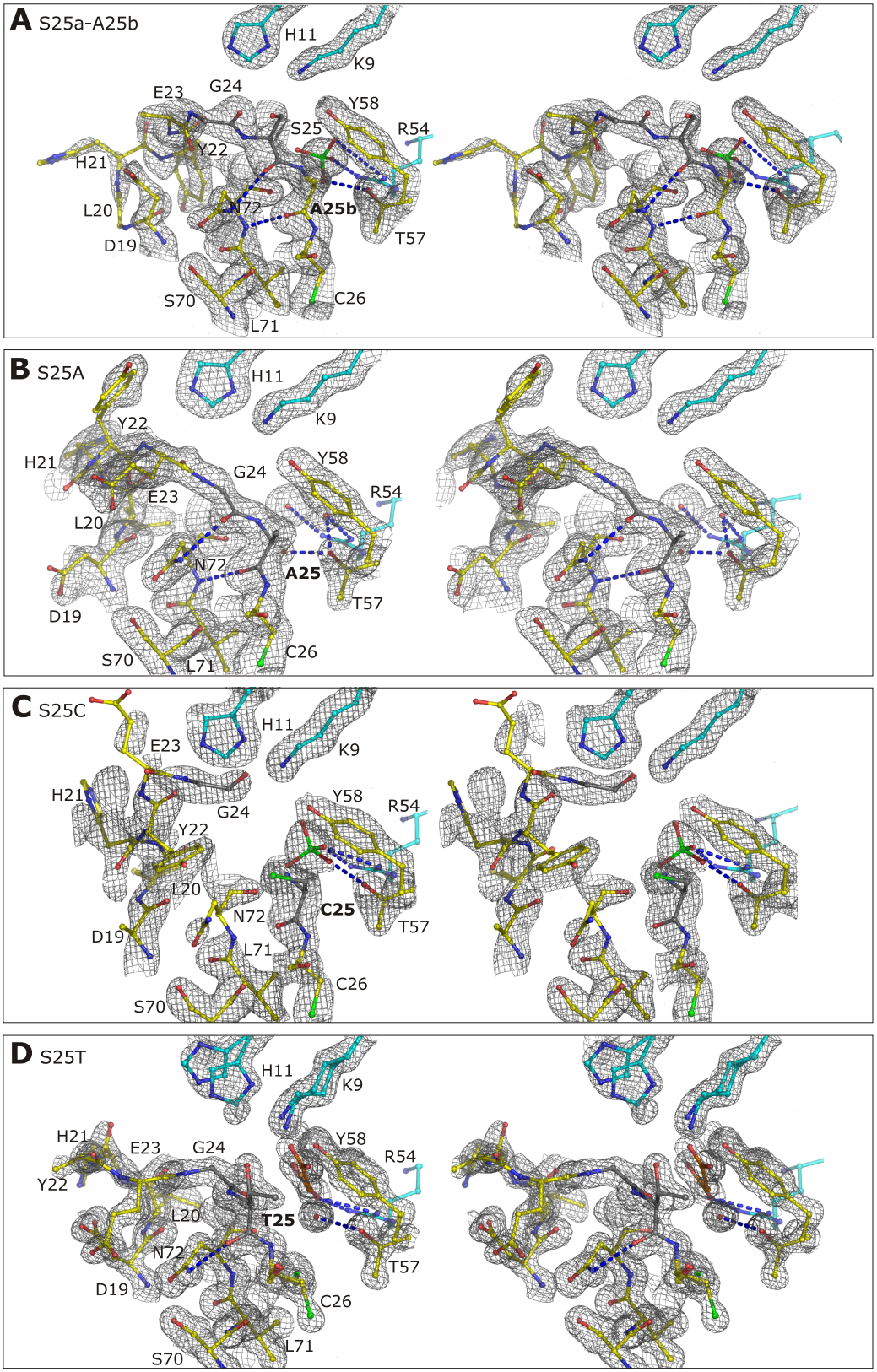

Figure 13.3 Stereo-illustration of the $2F_o$-F_c electron density maps (grey colour), of the self-cleavage site and its environment, contoured at 1.2 σ. All figures are from subunit A, except in S25C, in a similar orientation and show Asp19–Cys26, Ser70–Asn72, Thr57, Tyr58 and Lys9, His11, Arg54 from an adjacent subunit. The mutations are shown in bold letters. (**A**) S25a-A25b. (**B**) S25A. (**C**) S25C. (**D**) S25T. See Figure 3.4, Figure 4.5, and Figure 4.3 for colour scheme. All Figures prepared with PYMOL (DeLano, 2002).

peptide bond was modelled as unprocessed. The structure of the cleavage site, comprising the evolutionarily completely conserved residues Gly24 and Ser25 (Figure 13.4), is almost identical in the two subunits (residues Gly24, Ser25, and Cys26 superpose with a main-chain r.m.s.d. of 0.12 Å). Interestingly, Gly24 makes a kink with a positive ϕ, and adopts a conformation with ϕ/ψ angles of 97/13° and 92/6°, respectively in the two subunits.

Residues His21, Tyr22, and Glu23 of the solvent accessible ψ-loop, preceding Gly24, were disordered in both subunits. The B-factors in this region of the loop are up to more than two times as high as the average (see Table 13.2). There are significant differences in the conformations for the two subunits in this region and also for the highly conserved 3_{10} helical region between Asn72 and Ala76 (Figure 13.5), which lies adjacent to the loop.

The 3_{10} helical region of subunit B of pro-ADC superposes better with both subunits of processed ADC (all atom r.m.s.d. of 0.45 Å and 0.52 Å, respectively for residues Asn 72–His 77) and so subunit B probably presents a more realistic model. Consequently, for the comparison with the other ADC mutant structures this subunit was used as a reference (Table 13.1). However, the side-chains of Asn72 appear disordered in both subunits of pro-ADC. Whereas in subunit B, Asn72 could be modelled in a similar position as in processed ADC, in subunit A it is in a position further away from the Ser25 hydroxyl (see Figure 13.2 A). Instead, the hydroxyl group of Ser25 H-bonds to the main-chain carbonyl of Tyr22. The latter residue, lies in between Asn72 and Ser25, preventing interaction.

G24S Electron density was well defined for the Gly24→Ser mutation at 2σ contouring level in $2F_o$-F_c maps from the start of the refinement. Significant conformational differences of the amide bonds between both Glu23 and Ser24 as well as Ser24 and Ser25 with respect pro-ADC are observ-

Table 13.1 Comparative data of the structural analysis of the ADC enzymes 1

		Pro-ADC	A24a-G24b	G24S	H11A	S25a-A25b	S25A	S25C	S25T
Coordinate differences									
Main-chain r.m.s.d. for residues 1–16, 26–71, 77–115 with subunit B of pro-ADC (Å^2)[a]	A	0.30	0.25	0.36	0.31	0.28	0.30	0.28	0.36
	B	0.00	0.31	0.33	0.23	0.21	-	0.26	0.29
Main-chain r.m.s.d. for residue 24 (Å)[a,b]	A	0.29	0.83	0.60	0.98	2.14	2.15	8.04	3.53
	B	0.00	1.06	0.57	0.97	0.38	-	7.41	3.66
All-atom r.m.s.d. for residue 24 (Å)[a,b]	A	0.44	0.92	-	1.61	-	2.93	-	3.55
	B	0	1.02	-	1.43	-	-	-	3.66
Main-chain r.m.s.d. for residue 25 (Å)[a,b]	A	0.08	0.20	0.39	0.30	0.13	0.17	0.47	3.00
	B	0.00	0.30	0.23	0.17	0.43	-	0.40	3.04
All-atom r.m.s.d. for residue 25(Å)[a,b]	A	0.15	0.31	0.44	0.50	-	-	-	-
	B	0.00	0.39	0.25	0.17	-	-	-	-
ϕ/ψ angles of Thr57 (°)	A	-148/-145	-148/-146	-151/-152	-150/-152	-157/-146	-154/-149	-151/-148	-156/-155
	B	-151/-148	-154/-152	-156/-154	-146/-151	-153/-148/	-	-151/-151	-154/-155
Distance residue 24 O to Thr57 Oγ (Å)	A	2.6	2.6	4.1	4.7	8.6	6.2	9.8	6.1
	B	2.7	2.3	3.6	4.4	7.5	-	9.4	6.1
Distance Oγ of residue 25 to C of residue 24 (Å)	A	4.3	4.2	3.5	3.5	3.7	-	-	3.7
	B	4.1	4.3	3.7	3.5	4.0	-	-	3.7
Distance O/Sγ of residue 25 to Asn76 O/Nδ (Å)	A	10.4	3.0	4.5	3.0	12.7	-	4.2	6.9
	B	3.4	3.0	5.2	4.8	4.8	-	4.0	6.8

[a]Superposition of main-chain atoms of residues 1-16, 26-71, and 77-115 of subunits A and B, respectively with the corresponding residues of subunit B pro-ADC. A and B denote subunits A and B, respectively.
[b]Residues used for the superposition as in footnote **a**. Where possible (where the residue identity is the same and in the same position) an all-atom r.m.s.d. was calculated.

Table 13.2 Comparative data of the structural analysis of the ADC enzymes 2

		Pro-ADC	A24a-G24b	G24S	H11A	S25a-A25b	S25A	S25C	S25T
B-factor differences									
Average B-factor (Å^2) (all non-H atoms)[a]	protein	35	33	33	19	31	35	39	19
	water	44	42	38	19	31	35	39	19
Average B-factor for His21 (Å^2) (all non-H atoms)[a]	A	71	87	86	42	75	79	90	54
	B	112	80	99	62	92	-	58	56
Average B-factor for Tyr22 (Å^2) (all non-H atoms)[a]	A	45	95	74	35	60	95	114	63
	B	94	56	69	62	75	-	63	72
Average B-factor for Glu23 (Å^2) (all non-H atoms) [a]	A	59	95	56	29	56	75	129	37
	B	90	56	49	52	76	-	53	54

[a] Average B-factor includes the contribution from the TLS group as calculated with TLSANL (Howlin *et al.*, 1993).

able. The Ser24 carbonyl H-bonds with the side-chain hydroxyl of Tyr58 (Figure 13.2 C). The Glu23 carbonyl points in the opposite direction as it does in the pro-structure (Figure 13.6 A) and H-bonds to Ser25 Oγ, instead (Figure 13.2 C). Ser25 is in the generously allowed region of the Ramachandran plot with a mean ϕ-angle of 179° and a mean ψ-angle of 173°. The distance of the Nδ of Asn72 to the Oγ of Ser25 is larger than in pro-ADC (Table 13.1) and a water molecule was found in between the two residues. The two Ramachandran outliers (see Figure 12.5 B) are Arg16 from each of the subunits, the reason for which is presumably a sulphate molecule in the active site (see Figure 13.2 C). Although the sidechains of residues 21-23 are disordered (Table 13.2), the conformations for the modelled residues of the ψ-loop are very similar in between the two subunits.

S25A Similarly to the Gly24→Ser mutation, the Ser25→Ala mutation was clearly visible at 2σ contouring level, in $2F_o$-F_c electron density maps from the start of the refinement. The amino acid substitution resulted in some minor, but important, conformational changes. The carbonyl group of Gly24 H-bonds with the Nδ of Asn72 (distance 3.05 Å) (Figure 13.3 B). The conformation of Gly24 is now a more extended one, without the kink observed in the pro-ADC structure (Figure 13.6 A). In contrast, the conformation and position of the Ala25 are identical to that of Ser25 in pro-ADC (Table 13.1). Whereas residues in the 3_{10} helix and Asn72 are in identical positions as in subunit B of pro-ADC, there are differences in the coordinates for residues 19–24. Electron density for the ψ-loop region preceding the active site is better defined than for pro-ADC, except for the side-chains of His21, Tyr22, and Glu23 (Figure 13.3 B).

S25T The near-atomic resolution $2F_o$-F_c electron density map of the S25T ADC mutant showed the Ser25→Thr mutation clearly at 2σ contouring level in subunit A from the start of the refinement. However, in subunit B the side-chain is disordered, to a certain extent, and B-factors between 42

Table 13.3 Summary of self-processing activity (in solution) and structural features of the cleavage site of pro-ADC and the seven ADC mutants

Enzyme	Processing activity	Structural features
Pro-ADC	Yes	Apart from the differences in residues 17–25 and 72–77 the structure is essentially identical to the processed ADC structure (Albert *et al.*, 1998). Gly24 has ϕ/ψ angles of 97/13° and 92/6° in subunits A and B. Residues 21–23 in the ψ-loop preceding Gly24 are disordered. There are conformational differences in this region in the two subunits and also for the highly conserved 3_{10} helical region between Asn72 and Ala76, which lies adjacent to the loop.
A24a-G24b	Traces	The conformation of the Gly24b-Ser25 peptide bond is very similar to that in pro-ADC. However, there are significant differences in the ψ-loop.
G24S	Traces	The conformations of Glu23-Ser24 and Ser24-Ser25 peptide bonds are different from those in pro-ADC. The Gly24 carbonyl does not H-bond to Thr57.
H11A	Yes	The protein that crystallized was essentially the π-chain. The unprocessed cleavage site is shifted towards Ala11. Apart from the difference of the Gly24 carbonyl the structure is very similar to pro-ADC.
S25a-25b	No	S25a carbonyl in subunit A is rotated by $\sim$180° from its position in pro-ADC. Ala25b is in the same conformation and position as Ser25 in pro-ADC and Ala25 in S25A.
S25A	No	Gly24 carbonyl is rotated by $\sim$180° from its position in pro-ADC. Conformation and position of Ala25 are identical to that of Ser25 in pro-ADC.
S25C	Yes	The protein that crystallized was essentially the α- and β-chains. There is a free Cys25, instead of the pyruvoyl group, at the N-terminus of the α-chain.
S25T	Traces	The Gly24-Thr25 peptide bond is in a different position and conformation from that in the cleavage site in pro-ADC. There are no H-bonds between Gly24 carbonyl and Thr57 nor between Thr25 O and Asn72 N.

and 50 Å^2 for the side-chain atoms indicate possible rotational flexibility around the Cβ.

The S25T mutation caused the most dramatic change in Cα coordinates of all ADC mutants. The whole loop from Leu20 until Cys26 is in a position shifted towards the 3_{10} helix, relative to its position in pro-ADC (Figure 13.3 D). The Gly24 and Thr25 residues do not superpose with Gly24 and Ser25 in the pro-ADC structure (Figure 13.6 B). Also there are no H-bonds between Gly24 carbonyl and Thr57 Oγ nor between Thr25 Oγ and Asn72 Nδ. Cys26 appears to be forced out of its normal environment, adopting two different, alternate conformations. Also the S25T mutation causes a kink between residues Leu20 and Glu23, which are very tightly packed, so that the Asp19 Oδ H-bonds to the mainchain oxygen of Glu23. A malonic acid residue, H-bonding to Arg54, occupies the region that becomes the active site after self-processing (Figure 13.3 D).

A detailed description of the structural features of A24a-G24b (PDB-ID: 1PYQ), H11A (PDB-ID: 1PT1), S25a-A25b (PDB-ID: 1PTO), and S25C (PDB-ID: 1PYU) will be given elsewhere by Mairi Kilkenny and Carina Lobley. However, the respective electron density maps of the final models of A24a-S25b, H11A, S25C, and S25a-A25b are shown in Figure 13.2 and Figure 13.3. These will be relevant for the subsequent comparative analysis that was carried out and for the discussion of the results that follows below. A summary of the relevant structural features of the cleavage site that were analyzed in this study, together with processing activity in solution[1], of all ADC structures is given in Table 13.3.

[1]Information for self-processing activity, on prolonged storage, was provided by Michael Webb.

13.3 Interactions at the cleavage site

Specific H-bond interactions at the cleavage site have been identified in mutated precursors of the proteasome (Ditzel *et al.*, 1998) and glycosylasparaginase (Xu *et al.*, 1999). These were assumed to guarantee optimal conformation of the cleavage site and assist the self-processing catalysis by stabilization of reaction intermediates.

In this study, comparative analysis of the various ADC structures shows that residues in positions 24 and 25 in the two ADC subunits in the asymmetric unit superpose very well, with the exception of those in S25a-A25b, where two conformations are seen for the Ser25a carbonyl in the two subunits. However, the coincidence in position of residues 24 and 25 in the two subunits is in contrast to residues 21-23 in the ψ-loop preceding the cleavage site. The following interactions appear to be important for stabilizing the conformation of the Gly24-Ser25 peptide bonds (Figure 13.2 A):

- strong H-bonds of the Gly24 carbonyl to the Oγ of the invariant Thr57 (see Figure 13.4) and to a water molecule;
- an H-bond between the Ser25 carbonyl and the main-chain nitrogen of Asn72;
- an H-bond between the Glu23 main-chain carbonyl oxygen and the Nζ of the conserved Lys9 (see Figure 13.4) from an adjacent subunit.

The H-bond to the Thr57 Oγ appears to be important for orienting the carbonyl, and additionally may polarize the Gly24 carbonyl bond. Together with the well-ordered water molecule (Figure 13.2), Thr57 forms an oxyanion hole that is likely to stabilize the developing negative charge on the exo-cyclic oxygen (see Figure 13.1). The water molecule, whose

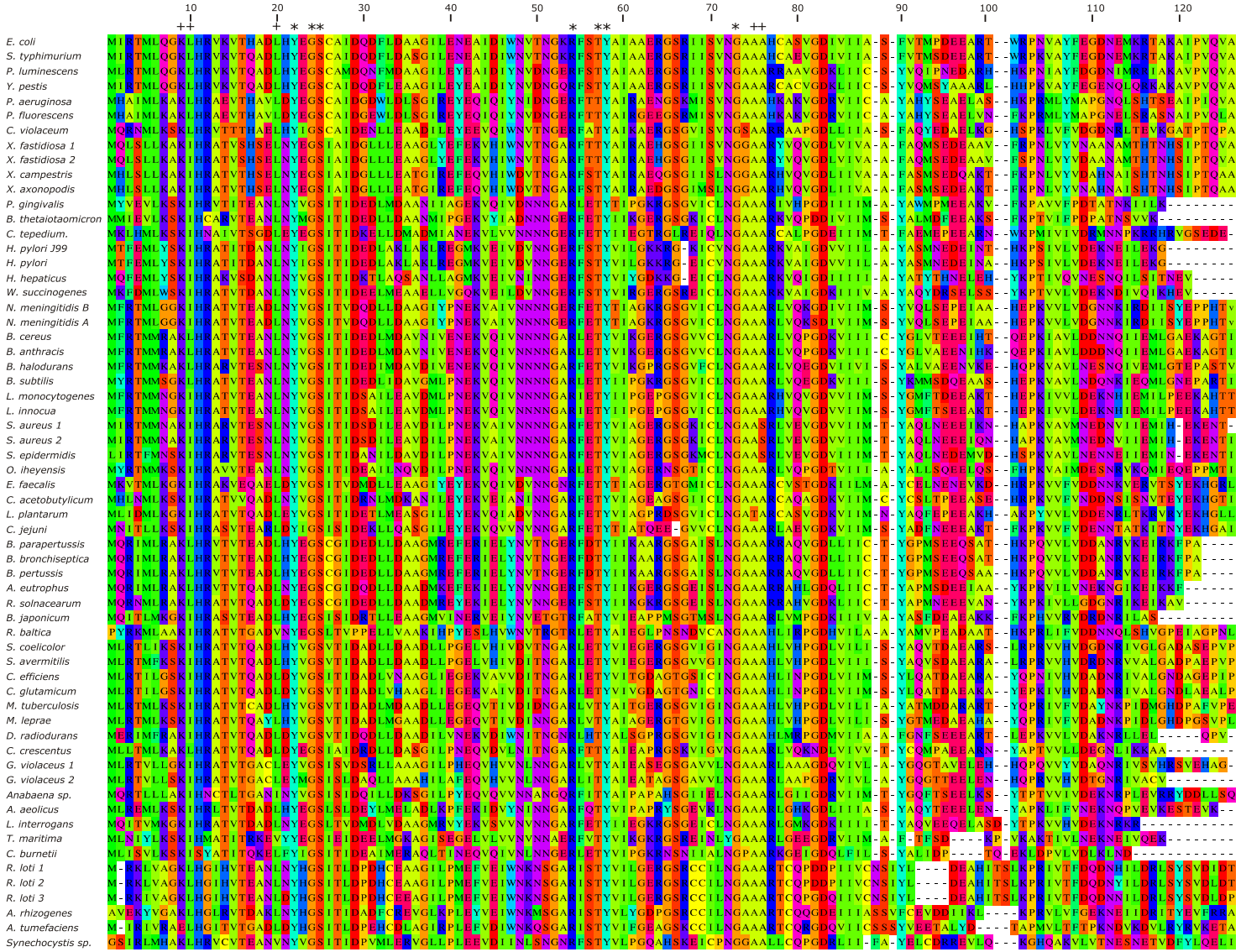

Figure 13.4 Representative amino acid sequence alignment of ADC orthologs. The alignment is numbered according to the full-length sequence of *E. coli* ADC. Several amino acid sequences have been truncated at the N- and C-termini. Note the invariant Gly24-Ser25 amino acid sequence motif. Invariant residues are indicated with a star and conserved residues with a +.

position is stabilized by two H-bonds to the Nη1 and Nη2 of the invariant Arg54 (Figure 13.4), is at a distance of 2.60 and 2.66 Å , for subunit A and B, respectively from the Gly24 carbonyl oxygen in the pro-ADC structure. It is consequently, at an equal distance from the oxygen as the Thr57 Oγ. This water molecule is also in the exactly same place in the processed ADC structure, and the crystal structures of the A24a-G24b and S25A ADC mutants. A sulfate molecule is located in the same position as the water molecule in the H11A, S25a-A25b and S25C structures, whereas in S25T a malonic acid residue can be found in this locus.

In all ADC structures, including all ADC mutants, pro-ADC and the processed ADC structure, Thr57 is in the same, generously allowed, region of the Ramachandran plot (see Table 13.1 and Figure 12.5), as is often found in catalytically important residues. Its conformation and position is maintained unchanged before and after the self-processing, as indicated by the structures of processed ADC (Albert *et al.*, 1998) and pro-ADC.

Asn72 is almost completely conserved (see Figure 13.4) and only in the thermophilic bacterium *T. maritima* is it a tyrosine. Although a closer look at the alignment shows that there is asparagine in the *T. maritima* sequence just two residues before the other aligned asparagine residue. Asn72 is followed by a sequence of conserved residues forming a 3_{10} helix (see Figure 13.5), whose role appears to be mainly structural and could serve to restrict the possible conformations of the ψ-loop. Disorder in the ψ-loop region in some ADC structures appears to be correlated with disorder in the 3_{10} helical region, which may be due to H-bond interactions.

13.4 Conformational constraints

Detailed knowledge of the differences in the conformations of the cleavage site in the various structures can be helpful in understanding the relevance

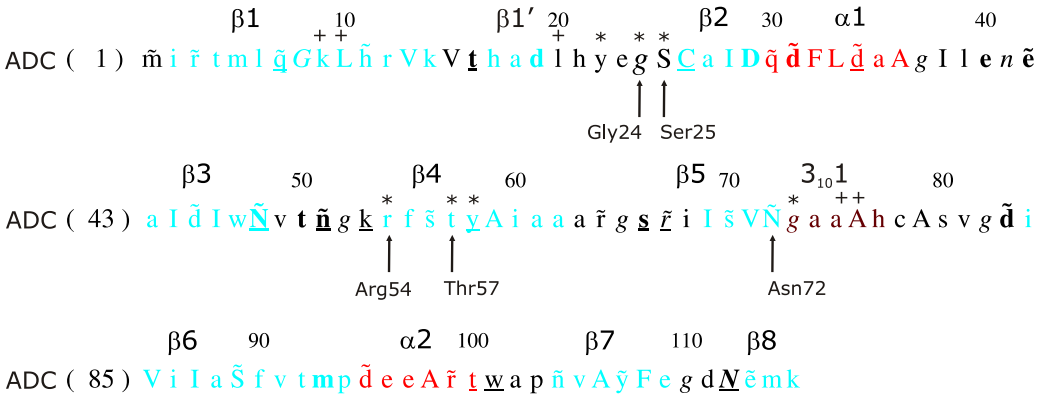

Figure 13.5 Structure-annotated amino acid sequence of ADC in JOY-format (Mizuguchi *et al.*, 1998a). See legend to Figure 11.4 for a key to the JOY-format.

of the conformation for self-processing. Because all ADC structures are globally isomorphous (Table 13.1) this model system could be particularly well suited for studying such conformational differences. Apart from the crystallization of S25T, the same precipitants for crystallization and the same cryoprotection conditions were used for all mutants. Of the residues around the cleavage site, only His11 is likely to have a pK_a between 4 and 8 , and consequently no major pH-dependent conformational changes can be expected in this region, in the pH range in which pro-ADC and the ADC mutants were crystallized (Section 12.3). Therefore, it is likely that the differences between the structures are predominantly due to the changes in the amino acid sequences and the corresponding differences in interactions in the protein.

In pro-ADC, Gly24 of the unprocessed cleavage site has a positive ϕ (Table 13.3) and is in a left-handed α-helical conformation. Any other residue in this position with the same conformation as this Gly would almost certainly introduce steric clashes. Consequently, in the structure of the inactive G24S mutant, the Ser24 carbonyl is in a different conformation (Figure 13.6 A), positioned away from Thr57.

However, introducing a residue before Gly24, as in A24a-G24b, does not

alter the conformation of the cleavage site significantly (Figure 13.6 B), and apparently leaves the geometry as well as the interactions of the cleavage site, observed in pro-ADC, intact (Figure 13.2 B). Nevertheless, this mutant interestingly does not process. As a consequence, in addition to the interactions and geometry of the cleavage site other factors appear to be influential for processing. Strain in the loop has been assumed to contribute to the self-processing reaction in pyruvoyl-dependent HisDC (Gallagher *et al.*, 1993), GyrA intein splicing (Klabunde *et al.*, 1998) and glycosylasparaginase (Xu *et al.*, 1999). In the latter case significant deviations in the ω- (the dihedral angle around the rotation of C-N bond) and τ-angles (the angle between N-C_α and C_α-C) of the peptide bond of the cleavage site were observed. In pro-ADC no evidence for strain in the form of significant deviations in ω- and τ-angles from the mean ideal values (Table 13.4), and/or unusual peptide conformations in the peptide bonds of Glu23-Gly24 and Gly24-Ser25 could be observed.

Table 13.4 ω- and τ-angles for Glu23 and Gly24 in subunits A and B of pro-ADC[a]

Residue		ω (°)	τ (°)
Glu23	A	177.9	114.9
	B	182.0	115.5
Gly24	A	181.7	114.7
	B	184.8	112.9

[a]The values were calculated with DANG (Rossmann *et al.*, 2001, p.729-731).

Nonetheless, in the eight structures of ADC presented here, side-chains of His21, Tyr22 and Glu23 in the ψ-loop are most often disordered (Figure 13.2 and Figure 13.3) and/or found in different conformations in the two subunits, even though the position of Gly24-Ser25 is constrained. Making the loop longer as in A24a-G24b and also in S25a-A25b lead to increased conformational heterogeneity, as indicated by the very different main-chain positions the residues in the ψ-loop adopt in the two subunits. Presumably, the added residue in A24a-G24b renders the loop incapable of stabilizing reaction intermediates of the self-processing reaction. Whereas a single defined conformation of the loop itself, does not seem to be essen-

tial, the results suggest that a certain degree of conformational flexibility may be an important factor in allowing the processing to proceed.

The second insertion mutant S25a-A25b, like A24a-G24b, is also inactive. However, in this mutant the cleavage site is moved out of its normal environment. Ala25b is in exactly the same place as Ser25 in pro-ADC and Ala25 in S25A (Figure 13.6 A). Significantly, removal of the side-chain hydroxyl of Ser25 in pro-ADC, as in both the inactive S25A mutant and, effectively also in the S25a-A25b mutant causes a change in the conformation of the Gly24 and S25a carbonyl, respectively. In both structures the carbonyl is rotated $\sim$180° from its position in pro-ADC (Figure 13.6 A), H-bonding to Asn72 instead of Thr57. This conformational change results in a more extended conformation of Gly24. If the Ser25 hydroxyl was present, as in pro-ADC, steric hindrance with the Gly24 carbonyl would most likely prevent such a conformation.

The more bulky side-chain of a threonine in the position of Ser25, as in the inactive S25T mutant, appears to force Gly24, Thr25, and Cys26 out of their normal environment and further to the left from the perspective in Figure 13.3 D. There are no H-bonds between the Gly24 carbonyl and the Thr57 Oγ. The structural differences in the mutant represent the most pronounced change in the position of the cleavage site in all mutant structures. However, it cannot be entirely excluded that the conformational change in S25T is partly caused by the malonic acid residue. This residue is located in the area that becomes the active site upon cleavage (see Figure 13.3 D).

In the structure of the active but unprocessed, H11A mutant, which is able to self-process, the conformation of the cleavage site is similar to that of pro-ADC, with the exception of the Gly24 carbonyl, which H-bonds to Oζ of Tyr58 instead of the Oγ of Thr57. The removal of the imidazole group created a cavity that is occupied by Glu23, whose Cα position is 1.75 Å further towards residue 11 compared with the position in pro-ADC.

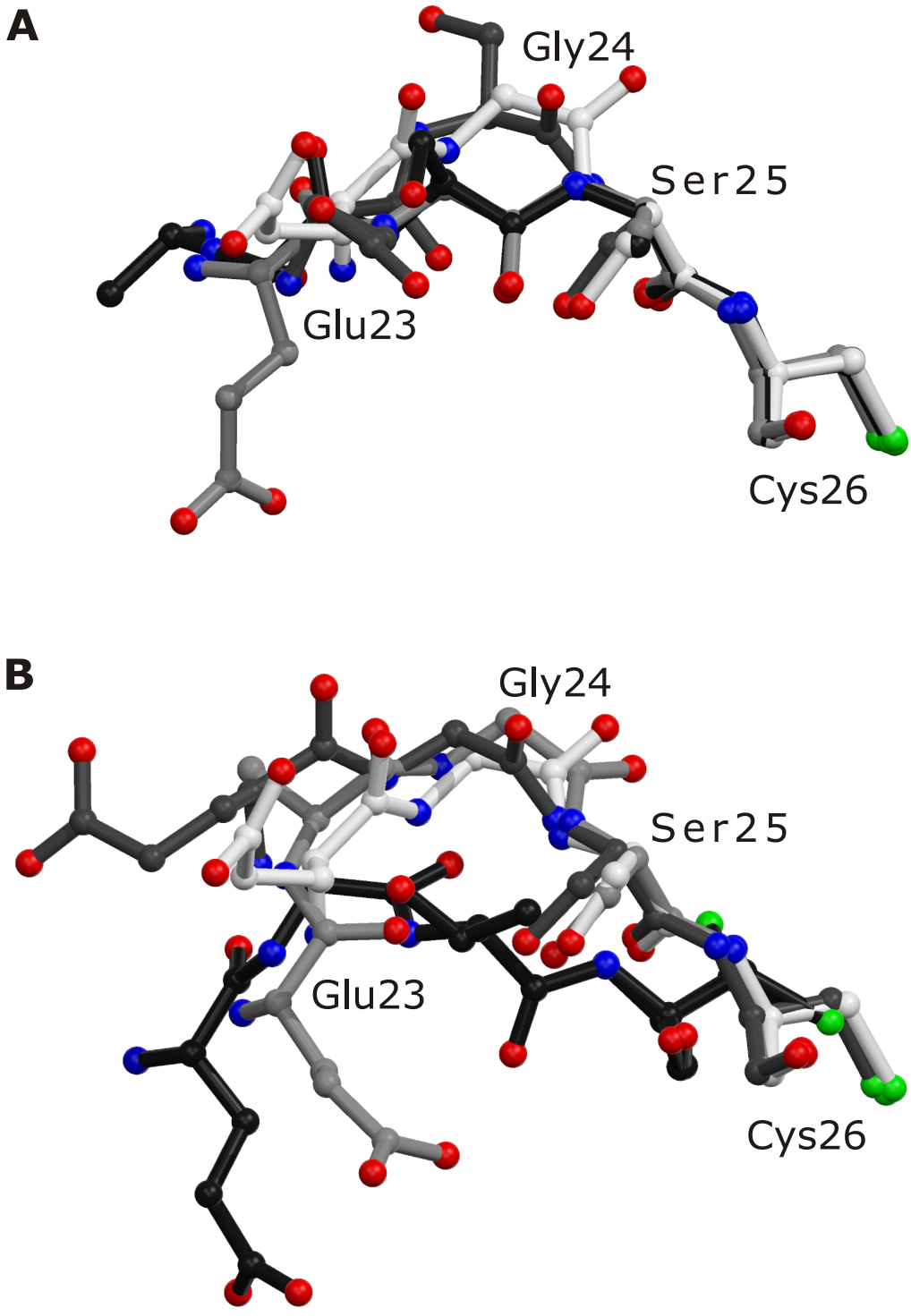

Figure 13.6 Illustration of the superposition of the three-dimensional coordinates of Glu23 to Cys26 of pro-ADC with the respective residues in the ADC mutants (superposition procedure as in the legend to Table 13.1). From all structures subunit A was used, except from pro-ADC. Colour scheme, in general, as before (Figure 3.4 and Figure 4.5); additional colour scheme as follows: (**A**) Carbon: pro-ADC, white; G24S, dark grey; S25a-A25b, black; S25A, light-grey. Note the almost identical position of the S25a and Gly24 carbonyl in S25a-A25b and S25A, respectively. (**B**) Carbon: pro-ADC, white; A24a-G24b, light-grey; H11A, dark grey; S25T, black. Note the shift in the cleavage site in H11A, relative to pro-ADC.

Glu23 is occupying the space of the removed imidazole ring. However, Gly24 is H-bonding to the Tyr58 (mean distance: 2.97 Å) instead of H-bonding to Thr57. A sulfate and a water molecule are occupying the space between Gly24 and Thr57. Nevertheless, the transition state for processing of this mutant could be similar to that for pro-ADC with the carbonyl of Gly24 H-bonded to Thr57. This could explain the normal self-processing activity in this mutant. Such a conformation appears to be readily accessible from the observed ground state structure without major rearrangement.

Other experimenters (Ditzel *et al.*, 1998; Tolbert *et al.*, 2003a,b) had before modelled the nucleophile (Ser or Thr) in place of the substituted amino acid and had based their hypotheses about self-processing on such structures. However, this analysis clearly demonstrates that deducing the conformation and geometry of the cleavage site from the structures of the inactive S25A and S25T mutants, by computationally modelling a serine in place of alanine or threonine, would have most likely led to incorrect models for the native, unprocessed peptide bond. As a consequence, the oxyanion hole would, presumably, not have been identified. Even very modest changes in the active site, as in H11A, can result in a different observed ground state of the protein. The differences in the structures indicate that small but important changes apparently account for the loss of processing activity. Tight conformational constraints, including H-bonds and steric effects, define the optimal geometry of the cleavage site. Seemingly conservative substitutions can lead to significant conformational changes, to which self-processing in ADC is very sensitive.

13.5 Molecular mechanism of self-processing in ADC

Based on the structure of the native cleavage site and knowledge of the three dimensional coordinates for the ester as well as the pyruvoyl group (Albert *et al.*, 2000), a mechanism for the cleavage reaction can be proposed. The N→O acyl shift is initiated by a nucleophilic attack of the Ser25 Oγ on the carbonyl carbon of Gly24. For this, the Gly24 carbonyl bond presumably is held in the conformation similar to the one observed in the pro-ADC structure in which it forms H-bonds to both the Oγ of Thr57 and the water molecule.

Molecular orbital constraints, which dictate the optimal trajectory for nucleophilic attack by the Ser25 hydroxyl, require the angle between the N_{Ser}—C_{Gly}=O_{Gly} peptide plane and the O_{Gly24}=C_{Gly24}—$O\gamma_{Ser25}$ plane to be approximately orthogonal (Figure 13.7 B). In conjunction with the orthogonal angle between the Gly24-Ser25 peptide plane and the O_{Gly24}=C_{Gly24}—$O\gamma_{Ser25}$ plane, the angle between the two vectors $O\gamma_{Ser25}$—C_{Gly24} and C_{Gly24}=O_{Gly24} ($O\gamma_{Ser25}$—C_{Gly24}=O_{Gly24} angle) ideally should lie between 100 and 110°(Figure 13.7 A). Bürgi *et al.* (1973) reported that this is the most favourable angle (Bürgi-Dunitz angle), at which the nucleophile approaches the carbonyl.

However, in pro-ADC the angle between the the N_{Ser}—C_{Gly}=O_{Gly} peptide plane and the O_{Gly24}=C_{Gly24}—$O\gamma_{Ser25}$ plane is less than 10°. In addition, the Ser25 Oγ is more than 4 Å away from the Gly24 carbonyl carbon (see Table 13.1). Consequently, considerable rearrangement of the Ser25 hydroxyl towards the Gly24 mainchain carbon is necessary in order to allow attack of the nucleophile on the carbon, and for the cleavage reaction to occur. Furthermore, a certain degree of conformational flexibility, described above, of the preceding loop appears to be required.

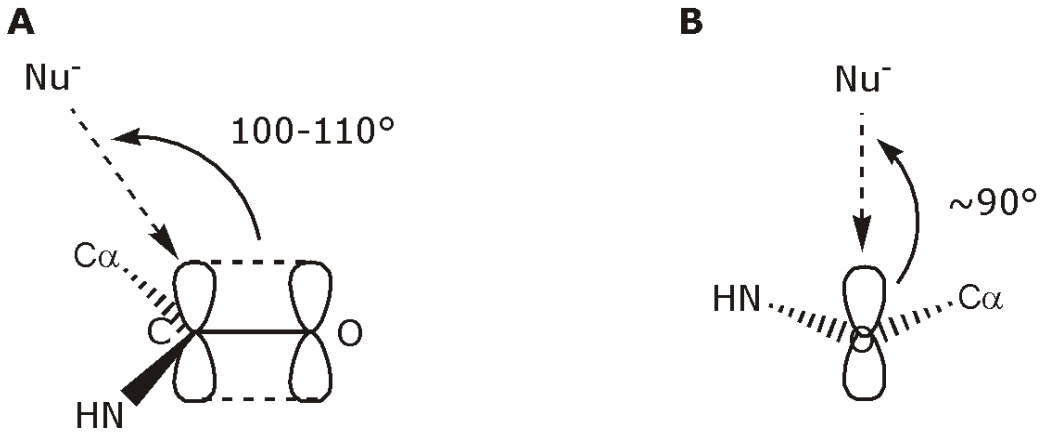

Figure 13.7 Optimal trajectory for nucleophilic attack. (**A**) Illustration of the Bürgi-Dunitz angle. (**B**) End-on view of the C—O bond indicating the orthogonal angle between the N_{Ser}—C_{Gly}=O_{Gly} peptide plane and the O_{Gly24}=C_{Gly24}—$O\gamma_{Ser25}$ plane.

It has been assumed that a specific residue or water molecule, base 1 of Figure 13.1, would act as a general base to enhance the nucleophilicity of the Ser25 Oγ. However, the residues in close vicinity that could directly or indirectly mediate this effect, Tyr58 and Asn72 (see Table 13.1, for distance comparison), have each been shown by mutagenesis to phenylalanine and alanine, respectively, not to be vital for self-processing (Michael Webb, personal communication). Both mutants were, apparently, capable of forming significant amounts of α- and β-chains on PAGE. Similarly, the hydroxyl group of the invariant Tyr58 (Figure 13.4) is a highly unlikely residue for the Hα proton abstraction, since the constructed Y58F mutation is able to form α- and β-chains in similar amounts as native ADC, as judged by tricine-PAGE (data not shown). Also, the Oη of Tyr58 is more than 4.4 Å away from the Ser25 Oγ in the pro-ADC structure. Likewise, although the Tyr22 side-chain is disordered in most of the ADC crystal structure, the likely position of the Oη of Tyr22 appears too far away in the structure models. Consequently, it is possible that in the hydrophilic environment of the Ser25 Oγ there is no specific residue for a general base. Instead, solvent, not observable in the electron density maps, may serve this role.

As described in Section 13.3, the Gly24 carbonyl forms H-bonds to both the Oγ of Thr57 and the water residue (see Figure 13.8 A), which may help self-processing by polarizing the carbonyl and making it more susceptible to a nucleophilic attack. Subsequent to the initial nucleophilic attack, a transient oxyoxazolidine intermediate with a tetrahedral Gly24 carbon is formed. In ADC the resulting negative charge on the oxygen could be stabilized by H-bonds to the water molecule and/or the Thr57 Oγ, which together effectively form the oxyanion hole. Assuming that the oxygen anion is in a position similar to that of the Gly24 carbonyl oxygen in pro-ADC, the oxyanion hole is ideally positioned to provide stabilization. Subsequent breakdown of the oxyoxazolidine intermediate necessitates protonation of the nitrogen by acid 1 (Figure 13.1), which may come from the water molecule or the Thr57 Oγ. The resulting partial negative charge at the Thr57 Oγ could be stabilized by either an H-bond to the main-chain amine of the invariant Tyr58 (Figure 13.4), which is at a distance of 2.79 Å to the Thr57 Oγ in the subunit with the ester intermediate of processed ADC (Albert *et al.*, 1998). Alternatively the water molecule, which in turn interacts with the Nη1 and Nη2 of Arg54, could stabilize the negative charge.

The Thr57 Oγ may perhaps then act as the base 2 (Figure 13.1), for the subsequent Hα proton abstraction of the ester intermediate (Figure 13.8 B). The Thr57 Oγ is at a distance of 3.51 Å from the assigned hydrogen in the ester form in the apo-ADC structure (Figure 13.8). Thr57→Ala and Thr57→Val mutants were completely unable to self-process (Webb, M.E., personal communication) consistent with its central role. Another possible base could be the mainchain carbonyl oxygen of Tyr58, which is 3.18 Å away from the assigned Hα hydrogen.

Apart from the main-chain H-bond between Asn72 and residue 25, none of the interactions identifiable in the cleavage site prior to processing is present any more in the structure of the ester intermediate in processed ADC (Albert *et al.*, 1998). This indicates the conformational rearrangements that have taken place. Upon breakdown of the ester, Gly24 at the

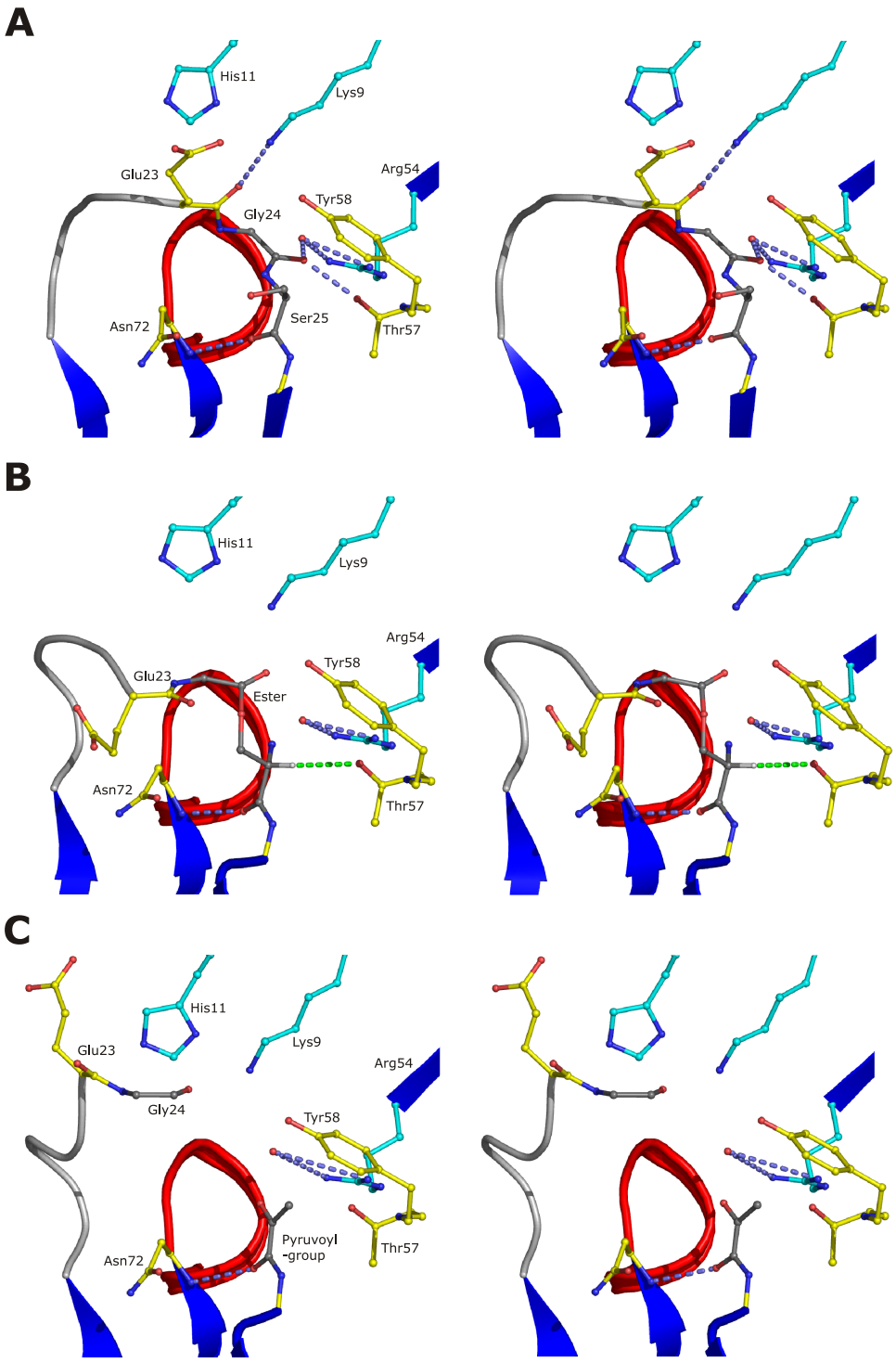

Figure 13.8 Stereo-illustration of the three states of self-processing ADC. (**A**) The unprocessed cleavage site of pro-ADC (subunit B). (**B**) The ester intermediate observed at half occupancy in the processed ADC structure (Albert *et al.*, 1998); indicated by a green dashed line is a possible line of the Hα proton abstraction by the Thr57 Oγ. (**C**) The cleavage site of processed ADC with the pyruvoyl group (Albert *et al.*, 1998). The Cα atom of Gly24 is now $\sim$8 Å away from its position in pro-ADC. All figures were prepared with PYMOL (DeLano, 2002).

carboxyl terminus of the β-chain, moves away (Figure 13.8 C). It has a strikingly different position in the pro-enzyme and active ADC enzyme structure with a Cα atom difference of 7.6 Å and 8.0 Å , in subunits A and B of processed ADC, respectively (subunit B of pro-ADC was superposed on subunits A and B of processed ADC). This difference illustrates the conformational change the loop undergoes upon self-processing. After the Gly24 residue had moved away from the pyruvoyl N-terminus of the α-chain, the active site for the decarboxylation of L-aspartate is formed.

13.6 Explanation for processing activity in the ADC mutants

The suggested model of the molecular mechanism for self-processing can provide a rationale for why S25A, G24S, S25T, A24a-G24b, and S25a-A25b do not process and why H11A and S25C do. The S25A mutant lacks the Oγ and so cannot process. In G24S the conformation of the Gly24 carbonyl is altered significantly, such that the H-bonds between the Gly24 carbonyl and Thr57 and the water molecule are not present. The same applies to S25T, where the Gly24 carbonyl oxygen is more than 5.8 Å away from the Thr57 Oγ. The insertion mutants A24a-G24b and S25a-A25b do not process because of the altered conformational freedom of the loop. Additionally, in S25a-A25b the Gly24 carbonyl is too far away to interact with Thr57.

The H11A structure appears to be able to access a similar transition state as in pro-ADC, in order to process. S25C does process in solution and the fraction that crystallized contained predominantly the processed α- and β-chains. The free Cys at the N-terminus of the α-chain of processed S25C, observed in the structure, arises, presumably, by non-productive hydrolysis of the more reactive thioester intermediate. The latter may be espe-

cially susceptible to hydrolysis under the acidic crystallization condition around pH 4. The lower pK_a of the thiol-group of cysteine compared to the hydroxyl-group of serine may explain why this structurally conservative substitution affects protein self-processing in ADC strongly.

13.7 Comparison with other self-processing systems

Protein self-processing occurs as a protein maturation event in evolutionary unrelated eukaryotic and prokaryotic enzymes. The cleavage site is comprised of Glu-Ser in pyruvoyl-dependent AdometDC and Ser-Ser in both pyruvoyl-dependent ArgDC and HisDC. In pyruvoyl-dependent phosphatidylserine decarboxylase the cleavage site is formed of Gly-Ser, whereas in subunits of pyruvoyl-dependent glycine-, proline-, and sarcosine reductase, it consists of Asn/Thr-Cys (Bednarski *et al.*, 2001; van Poelje and Snell, 1990). Ser→Cys and Ser→Thr mutants of the cleavage site Ser in HisDC, AdometDC, and phosphatidylserine decarboxylase were able to process at significantly reduced rates (Vanderslice *et al.*, 1988; Xiong *et al.*, 1997). Abortive chain cleavage without pyruvoyl-formation, which occurred in these mutants in HisDC, was most likely the result of hydrolysis of the ester, similar to S25C ADC. Instead of the pyruvoyl group, a cysteine or threonine was retained at the N-terminus of the α-chain, indicative that the ester-intermediate was hydrolyzed. A Ser81→Ala mutation in HisDC cleaved extremely slowly and mis-cleavage occurred between residues 80 and 81 (Gelfman *et al.*, 1991).

In the unprocessed structure of Ser53→Ala mutant of precursor ArgDC the modelled Ser53 side-chain was positioned to attack the *si*-face of the Ser52 carbonyl (Tolbert *et al.*, 2003a). It was proposed that the Ser52 side-chain could function as part of an oxyanion hole. Based on the analysis

in this study, described above (Section 13.4), an alternative conformation of this carbonyl in the native cleavage site H-bonding to the Oγ of the invariant Asn47 seems plausible. In order to H-bond to Asn47 the carbonyl would have to adopt a conformation pointing in a direction almost opposite to that in the S53A mutant structure. This would provide a role for Asn47, which is one of only three invariant residues in ArgDC (Tolbert *et al.*, 2003a), in the self-processing mechanism.

In AdometDC, the side-chain amine of a His is assumed to be base 2 for the Hα proton abstraction (Ekstrom *et al.*, 1999), but no oxyanion hole could be identified. In the precursor Ser68→Ala mutant structure the Glu carbonyl of the Glu-Ser cleavage site is H-bonding to the Sγ of the variant Cys82 (Tolbert *et al.*, 2003b). Based on this study on ADC and on comparison with other systems this conformation may not be considered representative of the native cleavage site and led to a different interpretation. A conformation with the carbonyl pointing in the opposite direction would provide a role for the Oγ of the invariant Ser229 as part of an oxyanion hole. In contrast to Cys82 (Xiong *et al.*, 1999), the hydroxyl group of Ser229 was shown by mutagenesis to be essential for processing (Xiong and Pegg, 1999). A Ser229→Cys mutant processed slowly, whereas a Ser229→Thr mutant processed normally (Xiong and Pegg, 1999). However, Ser66→Ala, Glu67→Gln, Ser69→Ala, mutants around the Glu67-Ser68 cleavage site had little effect on self-processing (Stanley and Pegg, 1991; Stanley *et al.*, 1989). Putrescine is known to stimulate the self-processing reaction in certain eukaryotic AdometDCs, although the putrescine binding site is well separated from the cleavage site (Ekstrom *et al.*, 2001).

The Oγ of an invariant Ser (Ser129) forms part of an oxyanion hole in the β-subunit of the proteasome (Ditzel *et al.*, 1998) and that potentially polarizes the Gly carbonyl. The Ser129→Ala mutation completely abolished processing, whereas Thr→Ser and Thr→Cys mutants were able to process at reduced rates (Seemuller *et al.*, 1996). The Gly of the conserved

Gly-Thr cleavage site makes a kink in the Thr1→Ala precursor mutant structure, similar to that in pro-ADC. The computationally modelled Thr Oγ is positioned at a distance of 3.1 Å from the Gly mainchain carbon.

Interestingly, a conserved Thr, whose Oγ can form part of an oxyanion hole, as in ADC, was also identified in the hedgehog protein, in inteins, and in the Ntn glycosylasparaginase. The completely conserved Thr326 is located in close proximity to the cleavage site in the hedgehog protein crystal structure (Hall *et al.*, 1997). In common with inteins, which are most likely evolutionary related to hedgehog proteins, this threonine forms part of a conserved (TxxH) amino acid sequence motif (see Subsection 4.2.2). The Oγ of the conserved Thr72 in the GyrA intein crystal structure, H-bonds to the carbonyl which is attacked by the nucleophile (Klabunde *et al.*, 1998) and is in a similar position both in the structures of *VMA29*-intein Cys1→Ala mutant (Poland *et al.*, 2000) and in the *VMA1*-derived intein precursor quadruple mutant (Mizutani *et al.*, 2002). However, again extrapolating from ADC, the carbonyl in the latter two may not be in its native conformation, since it either lacks any close H-bonding donors or else points away from the side-chains of the conserved Thr or His, which could contribute to an oxyanion hole.

In the unprocessed Trp11→Phe mutant crystal structure of *F. meningosepticum* glycosylasparaginase (Xu *et al.*, 1999) the Thr170 Oγ and a water residue H-bond to the carbonyl oxygen of Asp151 in the cleavage site, and appear to form an oxyanion hole. The relative position of the Thr170, the Asp151 carbonyl, and the oxyanion hole are strikingly similar to the respective arrangement in ADC (Figure 13.9). The Asp151 carbonyl oxygen forms two H-bonds, one to the water residue, which is at a distance of 2.9 Å, and another to the Thr170 Oγ, which is 2.7 Å away from the carbonyl oxygen. Similarly to ADC, a developing negative charge on the Thr170 could, in principle, be stabilized by H-bonding to a mainchain amine. The mainchain amines of Ser171 and Gly171 are both at a distance of 2.9 Å from the Thr170 Oγ.

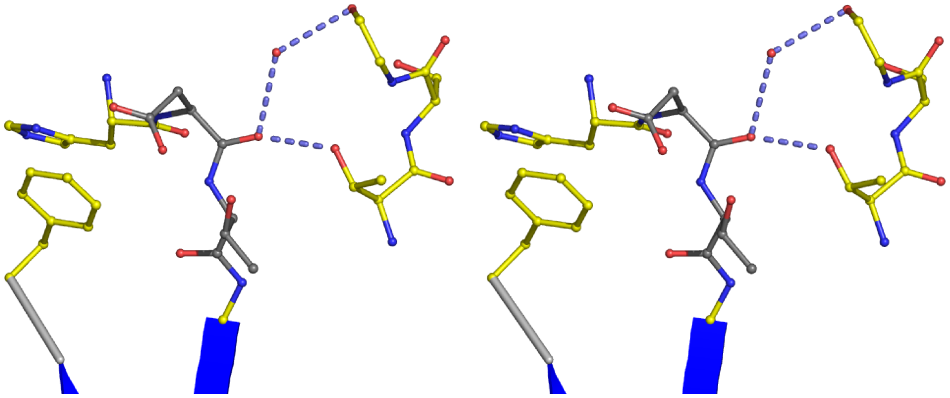

Figure 13.9 Stereo-illustration of the cleavage site and its environment of *F. meningosepticum* glycosylasparaginase Trp11→Phe mutant (PDB-ID: 9GAF) (Xu *et al.*, 1999). Shown are residues His150, the cleavage site residues Asp151 and Thr152 (carbon atom colour, grey, as before), Thr170-Gly172 (on the right) and the Trp11→Phe mutation. Figure prepared with PYMOL (DeLano, 2002).

The unprocessed T152C mutant structure of glycosylasparaginase indicated conformational strain within the peptide bond, that is normally cleaved. The peptide bond significantly deviates from planarity and has an ω-angle of 160° instead of the ideal 180°. Mutagenesis studies of Thr170→Ala/Cys/Ser showed a strong decrease in self-processing activity for the Thr170→Ala and Thr170→Cys mutants and confirmed the key role of this residue for self-processing (Guan *et al.*, 1998). Asp151 in the conserved Asp151-Thr152 cleavage site makes a kink that has similarity to that of Gly24 in ADC. Thr152→Ser and Thr152→Ser mutants were able to process at reduced rates (Guo *et al.*, 1998; Liu *et al.*, 1998), whereas Asp151 appears to be essential for processing. Of the site-directed mutants Asp151→Ala/Arg/Asn/Gly/Gln/Glu/His/Leu/Val/Met/Pro/Ser/Thr only the Asp151→Gly mutant was capable of self-processing (Guan *et al.*, 1998). Nonetheless, Asp151 was suggested to be able to act as the base (base 1 in Figure 13.1) (Xu *et al.*, 1999). However, results on *H. sapiens* glycosylasparaginase, indicate that a water molecule could act as the general base (Saarela *et al.*, 2004).

It was found that small amino acids, such as glycine, reversibly inhibit

the reaction of the T152C and W11F precursor mutants (Guan *et al.*, 1998; Xu *et al.*, 1999). In contrast, no inhibition was observed for the wildtype enzyme form. The glycine H-bonds to the Nη1 and Nη2 of the conserved Arg180 *via* its carboxyl group. The corresponding arginine (Arg211) in the *H. sapiens* glycosylasparaginase crystal structure (Oinonen *et al.*, 1995) also H-bonds to the two oxygens of the γ-carboxyl of L-aspartate (the mode of binding is presumably the same with the natural substrate asparagine). Notably, a similar interaction exists between the γ-carboxyl group and Arg54 in the structure of ADC in complex with β-alanine (see Figure 4.3).

The structures of Ser→Ala mutants of the Gly-Ser cleavage site in the Ntn cephalosporin acylase (Kim *et al.*, 2002) and the Ntn glutaryl-7-aminocephalosporanic acid acylase (Kim *et al.*, 2003) indicated that a water molecule could function as the base to increase the nucleophilicity of the Ser Oγ. In both enzymes the main-chain nitrogen of a histidine could stabilize a developing negative charge. Site-directed mutagenesis and activity studies showed that changing the Gly-1 prohibited the enzyme from self-processing (Li *et al.*, 1999a). A Ser→Thr replacement did not affect the processing rate in glutaryl-7-aminocephalosporanic acid acylase, whereas a Ser→Cys mutant showed reduced processing (Lee and Park, 1998). Whereas a Ser→Thr mutant did self-process in cephalosporin acylase, a Ser→Cys mutant was not able to self-process at a significant rate (Li *et al.*, 1999b).

Similarly as in ADC, in the trimeric ArgDC and HisDC the active site is located at the subunit interface. This is in contrast to the dimeric AdoMetDC, where the cleavage site is located far away from the interface. Although multimeric assembly does not appear to be a general requirement for self-processing, it was shown to be necessary for the proteasome (Seemuller *et al.*, 1996) and glycosylasparaginase (Tikkanen *et al.*, 1996).

13.8 Conclusions

Based on comparison with other self-processing systems, there does appear to be a preference for the location of the self-cleavage site in between two β-strands or at the beginning of a β-strand. Although no single structural motif is used by the self-processing enzymes for the cleavage site, glycine is the most common residue preceding the nucleophile in self-processing systems, presumably due to its conformational adaptability. In contrast to glycosylasparaginase and the GyrA intein, no evidence for the role of strain in self-processing could be found in ADC. However, the analysis described above, sounds a cautionary note about both the interpretation and the relevance of models and mechanisms, based solely on the structure of Ser→Ala or Ser→Thr mutants, in other self-processing systems. As shown in this study, the cleavage site of such models may not adopt the conformation it does in the native cleavage site, prior to self-processing.

So far, a specific molecule or residue for a general base, base 1, could be identified in cephalosporin acylase, glutaryl-7-aminocephalosporanic acid acylase, in the proteosome, and in glycosylasparaginase, but not in any of the other self-processing enzymes. In all four structures, (ordered) water appears to be able to act as the general base for the proton abstraction of the nucleophile. No such ordered water molecule could be observed in the model for the pro-ADC crystal structure. This could suggest that general base catalysis by a specific residue may not be a requirement for the initial nucleophilic attack. More likely, however, solvent water could serve this function.

At physiological or neutral pH the equilibrium between amides and esters strongly favours the amides in a general open chemical system (Elliott, 1952). Nonetheless, chemical model systems have shown that the equilibrium between amide bond and ester can be on the side of the ester, when

the amine is protonated (Iwai and Ando, 1967). Self-processing enzymes may additionally facilitate the N→O/S rearrangement by polarization of the attacked carbonyl bond and by stabilizing the oxyoxazolidine intermediate in an oxyanion hole. Comparison with other self-processing systems implies that the oxyanion hole is a crucial, general requirement for self-processing. Therefore, stabilization of the oxyoxazolidine intermediate could be *the* important step to drive the reaction. Interestingly, this stabilization in most of the self-processing enzymes is different from the oxyanion hole identified in serine proteases, which is formed mainly by hydrogens from main-chain nitrogens rather than oxygens. This could be a consequence of the multiple roles the stabilizing groups have in self-processing systems, to both polarize the carbonyl bond in the first step in the reaction and stabilize the resulting negative charge in the second step. In addition to the electrochemical rules the oxyanion hole serves it may also ensure the correct orientation of the carbonyl bond, and the oxyoxazolidine intermediate, so that amino acid residues are optimally positioned for the subsequent Hα proton-abstraction.

Solvent, an ordered water molecule, and the catalytic Thr57 residue could be sufficient to catalyze the self-processing reaction in the context of the protein in ADC. There appear to be overall similarities of the mechanism of self-processing mechanism in ADC with that known from serine proteases, such as the tetrahedral intermediate and general acid-base catalysis. In contrast, other features, such as a charge-relay system with a catalytic triad seem to be absent.

The reason why Ser25 is invariant, may be explained by the fact that the thioester, presumably formed in the Ser25→Cys mutant, is too reactive. However, considering the slow-processing rate, and the possible misprocessing (see Subsection 4.2.1) of ADC, the self-processing reaction in ADC may not be particularly optimized, although ADC appears to have considerable conformational specificity for the reaction.

Self-processing of ADC in *E. coli* was reported to be significantly faster *in vivo* than *in vitro* (Michael Witty, personal communication). The reason for this may be an external self-processing factor. A small molecule could bind to the pocket H-bonding to the Nη1 and Nη2 of Arg54, instead of the water molecule (see Figure 13.8), possibly providing additional stabilization of reaction intermediates, such as the oxyoxazolidine intermediate. This cavity appears to be particularly attractive for small molecule, because a sulfate residue was found in the H11A, S25a-A25b, and S25C structures. In S25T, a malonic acid residue was positioned there. A similar cavity exists in glycosylasparaginase, and a glycine molecule binds to this region in the T152C and W11F crystal structures. Although, in the later case glycine inhibits rather than stimulates the reaction.

However, this study has shown that functionality of residues for the self-processing reaction and the catalysis carried out by the enzyme overlap. This appears to be the case for Arg54, which H-bonds to the water for the self-processing reaction, and to the γ-carboxyl group of L-aspartate for the decarboxylation reaction.

The results of this study on *E. coli* ADC appear to be relevant to ADCs from pathogenic bacteria, such as *H. pylori* (Kwon *et al.*, 2002) and *M. tuberculosis* (Chopra *et al.*, 2002), which are potential drug targets, since the key residues are invariant throughout all sequences (Figure 13.4). Further structural information on the cleavage reaction is useful in understanding the respective biochemistry and can aid in producing drugs to prevent protein maturation. The molecular mechanism of ADC is likely to have significant parallels in other self-processing systems. However, high resolution crystal structures of intermediates and true precursors from other systems will be necessary to determine whether the geometrical restraints and principle mechanism that were proposed in ADC are the same in other systems.

13.9 Future work

Although this study could potentially identify interactions that are key to the self-processing reaction, a more thorough understanding of the molecular mechanism will doubtlessly require additional data. The mechanistic model of the self-processing reaction presented in Section 13.5 is mainly based on crystallographic results of the structures of pro-ADC and the site-directed mutants. It is consequently limited by what can be reliably interpreted from such kind of data. The limitation particularly applies to the static representation of the model, that does not reflect the dynamic elements of the reaction.

From this study the following issues appear to be in need to be addressed: The proposed chemical and structural role both of Thr57 and the water molecule have to be confirmed and, importantly, discerned, in detail (the Thr57 Oγ and the water are equidistant from the oxygen of the Gly24 carbonyl group, (Section 13.3, Table 13.1)). Does the Thr57 Oγ indeed provide the acid (acid 1 in Section 13.5) and/or the base 2; or is its role merely a structural one to orient the carbonyl-group. Is there a chemical role for the water? Does it, as suggested (Section 13.3, Section 13.5), stabilize the exocyclic oxygen in the oxyoxazolidine intermediate, and/or provide the acid 1? And/or is its role a structural one, or a combination of both?

Based on the analysis in this study, an Asn72→Ala site-directed amino acid mutation of ADC was already constructed and the respective crystal structure solved (Carina Lobley and Michael Webb, personal communication). The structure model revealed a processed cleavage site with a Ser instead of a pyruvoyl-group at the N-terminus of the α-chain. This does imply a role for Asn72 in the self-processing reaction. Nonetheless, a direct functional role for the Oδ1 or Nδ2 of Asn72 as base 1 seems improbable. The Oδ1 is likely to carry a fractional negative charge and the Nδ2 is usually not a strong enough general base. However, it needs to be investigated

whether the role of Asn72 is mainly structural, as this study suggests, or whether it includes chemical effects. In addition, ongoing work, presently underway, attempts to solve crystal structures of Thr57 mutants, such as Thr57→Ala, Thr57→Val, and Thr57→Ser mutants. An Arg54→Ala mutant would be interesting to study, since the water molecule could be displaced from its position without the stabilizing interactions with the Nη1 and Nη2 of the invariant Arg54.

Although conventional site-directed mutagenesis studies, such as carried out for this study, can suggest a particular role for a residue, the evidence is not comparable with the kind of evidence an atom-specific mechanistic understanding would require. However, more subtle and systematic investigation may in principle be done by incorporation of unnatural amino acids. These have successfully been used as probes for enzyme function and protein structure, before (Cook *et al.*, 1995). Implementation of such amino acids in the protein is possible by nonsense suppression. The respective methodology for the incorporation of unnatural amino acids has been reviewed by Dougherty (2000). It is essentially based on site-directed mutagenesis of the codon, encoding the residue one intends to alter, to a stop-codon. An aminoacyl transfer-RNA, with the appropriate anticodon, is then chemically acylated with the unnatural amino acid and added to the translation system (often an *in vitro* system).

In the case of ADC it would be interesting to apply this technique in order to study the effect of polarity of the Thr57 sidechain on self-processing. One could exchange the Oγ of Thr57 to a less nucleophilic atom, such as a nitrogen, and carry out activity assays and crystallographic analysis of this mutant. Furthermore, a Thr57 mutant with an additional carbon on the oxygen could potentially be of interest, to test the implication of the Oγ hydrogens in acid-base catalysis. Likewise, an amino acid could be synthesized where the side-chain atoms Oδ1 or Nδ2 of Asn72 are both, or individually, changed to a carbon atom. One could possibly distinguish steric and chemical effects of such an isosteric mutant in more detail. Ad-

ditionally, the nonsense suppression technique could be applied to incorporate an ester, instead of the backbone amide, for Asn72 and Cys26, by acylating the respective α-hydroxy acids instead of the α-amino acids with aminoacyl transfer-RNAs. In this manner the importance of the backbone interactions of Asn72-Ser25 and also Cys26-Tyr58 for the conformation of Ser25, and, consequently, for self-processing could be determined.

However, the static, time-averaged representation of the state that conventional crystallography, with a monochromatic X-ray source, can provide may not be able to unequivocally answer the questions raised above. More direct, chemical evidence would be desirable. Following the events of proton donation and proton abstraction, as indicated in the chemical scheme of (Figure 13.1), directly, is difficult by any biophysical and biochemical means. Radioactive labelling experiments with deuterium do not seem to be feasible, as the deuterium would presumably quickly exchange with solvent. In contrast, obtaining a three-dimensional representation of the oxyoxazolidine intermediate (Figure 13.1) of the reaction could be amenable. Identification of the interactions of the intermediate could then provide more reliable evidence for a molecular mechanism.

For this purpose, a time-resolved crystallographic technique known as Laue diffraction (reviewed by Ren *et al.* (1999)) may help to collect X-ray data in a short enough time. This technique could be combined with a chemically protected Ser25 residue. Incorporation of the serine, the Oγ of which could be protected with for instance a *o*-nitrobenzyl group, would be carried out with nonsense suppression as described above. Photolysis by light then releases the nitrobenzyl group and reveals the previously protected functionality. This technique has indeed already been carried out with a serine residue, for elucidation of self-processing in inteins, by Cook *et al.* (1995). Protection of Ser25 in ADC may be necessary to initiate the reaction only after the protein has been crystallized and after the crystals have been brought to the X-ray data collection facility. The self-processing reaction would then be started at a specific time point.

Rapid reaction initiation (at the femtosecond scale), to deprotect the serine residue, can be accomplished by applying a short laser pulse. Laser systems with nanosecond and femtosecond pulses have been implemented at ID09 beamline, at the European synchrotron radiation facility, and synchronized with the X-ray pulses (see references in (Ren *et al.*, 1999)). Reaction initiation by this method and immediate Laue diffraction data collection have successfully been applied by Stoddard *et al.* (1990). In this case, an active-site serine residue of chymotrypsin was protected as a cinnamic acid ester, photolyzed, and X-ray data were collected. For ADC, an identical experiment as for the native enzyme could also be carried out with a Ser25→Cys ADC mutant, which may be more reactive with respect to the initial step. The higher electron density of the sulfur, may additionally assist in interpretation of the possible intermediate. Particularly, at medium resolution it is sometimes difficult to discern oxygen atoms from carbon atoms in electron density maps.

The results of such experiments, as implied by the parallels of the self-processing reaction of ADC with other systems (Section 13.7), may indeed have a far-reaching impact on the understanding of the molecular mechanism of self-processing.

Part IV

Concluding remarks

Crystallization of KPHMT with cofactor analogues Several crystallization conditions for KPHMT in the presence of two cofactor-analogues and/or α-kiva were obtained. However, for most of them the quality of the crystals was limited, as the X-ray diffraction extended to $\sim$4.0 Å on the in-house X-ray generator. Eventually, crystallization conditions were found that yielded KPHMT crystals that were larger in all three dimensions. Reproducing these crystals could improve the X-ray diffraction limit, and consequently may enable successful data collection. Calculating Fourier-maps will be imperative to determine the mode of binding of the cofactor to the enzyme.

Comparison of the KPHMT structure Based on the sequence analysis, and structure analysis of the KPHMT $(\beta\alpha)_8$-fold, it was possible to classify KPHMT into a superfamily in the SCOP database, to draw conclusions about the evolutionary relationship of the enzymes within this superfamily, and to gain insight into both the conservation of ligand binding to the catalytic Mg^{2+}, as well as similarities and differences in the molecular mechanism. It was shown that KPHMT is a member of the PEP/pyruvate superfamily, forming a family of its own within. Based on the analysis of the members of this superfamily, a subdivision into two groups in this superfamily appeared conceivable. This analysis completed the structural analysis of the four *E. coli* enzymes of pantothenate biosynthesis.

Structure solution of ADC enzymes Jointly with Michael Webb, a fast purification protocol for eight ADC mutants has been devised. This allowed crystallization of several of these ADC mutants. In this study, a full X-ray diffraction dataset was successfully collected on two ADC mutants, G24S and S25T, for the latter at near atomic resolution. Their crystal structures were solved by molecular replacement, and/or directly by using the previously solved ADC structure as a search model. Together with the already solved pro-ADC and S25A structures, these structures were refined

to comparably low R_{cryst} and R_{free} values, the structure models assessed, and deposited in the PDB.

Structural analysis of ADC enzymes The four ADC structures together with the ADC structures of A24a-G24b, H11A, S25a-A25b, and S25C, were the subject of a detailed comparative analysis. The results of the investigation imply that specific conformational constraints govern the self-processing reaction. Geometrical data from the cleavage site of the pro-ADC structure indicate that considerable rearrangement of the Ser25 hydroxyl towards the Gly24 carbonyl carbon is necessary for the self-processing reaction to occur. The conformational freedom in the loop preceding the cleavage site appears to play a determining role in the reaction. Comparison of the structural features with other self-processing systems shows significant similarity, particularly with the enzyme glycosylasparaginase. The results of the analysis suggest that the Thr57 Oγ and a water molecule form an oxyanion hole that could stabilize the oxyoxazolidine intermediate in the self-processing reaction. Furthermore, the Thr57 Oγ and this water could play key roles in the mechanism, by supporting acid-base catalysis during the reaction. In order to gain further insight into the mechanism, and, particularly, to discern steric and electrochemical effects, a series of experiments has been suggested.

Part V

Appendix

A Crystallization of morphine dehydrogenase

A.1 Introduction

Apart from structural research on the two enzymes in the pantothenate pathway, studies on the crystallization of morphine dehydrogenase were carried out. This project was in close collaboration with Dr Amrik Basran and Prof. Neil Bruce, whose laboratory (Institute of Biotechnology, University of Cambridge) the research on this enzyme was initiated in.

Morphine dehydrogenase is a monomeric enzyme with a molecular mass of ~32 kDa (Bruce *et al.*, 1991). It catalyzes the $NADP^+$-dependent oxidation of morphine and codeine to morphinone and codeinone, respectively. The enzyme was isolated from a *P. putida* M10 strain, found in opiate factory waste liquor (Bruce *et al.*, 1990). This bacterial strain apparently can grow on morphine as the sole source of carbon. The oxidation of morphine to morphinone (Figure A.1) by morphine dehydrogenase was observed to be the first step in the catabolism of morphine in this organism. The enzyme appears to have an absolute requirement for the $NADP^+$ cofactor for activity (Bruce *et al.*, 1991).

The gene, encoding morphine dehydrogenase, was found to be located on a large plasmid (165 kb), from which it appears to be constitutively expressed by the bacterium (Bruce *et al.*, 1990). The gene was cloned and sequenced (Willey *et al.*, 1993), and, subsequently, morphine dehydrogenase was recombinantly overexpressed in *E. coli* (Walker *et al.*, 2000). Amino acid sequence analysis indicated that the enzyme belongs to the aldo-ketoreductase family of alcohol dehydrogenases (Bruce *et al.*, 1994). It was shown to be closely homologous to eighteen proteins, and highest similarity was observed with 2,5-dioxo-D-gluconic acid reductases. The protein was purified to high homogeneity in a two-step protocol with affinity chromatography on Mimetic orange 3 A6X, a specific dye for $NADP^+$-binding enzymes, and gel-filtration (see Bruce *et al.* (1991) for a detailed purification procedure). During the whole purification procedure the protein solutions were supplemented with at least 2 mM DTT, and kept at 4°C and/or on ice. For crystallization trials in this study, the protein was purified and supplied by Dr Amrik Basran.

Morphine dehydrogenase was made the basis of a biochemical assay for the detection of opiates in urine (Holt *et al.*, 1995). Moreover, the enzyme certainly has a high potential for future biomedical applications, such as the production of semisynthetic opiate drugs. Rational engineering of

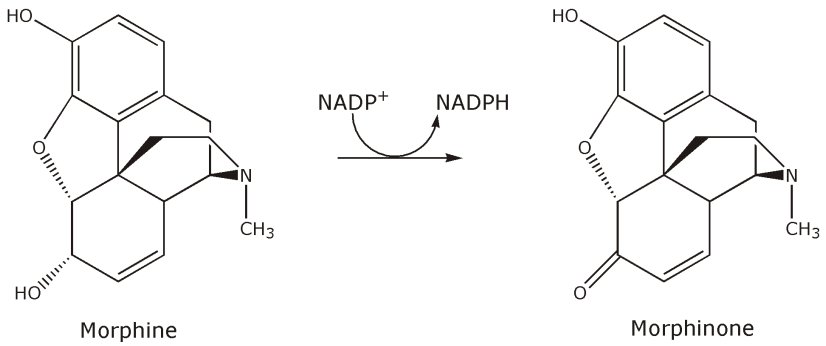

Figure A.1 Schematic illustration of the oxidation of morphine to morphinone catalyzed by morphine dehydrogenase.

the protein, based on structural information, could aid this aim. Consequently, the purpose of this study was to crystallize morphine dehydrogenase, either on its own and/or with its cofactor $NADP^+$.

A.2 Experimental procedures and results

Initially, crystallization screening focused on crystallization of morphine dehydrogenase on its own. The enzyme was dialyzed into 5 mM DTT, 25 mM Hepes pH 7.4 at 4°C, with a dialysis tubing with a molecular weight membrane cut-off of 25 kDa. The addition of DTT appeared important because the enzyme was found to rapidly lose activity if thiol-protecting agents were not added. Crystallization trials were carried out as soon as possible after the end of the dialysis, usually on the same day. Samples for later usage were frozen with 20%(v/v) glycerol at −80°C. The protein concentration was assessed as described in Subsection 7.3.1, using the theoretical absorption coefficient of morphine dehydrogenase at 280 nm (ϵ_{280} = 37 530 M^{-1} cm^{-1} or 1.173 ml mg^{-1} cm^{-1}) (caclculated with the Protparam utility[1] (Gill and von Hippel, 1989) from the Swissprot-Expasy internet server).

Morphine dehydrogenase was reported to be unstable at 4°C with respect to its activity (Bruce *et al.*, 1991). Therefore, all crystallization trials were carried out at 4°C as well as at 19°C, with a protein concentration of 14-38 mg/ml. However, in spite of extensive crystallization screening with a number of commercially available crystallization screens, no crystals could be obtained. Therefore, subsequent screening focused on crystallization of the protein, in the presence of $NADP^+$ and a non-hydrolysable analogue, thionicotinamide adenine dinucleotide phosphate.

The cofactor and cofactor analogue were initially dissolved in 1 M

[1] http://ca.expasy.org/tools/protparam.html

$NaCH_3CO_2$ pH 5.5 (adjusted with HCl) to potentially prolong the lifetime of the cofactor. $NADP^+$ is reported to be stable for a longer time at acidic pH. Later, the cofactor was dissolved in in MilliQ water, which did not have any apparent impact on the crystallization, compared with dissolution in $NaCH_3CO_2$. For crystallization trials the protein was incubated with the cofactor solution on ice, in a protein:cofactor ratio of 1:2-6, for for at least 45 min, prior to setting up crystallization trials.

Several crystallization conditions of morphine dehydrogenase in the presence of $NADP^+$ could be found. Crystals usually appeared after 3-4 weeks. The crystallization trays at 17°C were covered with foil to minimize exposure to light. Likewise, the trays at 4°C were in an incubator impenetrable to visible light.

The largest three-dimensional crystals (Figure A.2) could be obtained with a precipitant solution of 1.5 and 1.9 M sodium malonate, NaOH, pH: 5.0. The protein concentrations used were 17 and 29.8 mg/ml, with a protein: $NADP^+$ ratio of 1:3. Crystals grew, at 4°C only, as needles with a maximum length in one dimension of 170 µm (Figure A.2 A,B). Using a protein concentration of 17 mg/ml yielded fewer, but larger crystals.

A second promising precipitant condition was 30% PEG8000, 0.1 M Na-cacodylate, 0.2 M $(NH_4)_2SO_4$, pH 6.5 (condition 14 of the Molecular Dimensions Structure screen 1). The protein concentrations were 17 and 29.8 mg/ml, with a protein:$NADP^+$ ratio of 1:4. The crystals grew at 17°C only, as needles (Figure A.2 C). However, no significant difference in the size of the crystals could be observed by using a protein concentration of 17 or 29.8 mg/ml.

Both of the above mentioned crystallization conditions could be reproduced, however, as mentioned above, the response time from setting up crystallization trials to obtaining crystals was 3-4 weeks.

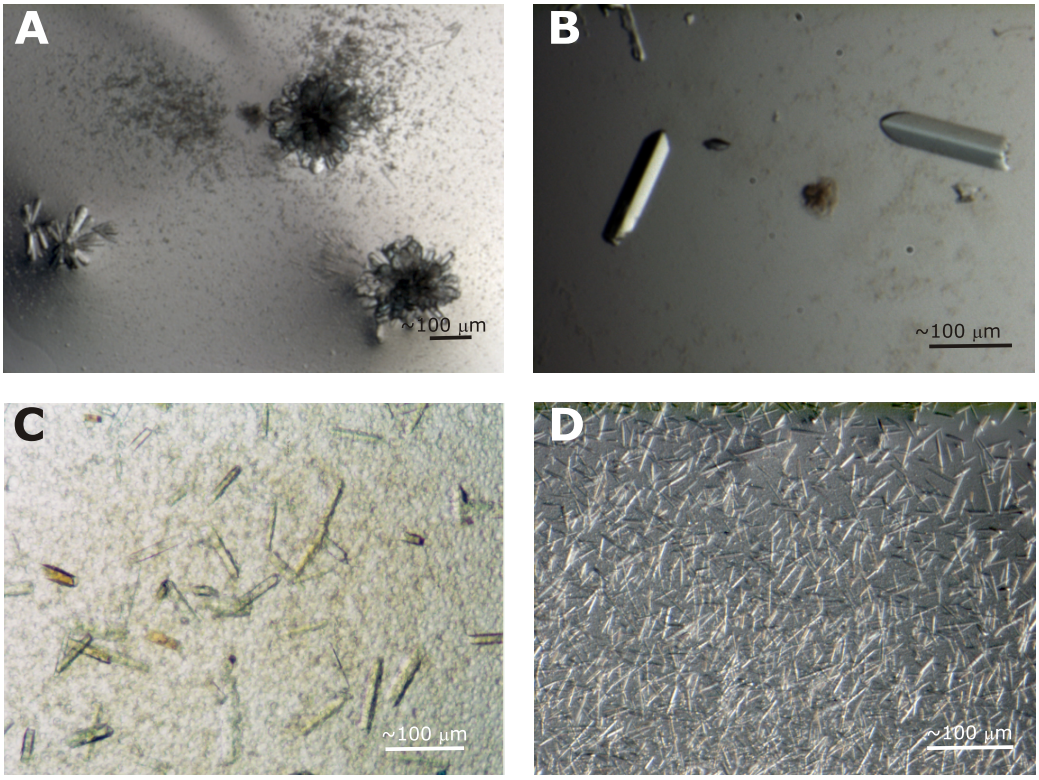

Figure A.2 Morphology of the morphine dehydrogenase crystals obtained in the presence of $NADP^+$. (**A**), (**B**) Crystals grown with sodium malonate as a precipitant, (**C**) Crystals grown with 30% PEG8000, 0.1 M Na-cacodylate, 0.2 M $(NH_4)_2SO_4$, pH 6.5 as precipitant solution. (**D**) Crystals grown with ammonium formate, 20% PEG3350, pH: 6.6.

Ammonium formate, 20% PEG3350, pH: 6.6 (solution 23 of the Hampton PEG/ion screen) was another successful precipitant to give crystals for morphine dehydrogenase. A protein: $NADP^+$ ratio of 1:4, with a protein concentration of 17 and 29.8 mg/ml, was used. Crystals grew as fine needles (Figure A.2 D) at 17°C only.

Several crystals, grown of morphine dehydrogenase in the presence of $NADP^+$, with sodium malonate (1.5+1.9 M, pH 5.0) as precipitant, have been cryoprotected with sodium malonate, and were frozen in liquid N_2. These await screening at a synchroton X-ray source. Further optimization of the obtained crystallization conditions is currently underway, and it is

hoped to increase the size of the crystals. One problem, so far, was the rather long response time for the crystallization experiments. Crystals appear to grow rather slowly and could only be observed 3-4 weeks after the crystallization trials were set up. Optimization of the crystallization conditions could possibly reduce the time until crystals emerge.

Part VI

References

Bibliography

Aberhart, D. and Russell, D. G. (1984) Steric course of Ketopantoate Hydroxymethyltransferase in E.coli. *J Am Chem Soc*, **106**:4902–4906. 26

Abiko, Y. (1975) *Metabolism of coenzyme A*, vol. 7 of *Metabolic pathways, 3rd ed.* Academic Press, New York. 15

Adams, P. D., Pannu, N. S., Read, R. J. and Brunger, A. T. (1997) Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc Natl Acad Sci U S A*, **94**(10):5018–23. 102

Afshar, K., Gonczy, P., DiNardo, S. and Wasserman, S. A. (2001) fumble encodes a pantothenate kinase homolog required for proper mitosis and meiosis in Drosophila melanogaster. *Genetics*, **157**(3):1267–76. 21

Aghajanian, S. and Worrall, D. M. (2002) Identification and characterization of the gene encoding the human phosphopantetheine adenylyltransferase and dephospho-CoA kinase bifunctional enzyme (CoA synthase). *Biochem J*, **365**(Pt 1):13–8. 16, 17

Albert, A., Dhanaraj, V., Genschel, U., Khan, G., Ramjee, M. K., Pulido, R., Sibanda, B. L., von Delft, F., Witty, M., Blundell, T. L., Smith, A. G. and Abell, C. (1998) Crystal structure of aspartate decarboxylase at 2.2 A resolution provides evidence for an ester in protein self-processing. *Nat Struct Biol*, **5**(4):289–93. 36, 37, 38, 48, 49, 50, 51, 55, 162, 167, 169, 173, 175, 182, 186, 194, 195

Albert, A., Martinez-Ripoll, M., Espinosa-Ruiz, A., Yenush, L., Culianez-Macia, F. A. and Serrano, R. (2000) The X-ray structure of the FMN-binding protein AtHal3 provides the structural basis for the activity of a regulatory subunit involved in signal transduction. *Structure Fold Des*, **8**(9):961–9. 17, 192

Almassy, R. J., Janson, C. A., Kan, C. C. and Hostomska, Z. (1992) Structures of apo and complexed Escherichia coli glycinamide ribonucleotide transformylase. *Proc Natl Acad Sci U S A*, **89**(13):6114–8. 117

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**(3):403–10. 112

Alves, R., Chaleil, R. A. and Sternberg, M. J. (2002) Evolution of enzymes in metabolism: a network perspective. *J Mol Biol*, **320**(4):751–70. 8

Arndt, U. W. and Wonacott, A. J. (1977) *The rotation method in crystallography*. North-Holland, Amsterdam. 94

Artymiuk, P. J. (1995) A sting in the (N-terminal) tail. *Nat Struct Biol*, **2**(12):1035–7. 41, 42

Babbitt, P. C. and Gerlt, J. A. (2000) New functions from old scaffolds: how nature reengineers enzymes for new functions. *Adv Protein Chem*, **55**:1–28. 7

Baigori, M., Grau, R., Morbidoni, H. R. and de Mendoza, D. (1991) Isolation and characterization of Bacillus subtilis mutants blocked in the synthesis of pantothenic acid. *J Bacteriol*, **173**(13):4240–2. 11

Bairoch, A. and Boeckmann, B. (1991) The SWISS-PROT protein sequence data bank. *Nucleic Acids Res*, **19 Suppl**:2247–9. 113

Bal, J., Kajtaniak, E. and Pieniazek, N. (1977) 4-nitroquinoline-1-oxide: a good mutagen for Aspergillus nidulans. *Mutation Res*, **56**:153–156. 14

Barbarat, B. and Podevin, R. A. (1986) Pantothenate-sodium cotransport in renal brush-border membranes. *J Biol Chem*, **261**(31):14455–60. 11

Beachy, P. A., Cooper, M. K., Young, K. E., von Kessler, D. P., Park, W. J., Hall, T. M., Leahy, D. J. and Porter, J. A. (1997) Multiple roles of cholesterol in hedgehog protein biogenesis and signaling. *Cold Spring Harb Symp Quant Biol*, **62**:191–204. 42

Bednarski, B., Andreesen, J. R. and Pich, A. (2001) In vitro processing of the proproteins GrdE of protein B of glycine reductase and PrdA of D-proline reductase from Clostridium sticklandii: formation of a pyruvoyl group from a cysteine residue. *Eur J Biochem*, **268**(12):3538–44. 38, 197

Begley, T. P., Kinsland, C. and Strauss, E. (2001) The biosynthesis of coenzyme A in bacteria. *Vitam Horm*, **61**:157–71. 15

Belfort, M. and Perlman, P. S. (1995) Mechanisms of intron mobility. *J Biol Chem*, **270**(51):30237–40. 43

Belfort, M. and Roberts, R. J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res*, **25**(17):3379–88. 43

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res*, **28**(1):235–42. 110

Biou, V., Dumas, R., Cohen-Addad, C., Douce, R., Job, D. and Pebay-Peyroula, E. (1997) The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog determined at 1.65 Å resolution. *EMBO J*, **16**(12):3405–15. 10

Blattner, F. R., Plunkett, G., r., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. (1997) The complete genome sequence of Escherichia coli K-12. *Science*, **277**(5331):1453–74. 4

Blow, D. (2002) *Outline of Crystallography for Biologists*. Oxford University Press, Oxford. 82

Blundell, T. L. and Johnson, L. N. (1976) *Protein crystallography*. Academic Press, New York. 82

Bompard-Gilles, C., Villeret, V., Davies, G. J., Fanuel, L., Joris, B., Frere, J. M. and Van Beeumen, J. (2000) A new variant of the Ntn hydrolase fold revealed by the crystal structure of L-aminopeptidase D-ala-esterase/amidase from Ochrobactrum anthropi. *Structure Fold Des*, **8**(2):153–62. 41, 44

Boyington, J. C., Gladyshev, V. N., Khangulov, S. V., Stadtman, T. C. and Sun, P. D. (1997) Crystal structure of formate dehydrogenase H: catalysis involving Mo, molybdopterin, selenocysteine, and an Fe4S4 cluster. *Science*, **275**(5304):1305–8. 50

Branden, C. I. (1991) The TIM-barrel the most frequently occurring folding motif in proteins. *Curr Opin Struct Biol*, **1**:978–983. 134

Branden, C. I. and Jones, T. A. (1990) Between objectivity and subjectivity. *Nature*, **343**:687–689. 104

Brannigan, J. A., Dodson, G., Duggleby, H. J., Moody, P. C., Smith, J. L., Tomchick, D. R. and Murzin, A. G. (1995) A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature*, **378**(6555):416–9. 42

Britton, K., Langridge, S., Baker, P. J., Weeradechapon, K., Sedelnikova, S. E., De Lucas, J. R., Rice, D. W. and Turner, G. (2000) The crystal structure and active site location of isocitrate lyase from the fungus Aspergillus nidulans. *Structure Fold Des*, **8**(4):349–62. 142

Britton, K. L., Abeysinghe, I. S., Baker, P. J., Barynin, V., Diehl, P., Langridge, S. J., McFadden, B. A., Sedelnikova, S. E., Stillman, T. J., Weeradechapon, K. and Rice, D. W. (2001) The structure and domain organization of Escherichia coli isocitrate lyase. *Acta Crystallogr D Biol Crystallogr*, **57**(Pt 9):1209–18. 142

Brown, G. M. (1959) The metabolism of pantothenic acid. *J Biol Chem*, **234**(2):370–8. 15

Brown, G. M. and Williamson, J. M. (1982) Biosynthesis of riboflavin, folic acid, thiamine, and pantothenic acid. *Adv Enzymol Relat Areas Mol Biol*, **53**:345–81. 10

Bruce, N. C., Willey, D. L., Coulson, A. F. and Jeffery, J. (1994) Bacterial morphine dehydrogenase further defines a distinct superfamily of oxidoreductases with diverse functional activities. *Biochem J*, **299 (Pt 3)**:805–11. 214

Bruce, N. C., Wilmot, C. J., Jordan, K. N., Stephens, L. D. and Lowe, C. R. (1991) Microbial degradation of the morphine alkaloids. Purification and characterization of morphine dehydrogenase from Pseudomonas putida M10. *Biochem J*, **274 (Pt 3)**:875–80. 213, 214, 215

Bruce, N. C., Wilmot, C. J., Jordan, K. N., Trebilcock, A. E., Gray Stephens, L. D. and Lowe, C. R. (1990) Microbial degradation of the morphine alkaloids: identification of morphine as an intermediate in the metabolism of morphine by Pseudomonas putida M10. *Arch Microbiol*, **154**(5):465–70. 213, 214

Brunger, A. T. (1992) The R-free value: a novel statistical quality for assessing the accuracy of crystal structures. *Nature*, **335**:472–474. 104

Brunger, A. T. (1993) Assessment of Phase Accuracy by Cross Validation: the Free R Value. Methods and Applications. *Acta Crystallogr D Biol Crystallogr*, **49**:24–36. 105

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, **54**(Pt 5):905–21. 101, 105

Bucovaz, E. T., Macleod, R. M., Morrison, J. C. and Whybrew, W. D. (1997) The coenzyme A-synthesizing protein complex and its proposed role in CoA biosynthesis in bakers' yeast. *Biochimie*, **79**(12):787–98. 17

Bürgi, M., Dunitz, J. and Shefter, E. (1973) Geometrical reaction coordinates Nucleophilic addition to a carbonyl group. *J Am Chem Soc*, **95**:5065–5067. 192

Calder, R. B., Williams, R. S., Ramaswamy, G., Rock, C. O., Campbell, E., Unkles, S. E., Kinghorn, J. R. and Jackowski, S. (1999) Cloning and characterization of a eukaryotic pantothenate kinase gene (panK) from Aspergillus nidulans. *J Biol Chem*, **274**(4):2014–20. 16, 18

Campbell, L. (1957) Mechanism of uracil degradation by Chlostridium uracilicum. *J Bacteriol*, **73**:225–229. 14

Canellakis, E. S. (1956) Pyrimidine metabolism. I. Enzymatic pathways of uracil and thymine degradation. *J Biol Chem*, **221**(1):315–22. 14

Carter, C. and Sweet, R. (1997a) *Macromolecular Crystallography-Part A*, vol. 276 of *Methods in ENZYMOLOGY*. Academic Press, San Diego. 81, 82

Carter, C. and Sweet, R. (1997b) *Macromolecular Crystallography-Part B*, vol. 277 of *Methods in ENZYMOLOGY*. Academic Press, San Diego. 82, 106, 107

Castillo, R. M., Mizuguchi, K., Dhanaraj, V., Albert, A., Blundell, T. L. and Murzin, A. G. (1999) A six-stranded double-psi beta barrel is shared by several protein superfamilies. *Structure Fold Des*, **7**(2):227–36. 48, 49, 50, 153

Chaudhuri, B. N., Sawaya, M. R., Kim, C. Y., Waldo, G. S., Park, M. S., Terwilliger, T. C. and Yeates, T. O. (2003) The crystal structure of the first

enzyme in the pantothenate biosynthetic pathway, ketopantoate hydroxymethyltransferase, from M tuberculosis. *Structure Camb*, **11**(7):753–64. 32

Chen, P. and Hochstrasser, M. (1996) Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell*, **86**(6):961–72. 41

Ching, K. H., Westaway, S. K., Gitschier, J., Higgins, J. J. and Hayflick, S. J. (2002) HARP syndrome is allelic with pantothenate kinase-associated neurodegeneration. *Neurology*, **58**(11):1673–4. 21

Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., Pelletier, J. J., Paulus, H. and Xu, M. Q. (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, **192**(2):271–81. 54

Chopra, S., Pai, H. and Ranganathan, A. (2002) Expression, purification, and biochemical characterization of Mycobacterium tuberculosis aspartate decarboxylase, PanD. *Protein Expr Purif*, **25**(3):533–40. 35, 37, 157, 204

Chothia, C., Gough, J., Vogel, C. and Teichmann, S. A. (2003) Evolution of the protein repertoire. *Science*, **300**(5626):1701–3. 7

Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**:760–763. 98, 99, 106

Cook, S., Jack, W., Xiong, X., Danley, L., Ellmann, J. and Schultz, P. (1995) Photochemically initiated protein splicing. *Angew Chem Int Ed Engl*, **34**:1629–1630. 206, 207

Cooper, S. J., Leonard, G. A., McSweeney, S. M., Thompson, A. W., Naismith, J. H., Qamar, S., Plater, A., Berry, A. and Hunter, W. N. (1996) The crystal structure of a class II fructose-1,6-bisphosphate aldolase shows a novel binuclear metal-binding active site embedded in a familiar fold. *Structure*, **4**(11):1303–15. 34

Copley, R. R. and Bork, P. (2000) Homology among (betaalpha)(8) barrels: implications for the evolution of metabolic pathways. *J Mol Biol*, **303**(4):627–41. 134, 135

Copley, S. D. (2000) Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. *Trends Biochem Sci*, **25**(6):261–5. 7

Cossu, G., Melis, M., Floris, G., Hayflick, S. J. and Spissu, A. (2002) Hallervorden Spatz syndrome (pantothenate kinase associated neurodegeneration) in two Sardinian brother with homozygous mutation in PANK 2 gene. *J Neurol*, **249**(11):1599–600. 21

Cowtan, K. D. and Zhang, K. Y. (1999) Density modification for macromolecular phase improvement. *Prog Biophys Mol Biol*, **72**(3):245–70. 106

Cronan, J. E. (1980) Beta-alanine synthesis in Escherichia coli. *J Bacteriol*, **141**(3):1291–7. 10, 11, 13, 35

Cronan, J. E., Littel, K. J. and Jackowski, S. (1982) Genetic and biochemical analyses of pantothenate biosynthesis in Escherichia coli and Salmonella typhimurium. *J Bacteriol*, **149**(3):916–22. 10, 11, 13, 19

Dalgaard, J., Moser, M., Hughey, R. and Mian, I. (1997) Statistical modeling, phylogenetic analysis and structure prediction of a protein splicing domain common to inteins and hedgehog proteins. *J Comput Biol*, **4**(2):193–214. 44

Dandekar, T., Schuster, S., Snel, B., Huynen, M. and Bork, P. (1999) Pathway alignment: application to the comparative analysis of glycolytic enzymes. *Biochem J*, **343 (Pt 1)**:115–24. 4, 5

Daugherty, M., Polanuyer, B., Farrell, M., Scholle, M., Lykidis, A., de Crecy-Lagard, V. and Osterman, A. (2002) Complete reconstitution of the human coenzyme A biosynthetic pathway via comparative genomics. *J Biol Chem*, **277**(24):21431–9. 16, 17

Dauter, Z. (1999) Data-collection strategies. *Acta Crystallogr D Biol Crystallogr*, **55 (Pt 10)**:1703–17. 92

Davies, G. J., Dodson, G. G., Hubbard, R. E., Tolley, S. P., Dauter, Z., Wilson, K. S., Hjort, C., Mikkelsen, J. M., Rasmussen, G. and Schulein, M. (1993) Structure and function of endoglucanase V. *Nature*, **365**(6444):362–4. 50

DeLano, W. L. (2002) The PyMOL Molecular Graphics system. **http://www.pymol.org**. 49, 115, 176, 177, 195, 200

Diaz, E. and Anton, D. L. (1991) Alkylation of an active-site cysteinyl residue during substrate- dependent inactivation of Escherichia coli S-adenosylmethionine decarboxylase. *Biochemistry*, **30**(16):4078–81. 48

Ding, Y., Xu, M. Q., Ghosh, I., Chen, X., Ferrandon, S., Lesage, G. and Rao, Z. (2003) Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J Biol Chem*, **278**(40):39133–42. 44

Ditzel, L., Huber, R., Mann, K., Heinemeyer, W., Wolf, D. H. and Groll, M. (1998) Conformational constraints for protein self-cleavage in the proteasome. *J Mol Biol*, **279**(5):1187–91. 184, 191, 198

Dodson, E. (1998) The role of validation in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **54**(Pt 6 Pt 1):1109–18. 109

Dodson, E. J., Davies, G. J., Lamzin, V. S., Murshudov, G. N. and Wilson, K. S. (1998) Validation tools: can they indicate the information content of macromolecular crystal structures? *Structure*, **6**(6):685–90. 109

Doherty, E. A. and Doudna, J. A. (2000) Ribozyme structures and mechanisms. *Annu Rev Biochem*, **69**:597–615. 43

Dougherty, D. A. (2000) Unnatural amino acids as probes of protein structure and function. *Curr Opin Chem Biol*, **4**(6):645–52. 206

Downs, D. M. and Roth, J. R. (1991) Synthesis of thiamine in Salmonella typhimurium independent of the purF function. *J Bacteriol*, **173**(20):6597–604. 25

Dreyer, M. K. and Schulz, G. E. (1996) Catalytic mechanism of the metal-dependent fuculose aldolase from Escherichia coli as derived from the structure. *J Mol Biol*, **259**(3):458–66. 150

Duan, X., Gimble, F. S. and Quiocho, F. A. (1997) Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell*, **89**(4):555–64. 44

Duggleby, H. J., Tolley, S. P., Hill, C. P., Dodson, E. J., Dodson, G. and Moody, P. C. (1995) Penicillin acylase has a single-amino-acid catalytic centre. *Nature*, **373**(6511):264–8. 44

Dusch, N., Puhler, A. and Kalinowski, J. (1999) Expression of the Corynebacterium glutamicum panD gene encoding L- aspartate-alpha-decarboxylase leads to pantothenate overproduction in Escherichia coli. *Appl Environ Microbiol*, **65**(4):1530–9. 11, 13, 22

Ekstrom, J. L., Mathews, I., Stanley, B. A., Pegg, A. E. and Ealick, S. E. (1999) The crystal structure of human S-adenosylmethionine decarboxylase at 2.25 A resolution reveals a novel fold. *Structure Fold Des*, **7**(5):583–95. 38, 198

Ekstrom, J. L., Tolbert, W. D., Xiong, H., Pegg, A. E. and Ealick, S. E. (2001) Structure of a human S-adenosylmethionine decarboxylase self-processing ester intermediate and mechanism of putrescine stimulation of processing as revealed by the H243A mutant. *Biochemistry*, **40**(32):9495–504. 36, 198

Elischewski, F., Puhler, A. and Kalinowski, J. (1999) Pantothenate production in Escherichia coli K12 by enhanced expression of the panE gene encoding ketopantoate reductase. *J Biotechnol*, **75**(2-3):135–46. 10, 22

Elliott, D. (1952) A search of specific chemical methods for fission of peptide bonds.1. The N-acyl to O-acyl transformation in the degradation of silk fibroin. *Biochemical Journal*, **50**:542–550. 202

Engh, R. and Huber, R. (1991) Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement. *Acta Crystallogr A*, **47**:392–400. 102, 109

Evans, P. R. (1999) Some notes on choices in data collection. *Acta Crystallogr D Biol Crystallogr*, **55 (Pt 10)**:1771–2. 92

Evans, T. C., Benner, J. and Xu, M. Q. (1999) The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *J Biol Chem*, **274**(26):18359–63. 54

Falk, K. L. and Guerra, D. J. (1993) Coenzyme A biosynthesis in plants: partial purification and characterization of pantothenate kinase from spinach. *Arch Biochem Biophys*, **301**(2):424–30. 18

Farber, G. K. and Petsko, G. A. (1990) The evolution of alpha/beta barrel enzymes. *Trends Biochem Sci*, **15**(6):228–34. 135

Fink, R. M., Fink, K. and Henderson, R. B. (1953) beta-amino acid formation by tissue slices incubated with pyrimidines. *J Biol Chem*, **201**(1):349–55. 14

Fisher, M. N., Robishaw, J. D. and Neely, J. R. (1985) The properties and regulation of pantothenate kinase from rat heart. *J Biol Chem*, **260**(29):15745–51. 18

Flechner, A., Gross, W., Martin, W. F. and Schnarrenberger, C. (1999) Chloroplast class I and class II aldolases are bifunctional for fructose-1,6-biphosphate and sedoheptulose-1,7-biphosphate cleavage in the Calvin cycle. *FEBS Lett*, **447**(2-3):200–2. 28

Frodyma, M., Rubio, A. and Downs, D. M. (2000) Reduced flux through the purine biosynthetic pathway results in an increased requirement for coenzyme A in thiamine synthesis in Salmonella enterica serovar typhimurium. *J Bacteriol*, **182**(1):236–40. 19

Frodyma, M. E. and Downs, D. (1998a) ApbA, the ketopantoate reductase enzyme of Salmonella typhimurium is required for the synthesis of thiamine via the alternative pyrimidine biosynthetic pathway. *J Biol Chem*, **273**(10):5572–6. 10, 13

Frodyma, M. E. and Downs, D. (1998b) The panE gene, encoding ketopantoate reductase, maps at 10 minutes and is allelic to apbA in Salmonella typhimurium. *J Bacteriol*, **180**(17):4757–9. 10, 13

Gallagher, T., Rozwarski, D. A., Ernst, S. R. and Hackert, M. L. (1993) Refined structure of the pyruvoyl-dependent histidine decarboxylase from Lactobacillus 30a. *J Mol Biol*, **230**(2):516–28. 38, 39, 188

Gallagher, T., Snell, E. E. and Hackert, M. L. (1989) Pyruvoyl-dependent histidine decarboxylase. Active site structure and mechanistic analysis. *J Biol Chem*, **264**(21):12737–43. 37

Gallet, A., Rodriguez, R., Ruel, L. and Therond, P. P. (2003) Cholesterol modification of hedgehog is required for trafficking and movement, revealing an asymmetric cellular response to hedgehog. *Dev Cell*, **4**(2):191–204. 42

Garman, E. F. and Schneider, T. (1997) Macromolecular Cryocrystallography. *J App Cryst*, **30**:211–237. 87

Geerlof, A., Lewendon, A. and Shaw, W. V. (1999) Purification and characterization of phosphopantetheine adenylyltransferase from Escherichia coli. *J Biol Chem*, **274**(38):27105–11. 16

Gelfman, C. M., Copeland, W. C. and Robertus, J. D. (1991) Site-directed alteration of four active-site residues of a pyruvoyl- dependent histidine decarboxylase. *Biochemistry*, **30**(4):1057–62. 197

Genschel, U., Powell, C. A., Abell, C. and Smith, A. G. (1999) The final step of pantothenate biosynthesis in higher plants: cloning and characterization of pantothenate synthetase from Lotus japonicus and Oryza sativum (rice). *Biochem J*, **341 (Pt 3)**:669–78. 14

Gerdes, S. Y., Scholle, M. D., D'Souza, M., Bernal, A., Baev, M. V., Farrell, M., Kurnasov, O. V., Daugherty, M. D., Mseeh, F., Polanuyer, B. M., Campbell, J. W., Anantha, S., Shatalin, K. Y., Chowdhury, S. A., Fonstein, M. Y. and Osterman, A. L. (2002) From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. *J Bacteriol*, **184**(16):4555–72. 16, 23

Gerlt, J. A. (2000) New wine from old barrels. *Nat Struct Biol*, **7**(3):171–3. 134

Gerlt, J. A. and Babbitt, P. C. (2001) Barrels in pieces? *Nat Struct Biol*, **8**(1):5–7. 34

Gerrard, J. A., Sparrow, A. D. and Wells, J. A. (2001) Metabolic databases - what next? *Trends Biochem Sci*, **26**(2):137–40. 2

Gill, S. and von Hippel, P. (1989) Calculation of protein extinction coefficients from amino-acid sequence data. *Analytical Biochemistry*, **182**:319–326. 76, 215

Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L. and Hilario, E. (2002) Inteins: structure, function, and evolution. *Annu Rev Microbiol*, **56**:263–87. 43

Groll, M., Heinemeyer, W., Jager, S., Ullrich, T., Bochtler, M., Wolf, D. H. and Huber, R. (1999) The catalytic sites of 20S proteasomes and their role in subunit maturation: a mutational and crystallographic study. *Proc Natl Acad Sci U S A*, **96**(20):10976–83. 41

Guan, C., Cui, T., Rao, V., Liao, W., Benner, J., Lin, C. L. and Comb, D. (1996) Activation of glycosylasparaginase. Formation of active N-terminal threonine by intramolecular autoproteolysis. *J Biol Chem*, **271**(3):1732–7. 41

Guan, C., Liu, Y., Shao, Y., Cui, T., Liao, W., Ewel, A., Whitaker, R. and Paulus, H. (1998) Characterization and functional analysis of the cis-autoproteolysis active center of glycosylasparaginase. *J Biol Chem*, **273**(16):9695–702. 41, 200, 201

Guo, H. C., Xu, Q., Buckley, D. and Guan, C. (1998) Crystal structures of Flavobacterium glycosylasparaginase. An N- terminal nucleophile hydrolase activated by intramolecular proteolysis. *J Biol Chem*, **273**(32):20205–12. 44, 200

Hall, D. R., Bond, C. S., Leonard, G. A., Watt, C. I., Berry, A. and Hunter, W. N. (2002) Structure of tagatose-1,6-bisphosphate aldolase. Insight into chiral discrimination, mechanism, and specificity of class II aldolases. *J Biol Chem*, **277**(24):22018–24. 150

Hall, D. R., Leonard, G. A., Reed, C. D., Watt, C. I., Berry, A. and Hunter, W. N. (1999) The crystal structure of Escherichia coli class II fructose-1, 6- bisphosphate aldolase in complex with phosphoglycolohydroxamate reveals details of mechanism and specificity. *J Mol Biol*, **287**(2):383–94. 150

Hall, T. M., Porter, J. A., Beachy, P. A. and Leahy, D. J. (1995) A potential catalytic site revealed by the 1.7-A crystal structure of the amino-terminal signalling domain of Sonic hedgehog. *Nature*, **378**(6553):212–6. 44

Hall, T. M., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A. and Leahy, D. J. (1997) Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell*, **91**(1):85–97. 43, 44, 199

Halvorsen, O. (1983) Effects of hypolipidemic drugs on hepatic CoA. *Biochem Pharmacol*, **32**(6):1126–8. 18, 20

Halvorsen, O. and Skrede, S. (1982) Regulation of the biosynthesis of CoA at the level of pantothenate kinase. *Eur J Biochem*, **124**(1):211–5. 18, 20

Hammerschmidt, M., Brook, A. and McMahon, A. P. (1997) The world according to hedgehog. *Trends Genet*, **13**(1):14–21. 42

Hayflick, S. J. (2003a) Pantothenate kinase-associated neurodegeneration (formerly Hallervorden-Spatz syndrome). *J Neurol Sci*, **207**(1-2):106–7. 21

Hayflick, S. J. (2003b) Unraveling the Hallervorden-Spatz syndrome: pantothenate kinase-associated neurodegeneration is the name. *Curr Opin Pediatr*, **15**(6):572–7. 21

Hayflick, S. J., Westaway, S. K., Levinson, B., Zhou, B., Johnson, M. A., Ching, K. H. and Gitschier, J. (2003) Genetic, clinical, and radiographic delineation of Hallervorden-Spatz syndrome. *N Engl J Med*, **348**(1):33–40. 21

Hegyi, H. and Gerstein, M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J Mol Biol*, **288**(1):147–64. 134

Herzberg, O., Chen, C. C., Kapadia, G., McGuire, M., Carroll, L. J., Noh, S. J. and Dunaway-Mariano, D. (1996) Swiveling-domain mechanism for enzymatic phosphotransfer between remote reaction sites. *Proc Natl Acad Sci U S A*, **93**(7):2652–7. 150

Herzberg, O., Chen, C. C., Liu, S., Tempczyk, A., Howard, A., Wei, M., Ye, D. and Dunaway-Mariano, D. (2002) Pyruvate site of pyruvate phosphate dikinase: crystal structure of the enzyme-phosphonopyruvate complex, and mutant analysis. *Biochemistry*, **41**(3):780–7. 150

Hoagland, M. B. and Novelli, G. D. (1954) Biosynthesis of coenzyme A from phospho-pantetheine and of pantetheine from pantothenate. *J Biol Chem*, **207**(2):767–73. 17

Hocker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A. and Sterner, R. (2001) Dissection of a (betaalpha)8-barrel enzyme into two folded halves. *Nat Struct Biol*, **8**(1):32–6. 135

Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol*, **233**(1):123–38. 114

Holt, P. J., Stephens, L. D., Bruce, N. C. and Lowe, C. R. (1995) An amperometric opiate assay. *Biosens Bioelectron*, **10**(6-7):517–26. 214

Hooft, R. W., Sander, C. and Vriend, G. (1996a) Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins*, **26**(4):363–76. 170

Hooft, R. W., Vriend, G., Sander, C. and Abola, E. E. (1996b) Errors in protein structures. *Nature*, **381**(6580):272. 110, 111

Hoppensack, A., Rehm, B. H. and Steinbuchel, A. (1999) Analysis of 4-phosphopantetheinylation of polyhydroxybutyrate synthase from Ralstonia eutropha: generation of beta-alanine auxotrophic Tn5 mutants and cloning of the panD gene region. *J Bacteriol*, **181**(5):1429–35. 13

Horowitz, N. (1945) On the evolution of biochemical synthesis. *Proc Natl Acad Sci U S A*, **31**:153–157. 8

Hortnagel, K., Prokisch, H. and Meitinger, T. (2003) An isoform of hPANK2, deficient in pantothenate kinase-associated neurodegeneration, localizes to mitochondria. *Hum Mol Genet*, **12**(3):321–7. 20

Howlin, B., Moss, D., Harris, G. and H.P.C., D. (1993) TLSANL:TLS parameter-analysis program for segmented anisotropic refinement of macromolecular structures. *J Appl Cryst*, **26**:622–624. 106, 180

Howlin, B., Moss, D. S. and Harris, G. W. (1989) Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model. *Acta Crystallogr A*, **45 (Pt 12)**:851–61. 106

Huang, K., Li, Z., Jia, Y., Dunaway-Mariano, D. and Herzberg, O. (1999) Helix swapping between two alpha/beta barrels: crystal structure of phosphoenolpyruvate mutase with bound Mg(2+)-oxalate. *Structure Fold Des*, **7**(5):539–48. 139, 143

Hugo, E. R. and Byers, T. J. (1993) S-adenosyl-L-methionine decarboxylase of Acanthamoeba castellanii (Neff): purification and properties. *Biochem J*, **295 (Pt 1)**:203–9. 38

Huynen, M. A., Dandekar, T. and Bork, P. (1999) Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol*, **7**(7):281–91. 4, 5

Huynh, Q. K., Recsei, P. A., Vaaler, G. L. and Snell, E. E. (1984a) Histidine decarboxylase of Lactobacillus 30a. Sequences of the overlapping peptides, the complete alpha chain, and prohistidine decarboxylase. *J Biol Chem*, **259**(5):2833–9. 37

Huynh, Q. K. and Snell, E. E. (1985a) Pyruvoyl-dependent histidine decarboxylases. Comparative sequences of cysteinyl peptides of the enzymes from Lactobacillus 30a, Lactobacillus buchneri, and Clostridium perfringens. *J Biol Chem*, **260**(5):2794–7. 38

Huynh, Q. K. and Snell, E. E. (1985b) Pyruvoyl-dependent histidine decarboxylases. Preparation and amino acid sequences of the beta chains of histidine decarboxylase from Clostridium perfringens and Lactobacillus buchneri. *J Biol Chem*, **260**(5):2798–803. 38

Huynh, Q. K. and Snell, E. E. (1986a) Histidine decarboxylase of Lactobacillus 30a. Hydroxylamine cleavage of the -seryl-seryl- bond at the activation site of prohistidine decarboxylase. *J Biol Chem*, **261**(4):1521–4. 37

Huynh, Q. K. and Snell, E. E. (1986b) Pyruvoyl-dependent histidine decarboxylase from Lactobacillus 30a. Covalent modifications of aspartic acid 191, lysine 155, and the pyruvoyl group. *J Biol Chem*, **261**(10):4389–94. 37

Huynh, Q. K., Vaaler, G. L., Recsei, P. A. and Snell, E. E. (1984b) Histidine decarboxylase of Lactobacillus 30a. Sequences of the cyanogen bromide peptides from the alpha chain. *J Biol Chem*, **259**(5):2826–32. 37

Hyvonen, M. and Saraste, M. (1997) *Expression of cDNAs of Escherichia coli using the T7 promoter.*, vol. 4 of *Cell biology - A laboratory manual*. Academic Press, 2nd edn. 65

Ikonen, E., Baumann, M., Gron, K., Syvanen, A. C., Enomaa, N., Halila, R., Aula, P. and Peltonen, L. (1991) Aspartylglucosaminuria: cDNA encoding human aspartylglucosaminidase and the missense mutation causing the disease. *EMBO J*, **10**(1):51–8. 41

Isupov, M. N., Obmolova, G., Butterworth, S., Badet-Denisot, M. A., Badet, B., Polikarpov, I., Littlechild, J. A. and Teplyakov, A. (1996) Substrate binding is required for assembly of the active conformation of the catalytic site in Ntn amidotransferases: evidence from the 1.8 A crystal structure of the glutaminase domain of glucosamine 6-phosphate synthase. *Structure*, **4**(7):801–10. 44

Iwai, K. and Ando, T. (1967) *N-O acyl shifts*, vol. 11 of *Methods IN ENZYMOLOGY*. Academic Press. 39, 173, 203

Izard, T. (2003) A novel adenylate binding site confers phosphopantetheine adenylyltransferase interactions with coenzyme A. *J Bacteriol*, **185**(14):4074–80. 17

Izard, T. and Blackwell, N. C. (2000) Crystal structures of the metal-dependent 2-dehydro-3-deoxy-galactarate aldolase suggest a novel reaction mechanism. *EMBO J*, **19**(15):3849–56. 16, 17, 145, 148

Izard, T. and Geerlof, A. (1999) The crystal structure of a novel bacterial adenylyltransferase reveals half of sites reactivity. *EMBO J*, **18**(8):2021–30. 16, 17

Jackowski, S. (1996) *Biosynthesis of pantothenic acid and coenzyme A*, vol. Escherichia coli and Salmonella: cellular and molecular biology, 2nd ed. of *Escherichia coli and Salmonella: cellular and molecular biology, 2nd ed.* American Society for Microbiology, Washington, D.C. 13, 15, 18

Jackowski, S. and Alix, J. H. (1990) Cloning, sequence, and expression of the pantothenate permease (panF) gene of Escherichia coli. *J Bacteriol*, **172**(7):3842–8. 11

Jackowski, S. and Rock, C. O. (1981) Regulation of coenzyme A biosynthesis. *J Bacteriol*, **148**(3):926–32. 11, 18, 19, 71

Jackowski, S. and Rock, C. O. (1984) Metabolism of 4'-phosphopantetheine in Escherichia coli. *J Bacteriol*, **158**(1):115–20. 19

Jansen, R. and Gerstein, M. (2000) Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins. *Nucleic Acids Res*, **28**(6):1481–8. 134

Jardine, O., Gough, J., Chothia, C. and Teichmann, S. A. (2002) Comparison of the small molecule metabolic enzymes of Escherichia coli and Saccharomyces cerevisiae. *Genome Res*, **12**(6):916–29. 5

Jensen, R. A. (1976) Enzyme recruitment in evolution of new function. *Annu Rev Microbiol*, **30**:409–25. 7

John, R. A. (1995) Pyridoxal phosphate-dependent enzymes. *Biochim Biophys Acta*, **1248**(2):81–96. 45

Jones, C. E., Brook, J. M., Buck, D., Abell, C. and Smith, A. G. (1993) Cloning and sequencing of the Escherichia coli panB gene, which encodes ketopantoate hydroxymethyltransferase, and overexpression of the enzyme. *J Bacteriol*, **175**(7):2125–30. 14, 25, 118, 119

Jones, C. E., Dancer, J., Abell, C. and Smith, A. G. (1994) Evidence for the pathway to pantothenate in plants. *Can J Chem*, **72**:261–263. 14

Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*, **47**(Pt 2):110–9. 108

Jurgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M. and Sterner, R. (2000) Directed evolution of a (beta alpha)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci U S A*, **97**(18):9925–30. 135

Kabisch, U. C., Grantzdorffer, A., Schierhorn, A., Rucknagel, K. P., Andreesen, J. R. and Pich, A. (1999) Identification of D-proline reductase from Clostridium sticklandii as a selenoenzyme and indications for a catalytically active pyruvoyl group derived from a cysteine residue by cleavage of a proprotein. *J Biol Chem*, **274**(13):8445–54. 38

Kabsch, W. (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr A*, **32**:922–923. 112

Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**(12):2577–637. 114

Kaefer, E. (1958) An 8-chromosome map of Aspergillus nidulans. *Adv Genet*, **9**:105–145. 14

Kai, Y., Matsumura, H., Inoue, T., Terada, K., Nagara, Y., Yoshinaga, T., Kihara, A., Tsumura, K. and Izui, K. (1999) Three-dimensional structure of phosphoenolpyruvate carboxylase: a proposed mechanism for allosteric inhibition. *Proc Natl Acad Sci U S A*, **96**(3):823–8. 150

Kallen, R. G. and Jencks, W. P. (1966) The mechanism of the condensation of formaldehyde with tetrahydrofolic acid. *J Biol Chem*, **241**(24):5851–63. 26

Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M. and Stevens, T. H. (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science*, **250**(4981):651–7. 43

Kanehisa, M., Goto, S., Kawashima, S. and Nakaya, A. (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res*, **30**(1):42–6. 5

Karasawa, T., Yoshida, K., Furukawa, K. and Hosoki, K. (1972) Feedback inhibition of pantothenate kinase by coenzyme A and possible role of the enzyme for the regulation of cellular coenzyme A level. *J Biochem Tokyo*, **71**(6):1065–7. 18

Karp, P. D., Paley, S. and Romero, P. (2002a) The Pathway Tools software. *Bioinformatics*, **18 Suppl 1**:S225–32. 4

Karp, P. D., Riley, M., Paley, S. M. and Pellegrini-Toole, A. (2002b) The MetaCyc Database. *Nucleic Acids Res*, **30**(1):59–61. 5

Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Collado-Vides, J., Paley, S. M., Pellegrini-Toole, A., Bonavides, C. and Gama-Castro, S. (2002c) The EcoCyc Database. *Nucleic Acids Res*, **30**(1):56–8. 3

Kashiwagi, K., Taneja, S. K., Liu, T. Y., Tabor, C. W. and Tabor, H. (1990) Spermidine biosynthesis in Saccharomyces cerevisiae. Biosynthesis and processing of a proenzyme form of S-adenosylmethionine decarboxylase. *J Biol Chem*, **265**(36):22321–8. 38

Kataoka, M., Shimizu, S. and Yamada, H. (1992) Purification and characterization of a novel FMN-dependent enzyme. Membrane-bound L-(+)-pantoyl lactone dehydrogenase from Nocardia asteroides. *Eur J Biochem*, **204**(2):799–806. 13

Keefe, A. D., Newton, G. L. and Miller, S. L. (1995) A possible prebiotic synthesis of pantetheine, a precursor to coenzyme A. *Nature*, **373**(6516):683–5. 9

Kelley, L. A., MacCallum, R. M. and Sternberg, M. J. (2000) Enhanced genome annotation using structural profiles in the program 3D- PSSM. *J Mol Biol*, **299**(2):499–520. 113

Kim, A. D., Graham, D. E., Seeholzer, S. H. and Markham, G. D. (2000a) S-Adenosylmethionine decarboxylase from the archaeon Methanococcus jannaschii: identification of a novel family of pyruvoyl enzymes. *J Bacteriol*, **182**(23):6667–72. 38

Kim, J. and Dunaway-Mariano, D. (1996) Phosphoenolpyruvate mutase catalysis of phosphoryl transfer in phosphoenolpyruvate: kinetics and mechanism of phosphorus-carbon bond formation. *Biochemistry*, **35**(14):4628–35. 139, 143

Kim, J. K., Yang, I. S., Rhee, S., Dauter, Z., Lee, Y. S., Park, S. S. and Kim, K. H. (2003) Crystal structures of glutaryl 7-aminocephalosporanic acid acylase: insight into autoproteolytic activation. *Biochemistry*, **42**(14):4084–93. 44, 201

Kim, Y., Kim, S., Earnest, T. N. and Hol, W. G. (2002) Precursor structure of cephalosporin acylase. Insights into autoproteolytic activation in a new N-terminal hydrolase family. *J Biol Chem*, **277**(4):2823–9. 201

Kim, Y., Yoon, K., Khang, Y., Turley, S. and Hol, W. G. (2000b) The 2.0 A crystal structure of cephalosporin acylase. *Structure Fold Des*, **8**(10):1059–68. 44

King, H. L., Dyar, R. E. and Wilken, D. R. (1974) Ketopantoyl lactone and ketopantoic acid reductases. Characterization of the reactions and purification of two forms of ketopantoyl lactone reductase. *J Biol Chem*, **249**(15):4689–95. 10, 14

King, H. L. and Wilken, D. R. (1972) Separation and Preliminary Studies on 2-Ketopantoyl Lactone and 2-Ketopantoic Acid Reductases of Yeast. *J Biol Chem*, **247**:4096–4098. 14

Kirschbaum, N., Clemons, R., Marino, K. A., Sheedy, G., Nguyen, M. L. and Smith, C. M. (1990) Pantothenate kinase activity in livers of genetically diabetic mice (db/db) and hormonally treated cultured rat hepatocytes. *J Nutr*, **120**(11):1376–86. 18

Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R., J. and Sacchettini, J. C. (1998) Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing. *Nat Struct Biol*, **5**(1):31–6. 188, 199

Kleinkauf, H. and von Dohren, H. (1995) Linking peptide and polyketide biosynthesis. *J Antibiot Tokyo*, **48**(7):563–7. 15

Kleinkauf, H. and Von Dohren, H. (1996) A nonribosomal system of peptide biosynthesis. *Eur J Biochem*, **236**(2):335–51. 15

Kleywegt, G. J. (2000) Validation of protein crystal structures. *Acta Crystallogr D Biol Crystallogr*, **56 (Pt 3)**:249–65. 109

Kleywegt, G. J. and Brunger, A. T. (1996) Checking your imagination: applications of the free R value. *Structure*, **4**(8):897–904. 105, 109

Kleywegt, G. J. and Jones, T. A. (1994) A super position. *CCP4ESF EACBM Newsletter on Protein Crystallography*, **31**:9–14. 112

Kleywegt, G. J. and Jones, T. A. (1995) Where freedom is given, liberties are taken. *Structure*, **3**(6):535–40. 105

Kleywegt, G. J. and Jones, T. A. (1998) Databases in protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **54**(Pt 6 Pt 1):1119–31. 108

Kleywegt, G. J. and Jones, T. A. (2002) Homo crystallographicus–quo vadis? *Structure Camb*, **10**(4):465–72. 109

Kohls, D., Sulea, T., Purisima, E. O., MacKenzie, R. E. and Vrielink, A. (2000) The crystal structure of the formiminotransferase domain of formiminotransferase-cyclodeaminase: implications for substrate channeling in a bifunctional enzyme. *Structure Fold Des*, **8**(1):35–46. 117

Kondrup, J. and Grunnet, N. (1973) The effect of acute and prolonged ethanol treatment on the contents of coenzyme A, carnitine and their derivatives in rat liver. *Biochem J*, **132**(3):373–9. 20

Kraulis, P. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst*, **24**:946–950. 33, 115

Kuhn, R. and Wieland, T. (1942) Title not retrievable. *Chem Ber*, **75**:121. 24

Kupke, T. (2001) Molecular characterization of the 4'-phosphopantothenoylcysteine decarboxylase domain of bacterial Dfp flavoproteins. *J Biol Chem*, **276**(29):27597–604. 16

Kupke, T. (2002) Molecular characterization of the 4'-phosphopantothenoylcysteine synthetase domain of bacterial dfp flavoproteins. *J Biol Chem*, **277**(39):36137–45. 16

Kupke, T., Hernandez-Acosta, P., Steinbacher, S. and Culianez-Macia, F. A. (2001) Arabidopsis thaliana flavoprotein AtHAL3a catalyzes the decarboxylation of 4'-Phosphopantothenoylcysteine to 4'-phosphopantetheine, a key step in coenzyme A biosynthesis. *J Biol Chem*, **276**(22):19190–6. 16

Kupke, T., Uebele, M., Schmid, D., Jung, G., Blaesse, M. and Steinbacher, S. (2000) Molecular characterization of lantibiotic-synthesizing enzyme EpiD reveals a function for bacterial Dfp proteins in coenzyme A biosynthesis. *J Biol Chem*, **275**(41):31838–46. 16

Kurtov, D., Kinghorn, J. R. and Unkles, S. E. (1999) The Aspergillus nidulans panB gene encodes ketopantoate hydroxymethyltransferase, required for biosynthesis of pantothenate and Coenzyme A. *Mol Gen Genet*, **262**(1):115–20. 14, 25

Kwon, A., Lee, B., Han, B., Ahn, H., Yang, J., Yoon, H. and Suh, S. (2002) Crystallization and preliminary x-ray crystallographic analysis of aspartate 1-decarboxylase from Helicobacter pylori. *Acta Crystallogr D Biol Crystallogr*, **58**:861–863. 35, 52, 204

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**(259):680–5. 62

Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. and Wilmanns, M. (2000) Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science*, **289**(5484):1546–50. 135

Larsen, T. M., Benning, M. M., Rayment, I. and Reed, G. H. (1998) Structure of the bis(Mg2+)-ATP-oxalate complex of the rabbit muscle pyruvate kinase at 2.1 A resolution: ATP binding over a barrel. *Biochemistry*, **37**(18):6247–55. 149

Laskowski, R. A., M.W., M., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst*, **26**:283–291. 111

Lee, J. J., Ekker, S. C., von Kessler, D. P., Porter, J. A., Sun, B. I. and Beachy, P. A. (1994) Autoproteolysis in hedgehog protein biogenesis. *Science*, **266**(5190):1528–37. 42, 44

Lee, Y. S., Kim, H. W. and Park, S. S. (2000) The role of alpha-amino group of the N-terminal serine of beta subunit for enzyme catalysis and autoproteolytic activation of glutaryl 7- aminocephalosporanic acid acylase. *J Biol Chem*, **275**(50):39200–6. 41

Lee, Y. S. and Park, S. S. (1998) Two-step autocatalytic processing of the glutaryl 7- aminocephalosporanic acid acylase from Pseudomonas sp. strain GK16. *J Bacteriol*, **180**(17):4576–82. 41, 201

Leys, D., Basran, J. and Scrutton, N. S. (2003) Channelling and formation of 'active' formaldehyde in dimethylglycine oxidase. *EMBO J*, **22**(16):4038–48. 117

Li, Q. X. and Dowhan, W. (1988) Structural characterization of Escherichia coli phosphatidylserine decarboxylase. *J Biol Chem*, **263**(23):11516–22. 38

Li, Q. X. and Dowhan, W. (1990) Studies on the mechanism of formation of the pyruvate prosthetic group of phosphatidylserine decarboxylase from Escherichia coli. *J Biol Chem*, **265**(7):4111–5. 38

Li, S., Smith, J. L. and Zalkin, H. (1999a) Mutational analysis of Bacillus subtilis glutamine phosphoribosylpyrophosphate amidotransferase propeptide processing. *J Bacteriol*, **181**(5):1403–8. 41, 201

Li, Y., Chen, J., Jiang, W., Mao, X., Zhao, G. and Wang, E. (1999b) In vivo post-translational processing and subunit reconstitution of cephalosporin acylase from Pseudomonas sp.130. *Eur J Biochem*, **262**:713–719. 41, 201

Lindley, P. F. (1999) Macromolecular crystallography with a third-generation synchrotron source. *Acta Crystallogr D Biol Crystallogr*, **55 (Pt 10)**:1654–62. 90

Liu, X. Q. (2000) Protein-splicing intein: Genetic mobility, origin, and evolution. *Annu Rev Genet*, **34**:61–76. 43, 44

Liu, Y., Guan, C. and Aronson, N. N., J. (1998) Site-directed mutagenesis of essential residues involved in the mechanism of bacterial glycosylasparaginase. *J Biol Chem*, **273**(16):9688–94. 41, 200

Lobley, C. M., Schmitzberger, F., Kilkenny, M. L., Whitney, H., Ottenhof, H. H., Chakauya, E., Webb, M. E., Birch, L. M., Tuck, K. L., Abell, C., Smith, A. G. and Blundell, T. L. (2003) Structural insights into the evolution of the pantothenate-biosynthesis pathway. *Biochem Soc Trans*, **31**(Pt 3):563–71. 153

Lopaschuk, G. D., Hansen, C. A. and Neely, J. R. (1986) Fatty acid metabolism in hearts containing elevated levels of CoA. *Am J Physiol*, **250**(3 Pt 2):H351–9. 19

Lopaschuk, G. D., Michalak, M. and Tsang, H. (1987) Regulation of pantothenic acid transport in the heart. Involvement of a Na+-cotransport system. *J Biol Chem*, **262**(8):3615–9. 11

Lovell, S. C., Davis, I. W., Arendall, W. B., r., De Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*, **50**(3):437–50. 111, 170, 172

Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. and Huber, R. (1995) Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. *Science*, **268**(5210):533–9. 44

Ludvigsen, S. and Poulsen, F. M. (1992) Three-dimensional structure in solution of barwin, a protein from barley seed. *Biochemistry*, **31**(37):8783–9. 50

Luo, Y., Samuel, J., Mosimann, S. C., Lee, J. E., Tanner, M. E. and Strynadka, N. C. (2001) The structure of L-ribulose-5-phosphate 4-epimerase: an aldolase-like platform for epimerization. *Biochemistry*, **40**(49):14763–71. 150

Maas, W. K. (1952a) The biosynthesis of pantothenic acid. In *4th International Congress of Biochemistry*, pp. 161–168. 10

Maas, W. K. (1952b) Pantothenate studies. III. Description of the extracted pantothenate-synthesizing enzyme of Escherichia coli. *J Biol Chem*, **198**(1):23–32. 11

Maas, W. K. and Vogel, H. J. (1953) alpha-Ketoisovaleric acid, a precursor of pantothenic acid in Escherichia coli. *J Bacteriol*, **65**(4):388–93. 24

Makarova, K. S., Aravind, L., Galperin, M. Y., Grishin, N. V., Tatusov, R. L., Wolf, Y. I. and Koonin, E. V. (1999) Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. *Genome Res*, **9**(7):608–28. 4

Manoj, N., Strauss, E., Begley, T. P. and Ealick, S. E. (2003) Structure of human phosphopantothenoylcysteine synthetase at 2.3 A resolution. *Structure Camb*, **11**(8):927–36. 17

Markham, G. D., Tabor, C. W. and Tabor, H. (1982) S-adenosylmethionine decarboxylase of Escherichia coli. Studies on the covalently linked pyruvate required for activity. *J Biol Chem*, **257**(20):12063–8. 46

Martin, D. P. and Drueckhammer, D. G. (1993) Separate enzymes catalyze the final two steps of coenzyme A biosynthesis in Brevibacterium ammoniagenes: purification of pantetheine phosphate adenylyltransferase. *Biochem Biophys Res Commun*, **192**(3):1155–61. 17

Martin, J., Slade, A., Aitken, A., Arche, R. and Virden, R. (1991) Chemical modification of serine at the active site of penicillin acylase from Kluyvera citrophila. *Biochem J*, **280 (Pt 3)**:659–62. 41

Matak-Vinkovic, D., Vinkovic, M., Saldanha, S. A., Ashurst, J. L., von Delft, F., Inoue, T., Miguel, R. N., Smith, A. G., Blundell, T. L. and Abell, C. (2001) Crystal structure of Escherichia coli ketopantoate reductase at

1.7 A resolution and insight into the enzyme mechanism. *Biochemistry*, **40**(48):14493–500. 55, 153

Matsumura, H., Terada, M., Shirakata, S., Inoue, T., Yoshinaga, T., Izui, K. and Kai, Y. (1999) Plausible phosphoenolpyruvate binding site revealed by 2.6 A structure of Mn2+-bound phosphoenolpyruvate carboxylase from Escherichia coli. *FEBS Lett*, **458**(2):93–6. 150

Mattevi, A., Valentini, G., Rizzi, M., Speranza, M. L., Bolognesi, M. and Coda, A. (1995) Crystal structure of Escherichia coli pyruvate kinase type I: molecular basis of the allosteric transition. *Structure*, **3**(7):729–41. 149

McDonough, M. A., Klei, H. E. and Kelly, J. A. (1999) Crystal structure of penicillin G acylase from the Bro1 mutant strain of Providencia rettgeri. *Protein Sci*, **8**(10):1971–81. 44

McIntosh, E., Purko, M. and Wood, W. A. (1957) Ketopantoate formation by a hydroxy-methylation enzyme from Escherichia coli. *J Biol Chem*, **228**(1):499–510. 24

McPherson, A. (1999) *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor. 84

McPherson, A. (2001) A comparison of salts for the crystallization of macromolecules. *Protein Sci*, **10**(2):418–22. 162

McRee, D. E. (1999) A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol*, **125**:156–165. 108

Merkamm, M., Chassagnole, C., Lindley, N. D. and Guyonvarch, A. (2003) Ketopantoate reductase activity is only encoded by ilvC in Corynebacterium glutamicum. *J Biotechnol*, **104**(1-3):253–60. 13

Merkel, W. K. and Nichols, B. P. (1996) Characterization and sequence of the Escherichia coli panBCD gene cluster. *FEMS Microbiol Lett*, **143**(2-3):247–52. 13

Merritt, E. A. and Bacon, D. (1997) *Raster3D: Photorealistic molecular graphics*, vol. 277 of *Methods in ENZYMOLOGY*. Academic Press, San Diego. 33, 115

Mishra, P., Park, P. K. and Drueckhammer, D. G. (2001) Identification of yacE (coaE) as the structural gene for dephosphocoenzyme A kinase in Escherichia coli K-12. *J Bacteriol*, **183**(9):2774–8. 16

Miyatake, K., Nakano, Y. and Kitaoka, S. (1976) Pantothenate synthetase of Escherichia coli B. I. Physicochemical properties. *J Biochem Tokyo*, **79**(3):673–8. 11

Miyatake, K., Nakano, Y. and Kitaoka, S. (1978) Enzymological properties of pantothenate synthetase from Escherichia coli B. *J Nutr Sci Vitaminol Tokyo*, **24**(3):243–53. 11

Miyatake, K., Nakano, Y. and Kitaoka, S. (1979) *Pantothenate synthetase from Escherichia coli [D-pantoate: beta-alanine ligase (AMP-forming), EC 6.3.2.1]*, vol. 62 of *Methods in ENZYMOLOGY*. Academic Press. 11

Mizuguchi, K., Deane, C. M., Blundell, T. L., Johnson, M. S. and Overington, J. P. (1998a) JOY: protein sequence-structure representation and analysis. *Bioinformatics*, **14**(7):617–23. 114, 144, 187

Mizuguchi, K., Deane, C. M., Blundell, T. L. and Overington, J. P. (1998b) HOMSTRAD: a database of protein structure alignments for homologous families. *Protein Sci*, **7**(11):2469–71. 114

Mizutani, R., Nogami, S., Kawasaki, M., Ohya, Y., Anraku, Y. and Satow, Y. (2002) Protein-splicing reaction via a thiazolidine intermediate: crystal structure of the VMA1-derived endonuclease bearing the N and C-terminal propeptides. *J Mol Biol*, **316**(4):919–29. 199

Mononen, I., Fisher, K. J., Kaartinen, V. and Aronson, N. N., J. (1993) Aspartylglycosaminuria: protein chemistry and molecular biology of the most common lysosomal storage disorder of glycoprotein degradation. *FASEB J*, **7**(13):1247–56. 41

Morse, D. E. and Horecker, B. L. (1968) The mechanism of action of aldolases. *Adv Enzymol Relat Areas Mol Biol*, **31**:125–81. 26

Muir, T. W., Sondhi, D. and Cole, P. A. (1998) Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A*, **95**(12):6705–10. 54

Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr D Biol Crystallogr*, **53**:240–255. 101, 105, 170

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. and Dodson, E. J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D Biol Crystallogr*, **55**(Pt 1):247–55. 102

Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*, **247**(4):536–40. 42, 114

Nagano, N., Hutchinson, E. G. and Thornton, J. M. (1999) Barrel structures in proteins: automatic identification and classification including a sequence analysis of TIM barrels. *Protein Sci*, **8**(10):2072–84. 134

Nagano, N., Porter, C. T. and Thornton, J. M. (2001) The (betaalpha)(8) glycosidases: sequence and structure analyses suggest distant evolutionary relationships. *Protein Eng*, **14**(11):845–55. 136

Navaza, J. (2001) Implementation of molecular replacement in AMoRe. *Acta Crystallogr D Biol Crystallogr*, **57**(Pt 10):1367–72. 99

Neuburger, M., Day, D. A. and Douce, R. (1984) Transport of coenzyme A in plant mitochondria. *Arch Biochem Biophys*, **229**(1):253–8. 19

Noren, C. J., Wang, J. and Perler, F. B. (2000) Dissecting the Chemistry of Protein Splicing and Its Applications. *Angew Chem Int Ed Engl*, **39**(3):450–466. 43, 44

Obmolova, G., Teplyakov, A., Bonander, N., Eisenstein, E., Howard, A. J. and Gilliland, G. L. (2001) Crystal structure of dephospho-coenzyme A kinase from Haemophilus influenzae. *J Struct Biol*, **136**(2):119–25. 17

Oinonen, C. and Rouvinen, J. (2000) Structural comparison of Ntn-hydrolases. *Protein Sci*, **9**(12):2329–37. 42

Oinonen, C., Tikkanen, R., Rouvinen, J. and Peltonen, L. (1995) Three-dimensional structure of human lysosomal aspartylglucosaminidase. *Nat Struct Biol*, **2**(12):1102–8. 44, 201

O'Toole, N., Barbosa, J. A., Li, Y., Hung, L. W., Matte, A. and Cygler, M. (2003) Crystal structure of a trimeric form of dephosphocoenzyme A kinase from Escherichia coli. *Protein Sci*, **12**(2):327–36. 16, 17

Ottenhof, H. H., Ashurst, J. L., Whitney, H. M., Saldanha, S. A., Schmitzberger, F., Gweon, H. S., Blundell, T. L., Abell, C. and Smith, A. G. (2004) Organisation of the pantothenate (vitamin B5) biosynthesis pathway in higher plants. *Plant J*, **37**(1):61–72. 14, 25

Otwinowski, Z. (1991) *Processing of X-ray diffraction data in oscillation mode*, vol. 276 of *Methods in ENZYMOLOGY*. Academic Press, San Diego. 94, 95, 97, 170

Ouzounis, C. A. and Karp, P. D. (2000) Global properties of the metabolic map of Escherichia coli. *Genome Res*, **10**(4):568–76. 4, 5

Overbeek, R., Larsen, N., Pusch, G. D., D'Souza, M., Selkov, E., J., Kyrpides, N., Fonstein, M., Maltsev, N. and Selkov, E. (2000) WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res*, **28**(1):123–5. 5

Pannu, N. S. and Read, R. J. (1996) Improved Structure Refinement Through Maximum Likelihood. *Acta Crystallogr A*, **52**:659–668. 101

Pant, K., Bilwes, A. M., Adak, S., Stuehr, D. J. and Crane, B. R. (2002) Structure of a nitric oxide synthase heme protein from Bacillus subtilis. *Biochemistry*, **41**(37):11071–9. 117

Paulus, H. (2000) Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem*, **69**:447–96. 39, 43

Pegg, A. E., Xiong, H., Feith, D. J. and Shantz, L. M. (1998) S-adenosylmethionine decarboxylase: structure, function and regulation by polyamines. *Biochem Soc Trans*, **26**(4):580–6. 38

Pei, J. and Grishin, N. V. (2003) Peptidase family U34 belongs to the superfamily of N-terminal nucleophile hydrolases. *Protein Sci*, **12**(5):1131–5. 42

Perler, F. B. (2002) InBase: the Intein Database. *Nucleic Acids Res*, **30**(1):383–4. 44

Perler, F. B., Xu, M. Q. and Paulus, H. (1997) Protein splicing and autoproteolysis mechanisms. *Curr Opin Chem Biol*, **1**(3):292–9. 43

Perrakis, A., Morris, R. and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol*, **6**(5):458–63. 107

Perry, K. M., Carreras, C. W., Chang, L. C., Santi, D. V. and Stroud, R. M. (1993) Structures of thymidylate synthase with a C-terminal deletion: role of the C-terminus in alignment of 2'-deoxyuridine 5'-monophosphate and 5,10-methylenetetrahydrofolate. *Biochemistry*, **32**(28):7116–25. 117

Phillips, W. C., Stewart, A., Stanton, M., Naday, I. and Ingersoll, C. (2002) High-sensitivity CCD-based X-ray detector. *J Synchrotron Radiat*, **9**(Pt 1):36–43. 92

Pietrokovski, S. (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci*, **3**(12):2340–50. 44

Plater, A. R., Zgiby, S. M., Thomson, G. J., Qamar, S., Wharton, C. W. and Berry, A. (1999) Conserved residues in the mechanism of the E. coli Class II FBP- aldolase. *J Mol Biol*, **285**(2):843–55. 150

Poland, B. W., Xu, M. Q. and Quiocho, F. A. (2000) Structural insights into the protein splicing mechanism of PI-SceI. *J Biol Chem*, **275**(22):16408–13. 199

Porcelli, M., Cacciapuoti, G., Fusco, S., Iacomino, G., Gambacorta, A., De Rosa, M. and Zappia, V. (1993) S-adenosylhomocysteine hydrolase from the thermophilic archaeon Sulfolobus solfataricus: purification, physico-chemical and immunological properties. *Biochim Biophys Acta*, **1164**(2):179–88. 38

Porter, J. A., Ekker, S. C., Park, W. J., von Kessler, D. P., Young, K. E., Chen, C. H., Ma, Y., Woods, A. S., Cotter, R. J., Koonin, E. V. and Beachy, P. A. (1996) Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell*, **86**(1):21–34. 42

Porter, J. A., von Kessler, D. P., Ekker, S. C., Young, K. E., Lee, J. J., Moses, K. and Beachy, P. A. (1995) The product of hedgehog autoproteolytic cleavage active in local and long-range signalling. *Nature*, **374**(6520):363–6. 42, 43

Powers, S. G. and Snell, E. E. (1976) Ketopantoate hydroxymethyltransferase. II. Physical, catalytic, and regulatory properties. *J Biol Chem*, **251**(12):3786–93. 24, 28, 29, 30, 31, 85

Primerano, D. A. and Burns, R. O. (1983) Role of acetohydroxy acid isomeroreductase in biosynthesis of pantothenic acid in Salmonella typhimurium. *J Bacteriol*, **153**(1):259–69. 10

Pujadas, G. and Palau, J. (1999) TIM barrrel fold: structural, functional and evolutionary characteristics in natural and designed molecules. *Biologia Bratislava*, **54**:231–254. 134

Ramakrishnan, C. and Ramachandran, G. N. (1965) Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys J*, **5**(6):909–33. 109

Ramaswamy, G., Karim, M. A., Murti, K. G. and Jackowski, S. (2003) PPA-Ralpha controls the intracellular coenzyme A concentration via regulation of PANK1alpha gene expression. *J Lipid Res*. 20

Ramjee, M. K., Genschel, U., Abell, C. and Smith, A. G. (1997) Escherichia coli L-aspartate-alpha-decarboxylase: preprotein processing and observation of reaction intermediates by electrospray mass spectrometry. *Biochem J*, **323**(Pt 3):661–9. 35, 36, 37, 46, 48, 70, 71, 155, 156, 157, 158

Read, R. J. (1986) Improved Fourier Coefficients for maps using phases from partial structures with errors. *Acta Crystallogr A*, **42**:140–149. 107, 109

Recsei, P. A., Huynh, Q. K. and Snell, E. E. (1983a) Conversion of prohistidine decarboxylase to histidine decarboxylase: peptide chain cleavage by nonhydrolytic serinolysis. *Proc Natl Acad Sci U S A*, **80**(4):973–7. 36, 37

Recsei, P. A., Moore, W. M. and Snell, E. E. (1983b) Pyruvoyl-dependent histidine decarboxylases from Clostridium perfringens and Lactobacillus buchneri. Comparative structures and properties. *J Biol Chem*, **258**(1):439–44. 38

Recsei, P. A. and Snell, E. E. (1970) Histidine decarboxylase of Lactobacillus 30a. VI. Mechanism of action and kinetic properties. *Biochemistry*, **9**(7):1492–7. 37, 46

Recsei, P. A. and Snell, E. E. (1973) Prohistidine decarboxylase from Lactobacillus 30a. A new type of zymogen. *Biochemistry*, **12**(3):365–71. 37

Recsei, P. A. and Snell, E. E. (1982) Histidine decarboxylase of Lactobacillus 30a. Comparative properties of wild type and mutant proenzymes and their derived enzymes. *J Biol Chem*, **257**(12):7196–202. 37

Recsei, P. A. and Snell, E. E. (1985) Pyruvoyl-dependent histidine decarboxylases. Mechanism of cleavage of the proenzyme from Lactobacillus buchneri. *J Biol Chem*, **260**(5):2804–6. 38

Reibel, D. K., Wyse, B. W., Berkich, D. A. and Neely, J. R. (1981a) Regulation of coenzyme A synthesis in heart muscle: effects of diabetes and fasting. *Am J Physiol*, **240**(4):H606–11. 20

Reibel, D. K., Wyse, B. W., Berkich, D. A., Palko, W. M. and Neely, J. R. (1981b) Effects of diabetes and fasting on pantothenic acid metabolism in rats. *Am J Physiol*, **240**(6):E597–601. 20

Ren, Z., Bourgeois, D., Helliwell, J., Moffat, K., Srajer, V. and Stoddard, B. L. (1999) Laue crystallography: coming of age. *J Syn Rad*, **6**:891–917. 207, 208

Rendina, G. and Coon, M. J. (1957) Enzymatic hydrolysis of the coenzyme a thiol esters of beta-hydroxypropionic and beta-hydroxyisobutyric acids. *J Biol Chem*, **225**(1):523–34. 14

Rever, B. (1965) *Biochemical and genetic studies of inorganic nitrogen metabolism in Aspergillus nidulans*. Ph.D. thesis, University of Cambridge. 14

Rice, L. M. and Brunger, A. T. (1994) Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins*, **19**(4):277–90. 104

Riikonen, A., Tikkanen, R., Jalanko, A. and Peltonen, L. (1995) Immediate interaction between the nascent subunits and two conserved amino acids Trp34 and Thr206 are needed for the catalytic activity of aspartylglucosaminidase. *J Biol Chem*, **270**(9):4903–7. 41

Riley, M. (1998) Genes and proteins of Escherichia coli K-12. *Nucleic Acids Res*, **26**(1):54. 4

Rison, S. C., Teichmann, S. A. and Thornton, J. M. (2002) Homology, pathway distance and chromosomal localization of the small molecule metabolism enzymes in Escherichia coli. *J Mol Biol*, **318**(3):911–32. 6, 7

Rison, S. C. and Thornton, J. M. (2002) Pathway evolution, structurally speaking. *Curr Opin Struct Biol*, **12**(3):374–82. 3

Robishaw, J. D., Berkich, D. and Neely, J. R. (1982) Rate-limiting step and control of coenzyme A synthesis in cardiac muscle. *J Biol Chem*, **257**(18):10967–72. 18, 20

Robishaw, J. D. and Neely, J. R. (1984) Pantothenate kinase and control of CoA synthesis in heart. *Am J Physiol*, **246**(4 Pt 2):H532–41. 18

Robishaw, J. D. and Neely, J. R. (1985) Coenzyme A metabolism. *Am J Physiol*, **248**(1 Pt 1):E1–9. 18

Rock, C. O., Calder, R. B., Karim, M. A. and Jackowski, S. (2000) Pantothenate kinase regulation of the intracellular concentration of coenzyme A. *J Biol Chem*, **275**(2):1377–83. 16, 18

Rock, C. O., Park, H. W. and Jackowski, S. (2003) Role of feedback regulation of pantothenate kinase (CoaA) in control of coenzyme A levels in Escherichia coli. *J Bacteriol*, **185**(11):3410–5. 18

Rossmann, M., Arnold, E. and Eds. (2001) *International Tables for Crystallography Crystallography of Biological Macromolecules*, vol. F. The International Union of Crystallography, Dordrecht/Boston/London. 82, 89, 91, 92, 99, 101, 102, 103, 106, 112, 188

Rubio, A. and Downs, D. M. (2002) Elevated levels of ketopantoate hydroxymethyltransferase (PanB) lead to a physiologically significant coenzyme A elevation in Salmonella enterica serovar Typhimurium. *J Bacteriol*, **184**(10):2827–32. 11, 19, 25

Saarela, J., Laine, M., Tikkanen, R., Oinonen, C., Jalanko, A., Rouvinen, J. and Peltonen, L. (1998) Activation and oligomerization of aspartylglucosaminidase. *J Biol Chem*, **273**(39):25320–8. 41

Saarela, J., Oinonen, C., Jalanko, A., Rouvinen, J. and Peltonen, L. (2004) Autoproteolytic activation of human aspartylglucosaminidase. *Biochem J*, **378**(Pt 2):363–71. 41, 200

Sahm, H. and Eggeling, L. (1999) D-Pantothenate synthesis in Corynebacterium glutamicum and use of panBC and genes encoding L-valine synthesis for D-pantothenate overproduction. *Appl Environ Microbiol*, **65**(5):1973–9. 13, 22

Saldanha, S. A. (2002) *Studies on pantothenate enzymes*. Ph.D. thesis, University of Cambridge. 133

Saldanha, S. A., Birch, L. M., Webb, M. E., Nabbs, B. K., von Delft, F., Smith, A. G. and Abell, C. (2001) Identification of Tyr58 as the proton donor in the aspartate-alpha-decarboxylase reaction. *Chem Commun Camb*, (18):1760–1. 46

Sali, A. and Blundell, T. L. (1990) Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. *J Mol Biol*, **212**(2):403–28. 114, 141

Saliba, K. J. and Kirk, K. (2001) H+-coupled pantothenate transport in the intracellular malaria parasite. *J Biol Chem*, **276**(21):18115–21. 11

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2nd edn. 65, 158

Saridakis, V., Christendat, D., Thygesen, A., Arrowsmith, C. H., Edwards, A. M. and Pai, E. F. (2002) Crystal structure of Methanobacterium thermoautotrophicum conserved protein MTH1020 reveals an NTN-hydrolase fold. *Proteins*, **48**(1):141–3. 42

Satre, M. and Kennedy, E. P. (1978) Identification of bound pyruvate essential for the activity of phosphatidylserine decarboxylase of Escherichia coli. *J Biol Chem*, **253**(2):479–83. 46

Savage, A., King, J. and Gamborg, O. (1979) Title not retrievable. *Plant Sci Letters*, **16**:367–376. 14

Sawaya, M. R. and Kraut, J. (1997) Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence. *Biochemistry*, **36**(3):586–603. 117

Scarsdale, J. N., Radaev, S., Kazanina, G., Schirch, V. and Wright, H. T. (2000) Crystal structure at 2.4 A resolution of E. coli serine hydroxymethyltransferase in complex with glycine substrate and 5-formyl tetrahydrofolate. *J Mol Biol*, **296**(1):155–68. 30

Schagger, H. and Von Jagow, G. (1987) Tricine-Sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Bioch*, **166**:368–379. 62

Schilling, C. H., Edwards, J. S., Letscher, D. and Palsson, B. O. (2000) Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol Bioeng*, **71**(4):286–306. 2

Schilling, C. H. and Palsson, B. O. (1998) The underlying pathway structure of biochemical reaction networks. *Proc Natl Acad Sci U S A*, **95**(8):4193–8. 2

Schindelin, H., Kisker, C., Hilton, J., Rajagopalan, K. V. and Rees, D. C. (1996) Crystal structure of DMSO reductase: redox-linked changes in molybdopterin coordination. *Science*, **272**(5268):1615–21. 50

Schirch, V. and Strong, W. B. (1989) Interaction of folylpolyglutamates with enzymes in one-carbon metabolism. *Arch Biochem Biophys*, **269**(2):371–80. 30

Schmitzberger, F., Kilkenny, M. L., Lobley, C. M., Webb, M. E., Vinkovic, M., Matak-Vinkovic, D., Witty, M., Chirgadze, D. Y., Smith, A. G., Abell, C. and Blundell, T. L. (2003a) Structural constraints on protein self-processing in L-aspartate-alpha-decarboxylase. *EMBO J*, **22**(23):6193–204. 156, 157

Schmitzberger, F., Smith, A. G., Abell, C. and Blundell, T. L. (2003b) Comparative analysis of the Escherichia coli ketopantoate hydroxymethyltransferase crystal structure confirms that it is a member of the (betaalpha)8 phosphoenolpyruvate/pyruvate superfamily. *J Bacteriol*, **185**(14):4163–71. 136

Schomaker, V. and Trueblood, K. (1998) Correlation of Internal Torsional Motion with Overall Molecular Motion in Crystals. *Acta Crystallogr B*, **54**:507–514. 106

Schroder, G. and Schroder, J. (1995) cDNAs for S-adenosyl-L-methionine decarboxylase from Catharanthus roseus, heterologous expression, identification of the proenzyme-processing site, evidence for the presence of both subunits in the active enzyme, and a conserved region in the 5′ mRNA leader. *Eur J Biochem*, **228**(1):74–8. 38

Schurmann, M. and Sprenger, G. A. (2001) Fructose-6-phosphate aldolase is a novel class I aldolase from Escherichia coli and is related to a novel group of bacterial transaldolases. *J Biol Chem*, **276**(14):11055–61. 28

Schuster, S., Fell, D. A. and Dandekar, T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol*, **18**(3):326–32. 3

Seemuller, E., Lupas, A. and Baumeister, W. (1996) Autocatalytic processing of the 20S proteasome. *Nature*, **382**(6590):468–71. 41, 198, 201

Selkov, E., Grechkin, Y., Mikhailova, N. and Selkov, E. (1998) MPW: the Metabolic Pathways Database. *Nucleic Acids Res*, **26**(1):43–5. 3

Seto, B. (1978) A pyruvate-containing peptide of proline reductase in Clostridium sticklandii. *J Biol Chem*, **253**(13):4525–9. 38

Sharma, V., Sharma, S., Hoener zu Bentrup, K., McKinney, J. D., Russell, D. G., Jacobs, W. R., J. and Sacchettini, J. C. (2000) Structure of isocitrate lyase, a persistence factor of Mycobacterium tuberculosis. *Nat Struct Biol*, **7**(8):663–8. 142

Shi, J., Blundell, T. L. and Mizuguchi, K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol*, **310**(1):243–57. 113

Shimizu, S., Kataoka, M., Chung, M. C. and Yamada, H. (1988) Ketopantoic acid reductase of Pseudomonas maltophilia 845. Purification, characterization, and role in pantothenate biosynthesis. *J Biol Chem*, **263**(24):12077–84. 13

Simeonidis, E., Rison, S. C., Thornton, J. M., Bogle, I. D. and Papageorgiou, L. G. (2003) Analysis of metabolic networks using a pathway distance metric through linear programming. *Metab Eng*, **5**(3):211–9. 4

Skrede, S. and Halvorsen, O. (1979) Increased biosynthesis of CoA in the liver of rats treated with clofibrate. *Eur J Biochem*, **98**(1):223–9. 18

Skrede, S. and Halvorsen, O. (1983) Mitochondrial pantetheinephosphate adenylyltransferase and dephospho-CoA kinase. *Eur J Biochem*, **131**(1):57–63. 19

Slade, A., Horrocks, A. J., Lindsay, C. D., Dunbar, B. and Virden, R. (1991) Site-directed chemical conversion of serine to cysteine in penicillin acylase from Escherichia coli ATCC 11105. Effect on conformation and catalytic activity. *Eur J Biochem*, **197**(1):75–80. 41

Slotnick, I. J. (1956) Dihydrouracil as a growth factor for a mutant strain of Escherichia coli. *J Bacteriol*, **72**(2):276–7. 14

Smith, C. M., Cano, M. L. and Potyraj, J. (1978) The relationship between metabolic state and total CoA content of rat liver and heart. *J Nutr*, **108**(5):854–62. 20

Smith, C. M. and Savage, C. R., J. (1980) Regulation of coenzyme A biosynthesis by glucagon and glucocorticoid in adult rat liver parenchymal cells. *Biochem J*, **188**(1):175–84. 20

Smith, J. L., Zaluzec, E. J., Wery, J. P., Niu, L., Switzer, R. L., Zalkin, H. and Satow, Y. (1994) Structure of the allosteric regulatory enzyme of purine biosynthesis. *Science*, **264**(5164):1427–33. 44

Smith, R. (1988) *The structure and molecular characterisation of L-aspartate-a-decarboxylase from E.coli*. Ph.D. thesis, Massachusetts Institute of Technology. 35, 46, 48

Snell, E. E. and Huynh, Q. K. (1986) *Prohistidine decarboxylase from Lactobacillus 30a*, vol. 122 of *Methods in ENZYMOLOGY*. Academic Press. 37

Song, W. J. and Jackowski, S. (1992a) Cloning, sequencing, and expression of the pantothenate kinase (coaA) gene of Escherichia coli. *J Bacteriol*, **174**(20):6411–7. 16

Song, W. J. and Jackowski, S. (1992b) coaA and rts are allelic and located at kilobase 3532 on the Escherichia coli physical map. *J Bacteriol*, **174**(5):1705–6. 16

Song, W. J. and Jackowski, S. (1994) Kinetics and regulation of pantothenate kinase from Escherichia coli. *J Biol Chem*, **269**(43):27051–8. 16, 18

Sorokin, A., Azevedo, V., Zumstein, E., Galleron, N., Ehrlich, S. D. and Serror, P. (1996) Sequence analysis of the Bacillus subtilis chromosome region between the serA and kdg loci cloned in a yeast artificial chromosome. *Microbiology*, **142 (Pt 8)**:2005–16. 13

Spitzer, E. D., Jimenez-Billini, H. E. and Weiss, B. (1988) beta-Alanine auxotrophy associated with dfp, a locus affecting DNA synthesis in Escherichia coli. *J Bacteriol*, **170**(2):872–6. 16

Stanley, B. A. and Pegg, A. E. (1991) Amino acid residues necessary for putrescine stimulation of human S-adenosylmethionine decarboxylase proenzyme processing and catalytic activity. *J Biol Chem*, **266**(28):18502–6. 38, 198

Stanley, B. A., Pegg, A. E. and Holm, I. (1989) Site of pyruvate formation and processing of mammalian S- adenosylmethionine decarboxylase proenzyme. *J Biol Chem*, **264**(35):21073–9. 38, 198

Stanley, B. A., Shantz, L. M. and Pegg, A. E. (1994) Expression of mammalian S-adenosylmethionine decarboxylase in Escherichia coli. Determination of sites for putrescine activation of activity and processing. *J Biol Chem*, **269**(11):7901–7. 38

Steller, I., Bolotovsky, R. and Rossmann, M. (1997) An Algorithm for Automatic Indexing of Oscillation Images using Fourier Analysis. *J App Cryst*, **30**:1036–1040. 95

Stoddard, B. L., Bruhnke, J., Koenigs, P., Porter, N., Ringe, D. and Petsko, G. A. (1990) Photolysis and deacylation of inhibited chymotrypsin. *Biochemistry*, **29**(35):8042–51. 208

Stolz, J. and Sauer, N. (1999) The fenpropimorph resistance gene FEN2 from Saccharomyces cerevisiae encodes a plasma membrane H+-pantothenate symporter. *J Biol Chem*, **274**(26):18747–52. 11

Stover, P. and Schirch, V. (1992) Enzymatic mechanism for the hydrolysis of 5,10-methenyltetrahydropteroylglutamate to 5-formyltetrahydropteroylglutamate by serine hydroxymethyltransferase. *Biochemistry*, **31**(7):2155–64. 29

Strauss, E. and Begley, T. P. (2001) Mechanistic studies on phosphopantothenoylcysteine decarboxylase. *J Am Chem Soc*, **123**(26):6449–50. 16

Strauss, E., Kinsland, C., Ge, Y., McLafferty, F. W. and Begley, T. P. (2001) Phosphopantothenoylcysteine synthetase from Escherichia coli. Identification and characterization of the last unidentified coenzyme A biosynthetic enzyme in bacteria. *J Biol Chem*, **276**(17):13513–6. 16

Stryer, L. (1995) *Biochemistry*. W.H. Freeman and Company, New York, 4th edn. 2, 3, 15

Sugantino, M., Zheng, R., Yu, M. and Blanchard, J. S. (2003) Mycobacterium tuberculosis ketopantoate hydroxymethyltransferase: tetrahydrofolate-independent hydroxymethyltransferase and enolization reactions with alpha-keto acids. *Biochemistry*, **42**(1):191–9. 25, 28

Suresh, C. G., Pundle, A. V., SivaRaman, H., Rao, K. N., Brannigan, J. A., McVey, C. E., Verma, C. S., Dauter, Z., Dodson, E. J. and Dodson, G. G. (1999) Penicillin V acylase crystal structure reveals new Ntn-hydrolase family members. *Nat Struct Biol*, **6**(5):414–6. 44

Sutcliffe, M. J., Haneef, I., Carney, D. and Blundell, T. L. (1987) Knowledge based modelling of homologous proteins, Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng*, **1**(5):377–84. 114

Suzuki, T., Abiko, Y. and Shimizu, M. (1967) Investigations on pantothenic acid and its related compounds. XII. Biochemical studies (7). Dephospho-coA pyrophosphorylase and dephospho-coA kinase as a possible bifunctional enzyme complex. *J Biochem Tokyo*, **62**(6):642–9. 17

Tahiliani, A. G. (1989) Dependence of mitochondrial coenzyme A uptake on the membrane electrical gradient. *J Biol Chem*, **264**(31):18426–32. 19

Tahiliani, A. G. (1991) Evidence for net uptake and efflux of mitochondrial coenzyme A. *Biochim Biophys Acta*, **1067**(1):29–37. 19

Tahiliani, A. G. and Beinlich, C. J. (1991) Pantothenic acid in health and disease. *Vitam Horm*, **46**:165–228. 9, 19, 20

Tahiliani, A. G. and Neely, J. R. (1987) Mitochondrial synthesis of coenzyme A is on the external surface. *J Mol Cell Cardiol*, **19**(12):1161–7. 19

Tanase, S., Guirard, B. M. and Snell, E. E. (1985) Purification and properties of a pyridoxal 5'-phosphate-dependent histidine decarboxylase from Morganella morganii AM-15. *J Biol Chem*, **260**(11):6738–46. 45

Teichmann, S. A., Rison, S. C., Thornton, J. M., Riley, M., Gough, J. and Chothia, C. (2001a) Small-molecule metabolism: an enzyme mosaic. *Trends Biotechnol*, **19**(12):482–6. 6

Teichmann, S. A., Rison, S. C., Thornton, J. M., Riley, M., Gough, J. and Chothia, C. (2001b) The evolution and structural anatomy of the small molecule metabolic pathways in Escherichia coli. *J Mol Biol*, **311**(4):693–708. 3, 6, 7, 152

Teller, J. H., Powers, S. G. and Snell, E. E. (1976) Ketopantoate hydroxymethyltransferase. I. Purification and role in pantothenate biosynthesis. *J Biol Chem*, **251**(12):3780–5. 10, 24

Teng, T. (1990) Mounting of crystals for macromolecular crystallography in a freestanding thin-film. *J App Cryst*, **23**:387–391. 88

The Arabidopsis genome initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, **408**(6814):796–815. 154

The Yeast genome directory (1997) Yeast genome directory. *Nature*, **387**(6632 Suppl):5. 154

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997) The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, **25**(24):4876–82. 113

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**(22):4673–80. 113

Tickle, I. J., Laskowski, R. A. and Moss, D. S. (2000) Rfree and the rfree ratio. II. Calculation Of the expected values and variances of cross-validation statistics in macromolecular least-squares refinement. *Acta Crystallogr D Biol Crystallogr*, **56 (Pt 4)**:442–50. 109

Tikkanen, R., Riikonen, A., Oinonen, C., Rouvinen, R. and Peltonen, L. (1996) Functional analyses of active site residues of human lysosomal aspartylglucosaminidase: implications for catalytic mechanism and autocatalytic activation. *EMBO J*, **15**(12):2954–60. 41, 201

Tolbert, W. D., Graham, D. E., White, R. H. and Ealick, S. E. (2003a) Pyruvoyl-Dependent Arginine Decarboxylase from Methanococcus jannaschii. Crystal Structures of the Self-Cleaved and S53A Proenzyme Forms. *Structure Camb*, **11**(3):285–94. 38, 191, 197, 198

Tolbert, W. D., Zhang, Y., Cottet, S. E., Bennett, E. M., Ekstrom, J. L., Pegg, A. E. and Ealick, S. E. (2003b) Mechanism of human S-adenosylmethionine decarboxylase proenzyme processing as revealed by the structure of the S68A mutant. *Biochemistry*, **42**(8):2386–95. 191, 198

Toney, M. D. (2001) Computational studies on nonenzymatic and enzymatic pyridoxal phosphate catalyzed decarboxylations of 2-aminoisobutyrate. *Biochemistry*, **40**(5):1378–84. 45

Trivedi, V., Gupta, A., Jala, V. R., Saravanan, P., Rao, G. S., Rao, N. A., Savithri, H. S. and Subramanya, H. S. (2002) Crystal structure of binary and ternary complexes of serine hydroxymethyltransferase from Bacillus stearothermophilus: insights into the catalytic mechanism. *J Biol Chem*, **277**(19):17161–9. 117

Trotter, P. J., Pedretti, J., Yates, R. and Voelker, D. R. (1995) Phosphatidylserine decarboxylase 2 of Saccharomyces cerevisiae. Cloning and mapping of the gene, heterologous expression, and creation of the null allele. *J Biol Chem*, **270**(11):6071–80. 38

Tsoka, S. and Ouzounis, C. A. (2001) Functional versatility and molecular diversity of the metabolic map of Escherichia coli. *Genome Res*, **11**(9):1503–10. 3, 4, 5

Vallari, D. S. and Jackowski, S. (1988) Biosynthesis and degradation both contribute to the regulation of coenzyme A content in Escherichia coli. *J Bacteriol*, **170**(9):3961–6. 19

Vallari, D. S., Jackowski, S. and Rock, C. O. (1987) Regulation of pantothenate kinase by coenzyme A and its thioesters. *J Biol Chem*, **262**(6):2468–71. 18

Vallari, D. S. and Rock, C. O. (1985) Isolation and characterization of Escherichia coli pantothenate permease (panF) mutants. *J Bacteriol*, **164**(1):136–42. 11

Valle, F., Balbas, P., Merino, E. and Bolivar, F. (1991) The role of penicillin amidases in nature and in industry. *Trends Biochem Sci*, **16**(1):36–40. 42

Van Broekhoven, A., Peeters, M. C., Debeer, L. J. and Mannaerts, G. P. (1981) Subcellular distribution of coenzyme A: evidence for a separate coenzyme A pool in peroxisomes. *Biochem Biophys Res Commun*, **100**(1):305–12. 19

van Poelje, P. D. and Snell, E. E. (1990) Pyruvoyl-dependent enzymes. *Annu Rev Biochem*, **59**:29–59. 37, 45, 197

Vandamme, E. J. (1992) Production of vitamins, coenzymes and related biochemicals by biotechnological processes. *J Chem Technol Biotechnol*, **53**(4):313–27. 22

Vanderslice, P., Copeland, W. C. and Robertus, J. D. (1988) Site-directed alteration of serine 82 causes nonproductive chain cleavage in prohistidine decarboxylase. *J Biol Chem*, **263**(22):10583–6. 37, 197

Vieira, J. and Messing, J. (1991) New pUC-derived cloning vectors with different selectable markers and DNA replication origins. *Gene*, **100**:189–94. 65

Vincent, S., Thomas, A., Brasher, B. and Benson, J. D. (2003) Targeting of proteins to membranes through hedgehog auto-processing. *Nat Biotechnol*, **21**(8):936–40. 54

Virtanen, A. and Laine, T. (1937) Title not retrievable. *Enzymologia*, **3**:266. 35

Voelker, D. R. (1997) Phosphatidylserine decarboxylase. *Biochim Biophys Acta*, **1348**(1-2):236–44. 38

Vollhardt, K. and Schore, N. (1995) *Organische Chemie*. VCH, Weinheim, 2nd edn. 25

Voltti, H., Savolainen, M. J., Jauhonen, V. P. and Hassinen, I. E. (1979) Clofibrate-induced increase in coenzyme A concentration in rat tissues. *Biochem J*, **182**(1):95–102. 20

von Delft, F. (2000) *Structural studies of three enzymes of pantothenate biosynthesis in Escherichia coli*. Ph.D. thesis, University of Cambridge. 30, 46, 52, 118, 121, 122, 161

von Delft, F., Inoue, T., Saldanha, S. A., Ottenhof, H. H., Schmitzberger, F., Birch, L. M., Dhanaraj, V., Witty, M., Smith, A. G., Blundell, T. L. and Abell, C. (2003) Structure of E. coli ketopantoate hydroxymethyl transferase complexed with ketopantoate and Mg2+, solved by locating 160 selenomethionine sites. *Structure Camb*, **11**(8):985–96. 30, 31, 32, 33, 55, 87, 136

von Delft, F., Lewendon, A., Dhanaraj, V., Blundell, T. L., Abell, C. and Smith, A. G. (2001) The crystal structure of E. coli pantothenate synthetase confirms it as a member of the cytidylyltransferase superfamily. *Structure Camb*, **9**(5):439–50. 55, 153

Walker, E. H., French, C. E., Rathbone, D. A. and Bruce, N. C. (2000) Mechanistic studies of morphine dehydrogenase and stabilization against covalent inactivation. *Biochem J*, **345 (Pt 3)**:687–92. 214

Wang, D. W., Driessen, H. P. and Tickle, I. J. (1991) MOLPACK: molecular graphics for studying the packing of protein molecules in the crystallographic unit cell. *J Mol Graph*, **9**(1):50–2, 38. 100

Wang, H., Huang, W., Fei, Y. J., Xia, H., Yang-Feng, T. L., Leibach, F. H., Devoe, L. D., Ganapathy, V. and Prasad, P. D. (1999) Human placental Na+-dependent multivitamin transporter. Cloning, functional expression, gene structure, and chromosomal localization. *J Biol Chem*, **274**(21):14875–83. 11

Watanabe, T., Nakamura, H., Liang, L. Y., Yamatodani, A. and Wada, H. (1979) Partial purification and characterization of L-histidine decarboxylase from fetal rats. *Biochem Pharmacol*, **28**(7):1149–55. 45

Webb, E. (1992) *Enzyme nomenclature*. Academic Press, San Diego. 15

Wierenga, R. K. (2001) The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett*, **492**(3):193–8. 135

Wilken, D. R., King, H. L., J. and Dyar, R. E. (1975) Ketopantoic acid and ketopantoyl lactone reductases. Stereospecificity of transfer of hydrogen from reduced nicotinamide adenine dinucleotide phosphate. *J Biol Chem*, **250**(6):2311–4. 10

Willey, D. L., Caswell, D. A., Lowe, C. R. and Bruce, N. C. (1993) Nucleotide sequence and over-expression of morphine dehydrogenase, a plasmid-encoded gene from Pseudomonas putida M10. *Biochem J*, **290 (Pt 2)**:539–44. 214

Williamson, J. M. and Brown, G. M. (1979) Purification and properties of L-Aspartate-alpha-decarboxylase, an enzyme that catalyzes the formation of beta-alanine in Escherichia coli. *J Biol Chem*, **254**(16):8074–82. 35, 46

Winn, M. D., Isupov, M. N. and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr*, **57**(Pt 1):122–33. 106

Worrall, D. M., Lambert, S. F. and Tubbs, P. K. (1985) Limited proteolysis of pig liver CoA synthase: evidence for subunit identity. *FEBS Lett*, **187**(2):277–9. 17

Worrall, D. M. and Tubbs, P. K. (1983) A bifunctional enzyme complex in coenzyme A biosynthesis: purification of pantetheine phosphate adenylyltransferase and dephospho-CoA kinase. *Biochem J*, **215**(1):153–7. 17

Wyckoff, H., Hirs, C. and Timasheff, S. (1985) *Diffraction Methods for Biological Macromolecules-Part B*, vol. 115 of *Methods in ENZYMOLOGY*. Academic Press, London. 115

Wymer, N., Buchanan, L. V., Henderson, D., Mehta, N., Botting, C. H., Pocivavsek, L., Fierke, C. A., Toone, E. J. and Naismith, J. H. (2001) Directed evolution of a new catalytic site in 2-keto-3-deoxy-6-phosphogluconate aldolase from Escherichia coli. *Structure Camb*, **9**(1):1–9. 34

Xiong, H. and Pegg, A. E. (1999) Mechanistic studies of the processing of human S-adenosylmethionine decarboxylase proenzyme. Isolation of an ester intermediate. *J Biol Chem*, **274**(49):35059–66. 38, 198

Xiong, H., Stanley, B. A. and Pegg, A. E. (1999) Role of cysteine-82 in the catalytic mechanism of human S- adenosylmethionine decarboxylase. *Biochemistry*, **38**(8):2462–70. 38, 48, 198

Xiong, H., Stanley, B. A., Tekwani, B. L. and Pegg, A. E. (1997) Processing of mammalian and plant S-adenosylmethionine decarboxylase proenzymes. *J Biol Chem*, **272**(45):28342–8. 197

Xu, Q., Buckley, D., Guan, C. and Guo, H. C. (1999) Structural insights into the mechanism of intramolecular proteolysis. *Cell*, **98**(5):651–61. 39, 184, 188, 199, 200, 201

Yun, M., Park, C. G., Kim, J. Y., Rock, C. O., Jackowski, S. and Park, H. W. (2000) Structural basis for the feedback regulation of Escherichia coli pantothenate kinase by coenzyme A. *J Biol Chem*, **275**(36):28093–9. 16, 17, 18

Zhao, L., Allanson, N. M., Thomson, S. P., Maclean, J. K., Barker, J. J., Primrose, W. U., Tyler, P. D. and Lewendon, A. (2003) Inhibitors of phosphopantetheine adenylyltransferase. *Eur J Med Chem*, **38**(4):345–9. 23

Zheng, R. and Blanchard, J. S. (2000a) Identification of active site residues in E. coli ketopantoate reductase by mutagenesis and chemical rescue. *Biochemistry*, **39**(51):16244–51. 10

Zheng, R. and Blanchard, J. S. (2000b) Kinetic and mechanistic analysis of the E. coli panE-encoded ketopantoate reductase. *Biochemistry*, **39**(13):3708–17. 10

Zheng, R. and Blanchard, J. S. (2001) Steady-state and pre-steady-state kinetic analysis of Mycobacterium tuberculosis pantothenate synthetase. *Biochemistry*, **40**(43):12904–12. 11

Zheng, R. and Blanchard, J. S. (2003) Substrate specificity and kinetic isotope effect analysis of the Eschericia coli ketopantoate reductase. *Biochemistry*, **42**(38):11289–96. 10, 87

Zhou, B., Westaway, S. K., Levinson, B., Johnson, M. A., Gitschier, J. and Hayflick, S. J. (2001) A novel pantothenate kinase gene (PANK2) is defective in Hallervorden-Spatz syndrome. *Nat Genet*, **28**(4):345–9. 21

Zhyvoloup, A., Nemazanyy, I., Babich, A., Panasyuk, G., Pobigailo, N., Vudmaska, M., Naidenov, V., Kukharenko, O., Palchevskii, S., Savinska, L., Ovcharenko, G., Verdier, F., Valovka, T., Fenton, T., Rebholz, H., Wang, M. L., Shepherd, P., Matsuka, G., Filonenko, V. and Gout, I. T. (2002) Molecular cloning of CoA Synthase. The missing link in CoA biosynthesis. *J Biol Chem*, **277**(25):22107–10. 16

Zhyvoloup, A., Nemazanyy, I., Panasyuk, G., Valovka, T., Fenton, T., Rebholz, H., Wang, M. L., Foxon, R., Lyzogubov, V., Usenko, V., Kyyamova, R., Gorbenko, O., Matsuka, G., Filonenko, V. and Gout, I. T. (2003) Subcellular localization and regulation of coenzyme A synthase. *J Biol Chem*, **278**(50):50316–21. 17, 20