STRUCTURAL STUDIES INTO THE MECHANISM OF SUBSTRATE BINDING AND CATALYSIS IN A PROMISCUOUS HYPERTHERMOPHILIC ALDOLASE.

Alexander Theodossis

A thesis for the degree of Doctor of Philosophy

University of St Andrews

June 2006

DECLARATION

I, Alexander Theodossis, hereby certify that this thesis, which is approximately 63,000 words in length, has been written by me, is a record of work carried out by myself and has not been submitted in any previous application for a higher degree.

Signature: Date: 11-8-06

I was admitted as a research student at the University of St Andrews in October of 2002 and as a candidate for the degree of Ph.D in October 2003.

Signature Date: 11-8-06

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Signature: Date: 11/8/6

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for he time being in force, subject to any copyright vested in the work not being directly affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Signature: Date: 11-8-06

"...ὥρη μὲν πολέων μύθων, ὥρη δὲ καὶ ὕπνου..."

Ομήρου Οδύσσεια

(Ραψωδία Λ, στοίχος 379)

Scientific happiness is…

"…to have an experiment that works, and keep doing it all the time."

Alfred Hershey (1908 – 1997)

"…Everything is in terms of pieces and parts and components and relationships. Nothing is figured out until it's run through the computer a dozen times …measured and proved…

…within the lines and shapes and symbols is a tremendous richness of underlying form."

Robert M. Pirsig
(Zen and the Art of Motorcycle Maintenance)

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Garry Taylor, for his guidance and encouragement over the last four years as well as the other members of the research group, past and present, for support and assistance. They are, in no particular order: Charlotte Ryan, Slava Zaitsev, Helen Connaris, Jane Potter, Christine Milburn, Simon Newstead, Ibrahim Mustafa and Rupert Russell. I would also like to thank Dr Henry Lamble, Professor Michael Danson, Dr David Hough and Dr Steven Bull for a very stimulating and rewarding collaboration on this project. Additional acknowledgements go to Dr Catherine Botting, Dr Sally Shirran and Mr Alex Houston for running the mass spectrometry facility and always having time to answer questions. Also, all other technical, administrative and support staff at the Centre for Biomolecular Sciences, without whom everything would grind to a halt.

Finally, I am grateful to the McDermotts for letting me share their home for the last nine months, to my friends in St Andrews and elsewhere for being there and most of all to my family for everything.

ABBREVIATIONS

Å	Ångström (0.1 nm)
Ab	antibody
Amp	ampicillin
apo	apoenzyme
ASA	aspartate-β-semialdehyde
a.s.u.	asymmetric unit
ATP	adenosine triphosphate
ATPase	ATP synthase
Bf	thermal B-factor
bp	base pair
CAC	citric acid cycle
$CaCl_2$	calcium chloride
C-C	carbon-carbon
C-H	carbon-hydrogen
Da	daltons
DERA	deoxyribose-5-phosphate aldolase
$\Delta G^{\ddagger}$	Gibbs energy of activation
ΔG°	Gibbs energy of reaction
DHAP	dihydroxyacetone phosphate
DHDPS	dihydrodipicolinate synthase
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DRP	2-deoxyribose-5-phosphate
EcDERA	*E. coli* DERA
EcDHDPS	*E. coli* DHDPS
EcFPA	*E. coli* FPA
EcKDPGA	*E. coli* KDPG aldolase
EcNAL	*E. coli* NAL
E. coli	*Escherichia coli*
ED	Entner-Doudoroff
EDTA	ethylene diamine tetraacetate

EM	Embden-Meyerhof
E:S	enzyme-substrate
ESI	electrospray ionisation
FBP	fructose-1,6-bisphosphate
FBPA	fructose-1,6-bisphosphate aldolase
FBPA Ia	archaeal type I FBPA
FBPA Ie	eukaryotic type I FBPA
FPA	fuculose-1-phosphate aldolase
G-6-P	glucose-6-phosphate
GAP	glyceraldehyde-3-phosphate
GAPN	GAP dehydrogenase
GD	gluconate dehydratase
GDH	glucose dehydrogenase
GOL	glycerol
HB	hydrogen bond
HCl	hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HiNAL	*H. influenzae* NAL
H. influenzae	*Haemophilus influenzae*
H. sapiens	*Homo sapiens*
IPTG	isopropyl-β-D-thiogalactopyranoside
k	kilo (1,000)
kb	1,000 bases
k_{cat}	catalytic constant
kDa	1,000 Da
KDG	2-keto-3-deoxygluconate
KDGA	2-keto-3-deoxygluconate aldolase
KDGal	2-keto-3-deoxygalactonate
KDG(al)	KDG and KDGal
KDPG	2-keto-3-deoxy-6-phosphogluconate
KDPGA	2-keto-3-deoxy-6-phosphogluconate aldolase
KDPGal	2-keto-3-deoxy-6-phosphogalactonate
KDPG(al)	KDPG and KDPGal

KD(P)G(al)	KDG, KDGal, KDPG and KDPGal
K_m	Michaelis constant
K_3PO_4	potassium phosphate
LB	Luria-Bertani
LBHB	low barrier hydrogen bond
M_r	relative molecular mass
MAD	multiple wavelength anomalous dispersion
MALDI-TOF	matrix-assisted laser desorption/ionisation-time of flight
ManNAc	N-acetyl-D-mannosamine
M/c	main chain
MCS	multiple cloning site
Mg	magnesium
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulfate
MPD	2-methyl-2,4-pentanediol
MR	molecular replacement
MS	mass spectrometry
MtDHDPS	*M. tuberculosis* DHDPS
M. tuberculosis	*Mycobacterium tuberculosis*
$NaBH_4$	sodium borohydride
NaCl	sodium chloride
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NAD(P)	NAD and NADP
NAL	N-acetylneuraminate lyase (sialic acid aldolase)
NaOH	sodium hydroxide
Na_3PO_4	sodium phosphate
Neu5Ac	N-acetylneuraminic acid (sialic acid)
$(NH_4)_2SO_4$	ammonium sulfate
NsDHDPS	*N. sylvestris* DHDPS
N. sylvestris	*Nicotiana sylvestris*
O. cuniculus	*Oryctolagus cuniculus*

OD_{600}	optical density at 600 nm
O-H	oxygen-hydrogen
O/N	overnight
O-O	oxygen-oxygen
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
Pfu	*Pyrococcus furiosus*
PGH	phosphoglycolohydroxamate
pKa	negative log of the acid ionisation constant
PMSF	phenylmethanesulfonyl fluoride
PpKDPGA	*Pseudomonas putida* KDPGA
PYR	pyruvate
RAMA	rabbit muscle FBP aldolase
RBr	rigid body refinement
Rf	R-factor
R-free	Free R-factor
rmsd	root mean square deviation
RNA	ribonucleic acid
Rr	restrained refinement
rRNA	ribosomal RNA
RT	room temperature
s.a.	surface area
SAS	succinate β-semialdehyde
SDS	sodium dodecyl sulfate
SeMet	selenomethionine
S. acidocaldarius	*Sulfolobus acidocaldarius*
SsGAPN	*S. solfataricus* GAPN
SsGD	*S. solfataricus* GD
SsGDH	*S. solfataricus* GDH
SsKDGA	*S. solfataricus* KDG aldolase
SsKDGK	*S. solfataricus* KDG kinase

S. solfataricus	*Sulfolobus solfataricus*
S. tokodaii	*Sulfolobus tokodaii*
T. acidophilum	*Thermoplasma acidophilum*
Taq pol.	*Thermus aquaticus* polymerase
TBA	2-thiobarbituric acid
TBP	tagatose-1,6-bisphosphate
TIM	triose phosphate isomerase
T. maritima	*Thermotoga maritima*
TmDHDPS	*T. maritima* DHDPS
TmKDPGA	*T. maritima* KDPG aldolase
Tris	tris(hydroxymethyl)aminomethane
TS	transition state
T. tenax	*Thermoproteus tenax*
TtFBPA	*T. tenax* FBPA
T. thermophilus	*Thermus thermophilus*
UV	ultraviolet
V.	volume
V_{max}	maximum velocity (limiting rate)
v/v	volume/volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside
Zn	zinc

CONTENTS

LIST OF FIGURES

SECTION 1 – INTRODUCTION

Part 1 – The thermoacidophilic Archaeaon *Sulfolobus solfataricus*

Part 2 – The 2-keto-3-deoxygluconate aldolase of *S. solfataricus*

SECTION 2 – RESULTS

Part 1 – Extending the resolution of the *S. solfataricus* KDGA structure

Part 2 – Substrate recognition and catalysis by *S. solfataricus* KDGA

Part 3 – Further investigations of SsKDGA

LIST OF TABLES

SECTION 2 – RESULTS

Part 1 – Extending the resolution of the *S. solfataricus* KDGA structure

Part 2 – Substrate recognition and catalysis by *S. solfataricus* KDGA

ABSTRACT

Sulfolobus solfataricus is a hyperthermophilic archaeon that grows optimally between 80 and 85°C. It metabolises glucose and galactose by a promiscuous variant of the Entner-Doudoroff pathway, in which both a non-phosphorylative and a part-phosphorylative branch have been proposed to operate in parallel. A single, broad specificity aldolase (SsKDGA) has been shown to catalyse the C-C bond cleavage of intermediates arising from both sugars, in either of the branches. The structure of this enzyme has previously been solved by MAD-phasing of a SeMet derivative, confirming that it belongs to the N-acetylneuraminate lyase (NAL) subfamily of type I aldolases. The current study has focused on the investigation of SsKDGA's mechanisms of substrate binding and catalysis. The resolution of the apoenzyme model has been extended from 2.5 to 1.7 Å, using native crystals grown at physiological pH, and a detailed characterisation of the active site region has been carried out. Also, high resolution complexes have been obtained of the aldolase with its natural substrates as Schiff base intermediates. These structures have afforded a better understanding of SsKDGA's promiscuous activity and permitted mechanistic comparisons to be made with other type I aldolases. The insights gained are providing the means by which a rational design approach can be used to tailor the enzyme's selectivity. So far, work has focused on dissecting SsKDGA's catalytic mechanism and inducing stereocontrol in the condensation reaction. However, the potential for generating novel aldol products has also being examined.

SECTION 1 - INTRODUCTION

Part 1

The thermoacidophilic Archaeon *Sulfolobus solfataricus*

1.1.1 Extremophiles and the third domain of life

Over the last four decades molecular and structural biology have revolutionised the investigation of biological processes and evolutionary relationships between organisms (Zuckerkandl and Pauling, 1965; Woese *et al.*, 1990). It has become possible to reduce biological problems to their fundamental components: processing of genetic information considered at the level of individual nucleotides; protein function at the level of atomic interactions. Moreover, studies of molecular sequences and structures have permitted the meaningful characterisation of microorganisms and made possible our current understanding of the phylogeny and diversity of microbial lineages.

The archaeal domain:

It is only through investigations at the molecular level that some of the most important characteristics of micro-organisms have become apparent (Woese *et al.*, 1990). Research into the phylogeny of 16S ribosomal RNA (rRNA) gene sequences in the late 1970's and the 1980's revealed an early divergence of prokaryotes into two distinct groups and challenged the accepted bipartite organisation of life based on the prokaryotic-eukaryotic division (Woese and Fox, 1977; Fox *et al.*, 1980). An increasing body of evidence has subsequently led to greater acceptance of a natural

classification of organisms based on three primary domains: the Eukarya, Bacteria and Archaea (Brown and Doolittle, 1997; Figure 1.1.1).

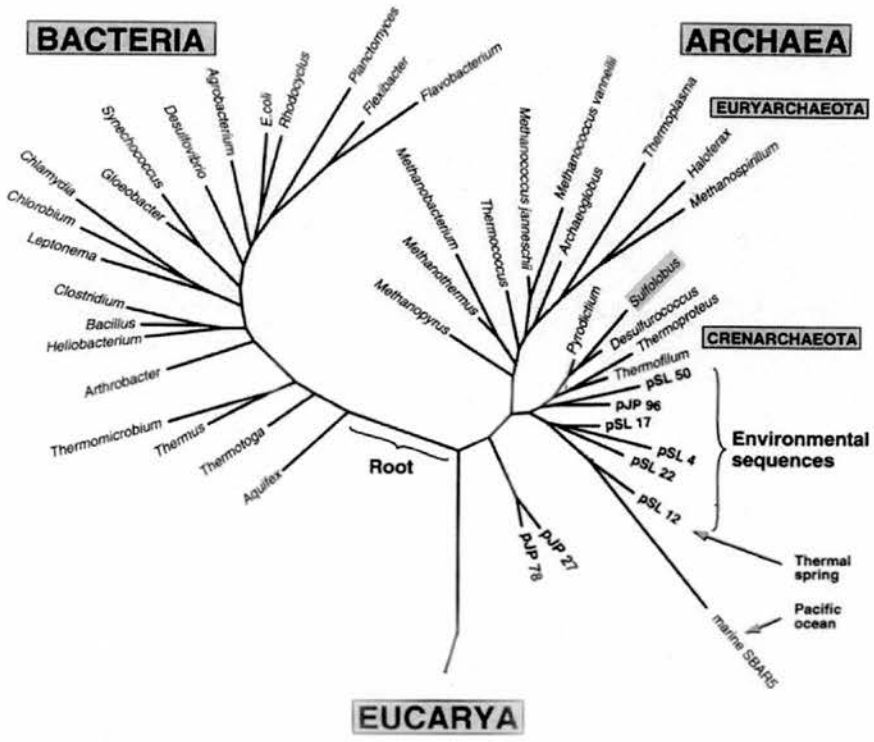


Figure 1.1.1 – The Universal Tree of Life. Phylogenetic tree of the three domains of life (Bacteria, Archaea, and Eukarya) based on the evolutionary distance of the 16S rRNA molecule (adapted from Barns *et al.*, 1996). Members of the Eukarya are omitted for simplicity. The genera listed are representative of major lineages in the two domains. The genus *Sulfolobus* is highlighted in red.

The Archaea represent the third domain of living organisms. First identified as a distinct group of methanogenic prokaryotes only in 1977 (Fox *et al.*, 1977), knowledge of their diversity has since increased greatly, while their status has been elevated to the highest taxonomic level alongside Eukarya and Bacteria (Fox *et al.*, 1980). Although generally associated with extremophily, many mesophilic representatives of the Archaea are now known and it is accepted that the domain is

widely distributed in the environment, typically accounting for ~10% of the total microbiota present (Robertson *et al.*, 2005).

The domain is divided into two main kingdoms: Euryarchaeota span a broad ecological range that includes hyperthermophiles, methanogens and halophiles (Woese *et al.*, 1990). The second kingdom, Crenarchaeota, consists mainly of hyperthermophiles and thermoacidophiles. Members of the latter kingdom include some of the best studied Archaea and the thermoacidophilic genus *Sulfolobus,* in particular, has been used extensively as a model system for genetic, biochemical and structural studies (Ciaramella *et al.*, 2002).

Research on these organisms has revealed unexpected relationships between the archaeal and eukaryotic systems of information processing (DNA replication, transcription, translation) and consequently, understanding of complex eukaryotic processes is being aided by investigations of their simpler archaeal equivalents (Langer *et al.*, 1995; Brown and Doolittle, 1997). Despite some similarities with the other two domains, however, the Archaea are characterized mainly by a large number of unique features, which reflect their evolutionary divergence and adaptation to a diverse range of habitats.

<u>At the extremes of life:</u>

Many environments on this planet are characterised by conditions once considered too hostile to permit the survival of living organisms. It is now recognised, however, that such environments constitute the natural habitats of an expanding group of organisms, primarily consisting of microbes, and known as extremophiles. These organisms

flourish under conditions of extreme temperature, pH, salinity, pressure, or combinations of the above (Ciaramella *et al.*, 2002).

Extreme habitats now known to sustain life have temperatures ranging from -2 to 110 °C; ionic strength of 2 to 5 M NaCl; pH of below 4 or above 9; (Hough and Danson, 1999) and pressures of up to 100 MPa (1,000 atm) (Abe and Horikoshi, 2001). With organisms now known that can survive autoclaving for up to one hour at 121 °C (Blochl *et al.*, 1997), the extreme limits of life may still be unknown. Certainly the diversity of extremophiles is now believed to be far greater than was initially suspected, while the majority of known species still remain uncharacterised due to the difficulties associated with growing them under laboratory conditions (Hough and Danson, 1999).

The extremophiles are represented in each of the three domains (Brown and Doolittle, 1997), though the majority are members of the Archaea (Fox *et al.*, 1980; Ciaramella *et al.*, 2002). The latter are found in all the extreme environments, ranging from Yellowstone hot springs to hypersaline mats and the deep sea. Moreover, they have come to dominate total microbial presence in some environments, such as those characterised by high temperatures and low pH (Robertson *et al.*, 2005).

The adaptations that allow extremophiles to thrive under some of the most hostile conditions on this planet are in many cases still poorly understood and their study has become a subject of great interest. Research has revealed a broad range of unique cellular, molecular and biochemical features (Ciaramella *et al.*, 2002). Moreover, extremophiles and particularly those from the archaeal domain have been shown to possess novel metabolic pathways and enzymes. As well as being stable under extreme conditions, many of these proteins have been shown to have unique

specificities and activities, offering the potential for exploitation in a range of applications (Hough and Danson, 1999).

Sulfolobus – a model organism:

The crenarchaeal genus *Sulfolobus* comprises a number of thermoacidophilic sulfur-oxidising species (Brock *et al.*, 1972), with variable growth requirements. They have been isolated from solfataric fields all over the world at temperatures ranging from between 55 °C and 90 °C and at a pH of 2 to 4, though it has been shown that they maintain their cytoplasmic pH at around 6.5 (Moll and Schäfer, 1988). The members of this genus have been the subject of extensive phylogenetic, molecular and biochemical investigations that have been greatly aided by the whole genome sequencing projects of recent years.

The last five years have seen the publication of the complete genomes for three *Sulfolobus* species; *S. solfataricus* (She *et al.*, 2001), *S. tokodaii* (Kawarabayasi *et al.*, 2001) and *S. acidocaldarius* (Chen *et al.*, 2005). Comparative analyses have revealed that a high proportion of predicted proteins (at least 40%) in these three genomes are specific to *Sulfolobus* and/or Archaea. In the most recently completed, *S. acidocaldarius* DSM639 (2005), 70% of protein-coding genes showed no detectable homologues in the Eukarya and Bacteria, of which 50% were exclusive to the genus and 13% unique to the organism (Chen *et al.*, 2005).

The species *S. solfataricus* has a temperature optimum of 80-85°C (Zillig *et al.*, 1980) and, in addition to being lithotrophic, can grow heterotrophically under aerobic conditions to high cell densities (Grogan, 1989). It has become one of the most comprehensively researched model organisms of archaeal metabolism and

bioenergetics (Schafer, 1996). The species is represented by several strains, including MT4, P1 and P2, which display characteristic differences with respect to their precise pH and temperature tolerance, as well as their ability to utilise different carbon and energy sources (Grogan, 1989).

Strains P1 and P2 have been shown to utilise a broad range of sugars, while strain MT4 is relatively limited in this respect. MT4 and P2, on the other hand, display a higher temperature tolerance and lower pH optimum than P1. The latter differences are small, however, amounting to less than 5 °C and 0.5 pH units, respectively. While early studies focused on *S. solfataricus* MT4 (De Rosa *et al.*, 1984; Giardina *et al.*, 1986), most of the subsequent work has been done using strains P1 and P2 and the enzymes relevant to this PhD project were initially cloned from strain P1 (DSM1616).

1.1.2 Central metabolism in *Sulfolobus solfataricus*

In all three domains polysaccharides are a major source of carbon and energy for heterotrophic growth. Their utilisation generally involves extracellular hydrolysis of polysaccharides, uptake of oligosaccharides and monosaccharides by specific transporters and further intracellular hydrolysis. Monosaccharides are subsequently catabolised via a well-conserved set of central metabolic pathways (Verhees *et al.*, 2003).

The first stage of central metabolism involves the oxidation of monosaccharides to pyruvate, carried out by glycolytic pathways. In fermentation pathways pyruvate is subsequently reduced to compounds such as acetate, or lactate. In respiration, on the other hand, pyruvate is first converted to acetyl-CoA, which is then fully oxidised via

the citric acid cycle (CAC). An electron transport system and membrane bound ATPase couple the CAC to oxidative phosphorylation. These metabolic processes in the Archaea are characterised by deviations and unique characteristics not observed in the other two domains (Danson, 1993; Brown and Doolittle, 1997).

Sugar metabolism in *S. solfataricus*:

S. solfataricus can utilise a broad range of sugars as sole sources of carbon and energy, ranging from monosaccharides to polysaccharides (Grogan, 1989; Elferink *et al.*, 2001). Assimilation of these sugar substrates relies on a variety of hydrolase activities, a number of which have been identified in the organism. They include extracellular α-amylase for the conversion of starch and dextrin to maltose and maltodextrin (Haseltine *et al.*, 1996), as well as the cytoplasmic enzymes β-glycosidase, α-fucosidase, β-xylosidase, α-xylosidase and β-glucuronidase for the hydrolysis of oligo- and disaccharides (She *et al.*, 2001).

Sugar transport into the cell is mediated predominantly by binding protein-dependent transport (Elferink *et al.*, 2001), the high affinity of such transporters believed to be essential in the organism's substrate-poor natural habitat. ABC-type transporters have been identified for the uptake of a number of monosaccharides and oligosaccharides known to support growth: glucose, galactose and mannose by the GlcS transporter (Albers *et al.*, 1999); arabinose, fructose and xylose (AraS); trehalose (TreS); maltose and maltodextrin (MalE); and cellobiose (CbtA) (Elferink *et al.*, 2001). It is also believed that secondary transport systems, driven by the pH gradient across the plasma membrane, may play a role and genes encoding several of the relevant proteins have been identified (She *et al.*, 2001).

Central metabolism in *S. solfataricus* involves glycolysis by a novel variant of the classical Entner-Doudoroff pathway (De Rosa *et al.*, 1984); production of acetyl-CoA by pyruvate:ferredoxin oxidoreductase in place of the conventional pyruvate dehydrogenase multienzyme complex (Kerscher *et al.*, 1982); and a citric acid cycle coupled to oxidative phosphorylation, (Danson *et al.*, 1985). It has been shown that the generation of a proton-motive force by respiration-coupled proton ejection, as well as driving ATPase activity in the direction of oxidative phosphorylation, contributes to the maintenance of a near neutral pH in the cytoplasm (Moll and Schäfer, 1988).

A modified Entner-Doudoroff pathway:

A variety of pathways are involved in sugar catabolism, the most common being Embden-Meyerhof (EM) glycolysis and the Entner-Doudoroff (ED) pathway. EM glycolysis is the general route of glucose degradation in all domains of life and the pathway most familiar to biochemists (Cooper, 1978). The ED pathway (Entner and Doudoroff, 1952) was identified in the 1950's as an alternative route for glucose dissimilation in microorganisms and is now known to be widely distributed in nature. Only the ED pathway will be considered here in detail as it is of most relevance to the PhD project.

In the classical ED pathway (Figure 1.1.2) glucose is first converted to glucose-6-phosphate (G-6-P) at the expense of one molecule of ATP. G-6-P is then oxidised to 6-phosphogluconate and dehydrated to 2-keto-3-deoxy-6-phosphogluconate (KDPG). KDPG is subsequently cleaved to pyruvate and glyceraldehyde-3-phosphate (GAP), the latter being converted to a second molecule of pyruvate by a five-step reaction

sequence: single step phosphorylation/oxidation of GAP to 1,3-bisphosphoglycerate;

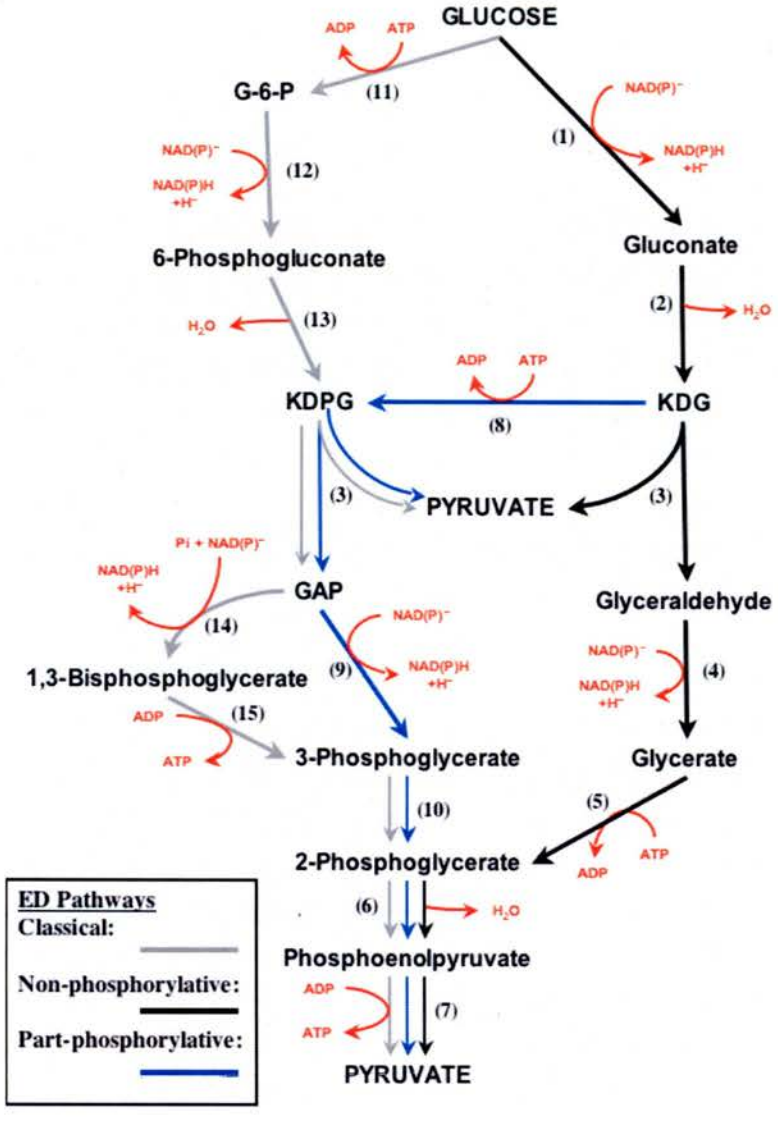


Figure 1.1.2 – The Entner-Doudoroff pathway. The classical ED pathway is shown alongside the variants in *S. solfataricus*, each corresponding to a different coloured arrow (see key). Cofactor usage is shown in red. Abbreviations: G-6-P, glucose-6-phosphate; KDG, 2-keto-3-deoxygluconate; KDPG, 2-keto-3-deoxy-6-phosphogluconate; GAP, glyceraldehyde-3-phosphate. Enzymes are denoted by numbers: **1** = glucose dehydrogenase; **2** = gluconate dehydratase; **3** = KD(P)G aldolase; **4** = glyceraldehyde dehydrogenase; **5** = glycerate kinase; **6** = enolase; **7** = pyruvate kinase; **8** = KDG kinase; **9** = non-phosphorylating GAP dehydrogenase; **10** = phosphoglycerate mutase; **11** = glucokinase; **12** = G-6-P dehydrogenase; **13** = 6-phosphogluconate dehydratase; **14** = phosphorylating GAP dehydrogenase; **15** = phosphoglycerate kinase.

dephosphorylation to give ATP and 3-phosphoglycerate; isomerisation to 2-phosphoglycerate; conversion to phosphoenolpyruvate; and a final dephosphorylation step to give pyruvate and a second ATP molecule. The overall process produces two molecules of pyruvate for every molecule of glucose, with a net ATP yield of one.

Though this classical pathway appears to be restricted to Bacteria, modifications have been identified in all three domains (Conway, 1992). ED-like pathways found in the Archaea are modified in two ways. A 'part-phosphorylative' variant of the ED pathway was first detected in the halophiles (Hochstein, 1974; Johnsen *et al.*, 2001), in which phosphorylation occurs at the level of 2-keto-3-deoxygluconate (KDG), which is converted by a KDG kinase to KDPG, rather than the general conversion of glucose to glucose-6-phosphate. In an alternative 'non-phosphorylative' variant, observed in hyperthermophilic Archaea, the phosphorylation step is omitted altogether (De Rosa *et al.*, 1984; Budgen and Danson, 1986).

This latter variant, which was first observed in *S. solfataricus* (De Rosa *et al.*, 1984), proceeds with no net yield of ATP and offers the opportunity of obtaining one pyruvate molecule per molecule of glucose without the need for any ATP consumption. It has subsequently been demonstrated that in addition to this non-phosphorylative pathway a part-phosphorylative ED variant also operates in *S. solfataricus*, an observation that has so far only been made in one other Archaeon, the hyperthermophile *Thermoproteus tenax* (Siebers *et al.*, 2004; Ahmed *et al.*, 2005). Due to variations in the component enzymes the part-phosphorylative pathway observed in *S. solfataricus* and *T. tenax* has a net ATP yield of zero, while its counterpart in the halophilic Archaea has a net yield of one molecule of ATP per molecule of glucose. This important difference is illustrative of the metabolic diversity in this third domain of life.

The part- and non-phosphorylative pathways of *S. solfataricus* share all but four enzymes in common and can be viewed as two branches in a novel ED-variant (Figure 1.1.2). Both branches begin with the oxidation of glucose to gluconate, catalysed by an NAD(P)-dependent glucose dehydrogenase (SsGDH) and subsequent dehydration of gluconate to 2-keto-3-deoxygluconate (KDG), catalysed by gluconate dehydratase (SsGD). At this point the two branches separate. In the non-phosphorylative branch KDG undergoes aldolate cleavage to pyruvate and glyceraldehyde, while in the part-phosphorylative branch KDG is first converted to 2-keto-3-deoxy-6-phosphogluconate (KDPG), catalysed by KDG kinase (SsKDGK), before undergoing aldolate cleavage to pyruvate and glyceraldehyde-3-phosphate (GAP). The same aldolase has been implicated in the aldolate cleavage step of both branches and thus represents a bifunctional KDG/KDPG aldolase (SsKDGA). This dual specificity has also been observed in the aldolase catalysing the equivalent steps in the two branches of the *T. tenax* pathway.

Following aldolate cleavage in the non-phosphorylative branch glyceraldehyde dehydrogenase catalyses the oxidation of glyceraldehyde to glycerate, which is then phosphorylated by glycerate kinase to give 2-phosphoglycerate. In the part-phosphorylative pathway GAP is first converted to 3-phosphoglycerate by the action of a non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (SsGAPN). Subsequent isomerisation to give 2-phosphoglycerate is catalysed by phosphoglycerate mutase (She *et al.*, 2001; Potters *et al.*, 2003). Two common steps complete both branches of the pathway. Enolase catalyses the conversion of 2-phosphoglycerate to phosphoenolpyruvate, from which a second molecule of pyruvate is generated by the action of pyruvate kinase.

There are currently a number of questions and speculations concerning the regulation and physiological significance of this 'branched' ED pathway. The genes encoding SsGD, SsKDGA, SsKDGK and SsGAPN are located in a cluster and all the relevant enzyme activities have been detected in cell extracts, suggesting that the two branches operate in parallel (Ahmed *et al.*, 2005). However, SsKDGA has been shown to have significantly higher catalytic efficiency with phosphorylated substrates (Ahmed *et al.*, 2005; Lamble *et al.*, 2005), indicating that glucose catabolism will occur via the part-phosphorylative branch in the presence of SsKDGK activity.

The relatively low affinity of KDG kinase for ATP could have a regulatory role, altering its activity in response to changing energy levels in the cell. Moreover, consensus promoter and termination sequences associated with the KDG kinase gene of *S. solfataricus* have been identified that are consistent with its production as a single transcript and imply that its expression can be independently regulated, acting as a switch between the two branches of the ED pathway (Lamble *et al.*, 2005).

As the non-phosphorylative branch can generate pyruvate without ATP involvement, inhibition of kinase induction could be linked to energy stress. On the other hand, it has also been suggested that the non-phosphorylative pathway might be more appropriate for growth at the upper temperature range, because of the lower stability of intermediates specific to the part-phosphorylative pathway (Ahmed *et al.*, 2004).

Promiscuity in the modified Entner-Doudoroff pathway of *S. solfataricus*:

It was shown early in the study of *S. solfataricus* sugar metabolism that glucose dehydrogenase, the first enzyme of the organism's ED pathway, was able to catalyse the NAD(P)-dependent oxidation of several monosaccharides (Giardina *et al.*, 1986).

The substrate-specificity of SsGDH isolated from cell extracts of strain MT4 included hexoses like D-glucose and D-galactose, as well as the pentose D-xylose. Based on these results it was speculated that multiple sugars might be metabolised via the Entner-Doudoroff pathway.

Extensive characterisation of the recombinant SsGDH enzyme cloned from strain P1, more than a decade later, produced similar results (Heyer, 1999). In the meantime, moreover, a β-glucosidase from the organism had been shown to also be responsible for β-galactosidase activity (Grogan, 1991), while a high-affinity transporter also characterised was shown to have specificity for both glucose and galactose (Albers *et al.*, 1999). These observations led to the suggestion that the ability of *S solfataricus* to utilise a number of different sugars (Grogan, 1989) might not rely on separate pathways, or enzymes with different specificities, but rather on the broad substrate specificities of a single set of proteins.

Systematic research over the last few years has demonstrated that *S. solfataricus* can metabolise D-glucose and its C_4 epimer, D-galactose via both branches of its Entner-Doudoroff pathway, using a single set of enzymes. Kinetic analysis has been carried out on all the enzymes of the upper ED pathway and has shown that SsGDH, SsGD, SsKDGA and SsKDGK all catalyse both glucose and galactose derived substrates with similar efficiency (Figure 1.1.3; Lamble *et al.*, 2003; Lamble *et al.*, 2004; Lamble *et al.*, 2005). These results, considered alongside the previously described evidence support the idea that the organism does not distinguish between the two sugars at any point during uptake and catabolism.

The ability of the *S. solfataricus* ED pathway to metabolise both glucose and galactose, has been described as metabolic promiscuity and there is evidence that it may extend to the catabolism of a number of other sugars (unpublished data). The

high relative activity of *Thermoplasma acidophilum* glucose dehydrogenase with galactose (Smith *et al.*, 1989) has led to the suggestion that such metabolic promiscuity may also occur in other thermoacidophilic Archaea closely related to *Sulfolobus*. The phenomenon, however, does not appear to be widespread.

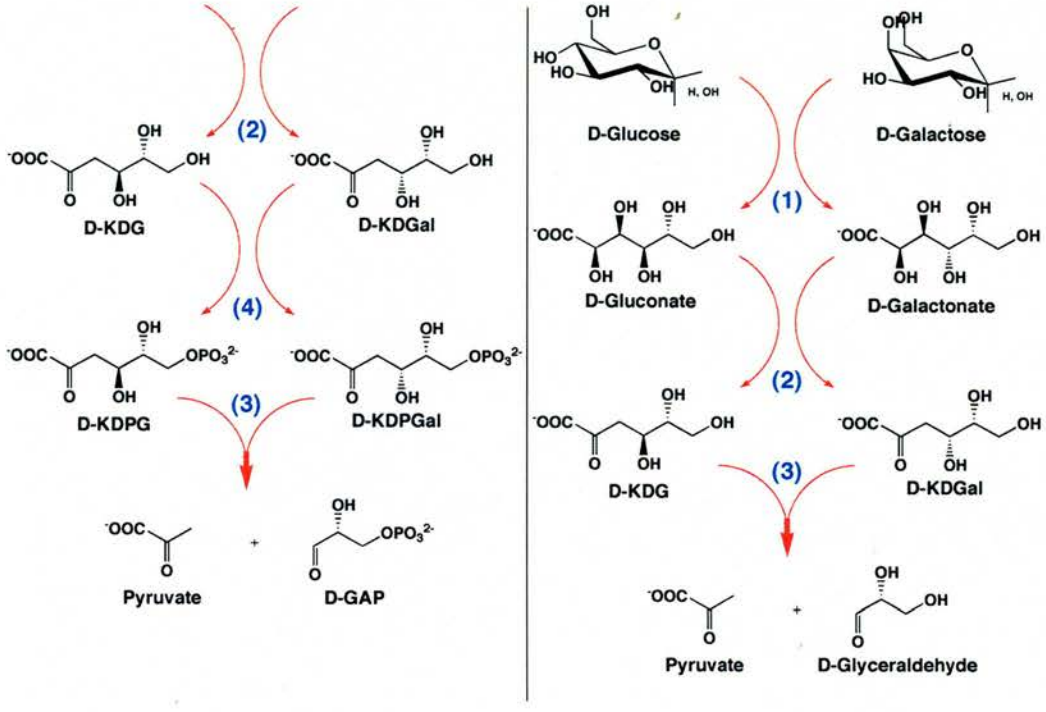


Figure 1.1.3 – Metabolic pathway promiscuity. Catalysis of D-glucose and D-galactose by the enzymes of the upper ED pathway of *S. solfataricus*. The reactions of the non-phosphorylative branch are shown on the right, while those of the part-phosphorylative branch are on the left. In the latter case, the substrates and products of the glucose dehydrogenase reaction are not shown for economy. Abbreviations: D-KDG, 2-keto-3-deoxy-D-gluconate; D-KDGal, 2-keto-3-deoxy-D-galactonate; D-KDPG, 2-keto-3-deoxy-6-phospho-D-gluconate; D-KDPGal, 2-keto-3-deoxy-6-phospho-D-galactonate; D-GAP, D-glyceraldehyde-3-phosphate. Enzymes are denoted by numbers: **1** = glucose dehydrogenase; **2** = gluconate dehydratase; **3** = KD(P)G aldolase; **4** = KDG kinase.

Although *Aspergillus* fungi also metabolise glucose via a non-phosphorylative ED pathway (Elzainy *et al.*, 1973), it has been demonstrated that they have a separate inducible pathway for galactose metabolism, and different aldolases with specificities for either KDG or KDGal (Auge and Delest, 1993; Auge and Delest, 1995).

Moreover, studies of the classical ED pathways in various bacteria have shown that they possess aldolases specific for KDPG (Shelton *et al.*, 1996). In these organisms, a DeLey-Doudoroff pathway often exists as a separate inducible route for the catabolism of galactose (De Ley and Doudoroff, 1957).

Enzyme substrate promiscuity has been predicted to play a critical role in the evolution of new activities (O'Brien and Herschlag, 1999) and the promiscuity of an entire metabolic pathway may be indicative of an early evolutionary state. It is unclear, however, whether the promiscuous ED pathway of *S. solfataricus* represents a primitive metabolic route or whether it is an adaptation to the organism's hostile natural habitat, allowing it to scavenge efficiently for energy substrates.

Part 2

The 2-keto-3-deoxygluconate aldolase of *Sulfolobus solfataricus*

1.2.1 Enzymes and catalysis

In all chemical reactions electrons from bonds broken flow to atoms where bonds will be made (Jakubowski, 2002). Consequently, the conversion of reactants to products is characterised by a build-up of partial or whole positive and negative charges (Figure 1.2.1**a**), and proceeds via an energy maximum represented by an unstable transition state (TS). Formation of the transition state is the rate-limiting step in any reaction. Charged intermediates may also be formed during a reaction as the result of localised energy minima along its coordinate. Although the TS cannot be observed directly, it can be predicted from such intermediates, which it resembles more closely than either the substrates or the products (Fersht, 1985; Clayden *et al.*, 2001; Wong and Whitesides, 1994). In the absence of intermediates the degree to which a TS resembles the products relative to reactants has been shown to be proportional to the height of the energy barrier (Morrison and Boyd, 1992).

Principles of chemical catalysis:

The thermodynamics of a reaction are dictated by the difference in free energy of the reactants and products ΔG°, while the kinetics are determined by the energy barrier $\Delta G^{\ddagger}$ between reactants and the TS (Figure 1.2.2). Even for a thermodynamically favourable reaction to take place (negative ΔG°) the energy threshold associated with transition state formation must be overcome. Moreover, having reached the activation energy the TS may decompose back to the reactants instead of being converted to

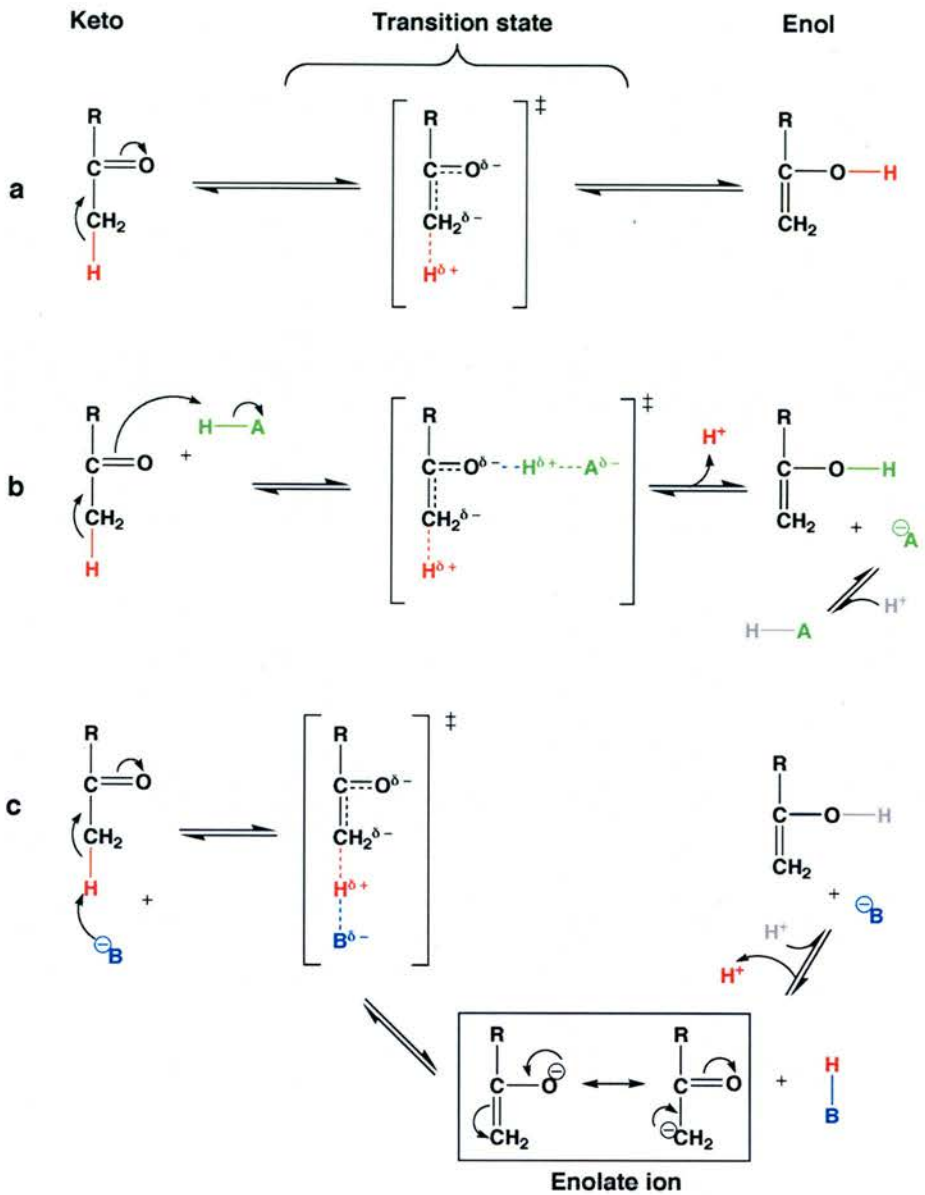


Figure 1.2.1 – Keto-enol tautomerisation. Uncatalysed enolisation of a carbonyl compound proceeds via an unstable TS, in which a negative charge builds up on the carbonyl oxygen and charge separation accompanies breaking of the C-H bond (**a**). The TS can be stabilised by an acid catalyst that donates a proton to the carbonyl oxygen (**b**). In the base catalysed reaction (**c**), breaking of the C-H bond is facilitated by stabilisation of the positive charge building up on the hydrogen. Base catalysis involves an enolate intermediate that is stabilised by delocalisation. In all cases (**a**, **b** and **c**) the tautomers are in equilibrium, which strongly favours the keto form. The rate of enolisation is, however, enhanced in the presence of catalysts, which in both **b** and **c** are regenerated at the end of the reaction. The red and grey free protons in **b** and **c** may not necessarily represent distinct hydrogen atoms. Adapted from Whitford, 2005.

products and only a certain proportion of the TS formed during a reaction will follow a productive trajectory (Garcia-Viloca *et al.*, 2004; Agarwal, 2005). Catalysts lower the activation energy of a reaction by stabilising the TS and may also induce an increase in the proportion of productive trajectories. In any case the energy difference between substrates and products is not affected and catalysis only serves to alter the kinetics of a reaction and not its thermodynamics (Jakubowski, 2002).

Although the effect is always a rate increase, there are several means by which a catalyst can function in a reaction. General acid or base catalysis involves the donation or abstraction of protons, respectively and is a frequently employed strategy for stabilising the developing charges in the TS (Figure 1.2.1**b**, **c**). This can also be achieved by electrostatic interactions and metal ion catalysts are commonly used in order to stabilise negative charges forming on the TS, as well as in enhancing the nucleophilic character of reactants (Fersht, 1985).

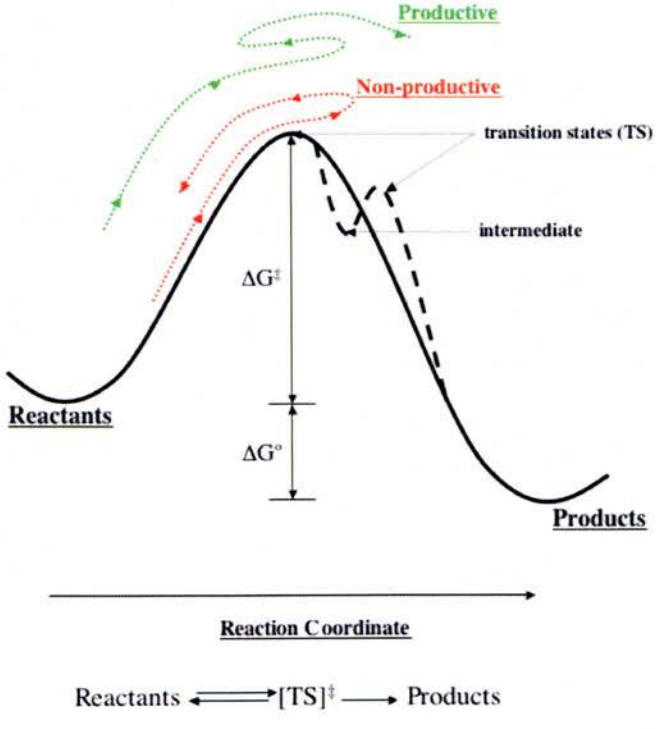


Figure 1.2.2 – Free energy profile of a chemical reaction. The solid and dashed curves represent a one-step and two-step reaction, respectively. TS's occupy 'peaks' along the reaction coordinate, while any intermediates occur at high energy 'troughs'. The reaction's activation barrier ($\Delta G^{\ddagger}$), is the energy difference between the reactants and the highest energy TS. The thermodynamic preference for the product state is determined by ΔG°. A reaction trajectory may cross the TS and then return to the reactant side (non-productive), or re-cross the barrier several times before reaching the product state (productive). Adapted from Agarwal, 2005.

An alternative approach to catalysis involves alterations to the reaction pathway, such as the formation of transient covalent bonds to the catalyst. These modifications can stabilise the TS, or serve as reactive intermediates that help overcome the energy barrier. Finally, a considerable contribution to the energy threshold of a reaction is made by the loss of entropy associated with the correct positioning and the formation of interactions between reactants (and catalyst). Control of their proximity and orientation, as observed in intramolecular reactions, can lead to significant rate enhancements by reducing this entropic term for TS formation (Whitford, 2005; Fersht, 1985).

Transition state stabilisation by enzymes:

Chemical reactions involve the formation of unstable positive and negative charges in the TS, and frequently result in the loss of entropy. Using many of the same catalytic principles described above enzymes can achieve rate increases of up to 10^{17} or 10^{19} (Radzicka and Wolfenden, 1995; Wolfenden and Snider, 2001) relative to the uncatalysed reactions in aqueous solution. The high catalytic efficiency of enzymes is attributed to many additional factors, including electrostatic pre-organisation, conformational fluctuations, desolvation effects, reactant destabilisation and quantum mechanical tunneling (Garcia-Viloca *et al.*, 2004). It is therefore necessary to consider the chemistry, structure and dynamics of an enzyme in order to fully comprehend the mechanism by which it (i) lowers the free energy of activation by stabilising the TS state relative to the reactants, and (ii) promotes (to a lesser extent) more productive barrier crossing (Garcia-Viloca *et al.*, 2004).

Precise molecular recognition is at the heart of enzymatic catalysis and active site pre-

organisation makes numerous contributions to this. Hydrophobic and hydrophilic interactions align substrates within the active site for optimal catalysis, while suitable protein or cofactor groups are employed to stabilize the TS by mechanisms of acid/base, nucleophilic, covalent, or electrostatic catalysis (Warshel, 1998; Garcia-Viloca *et al.*, 2004). In addition, the overall organisation of the active site may allow enzymes to preferentially select catalytically competent conformations of the substrate (i.e. those that most resemble the transition state) (Benkovic and Hammes-Schiffer, 2003). TS complementarity in effect induces 'straining' of the reactants towards the product structure. It results in unfavourable interactions and steric effects between enzyme and substrate being relieved in the TS, while other interactions are optimised and binding modes not available to the substrate are realised (Fersht, 1985).

Protein structure and function are determined by the properties of the component amino acids, their arrangement and their interactions. Of particular importance in catalysis are the ionisation properties of the amino acid side chains, which are determined by their pKa's and dictate their acid/base and electrophilic/nucleophilic behaviour. The ionisation constant (pKa) of a functional group is a measure of its acidity, as it is defined as the pH at which a given hydrogen atom on that group is half dissociated (Clayden *et al.*, 2001). Because key residues in the catalytic mechanism of an enzyme are often able to fulfill their function in only one ionisation state, their pKa's make a significant contribution to the pH dependence on the enzyme's activity (Fersht, 1985; Nielsen and McCammon, 2003).

It is integral to enzyme function that catalytic processes, such as nucleophilic attack or proton transfer, occur efficiently at near-physiological pH and this requires that key residues be in a specific protonic form under physiological conditions (Nielsen and

McCammon, 2003). Moreover, they must have suitable pKa's relative to other catalytic residues and to the reactants, the ionisation properties of which can change along the reaction coordinate (Cleland *et al.*, 1998). Consequently, while the ionisation properties of most residues in an enzyme remain unaltered relative to their solution free state, in the case of a small number of side chain functional groups the pKa's can be highly perturbed and may even change during the course of the reaction (McIntosh *et al.*, 1996). This tuning is made possible because the solvation, electrostatic environment and specific interactions of such residues can be regulated within the protein structure (Fersht, 1985).

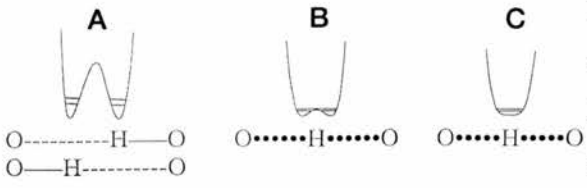


Figure 1.2.3 – Hydrogen bond energy diagrams. The energy curves define the barrier to changes in the O-H bond lengths. In a normal electrostatic hydrogen bond (**A**) the O-O distance is ~2.8 Å and the proton occupies one of two clearly defined positions. In a low barrier hydrogen bond (LBHB) of length 2.55 Å, the proton is diffusely distributed, with an average position in the center (**B**). As the O-O distance approaches 2.3 Å it becomes possible for the proton to move freely between the two oxygen atoms and the hydrogen bond (HB) develops a greater covalent character (**C**). Reproduced from Cleland *et al.*, 1998.

Hydrogen bonds (HB's) play an important role in many of the processes associated with enzyme function. Of particular interest to catalysis are low barrier (i.e. very short, very strong) hydrogen bond's (LBHB's) (Cleland and Kreevoy, 1994), which can occur in low dielectric environments between heteroatoms with well matched pKa's. In contrast to a regular electrostatic HB (e.g. O-O distance of ~2.8 Å for water) where the hydrogen is asymmetrically associated with the heavy atoms (the oxygens), LBHB's are largely covalent in character (O-O distance of 2.5 - 2.3 Å), with the proton able to move freely between the two heavy atoms (Cleland *et al.*,

1998). TS formation can be accompanied by pKa shifts that promote the transition of an electrostatic HB to an LBHB and in this way LBHB's can help stabilise the TS and facilitate proton transfer in acid/base catalysis (Gerlt and Gassman, 1993).

Enzyme-catalysed reactions can be viewed as essentially intramolecular, with all reaction steps taking place after the formation of an enzyme-substrate (E:S) complex (Jakubowski, 2002). Consequently any significant loss in entropy is associated with the formation of this complex and is compensated for by the enzyme's substrate binding energy (Fersht, 1985). Formation of the E:S complex is the result of van der Waals, hydrogen bonding, electrostatic and hydrophobic interactions within the active site and is accompanied by displacement of water molecules. The process of desolvation (or solvent substitution) plays an important role in subsequent catalysis, as the low dielectric and highly polar pre-organised environment of the protein exposes key functional groups and facilitates interactions that preferentially stabilise the TS (Benkovic and Hammes-Schiffer, 2003; Sievers *et al.*, 2004; Cannon and Benkovic, 1998).

The high binding energies that are observed reflect the increased stability of both substrate and enzyme in complex, compared to when they are individually bound to solvent. Formation of these E:S interactions can result in displacement of multiple water molecules that normally interact with both the enzyme and the substrate when they are free in solution. The entropic advantage of replacing interactions between a large number of molecules with interactions between only two (substrate and enzyme) can make a significant contribution to binding energy (Fersht, 1985). Enzymes, however, exploit only part of the 'available' binding energy to form the E:S complex, utilising potential interactions to induce greater complementarity for the TS. In doing

so they sacrifice a degree of affinity towards their substrate in favour of enhanced catalysis.

Protein Dynamics:

Proteins are not rigid structures, with a static arrangement of atoms in space. They have a dynamic nature, exhibiting internal motions over a broad range of time–scales. At one end of the spectrum are fast vibrations, consisting of the harmonic movements of bonds, angles and individual atoms, driven by the inherent kinetic energy of the protein. On the other, are slower concerted conformational motions that can involve not only surface loops and side chains but also large parts of the core structure. These coherent, collective and repeated movements differ from the random conformational fluctuations also observed (Agarwal, 2006; Clore and Schwieters, 2006).

Dynamic processes are implicit in the function of enzymes, often occurring at time-scales that coincide with substrate turnover (Eisenmesser *et al.*, 2002; Agarwal, 2005; McElheny *et al.*, 2005) and can be triggered by substrate binding, or be an intrinsic property of the enzyme, already present in its substrate-free state. In some cases the catalytic cycle involves the enzyme's interconversion between alternative structural sub-states and such transitions can limit the overall turnover rate (Wolf-Watz *et al.*, 2004; Eisenmesser *et al.*, 2005). Internal dynamics of enzymes have been implicated in altering active site accessibility to solvent, in events such as the binding and release of substrates, products and cofactors, as well as in allosteric regulation (Mulder *et al.*, 2001; McElheny *et al.*, 2005). The actual catalytic step has also been linked to dynamic events, with motions taking place along the reaction coordinate (McIntosh *et*

al., 1996; Benkovic and Hammes-Schiffer, 2003; Liang *et al.*, 2004; Garcia-Viloca *et al.*, 2004).

These motions can extend to large distances from the active site via dynamic networks of residues, allowing local changes to be propagated throughout the protein and energy transmitted from the surface to the core (Eisenmesser *et al.*, 2005). There are dynamic motions at fast time-scales that promote enzyme-substrate interactions and stabilise the TS, while slower conformational fluctuations alter the dynamic behaviour of reaction trajectories to induce more productive TS barrier crossing (Garcia-Viloca *et al.*, 2004; Agarwal, 2006). Dynamic coupling of the solvent to flexible surface loop regions allows the transfer of energy from solvent to enzyme, while networks of motions and vibrations enable it to be transmitted to the active site. In this way, thermodynamic fluctuations of the hydration-shell and bulk solvent can control internal protein dynamics and provide energy to promote catalysis (Agarwal, 2006).

1.2.2 Aldolases and the aldol reaction

Metabolism can be viewed as the sum of biochemical transformations within a cell, involving the processing of chemical bonds within and between organic molecules. In the case of energy metabolism and biosynthesis, the making and breaking of bonds between carbon atoms plays a central role. Different types of organic reactions are employed by biological systems for this purpose, of which the reversible aldol addition/condensation is one of the most frequently encountered. It occurs in the anabolic pathways of gluconeogenesis, the Calvin cycle (Stryer, 1981) and lysine

biosynthesis (Shedlarski and Gilvarg, 1970), while in its reverse, cleavage direction, it is central to carbohydrate metabolism (Grazi *et al.*, 1962; Vimr and Troy, 1985). There are several other reactions closely related to the aldol condensation (Morrison and Boyd, 1992). The Claisen ester condensation, in particular, is of great biological relevance as it is essential for fatty acid and polyketal biosynthesis (Heath and Rock, 2002; Haapalainen *et al.*, 2006).

The aldol reaction:

The aldol reaction, like the Claisen condensation, relies on the unique chemistry of the carbonyl group, specifically its reactivity and its ability to behave either as an electrophile, or as a nucleophile. The higher electronegativity of the carbonyl oxygen polarises the bond, so that the carbon atom has a partial positive charge that makes it electrophilic. A carbonyl compound with an α-carbon bonded to a proton can also tautomerise into an enol or enolate (Figure 1.2.1**a**, **c**), which behave as nucleophiles. Enolisation occurs because the carbonyl group increases the acidity of the α-hydrogen, facilitating proton transfer and shift of the double bond. Although this is normally a slow process in neutral solution it can be catalysed by acid or base (Figure 1.2.1**b**, **c**; Figure 1.2.4**i**; Clayden *et al.*, 2001).

The aldol reaction involves aldehydes and ketones. It occurs by nucleophilic attack of an enol or enolate onto a carbonyl group, giving rise to a β-hydroxy carbonyl compound, also referred to as an aldol (Figure 1.2.4**ii**). As the concentration of catalyst (eg base) increases, the rate limiting step shifts from enolisation to the addition step (Clayden *et al.*, 2001). In some cases (eg in the presence of a strong catalyst) the aldol product can undergo elimination of water to form an α,β-unsaturated carbonyl compound (Figure 1.2.4**iii**). The combination of aldol addition

followed by elimination is an aldol condensation, although the term is often used in reference to reactions in which the dehydration step does not actually occur (Clayden *et al.*, 2001).

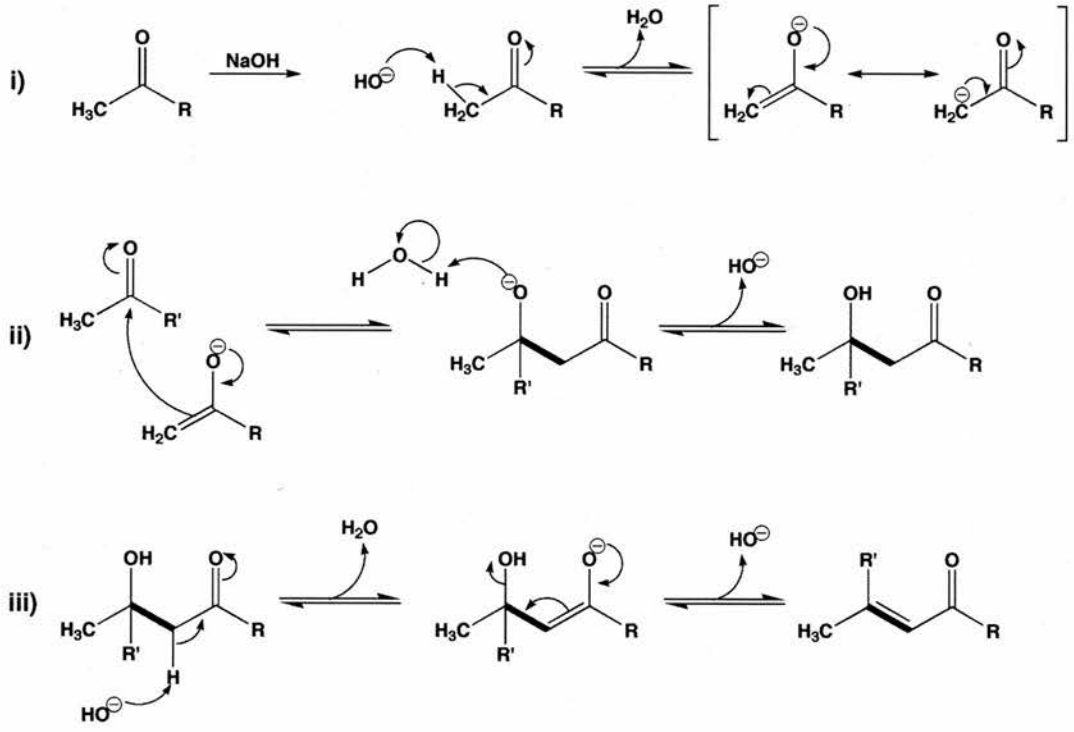


Figure 1.2.4 – The aldol reaction. The aldol condensation of two carbonyl compounds (R and R') catalysed by NaOH can be divided into three steps. **i)** Enolisation: base catalysed tautomerism gives rise to the enolate ion of R. **ii)** Aldol addition: the enolate carries out nucleophilic attack onto the carbonyl of R' to give a β-hydroxy (aldol) product. The C-C bond formed is shown in bold. **iii)** Dehydration: a second base catalysed enolisation is accompanied by E1cB elimination of a hydroxide ion and formation of the α,β-unsaturated carbonyl compound. Adapted from Clayden *et al.*, 2001.

A theoretical study involving density functional calculations was recently carried out in order to understand the process by which certain reactions requiring acid or base catalysis may occur slowly in water (Andrews *et al.*, 1973; Zhang and Houk, 2005). A three-step mechanism for the uncatalysed aldol reaction in pure water was described, using as model systems the reaction of acetaldehyde with acetaldehyde and acetaldehyde with acetone (Figure 1.2.3; Zhang and Houk, 2005). The mechanism

was determined as being the lowest available energy path for the uncatalysed reaction at room temperature. It relies on the spontaneous auto-ionisation of water (Geissler *et al.*, 2001; Bakker and Nienhuys, 2002) to give hydroxide and hydronium ions, which can then facilitate efficient enolisation of the donor species (aldehyde or ketone) by a concerted acid/base process.

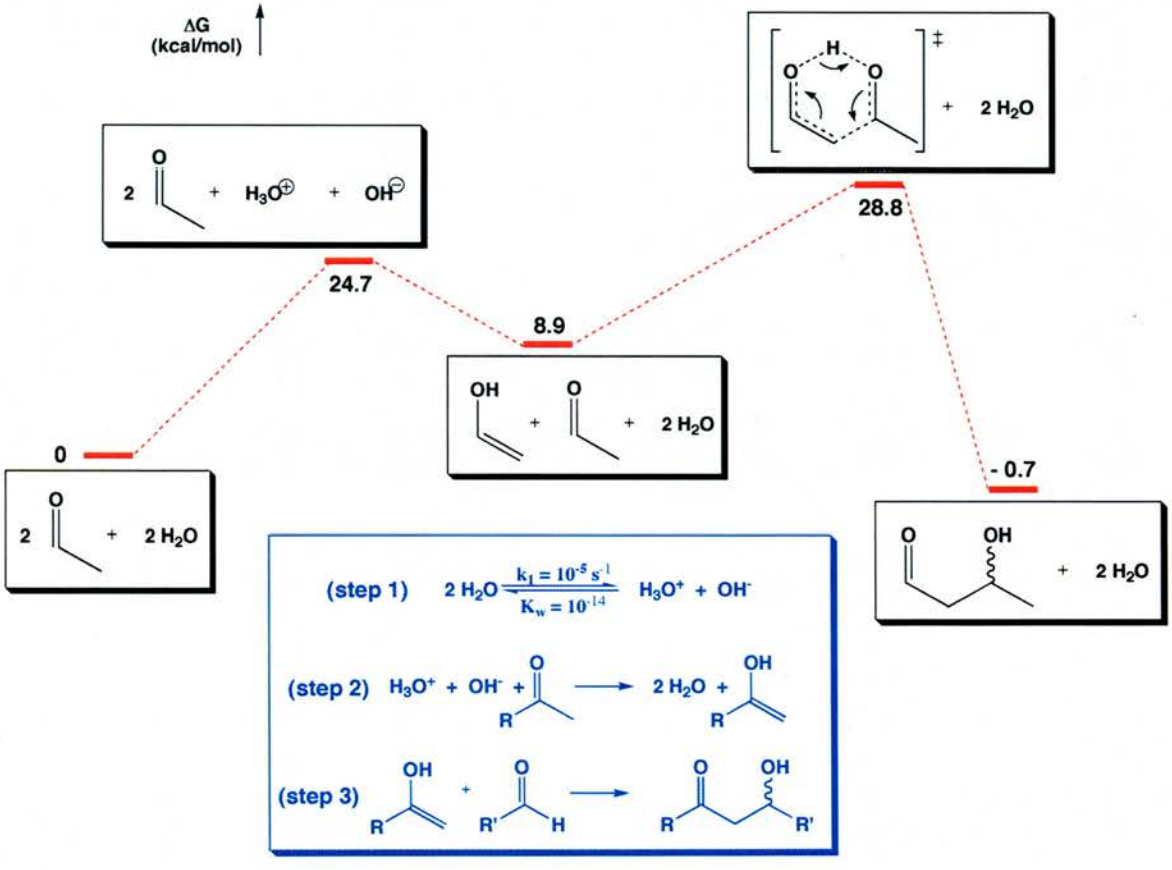


Figure 1.2.5 – Mechanism for the aldol reaction in pure water. Free energy diagram representing the pathway for the aqueous aldol reaction between two molecules of acetaldehyde. Inset in blue, the proposed three-step mechanism. The activation energy for the reaction ($\Delta G^{\ddagger}$) is 29 kcal/mol and proceeds via a concerted TS involving concurrent C-C bond formation and proton transfer between acetaldehyde and the enol. The standard Gibbs energy of the reaction (ΔG°) is 0.7 kcal/mol in favour of the aldol product. Adapted from Zhang and Houk, 2005.

The energy needed for water auto-ionisation in the first step was determined as the energy difference between isolated and solvated hydroxide and hydronium.

Interestingly, the third step was found to be rate determining, proceeding via a transition state between the enol and the acetaldehyde acceptor, with C-C bond formation and proton transfer occurring concurrently. Similar theoretical investigations previously carried out for the equivalent uncatalysed aldol reaction in organic solvent (Dickerson *et al.*, 2004) predicted that the addition step involves an identical transition state, though the mechanism for enolisation was not specified.

Aldolase enzymes:

The reversible aldol reaction in living organisms is facilitated by the diverse family of aldolase enzymes, which catalyse the reaction by stabilising an enolate intermediate through increased electron delocalisation (Takayama *et al.*, 1997). Over 30 different aldolases have been identified to date and they are divided into two distinct classes according to the mechanism they employ to achieve this delocalisation (Voet and Voet, 1995).

Type I aldolases utilise an enamine mechanism. Nucleophilic attack of the donor substrate's carbonyl by a catalytic lysine leads to the formation of a Schiff base (imine/iminium) intermediate (Grazi *et al.*, 1962; Grazi *et al.*, 1963), which tautomerises to give the activated enamine (the nitrogen derivative of an enol; (Clayden *et al.*, 2001), able then to attack the aldehyde electrophile (Horecker *et al.*, 1972). Type II aldolases, on the other hand, require a divalent metal-ion cofactor (mainly Zn^{2+}) in the active site. They are not Schiff base-forming, rather their cofactor acts as a Lewis acid, polarising the donor carbonyl in order to facilitate deprotonation of the α-acidic position and stabilise the transition state (Dreyer and Schulz, 1996; Fessner *et al.*, 1996). In both types of aldolase the formation of the

enolate (i.e. the deprotonation step) is rate determining (Machajewski and Wong, 2000).

The archetypal member of the aldolase family is D-Fructose-1,6-bisphosphate (FBP) aldolase, discovered during early work on the glycolytic pathway using muscle extracts and now known to occur in all domains of life as a repertoire of molecular forms that includes the two mechanistically distinct classes; type I (typified by the rabbit muscle enzyme) and type II (*E. coli* and yeast enzymes) (Rutter, 1964). Despite a lack of significant sequence similarity and their distinct catalytic mechanisms, the class I and II forms of FBP aldolase [both $(\alpha/\beta)_8$-TIM-barrel fold proteins with a conserved substrate phosphate binding motif] are believed to share a common evolutionary origin (Copley and Bork, 2000; Nagano *et al.*, 2002; Lorentzen *et al.*, 2004).

Type I and II activities have been observed in each of the three domains of life and, within different organisms, are involved in either direction of carbohydrate metabolism. In most contemporary life forms only one class of FBPA (either I or II) is constitutively present for carrying out hexose synthesis and breakdown, while some organisms also possess the ability to express the other enzyme class when grown under specific metabolic conditions. Whether retention of these two enzyme forms confers significant advantages is unclear, however, given the apparently overlapping physiological roles of the two types of aldolase, it has been suggested that their presence in a single organism may represent a case of catalytic redundancy (Marsh and Lebherz, 1992). Glycolysis is considered one of the earliest pathways of carbohydrate metabolism to have evolved, which would mean that its component enzymes are evolutionary quite primitive. It is therefore possible that both aldolase

activities were available in ancestral organisms and have since been eliminated by evolution from the great majority of contemporary life forms.

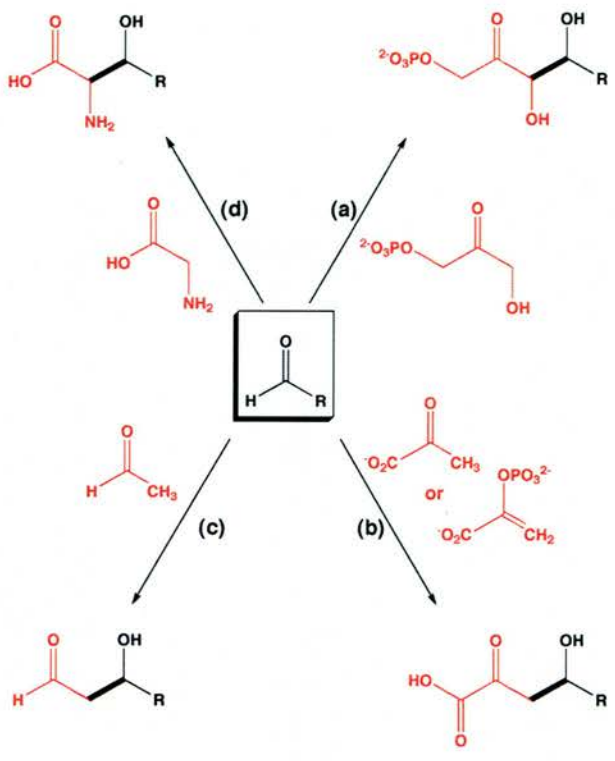


Figure 1.2.6 – Substrate specificity of aldolases. Aldolase enzymes are able to utilise a range of aldehyde acceptors, but can be categorised into four main groups according to their strict donor substrate specificity: dihydroxyacetone phosphate- (**a**), pyruvate/phosphoenol pyruvate- (**b**), acetaldehyde- (**c**) and glycine-dependent (**d**) aldolases. Adapted from Lamble, 2004.

Aldolases of both classes generally exhibit a high degree of "absolute" stereoselectivity (Chen *et al.*, 1992; Machajewski and Wong, 2000). Moreover, all aldolases are specific for their donor nucleophile, while accepting a range of aldehydes as the electrophilic substrate (Takayama *et al.*, 1997). Depending on their donor substrate specificity they are categorised into four main groups: **a)** dihydroxyacetone phosphate (DHAP), **b)** phosphoenol pyruvate and pyruvate, **c)** acetaldehyde and **d)** glycine dependent aldolases (Figure 1.2.6; Silvestri *et al.*, 2003). The dihydroxyacetone phosphate group includes FBP aldolase, D-tagatose-1,6-bisphosphate (TBP) aldolase and L-fuculose-1-phosphate aldolase (FPA). Members of the pyruvate-dependent aldolases include N-acetylneuraminate lyase (NAL) and dihydrodipicolinate synthase, as well as the KDPG(al), KDG(al) aldolases. 2-

Deoxyribose-5-phosphate aldolase is the only known member of the third group, while L-threonine aldolase is one of a number of representatives of the fourth group (Silvestri *et al.*, 2003).

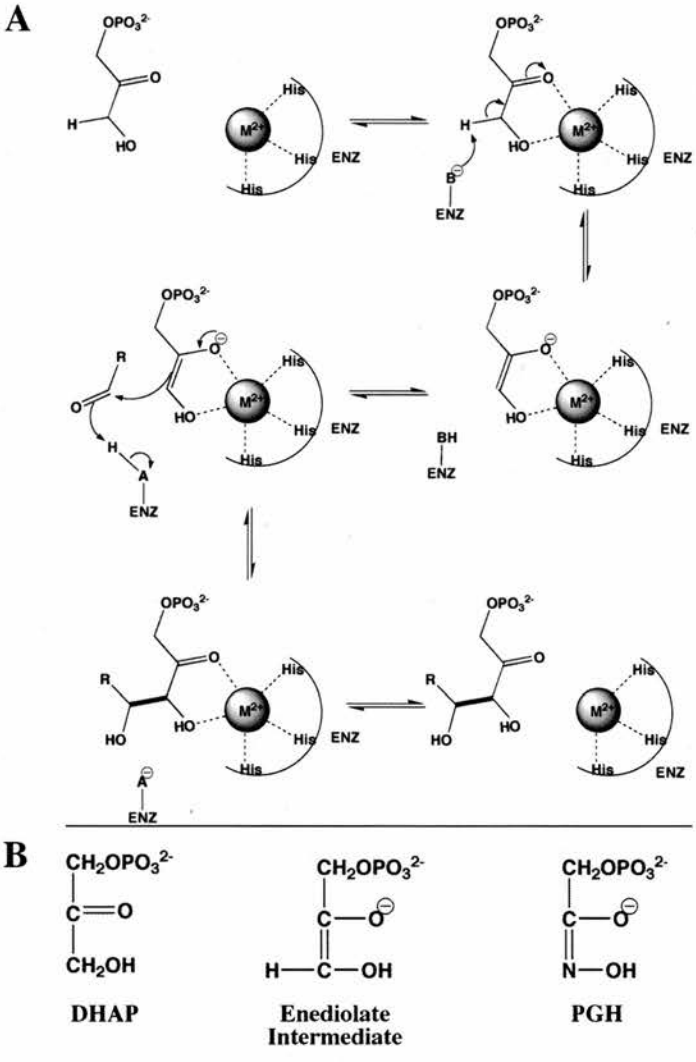


Figure 1.2.7 – Catalytic mechanism of type II aldolases. Representation of the type II aldolase catalysed reaction between DHAP and an aldehyde, as determined for *E. coli* FPA (**A**). Metal-dependent aldolases require a divalent metal ion for catalysis (usually Zn^{2+}), as well as the action of a general base (B^-) and a general acid (AH). Coordination of the donor substrate by the metal ion promotes general base catalysed proton abstraction. C-C bond formation by nucleophilic attack of the resulting enolate onto an acceptor carbonyl is facilitated by general acid mediated proton donation. The mechanism was determined by crystallographic studies of EcFPA in complex with the transition state analogue phosphoglycolohydroxamate (PGH), a potent inhibitor of DHAP dependent type II aldolases (**B**). Adapted from Fessner *et al.*, 1996.

The class II mechanism (Figure 1.2.7) as determined first for *E. coli* L-fuculose-1-phosphate aldolase (EcFPA) is initiated in the condensation direction by the coordination of the donor substrate, dihydroxyacetone phosphate (DHAP), to Zn^{2+} through both its hydroxyl and carbonyl oxygens. Polarisation of the carbonyl bond increases the acidity of the hydroxymethylene hydrogens and facilitates proton abstraction by a general base (Glu-73 in EcFPA). The nucleophilic enediolate that is

generated attacks an incoming aldehyde, assisted by a general acid (Tyr-113 in FPA)

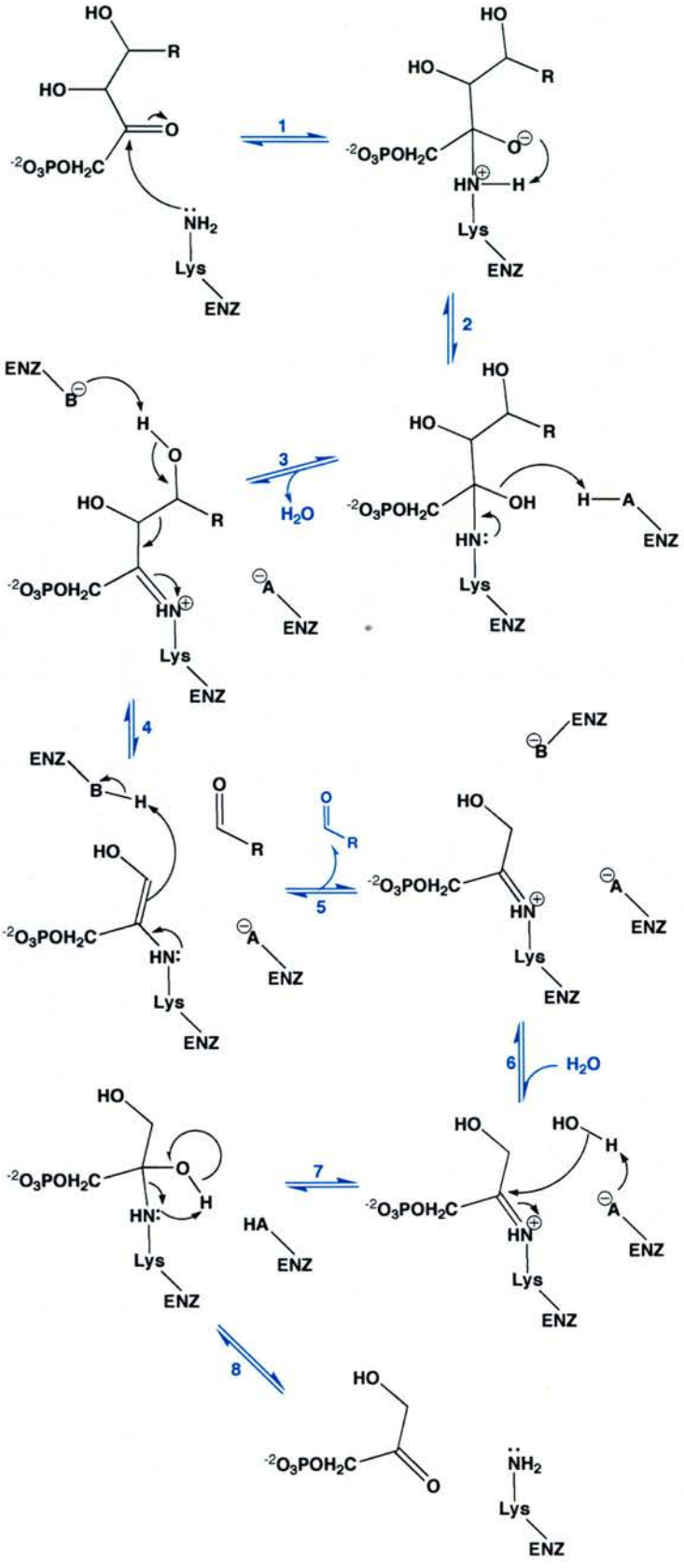


Figure 1.2.8 – Mechanism of type I aldolases. The catalytic cycle of Schiff base forming aldolases presented in the cleavage direction, as determined for the rabbit muscle FBPA enzyme. Nucleophilic attack onto the substrate carbonyl carbon by an active site lysine (**1**), results in formation of a neutral carbinolamine intermediate (**2**). Proton donation by a general acid results in dehydration to form an iminium intermediate (**3**). C-C bond cleavage is triggered by general base mediated proton abstraction (**4**). The deprotonated hydroxyl corresponds to the carbonyl of the acceptor aldehyde moiety. The aldehyde is released, while the donor substrate remains bound to the enzyme as an enamine intermediate. This tautomerises to give the iminium (**5**), which subsequently hydrolyses in the reverse of steps **2** and **1**, releasing DHAP (**6-8**). Adapted from Choi *et al.*, 2001; St-Jean *et al.*, 2005.

able to donate a proton to stabilise the developing charge. Liberation of the product from the catalyst may be preceded by ring closure involving a hydroxyl introduced by the aldehyde species (Fessner *et al.*, 1996; Dreyer and Schulz, 1996). This mechanism, determined from crystallographic studies of EcFPA in complex with the TS analogue phosphoglycolohydroxamate (PGH), has been shown to resemble that of EcFBPA (Hall *et al.*, 1999) and is believed to be generally applicable to other type II aldolases (Fessner, 1998; Izard and Blackwell, 2000).

The mechanism of type I aldolases is typified by the rabbit muscle FBP aldolase (RAMA). In the cleavage direction a reactive lysine residue in the active site carries out nucleophilic attack on the carbonyl of FBP and proceeds via a transient dipolar species to form a neutral carbinolamine intermediate. General acid mediated dehydration gives rise to the protonated iminium form of the Schiff base. Subsequent proton abstraction of the C_4 hydroxyl by general base catalysis initiates a rearrangement resulting in cleavage of the substrate C_3-C_4 bond, which gives rise to free GAP and an enamine. The enamine tautomerises to the iminium Schiff base upon protonation and proceeds via a carbinolamine intermediate to the free DHAP substrate (Maurady *et al.*, 2002; St-Jean *et al.*, 2005; Lorentzen *et al.*, 2005). While the role of general acid in the mechanism has been confidently assigned to a glutamic acid (Glu-187 in RAMA) the identity of the general base remains a matter of debate. Current evidence indicates that both catalytic roles could be fulfilled by the same glutamic acid, representing a multifunctional catalytic entity (St-Jean *et al.*, 2005). Alternatively, general base catalysis may be attributed to a conserved aspartate (Asp-33 in RAMA) (Choi *et al.*, 2001; Lorentzen *et al.*, 2005).

Catalytic antibodies and small peptides:

In recent years, catalytic-antibody technology has provided a method of developing new protein catalysts for a variety of reactions (Schultz and Lerner, 1995). Monoclonal antibodies elicited against a number of haptens designed to resemble the transition states of specific reactions have been shown to be capable of catalysing those reactions with remarkable rate accelerations. By appropriate design of the antigens, specific functional groups can be induced into the binding site of an antibody (Ab) to perform general acid/base and nucleophilic/electrophilic catalysis, as well as catalysis by strain or proximity effects. Aldolase catalytic antibodies have been developed that match the efficiency of natural aldolases while accepting a more diverse range of substrates (Machajewski and Wong, 2000).

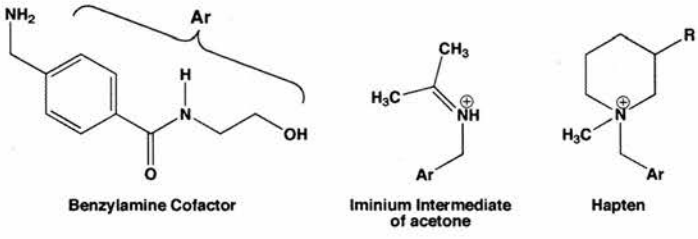


Figure 1.2.9 – Transition state complementarity. Aldolase Ab's can imitate the type I enzyme mechanism by utilising a benzylamine cofactor. They have been raised against a hapten that mimics the high-energy iminium intermediate between the cofactor and a donor carbonyl substrate. Adapted from Reymond and Chen, 1995.

The aldolase antibodies generated to date have been designed to mimic the class I enzymes. Initial success was achieved by employing a primary amine cofactor to effect enamine mediated aldol chemistry. Following the principle that transition state complementarity is paramount to enzyme catalysis, haptens were designed to mimic the high-energy iminium ion intermediate between the cofactor and a donor species (Figure 1.2.9). Antibodies elicited against these were found to possess active sites that were not only capable of binding the cofactor, but also of catalysing the aldol

condensation of acetone with aldehyde acceptors via the proposed covalent intermediate mechanism (Reymond and Chen, 1995).

Further progress was made in the field by switching the immune system's selection criteria from binding to chemical reactivity; a procedure referred to as the reactive immunisation (Figure 1.2.10**A**; Wagner *et al.*, 1995). The approach permitted the imprinting of the lysine-dependent type I aldolase mechanism in the active site and two aldolase antibodies generated by this process have demonstrated high efficiency (typical rate enhancements of 10^5-10^7 over the uncatalysed reactions) and stereoselectivity. Moreover, they have been found to possess broad substrate specificity and the ability to catalyse ketone-ketone, ketone-aldehyde and aldehyde-aldehyde aldol additions/condensations, as well as the respective retroaldol reactions (Barbas *et al.*, 1997; Hoffmann *et al.*, 1998).

Biochemical studies of the catalytic antibodies are consistent with an enamine mechanism, shared with the class I enzymes (Figure 1.2.10 **B**). Moreover, structural characterisation of one of the catalytic antibodies (Ab33F12) has revealed that the catalytic lysine is located at the bottom of a deep binding pocket in a hydrophobic environment. Its ε-amino group was found to be within hydrogen bonding distance of a water molecule that is, in turn, within hydrogen bonding distance of the hydroxyl group of a tyrosine residue. This tyrosine has been put forward as the general base catalyst in the proposed mechanism for these catalytic antibodies, its action facilitated by a presumed conformational change upon substrate binding (Barbas *et al.*, 1997; Hoffmann *et al.*, 1998).

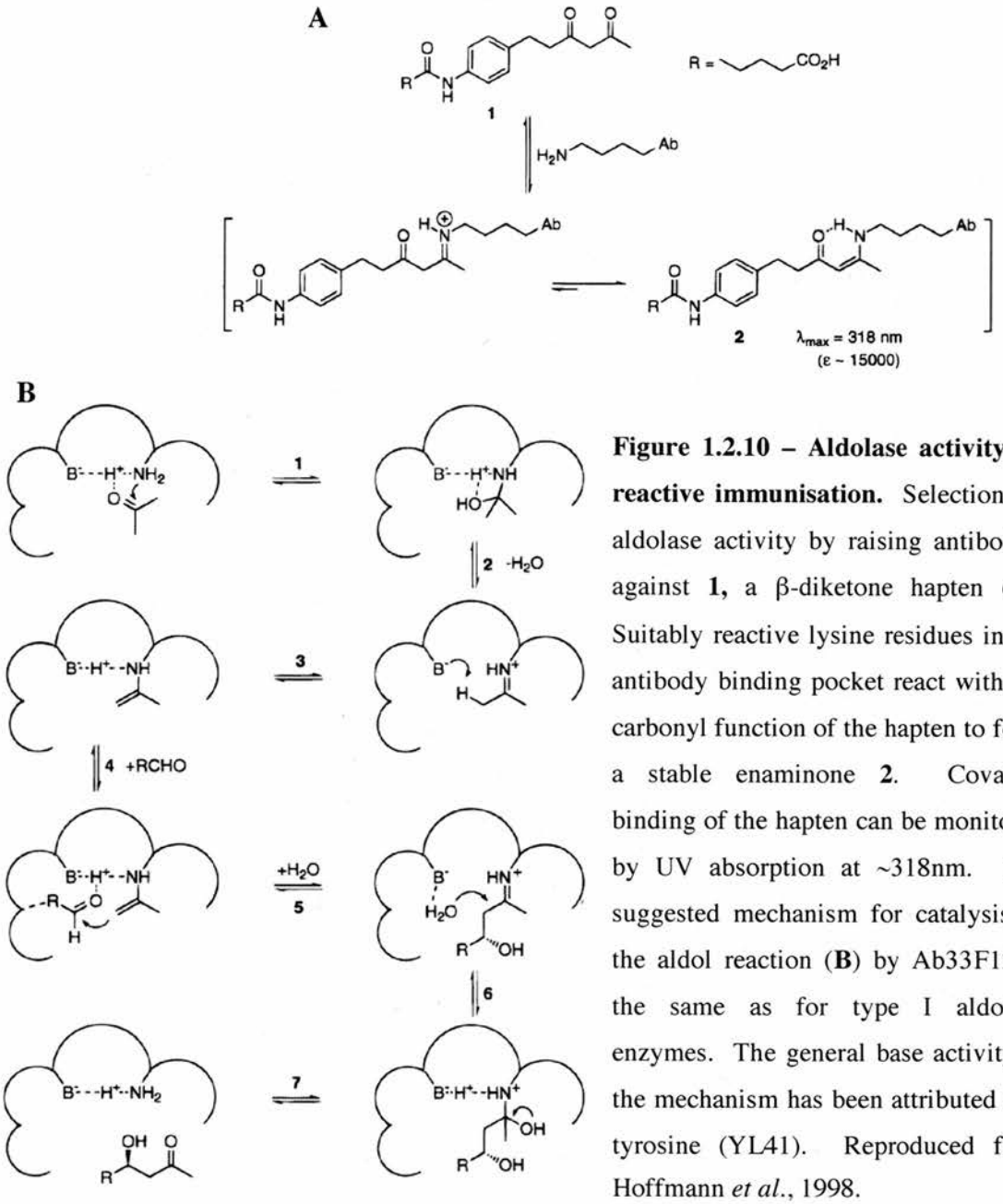


Figure 1.2.10 – Aldolase activity by reactive immunisation. Selection for aldolase activity by raising antibodies against **1,** a β-diketone hapten (**A**). Suitably reactive lysine residues in the antibody binding pocket react with the carbonyl function of the hapten to form a stable enaminone **2**. Covalent binding of the hapten can be monitored by UV absorption at ~318nm. The suggested mechanism for catalysis of the aldol reaction (**B**) by Ab33F12 is the same as for type I aldolase enzymes. The general base activity in the mechanism has been attributed to a tyrosine (YL41). Reproduced from Hoffmann *et al.*, 1998.

In addition to the catalytic antibodies described above, small designer aldolase enzymes (35 amino acids long) that operate via an enamine mechanism have also been developed. They were generated using reaction-based selection and then further improved with respect to substrate specificity by a modular assembly strategy (Tanaka *et al.*, 2005). The resulting peptide catalysts have been found to catalyse both the aldol and retro-aldol reactions via an enamine mechanism involving a catalytic lysine. Moreover, their activities correlate strongly with the increased

stability of their α-helical conformation, as well as their ability to self-associate. With selected substrates the catalytic peptides demonstrate rate accelerations of up to 1,900-fold relative to the uncatalysed reactions and achieve catalytic efficiencies per residue close to those of some larger protein catalysts.

Ribozymes:

According to the 'RNA world' hypothesis, very early life on earth consisted of macro-molecular or cellular assemblies, with RNA molecules rather than proteins catalysing chemical reactions. In support of this, a number of RNA enzymes (ribozymes) have been created or discovered that can carry out the sorts of chemical reactions that would have been necessary for life (Doudna and Cech, 2002; Fusz *et al.*, 2005). The most important example of this phenomenon and perhaps the most convincing evidence of a prehistoric 'RNA world' is the catalysis of peptide bond formation by the ribosome, which in all kingdoms of life is carried out by the RNA portion of the large subunit (Ban *et al.*, 2000).

An RNA world would have required some form of energy metabolism; a process in which aldol chemistry plays an important role within modern cells. In light of this, the recent in vitro evolution of the first known ribozyme able to catalyse an aldol reaction (Fusz *et al.*, 2005) has been considered very encouraging. From an initial library of 10^{15} RNA's, only a single sequence was obtained after 11 cycles of selection-amplification that could carry out the reaction efficiently. Nevertheless, it is plausible that such activity could have also been found by ribocytes (the presumed primordial RNA 'cells') using a more effective evolutionary search.

The ribozyme isolated is able to catalyse the aldol reaction with a 4,300-fold rate enhancement over the uncatalysed background reaction and is absolutely Zn^{2+} dependent, binding around 3 ions in a cooperative manner (Fusz *et al.*, 2005). Though the ions could simply be required to support the RNA molecule's folding and to stabilise its 'active' structure, it has been proposed that the ribozyme's catalytic mechanism may involve Zn^{2+} mediated deprotonation of the aldol donor. In such a case the ribozyme would share a fundamental mechanistic feature with the Zn^{2+} cofactor-dependent type II aldolase enzymes (Fessner *et al.*, 1996). The discovery of this aldolase ribozyme, which represents only the second carbon-carbon forming reaction for which a catalytic RNA has been found or generated (Jaschke, 2001), has lent support to the existence of RNA-originated biosynthetic pathways for metabolic sugar precursors and other biomolecules.

1.2.3 Type I aldolases

The Type I aldolases are a large and diverse group that includes the representative members fructose-1,6-biphosphate aldolase (FBPA), N-acetylneuraminate lyase (NAL), dihydrodipicolinate synthase (DHDPS), 2-keto-3-deoxy-6-phosphogluconate aldolase (KDPGA) and 2-deoxyribose-5-phosphate aldolase (DERA) (Figure 1.2.11). Although also noted as member of the family, transaldolase will not be considered here. Moreover, while the other five enzymes have been extensively characterised and display a number of important similarities and differences, these cannot be fully explored here. Instead, what follows is an outline of the most important features with respect to structure, catalytic mechanism and substrate specificity.

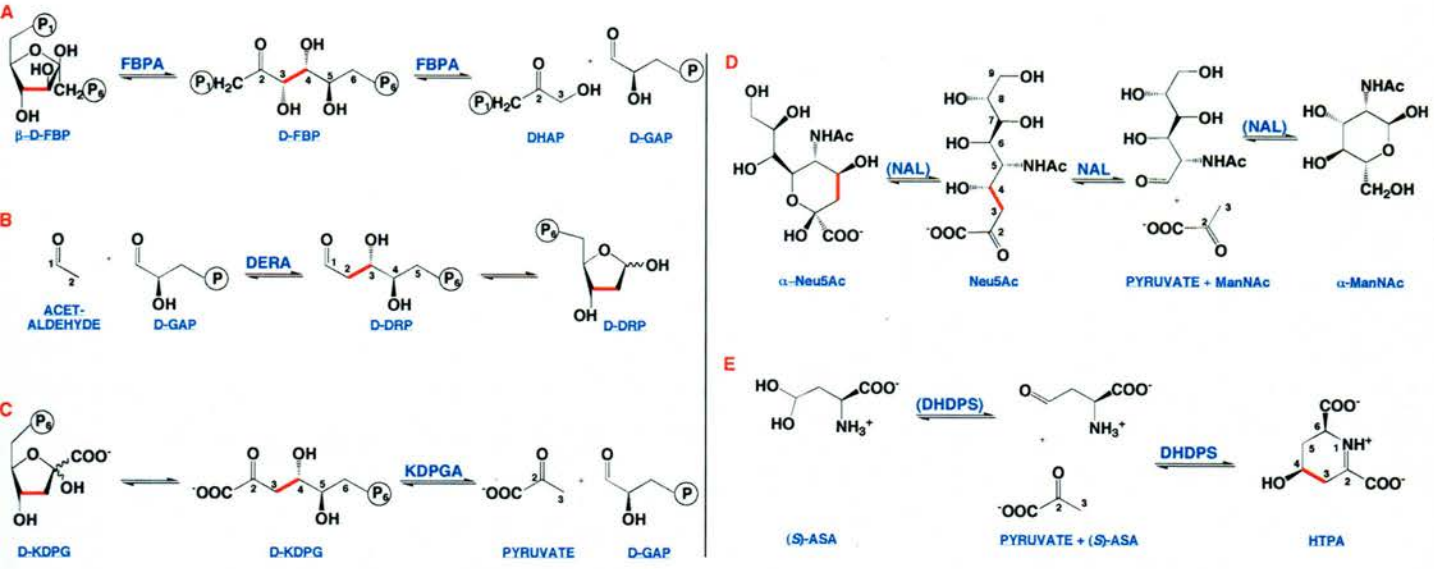


Figure 1.2.11 – Selected aldolase reactions. The reactions catalysed by five characteristic members of the class I aldolases: **A.** fructose-1,6-bisphopshate aldolase (FBPA); **B.** 2-deoxyribose-5-phosphate aldolase (DERA); **C.** 2-keto-3-deoxy-6-phosphogluconate aldolase (KDPGA); **D.** N-acetylneuraminate lyase (NAL); and **E.** dihydrodipicolinate synthase (DHDPS). Reactions are presented in their physiologically relevant direction. Brackets are used where there is uncertainty over whether the step is enzyme catalysed. Abbreviations: FBP, fructose-1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde-3-phosphate; DRP, 2-deoxyribose-5-phosphate; KDPG, 2-keto-3-deoxy-6-phosphogluconate; Neu5Ac, sialic acid; ManNAc N-acetyl-D-mannosamine; ASA, aspartate-β-semialdehyde. Adapted from Lamble, 2004.

Overall, these aldolases can be considered structurally homologous, adopting an $(\alpha/\beta)_8$-(TIM)-barrel fold, although they differ with respect to their preferred oligomeric state. In all cases, the active site is easily accessible to solvent and forms around the conserved catalytic lysine, which is located on strand β6 and points towards the barrel's centre. Moreover, those enzymes that are specific for phosphorylated substrates share a conserved binding site involving several residues, including one or two glycines positioned at the end of strand β7.

The aldolases described below vary in their choice of substrates and their ability to utilise unnatural donors and acceptors, as well as in their preferred stereoselectivity. All use an enamine mechanism involving a nucleophilic lysine, which proceeds via a

common and well characterised set of covalent intermediates and chemical transformations (Figure 1.2.8). However, the nature and identities of the general acids and general bases involved in catalysis vary between the enzymes and are less well understood.

FBP aldolase:

Fructose-1,6-bisphosphate aldolase (FBPA; EC 4.1.2.13) catalyses the aldol cleavage of D-fructose-1,6-bisphosphate (D-FBP) to dihydroxyacetone (DHAP) and D-glyceraldehyde-3-phosphate (D-GAP) (Figure 1.2.11**A**; Rutter, 1964; Horecker *et al.*, 1972). It is a ubiquitous enzyme, positioned at the metabolic junction between hexoses and trioses, where it fulfils an amphibolic function. Class I FBPA (FBPA I) enzymes are found in all three domains of life and are grouped into two separate sequence families; one with only eukaryotic members (FBPA Ie) and one with both bacterial and archaeal members (FBPA Ia) (Siebers *et al.*, 2001; Lorentzen *et al.*, 2004). Despite their lack of significant global sequence identity, the two FBPA I families share high overall structural homology and highly conserved active sites (Lorentzen *et al.*, 2003; Lorentzen *et al.*, 2005).

The eukaryotic FPBA I enzymes are homotetramers of ~160 kDa (Sygusch *et al.*, 1987; Gamblin *et al.*, 1990; Hester *et al.*, 1991; Kim *et al.*, 1998; Dalby *et al.*, 2001) and there is a high degree of sequence identity between species, FBPA from rabbit skeletal muscle (RAMA) differing from that of human muscle by only seven amino acids and sharing 70% identity with the *Drosophila melanogaster* enzyme. The archaeal family, on the other hand, is divergent with overall sequence identity between members as low as 20%. Moreover, the chain length of FBPA Ia enzymes is highly variable, ranging from 27-40 kDa (Thomson *et al.*, 1998; Siebers *et al.*, 2001)

and the quaternary structure consists of two oligomeric states in equilibrium; one most likely that of a highly stable pentamer, and the other a decamer that is observed in crystal structures and is formed by the loose association of two identical pentamers (Siebers *et al.*, 2001; Lorentzen *et al.*, 2003; Lorentzen *et al.*, 2004).

The structures of FBP I aldolase from a variety of tissues and organisms have been solved (Sygusch *et al.*, 1987; Gamblin *et al.*, 1990; Hester *et al.*, 1991; Kim *et al.*, 1998; Dalby *et al.*, 2001; Lorentzen *et al.*, 2003) and the mechanism of substrate binding and catalysis has been investigated biochemically (Grazi *et al.*, 1962; Rose and Warms, 1985; Kuo and Rose, 1985; Rose *et al.*, 1987; Morris and Tolan, 1993; Morris and Tolan, 1994; Maurady *et al.*, 2002), as well as by crystallographic analysis of covalent and non-covalent complexes between the enzyme and substrates, or substrate analogues (Blom and Sygusch, 1997; Dalby *et al.*, 1999; Choi *et al.*, 1999; Choi *et al.*, 2001; Lorentzen *et al.*, 2003; St-Jean *et al.*, 2005; Lorentzen *et al.*, 2005). Despite the presence of a number of suitable charged and polar residues in the active site (Figure 1.2.12), role assignment in catalysis has been complicated by the crystallographic trapping of substrates in non-productive or inconsistent binding modes, as well as the close proximity of these residues and the resulting susceptibility of their ionisation constants (pKa's) to electrostatic perturbation.

In RAMA (the most extensively studied FBPA Ie aldolase) Lys-229 is responsible for Schiff base formation and has also been implicated in the enzyme's ring opening activity (Figure 1.2.13; Lorentzen *et al.*, 2005). FBP exists in solution as an equilibrium between α- and β-anomers of its furanose ring form, with a ratio of 20:80 at 37 °C and only a very small percentage (~2%) is in the acyclic keto or hydrate forms (Koerner *et al.*, 1973; Midelfort *et al.*, 1976). Conversion of the cyclic FBP into its acyclic keto form requires proton donation to the bridging oxygen (O_5) and

proton removal from the 2'-hydroxyl. The enzyme has been shown to utilise only the β-anomer of FBP (β-FBP) and to catalyse ring opening as the first step in its reaction mechanism (Choi and Tolan, 2004; Lorentzen *et al.*, 2005).

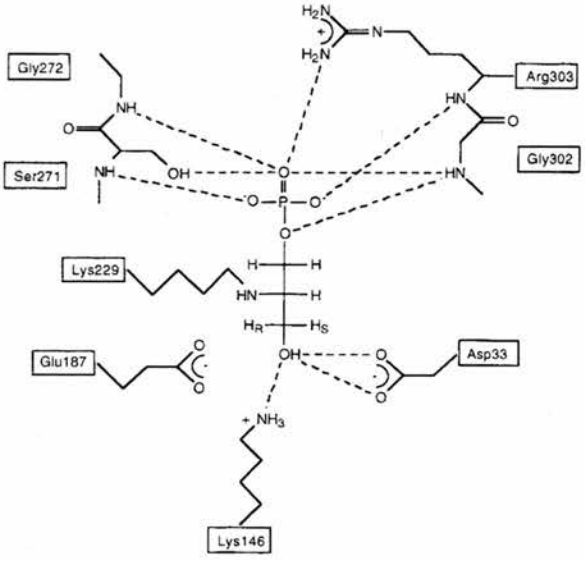


Figure 1.2.12 – FBP aldolase substrate binding. Schematic representation of the $NaBH_4$ reduced Schiff base intermediate of DHAP bound in the active site of RAMA. Lys-229 is responsible for nucleophilic attack onto the substrate's C_2 carbonyl. The enzyme's conserved phosphate binding site consists of residues Ser-271, Gly-272, Gly-302 and Arg-303 and is involved in stabilising DHAP (the P_1 phosphate of FBP). The P_6 phosphate of FBP (not shown) binds at a second site through interactions with Ser-35, Ser-38 and Lys-107. The substrate's C_3 hydroxyl makes hydrogen bonds to Lys-146 and Asp-33, while the C_4 hydroxyl of FBP (not shown) interacts with Lys-146 and Glu-187. Adapted from Choi *et al.*, 2001; St-Jean *et al.*, 2005.

Apart from Lys-229 the conserved residues Glu-187, Aps-33 and Lys-146 are also important in substrate binding and catalysis, and their mutation has been shown to significantly compromise activity (Morris and Tolan, 1993; Morris and Tolan, 1994; Morris *et al.*, 1996; Blonski *et al.*, 1997; Maurady *et al.*, 2002). The three residues make interactions with key substrate groups, correctly aligning them within the active site. Moreover, they form a network of interactions with each other and with other active site residues (Lys-107, Glu-189), promoting the appropriate protonation state of catalytic groups, as well as the substrate's electrostatic stabilisation during the proton transfers associated with catalysis (Blonski *et al.*, 1997; Maurady *et al.*, 2002; St-Jean *et al.*, 2005).

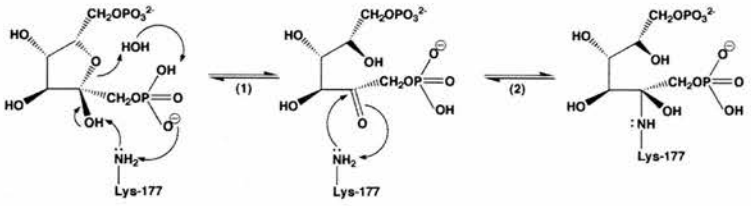


Figure 1.2.13 – Enzymatic ring opening activity. FBPA has been shown to facilitate the ring opening of its substrate. Moreover, the structure of the inactive *T. tenax* mutant, Tyr146Phe/Thr144Glu, in complex with β-FBP has enabled a rationalisation of this process. According to the proposed mechanism proton abstraction from the 2'-OH of the substrate is carried out by the catalytic lysine (Lys-177), while protonation of O_5 is attributed to an ordered water molecule. Moreover, the substrate's phosphate group is implicated in the proton shuffling. The same role has been assigned to Lys-229 of RAMA. Adapted from Lorentzen *et al.*, 2005.

The identities of the general acid and general base in catalysis are disputed. While there is significant evidence in support of Glu-187 fulfilling both functions in RAMA (Maurady *et al.*, 2002; St-Jean *et al.*, 2005), the conclusions of other studies have confirmed the residue's role as general acid, but have attributed general base catalysis to the strictly conserved Asp-33 (equivalent to Asp-24 in *T. tenax* FBPA I) (Morris and Tolan, 1993; Choi *et al.*, 2001; Lorentzen *et al.*, 2005). Moreover, although conserved in the eukaryal enzymes, the glutamic acid is observed in only two out of twenty seven archaeal enzyme sequences and is replaced by a tyrosine in another twenty. Structural and mutagenesis studies of TtFBPA have confirmed the importance of this tyrosine (Tyr-146) in catalysis and support its role as general acid in these enzymes (Lorentzen *et al.*, 2003; Lorentzen *et al.*, 2005).

It has been found that formation of the enzyme:substrate complex in RAMA is accompanied by a conformational change in the active site, involving small but significant localised backbone displacements in two adjacent α-helical structures (residues 34-65 and 302-329). This conformational change has been shown to displace the residues involved in binding the substrate phosphate groups, leading to improved interactions and improved substrate stability (St-Jean *et al.*, 2005).

Additional conformational variations have been observed in FBPA I enzymes, involving the flexible C-terminal region, and have been associated with the processes of substrate recognition and release (Hester *et al.*, 1991; Berthiaume *et al.*, 1993; Blom and Sygusch, 1997; Choi *et al.*, 1999; Choi *et al.*, 2001).

The rabbit muscle aldolase is highly specific for DHAP, though small variations can be tolerated and close analogues of DHAP have been identified that can act as weak substrates (≤10% of DHAP activity) (Bednarski *et al.*, 1989; Wong and Whitesides, 1994; Takayama *et al.*, 1997). Moreover, it has strict specificity for the P_1 phosphate of FBP, but is more tolerant with respect to P_6 and has been shown to catalyse the cleavage of fructose-1-phosphate, albeit with decreased efficiency. The enzyme is highly stereospecific, favouring D-enantiomeric substrates with a (*3S,4R*)-configuration (FBP numbering), although the stereoselectivity at C_4 and C_5 is less consistent (Rose *et al.*, 1965; Rose and Rieder, 1958; Takayama *et al.*, 1997). Moreover, while in the case of GAP the D-enantiomer is kinetically favoured by about 20-fold, lower enantioselectivity is observed with the non-phosphorylated aldehyde (Bednarski *et al.*, 1989; Wong and Whitesides, 1994).

D-2-deoxyribose-5-phosphate aldolase (DERA):

The enzyme DERA (EC 4.1.2.4) catalyses the reversible condensation of acetaldehyde with D-GAP (acting as donor and acceptor, respectively) to form 2-deoxy-D-ribose-5-phosphate (DRP) (Pricer and Horecker, 1960) and is unique amongst aldolases in using an aldehyde donor substrate (Figure 1.2.11**B**; Wong and Whitesides, 1994). DERA has been studied from a number of eukaryotes and

prokaryotes (Feingold and Hoffee, 1972) and structures have been determined for several of the microbial enzymes (Heine *et al.*, 2001; Sakuraba *et al.*, 2003; Lokanath *et al.*, 2004; Heine *et al.*, 2004).

E. coli DERA (EcDERA) is able to accept several donors, including acetone, although with much lower efficiency than acetaldehyde. In terms of its acceptor substrate specificity, the enzyme appears to react fastest with 2-hydroxyaldehydes and the D-isomers are better substrates than the L-isomers (Chen *et al.*, 1992). Moreover, a strong preference is observed for phosphorylated substrates. The chiral centre generated by the DERA reaction generally has the (*S*)-configuration (Barbas *et al.*, 1990; Wong and Whitesides, 1994; Machajewski and Wong, 2000). An interesting feature of DERA catalysis is that in many cases the product of one reaction cycle can serve as a substrate in the next and the enzyme has been shown to condense three aldehydes in a sequential and stereoselective manner, giving rise to 2,4-dideoxyhexoses. The absence of further reactions after the second condensation and the inability of certain substrates to behave in this way have been attributed to product cyclisation (Gijsen and Wong, 1994).

EcDERA consists of 259 residues and has a monomeric M_r of 28,000. Structures have been solved at ~1 Å resolution, both in its native form and in complex with DRP as a carbinolamine intermediate (Figure 1.2.14**B**). Moreover, in a third ultrahigh resolution structure, DRP has been found bound to the K201L mutant of EcDERA in the Schiff base intermediate form (Heine *et al.*, 2001; Heine *et al.*, 2004). Comparison of the native and complexed enzyme has revealed a significant deviation in the loop region between strand β8 and helix α9 (residues 237-240), which forms part of the substrate's phosphate binding site (Heine *et al.*, 2004). The conformational change

observed is consistent with a partial unwinding of the N-terminal of α9 upon phosphate binding.

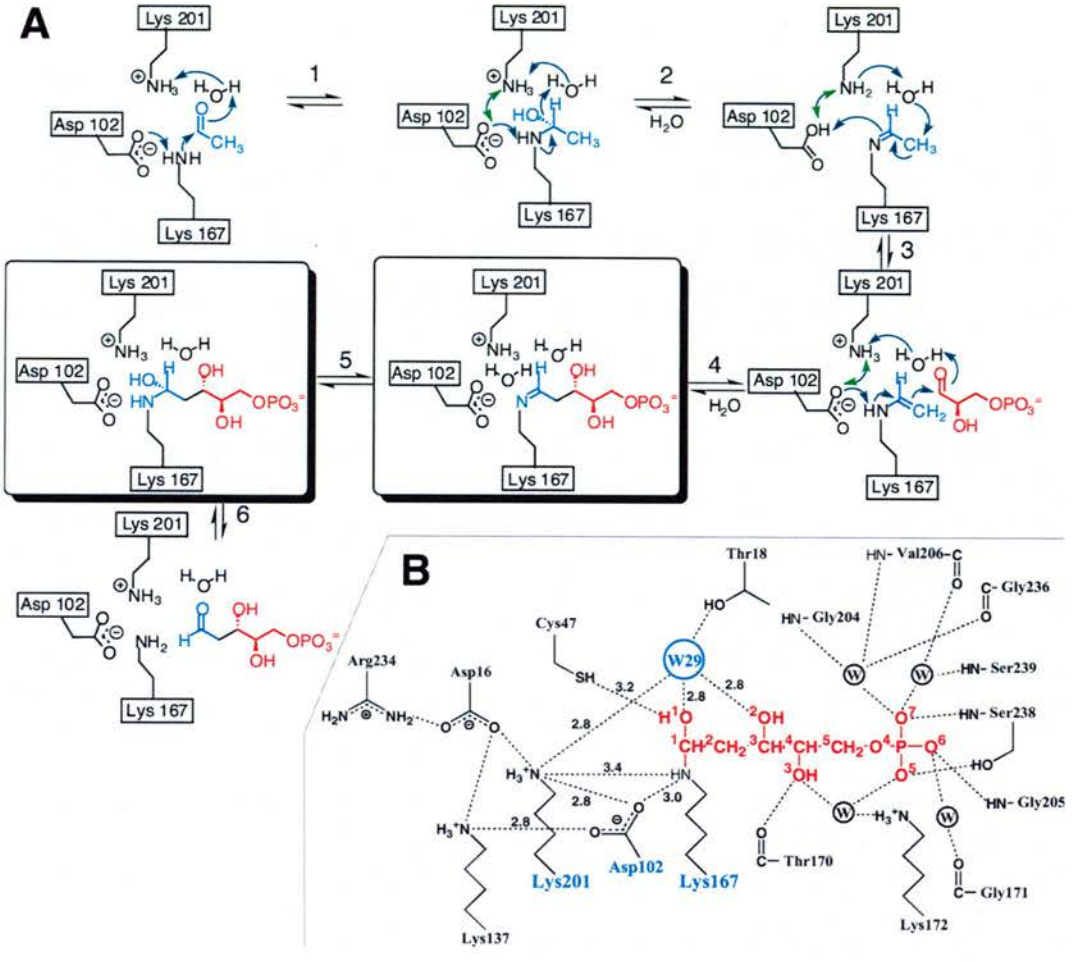


Figure 1.2.14 – DERA mechanism of substrate binding and catalysis. **A.** Proposed mechanism for the DERA catalysed condensation of acetaldehyde with GAP. Lys-167 is responsible for Schiff base formation. Asp-102, Lys-201 and an active site water molecule (W29 in **B**) are responsible for shuffling a proton between C_2 of the acetaldehyde imine and enamine, as well as subsequent C_3 hydroxyl protonation. The double arrows in green between Lys-201 and Asp-102 indicate rapid proton shuffling. **B.** In the substrate complex structure, DRP is observed as a carbinolamine intermediate. Its phosphate moiety is bound near the N-terminal of helix α9 and stabilised by direct interactions with residues Gly-205 and Ser-238 and water mediated interactions with Gly-171, Lys-172, Gly-204, Val-206, Ser-239. The intervening hydroxyl groups of the substrate interact directly with Cys-47 and via ordered water molecules with Thr-18, Thr-170, Lys-172 and Lys-201. Reproduced from Heine *et al.*, 2001.

The active site of EcDERA consists of the Schiff base forming lysine (Lys-167) and two networks of salt-bridges; one involving Lys-167, Lys-201 and the intervening residue Asp-102, and the other, Lys-201, Asp-16 and Arg-234 (Figure 1.2.14**B**). It has been suggested that the observed active site salt bridges are important for nucleophilic activation of the Schiff base forming lysine. Moreover, Lys-201 and Asp-102, along with an ordered water molecule, have been proposed to participate in a proton relay system, facilitating the donation and abstraction of hydrogen atoms during the course of the reaction (Heine *et al.*, 2001). Mutations of either Lys-201 or Asp-102 result in a 10^3-fold loss of enzyme activity, supporting the proposed mechanism (Figure 1.2.14**A**).

Crystallographic and other experimental evidence has demonstrated that the *E. coli* enzyme can exist as both a monomer and a dimer in solution (Valentin-Hansen *et al.*, 1982; Stura *et al.*, 1995). DERA's from thermophilic organisms, on the other hand, have been shown to exist as tetramers (Sakuraba *et al.*, 2003; Lokanath *et al.*, 2004). Despite the difference in their preferred oligomeric states, comparison of the DERA structures obtained from different organisms reveals that the key active site residues and the ordered water molecule proposed to participate in the proton relay system are all conserved (Heine *et al.*, 2004).

2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase:

KDPG aldolase (KDPGA; EC 4.1.2.14) catalyses the condensation between pyruvate and D-glyceraldehyde-3-phosphate (D-GAP), giving rise to 2-keto-3-deoxy-6-phospho-D-gluconate (D-KDPG) (Figure 1.2.11**C**; Conway, 1992). The enzymes from *E. coli* (EcKDPGA) and *Pseudomonas putida* (PpKDPGA) have been

extensively studied. KDPGA recognises the open chain form of KDPG, which exists at approximately 9% in aqueous solution at neutral pH (Midelfort *et al.*, 1977). Moreover, it is specific for phosphorylated D-enantiomeric substrates and operates exclusively under kinetic control in the synthetic direction, providing stereochemically pure products with the (*S*)-configuration at the C_4 position. Enzyme activity towards acceptor substrates requires a polar functionality at their C_2, C_3 or C_4 positions (Wong and Whitesides, 1994; Shelton *et al.*, 1996; Cotterill *et al.*, 1998).

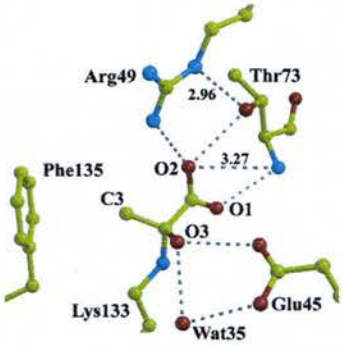


Figure 1.2.15 – Substrate binding in *E. coli* KDPG aldolase. Structure of EcKDPGA complexed with pyruvate. The substrate is covalently bound through its C_2 atom to Lys-133 in the carbinolamine intermediate form. The carboxylate of Glu-45 interacts with the carbinolamine O_3 hydroxyl both directly and via a water molecule (Wat35), while the substrate carboxylate is stabilised by Thr-73 and Arg-49. The pyruvate C_3 methyl makes a van der Waals contact with the aromatic ring of Phe-135, which has been suggested to contribute both to the stability of the intermediate and to the correct alignment of the pyruvate in the active site. Reproduced from Allard *et al.*, 2001.

Structures have been solved of PpKDPGA, EcKDPGA and most recently the *Thermotoga maritima* enzyme (TmKDPGA), and in all cases a trimeric assembly is adopted (Mavridis *et al.*, 1982; Allard *et al.*, 2001; Wymer *et al.*, 2001; Fullerton *et al.*, 2006). Covalent complexes of pyruvate trapped in the active sites of EcKDPGA and TmKDPGA have identified a set of conserved residues involved in substrate binding (Figure 1.2.15; Figure 1.2.16**A**). Moreover, sulfate and phosphate ions (components of the crystallisation medium) have been observed in all structures at a specific site predicted to be the binding pocket of the substrate phosphate group.

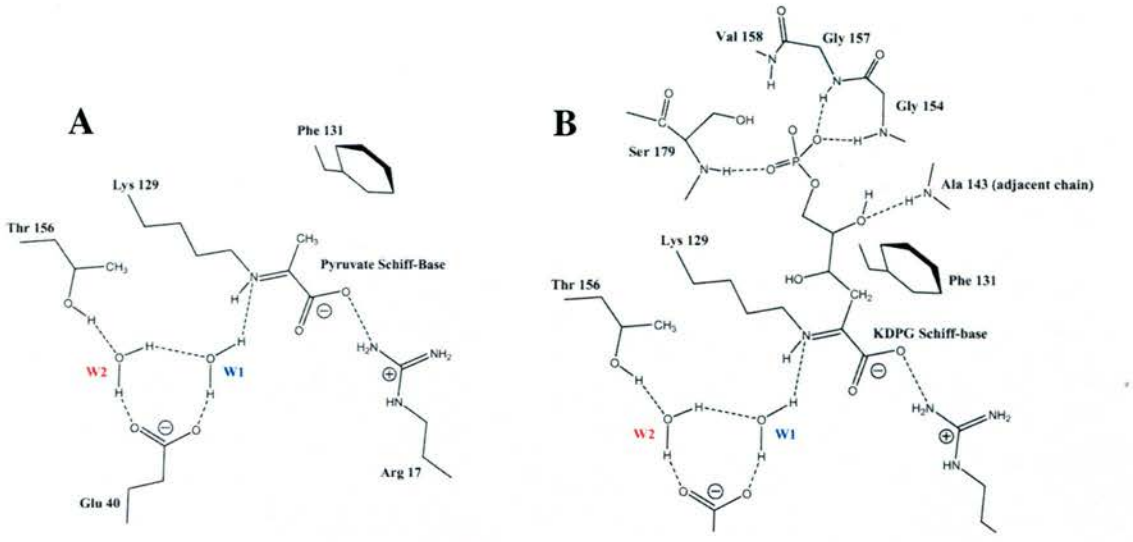


Figure 1.2.16 – Substrate binding in TmKDPGA. A. Structure of TmKDPGA in complex with pyruvate. The substrate is covalently bound to Lys-129 as a Schiff base intermediate and interacts with Arg-17, Thr-69 (not shown) and Phe-131, which are equivalent to Lys-133, Arg-49, Thr-73 and Phe-135 in *E. coli*. Glu-40, Lys-129 and the substrate C_2 are bridged by an ordered water molecule (W1). A second water (W2; equivalent to W35 in the EcKDPGA structure) forms hydrogen bonds with Glu-40, Thr-156 and W1. In the structure, a sulfate ion is located 7.7 Å from the pyruvate (not shown), in the proposed phosphate-binding pocket formed by Gly-157, Gly-158, Val-159 and Ser-179. **B.** Model of KDPG bound to the *T. maritima* enzyme based on the positions of pyruvate and sulfate molecules in the crystal structure. The residue numbers for the Val-Gly-Gly sequence were presumably intended as 159-158-157. Reproduced from Fullerton *et al.*, 2006.

Their positions, along with the coordinates of the covalent pyruvate adducts, were used in order to model D-KDPG into the active site of the *T. maritima* enzyme (Figure 1.2.16**B**; Fullerton *et al.*, 2006). Based on structural and mutagenesis evidence it has been proposed that the conserved active site glutamate (Glu-45 in *E. coli*) serves as general acid and base in KDPG aldolase catalysis (Allard *et al.*, 2001; Fullerton *et al.*, 2006). Moreover, according to the most recently proposed mechanism (Figure 1.2.17) catalysis require two mediating water molecules, corresponding to the ordered molecules W1 and W2 of the *T. maritima* crystal structure (Fullerton *et al.*, 2006).

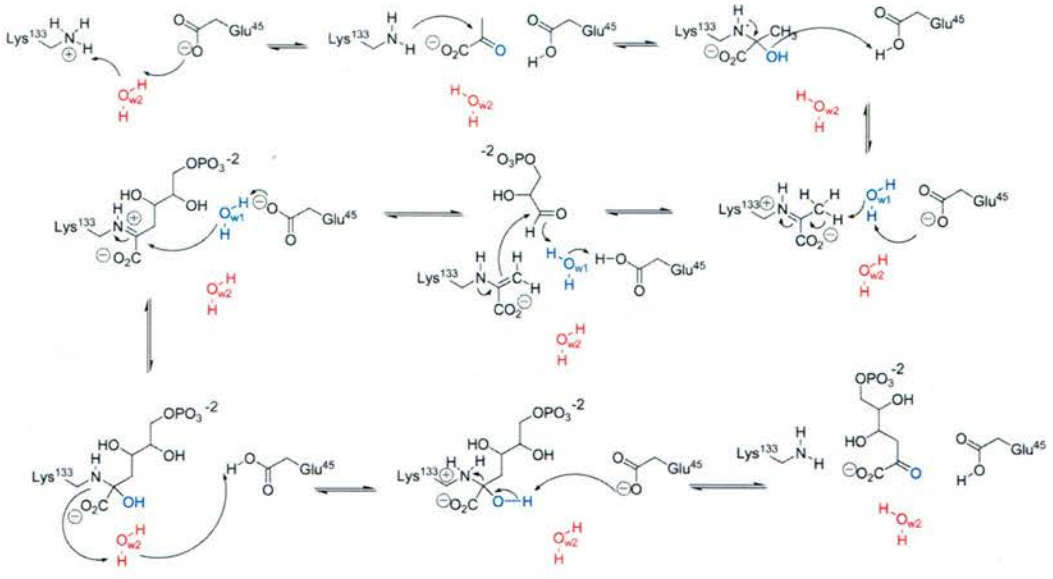


Figure 1.2.17 – Proposed catalytic mechanism for KDPG aldolases. Enzyme catalysed condensation of pyruvate with GAP to give KDPG. Catalysis is mediated by a series of proton transfers to and from a highly conserved glutamate that serves, as both a general acid and a general base. Two water molecules (W1 and W2) are crucial to the mechanism. The first is present throughout turnover, acting as a proton relay during both iminium and carbinolamine formation. The second water is transient, arising from the dehydration of the intermediate carbinolamine and consumed by the hydrolysis of the iminium, and relays protons during enolisation of the Schiff base. *E. coli* residue numbering used. W1 and W2 correspond to the ordered active site water molecules in the *T. maritima* enzyme structure. Reproduced from Fullerton *et al.*, 2006.

N-acetylneuraminate lyase (NAL):

NAL, or sialic acid aldolase (EC 4.1.3.3) catalyses the reversible condensation of pyruvate with N-acetyl-D-mannosamine (D-ManNAc) to from N-acetyl-D-neuraminic acid (D-sialic acid, or D-Neu5Ac; Figure 1.2.11**D**). The enzyme has been isolated and studied from a number of sources (Heimer and Meyer, 1956; Comb and Roseman, 1960; Uchida *et al.*, 1984; Aisaka *et al.*, 1991; Ferrero *et al.*, 1996; Lilley *et al.*, 1998). Consistent with its catabolic function in *vivo* the enzyme has an equilibrium constant that favours the retro-aldol reaction (Wong and Whitesides,

1994), although the equilibrium can be shifted towards the condensation product by providing pyruvate in excess (Uchida *et al.*, 1984).

NAL has been extensively characterised and structures have been solved for the *E. coli* (EcNAL) (Izard *et al.*, 1994; Lawrence *et al.*, 1997) and *Haemophilus influenzae* (HiNAL) enzymes (Barbosa *et al.*, 2000). They are both 130 kDa tetramers that share a high degree of structural homology, in addition to their 37% sequence identity. Crystallographic complexes of EcNAL and HiNAL with substrates and substrate analogues have also been obtained. In EcNAL complexes with pyruvate and 3-hydroxypyruvate, Lys-165 is responsible for nucleophilic attack at C_2 of the substrates' α-keto acid moiety. The substrate's carboxylate interacts with Ser-47 and Thr-48, as well as with the hydroxyl of Tyr-137 (Lawrence *et al.*, 1997). Moreover, crystal structures of HiNAL with three sialic acid analogues have permitted the prediction of additional residues involved in substrate binding (Figure 1.2.18; Barbosa *et al.*, 2000).

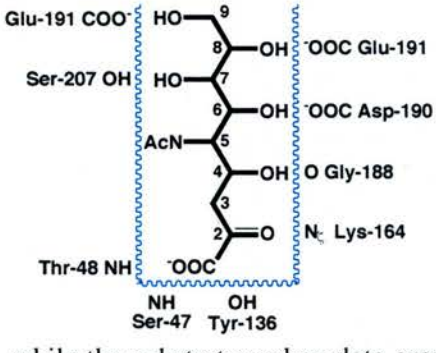


Figure 1.2.18 – NAL substrate binding. Schematic representation of the main active site interactions between HiNAL and sialic acid, predicted from the crystal structures of the enzyme in complex with Neu5Ac analogues. As observed in the EcNAL complexes, Schiff base formation occurs by nucleophilic attack from the ε-amino group of Lys-164 onto the substrate's carbonyl (C_2), while the substrate carboxylate group at C_1 is stabilised by interactions with Ser-47, Thr-48 and Tyr-136. Gly-188 hydrogen bonds to the hydroxyl at C_4 (O_4), while Asp-190 interacts with O_6 and O_8. Ser-207 also hydrogen bonds with O_6, as well as with the C_7 hydroxyl and Glu-191 stabilises the C_8 and C_9 hydroxyls. The N-acetyl group is oriented outwards from the active site and apart from a water mediated interaction with Ser-47 and Tyr-251, makes only van der Waal contacts with the protein. Adapted from Barbosa *et al.*, 2000.

The structural studies have shown that the active site of NAL lacks any acidic or basic residues that are suitably positioned to participate in proton shuffling and have lead to the proposal of a substrate assisted mechanism, with the carboxylate group of the α-keto acid moiety carrying out the function of the general acid/base in catalysis (Smith *et al.*, 1999). Moreover, it has been proposed that proton abstraction/donation catalysed by the substrate carboxylate is mediated by the hydroxyl of a conserved tyrosine (Tyr-137 in EcNAL) (Fessner and Walter, 1996; Barbosa *et al.*, 2000). This mechanism has been suggested to account for the lack of stereospecificity displayed by NAL at the C_4 position (Lin *et al.*, 1992), as it permits attack to occur from either of the opposite faces of the aldose carbonyl onto only one face of the enamine (Fitz *et al.*, 1995).

Although the β-anomers predominate in solution, it has been shown that it is the α-anomers of ManNAc and Neu5Ac that NAL recognises as substrates in the condensation and cleavage directions, respectively. The evidence supports initial ring opening within the active site and this is proposed to be the result of a shift in equilibrium between the isomeric forms (Deijl and Vliegenthart, 1983; Baumann *et al.*, 1989; Smith *et al.*, 1999; Barbosa *et al.*, 2000). There have, however, also been speculations that the enzyme might recognise only the minor open chain forms of its substrates, which would account for their apparent low binding constants.

The enzyme is strictly specific for pyruvate, but can utilise hexoses, pentoses, or tetroses as acceptor substrates and is selective for the D-configuration, although L-sugars are also tolerated (Kim *et al.*, 1988; Auge *et al.*, 1989; Auge *et al.*, 1990; Lin *et al.*, 1992; Kragl *et al.*, 1994; Fitz *et al.*, 1995). It does not, however, recognise two and three carbon acceptors (Wong and Whitesides, 1994). Moreover, the enzyme is

flexible with respect to stereochemistry and functionality at the acceptor's C_4, C_5 and C_6 positions, but is specific with respect to C_2 and C_3. The C_2 substituent is critical both in terms of size and stereochemistry, the (*S*)-configuration being favoured, while a free hydroxyl at C_3 is a prerequisite for successful aldol condensation (Auge *et al.*, 1990; Lin *et al.*, 1992; Wong and Whitesides, 1994; Kragl *et al.*, 1994; Fitz *et al.*, 1995; Takayama *et al.*, 1997).

An interesting feature of NAL catalysed reactions is that the product stereochemistry depends on the structure of the substrate and is not under strict enzymatic control, as in the case of most aldolases. While the acceptor carbonyls of most aldehyde substrates (ManNAc and D-mannose - Lin *et al.*, 1992; L-allose - Kragl *et al.*, 1994) are preferentially attacked from the *si*-face, giving rise to the (*S*)-configuration at the product's C_4 position (the sialic acid type), in some cases (L-mannose - Lin *et al.*, 1992; D-altrose and L-xylose - Kragl *et al.*, 1994) the stereochemistry can be reversed, resulting in the (*R*)-configuration at C_4. Moreover, the stereochemical course of these reactions has been correlated with the configuration of the hydroxyl at the acceptor's C_3 position, the choice of product configurations being determined by their thermodynamic stability (Lin *et al.*, 1992 - Fitz *et al.*, 1995; Gijsen *et al.*, 1996; Smith *et al.*, 1999).

Furthermore, the pathways involving *re*- and *si*-face attack have been proposed to proceed via transition states of different energies the latter being kinetically more favourable (Lin *et al.*, 1992; Kragl *et al.*, 1994; Fitz *et al.*, 1995). Consistent with this, where *re*-face attack is ultimately preferred, the product distribution changes during the course of the reaction because of the enzyme's reversibility and the thermodynamically more stable (*R*)-configured product accumulates over time (Lin *et al.*, 1992; Fitz *et al.*, 1995; Takayama *et al.*, 1997). For the reasons above, products

of most NAL-catalysed reactions possess a single stereochemistry, regardless of the preferred configuration. There have also been cases, however, where a lack of clear thermodynamic preference has been shown to result in the formation of diastereomeric mixtures under certain reaction conditions (Kragl *et al.*, 1994).

Dihydrodipicolinate synthase (DHDPS):

The enzyme DHDPS (EC 4.2.1.52) catalyses the condensation of pyruvate with (*S*)-aspartate-β-semialdehyde [(*S*)-ASA] to form (4*S*)-4-hydroxy-2,3,4,5-tetrahydro-(2*S*)-dipicolinic acid (HTPA) (Figure 1.2.11**E**; Yugari and Gilvarg, 1965). DHDPS catalyses the first key reaction common to the biosynthesis of lysine in plants and bacteria. Moreover, this amino acid has been shown to allosterically regulate the enzyme's activity by feedback inhibition (Shedlarski and Gilvarg, 1970; Blickling and Knablein, 1997; Dobson *et al.*, 2005b). *E. coli* DHDPS (EcDHDPS) been extensively studied and structures solved of the native enzyme, complexes with substrates, substrate analogues and inhibitors, as well as of several mutants (Mirwaldt *et al.*, 1995; Blickling *et al.*, 1997b; Dobson *et al.*, 2004; Dobson *et al.*, 2005b; Dobson *et al.*, 2005a). The structures have also been solved of the enzymes from *Nicotiana sylvestris* (Blickling *et al.*, 1997a), *T. maritima* (PDBid: 1O5K) and *Mycobacterium tuberculosis* (PDBid: 1XXX). All are homotetramers with an M_r 124-136,000.

The EcDHDPS structures show that Lys-161 is responsible for Schiff base formation (Shedlarski and Gilvarg, 1970). Moreover, the pyruvate carboxylate group interacts with Tyr-133, Thr-44 and Thr-45. The residues Tyr-133, Thr-44 and Tyr-107 (the latter tyrosine, provided by an adjacent subunit) are linked by hydrogen bonds between their side chain hydroxyl groups and have been proposed to serve as a proton

relay during catalysis, transferring protons to (and from) the active site *via* a water-filled channel leading to the (*S*)-lysine binding site and to the bulk solvent (Blickling *et al.*, 1997b; Dobson *et al.*, 2004). The three residues are conserved across all known DHDPS sequences and catalytic activity is substantially reduced upon mutation of any one of them (Dobson *et al.*, 2004). The currently accepted mechanism of DHDPS (Blickling *et al.*, 1997b; Hutton *et al.*, 2003) is presented in Figure 1.2.19.

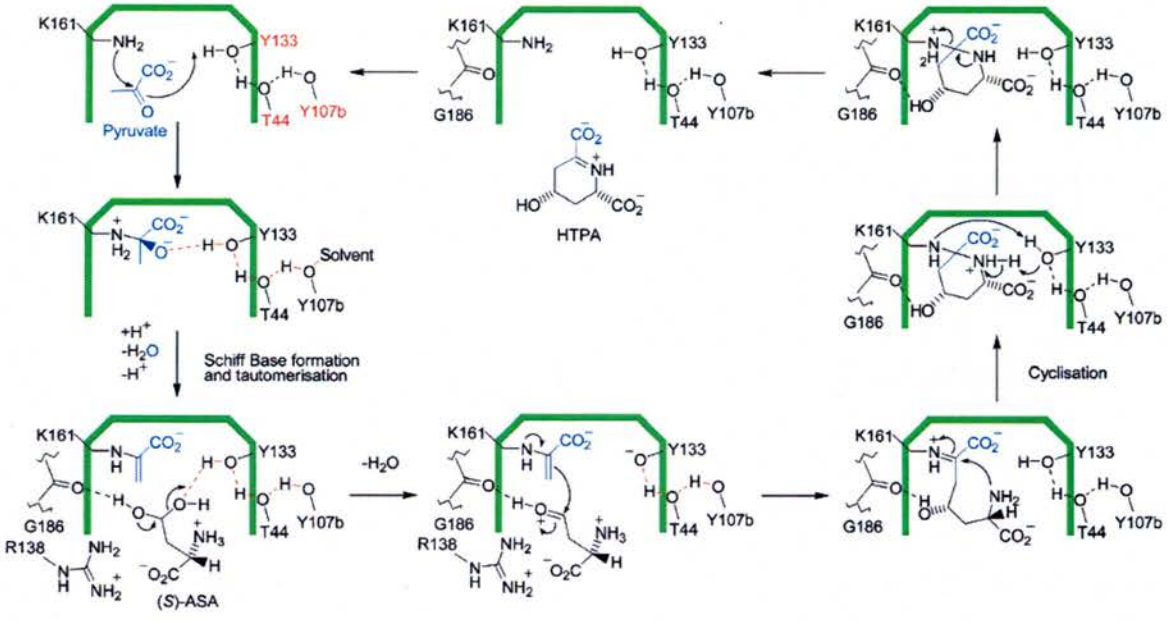


Figure 1.2.19 – DHDPS catalytic mechanism. Proposed aldol condensation of pyruvate and (*S*)-ASA catalysed by EcDHDPS, depicting key active site residues. Lys-161 forms the Schiff base with pyruvate and Arg-138 stabilises the acceptor substrate. Tyr-133, Tyr-44 and Tyr-107 (from an adjacent subunit; b), form a catalytic triad responsible for proton shuffling during the course of the reaction. In the proposed mechanism, Tyr-133 functions as both general acid and general base. Prior to nucleophilic attack from the enamine, (*S*)-ASA is dehydrated to its aldehyde form by proton donation from Y-133. Product release is preceded by formation of a cyclic imine intermediate. Reproduced from Dobson *et al.*, 2005a.

The acceptor substrate, (*S*)-ASA, is presumed to bind to the active site in its hydrated form, which predominates in solution (Tudor *et al.*, 1993; Coulter *et al.*, 1996), and to be converted to the aldehyde by the enzyme. Based on evidence from

crystallographic complexes with analogues of the substrate, its carboxylate group is stabilised through interactions with the side chain of Arg-138 (Blickling *et al.*, 1997b). This conserved residue has also been implicated in catalysis, and specifically in stabilising the catalytic triad residues, as well as modulating the ionisation properties of Tyr-133. Consistent with this assertion, mutation of Arg-138 seriously compromises enzyme activity (Dobson *et al.*, 2005a). A second residue, Asp-188, has also been implicated in stabilising the acceptor substrate through proposed interactions with its ammonium group (Blickling *et al.*, 1997b).

1.2.4 *S. solfataricus* 2-keto-3-deoxygluconate Aldolase

N-acetylneuraminate lyase (NAL) (Izard *et al.*, 1994) is the archetype for a subfamily of $(\beta/\alpha)_8$-barrel enzymes, discerned on the basis of their sequence similarity, common Schiff-base mechanism and known or predicted structural framework (Lawrence *et al.*, 1997). With the exception of D-5-keto-4-deoxyglucarate dehydratase (KDGDH; Jeffcoat *et al.*, 1969), the subfamily's members are pyruvate-dependent class I aldolases (Lawrence *et al.*, 1997). In addition to NAL, these are dihydrodipicolinate synthase (DHDPS; Mirwaldt *et al.*, 1995), trans-o-hydroxybenzylidenepyruvate hydrolase-aldolase (HBPHA; Eaton, 1994), trans-2'-carboxybenzalpyruvate hydratase-aldolase (CBPHA; Iwabuchi and Harayama, 1998) and 2-keto-3-deoxygluconate (KDG) aldolase (Buchanan *et al.*, 1999). Of these distinct enzymes, NAL and DHDPS have been the most extensively studied, and structurally characterised (Izard *et al.*, 1994; Mirwaldt *et al.*, 1995; Lawrence *et al.*, 1997; Blickling *et al.*, 1997b; Blickling *et al.*, 1997a; Barbosa *et al.*, 2000).

The NAL subfamily:

EcNAL and HiNAL share 37% sequence identity, while EcDHDPS has 27% and 25% identity with EcNAL and HiNAL, respectively. Moreover, their structures display a high degree of similarity. Both NAL and DHDPS exist as homotetramers with 222-point group symmetry, although their assembly can be more accurately described as that of a dimer of dimers (Mirwaldt *et al.*, 1995; Lawrence *et al.*, 1997). Their monomers have sequence lengths of ~295 residues and adopt $(\beta/\alpha)_8$-barrel folds, with an additional three α-helices at the C-terminus. Their active sites are situated in a pocket at the C-terminal mouth of the barrel. Within the monomer the consecutive β-strands are labelled β1-β8 and the α-helices α1-α11 (letters *a-h* and *A-K* were used by Barbosa *et al.*).

Study of the structures of NAL and DHDPS with substrate and analogue complexes by Lawrence and coworkers (Barbosa *et al.*, 2000) led to a proposed division of residues in the active sites of NAL subfamily members into two groups: (i) a conserved primary group of five amino acids that are involved in either binding of the α-keto acid moiety of the substrate, or in the aldol cleavage/condensation step; and (ii) a secondary group, associated with binding the rest of the substrate and with the catalytic steps peculiar to each particular enzyme (Barbosa *et al.*, 2000).

The primary group includes the Schiff base forming lysine and proposed catalytic tyrosine (Lys-164 and Tyr-136 in HiNAL), positioned at the C-termini of strands β6 and β5, respectively. In addition, there is the GxxGE motif situated in the loop between β2 and α2 (or on α2′), the second and third residues of which (Ser-47 and Thr-48 in HiNAL) stabilise the carboxylate on the substrate's α-keto acid moiety. A glycine residue located on the loop between β7 and α7 (loop β7-α7) has also been placed in the primary group, as it has been observed in crystal structures interacting

with the hydroxyl at C_4, from which proton abstraction occurs prior to C-C bond cleavage (Blickling *et al.*, 1997b; Barbosa *et al.*, 2000). It is worth noting that the tyrosine on β5 and the second residue of the GxxGE motif form part of a catalytic triad proposed for DHDPS. The third residue of the triad, a tyrosine on loop β4-α4 of the adjacent subunit forming the close dimer, is also conserved across the subfamily (Dobson *et al.*, 2005b), although a role has not been assigned to it in the other members.

It was predicted that the secondary group of residues would be modulated within the sub-family in order to achieve cognate substrate binding. In the case of NAL, they were identified as being an aspartate and a glutamate (Asp-190 and Glu-191 in HiNAL), located at the N-terminus of helix α7, as well as a serine on the β8-α8 loop (Ser-207 in HiNAL). Of these residues, however, only the aspartate was conserved in DHDPS. Moreover, the only residue confirmed by structural data as belonging to the secondary group of DHDPS, was an arginine (Arg-138 in EcDHDPS), located on helix α5′ and not observed in NAL. The aspartate on α7 (Asp-188 in EcDHDPS) was tentatively assigned to the group on the basis of predictions that it would interact with the ammonium group of (*S*)-ASA (Blickling *et al.*, 1997b; Barbosa *et al.*, 2000). Sequence alignments also permitted the prediction of the primary and secondary group residues in other sub-family members, including SsKDGA (Figure 1.2.20).

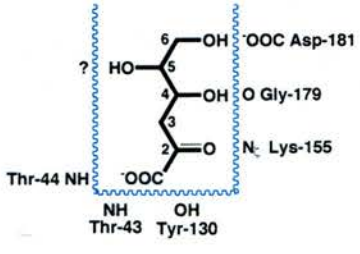


Figure 1.2.20 – The proposed active site of SsKDGA. Predicted primary and secondary group of active site residues in SsKDGA and their anticipated interactions with the substrate KDG. Residues were identified by sequence alignments with NAL and DHPDS. Adapted from Barbosa *et al.*, 2000.

2-keto-3-deoxygluconate aldolase:

The 2-keto-3-deoxygluconate aldolase of *S. solfataricus* (SsKDGA; EC 4.1.2.20) catalyses the reversible aldol cleavage step in the organism's modified Entner-Doudoroff pathway (Section 1.1.2; Figure 1.1.2; Figure 1.1.3). Initially cloned from *S. solfataricus* strain P1 (DSM1616), the SsKDGA gene encodes a 293 amino acid polypeptide with an M_r of 32,980. During initial characterisation the enzyme demonstrated high thermostability, having a half-life of 2.5 hr at 100 °C and 7.8 hr at 95 °C. Moreover, activity was detected over a wide temperature range, approximately doubling with every 10 °C rise between 30 and 80 °C.

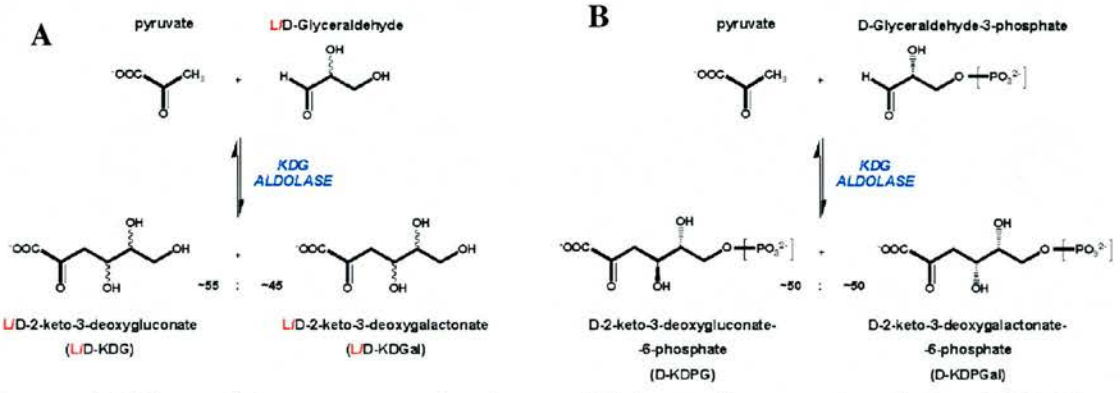


Figure 1.2.21 – Substrate promiscuity. Aldol reactions catalysed by SsKDGA. Condensation of pyruvate with D-glyceraldehyde produces equal proportions of the two diastereomers D-KDG and D-KDGal and in the cleavage direction the enzyme displays similar catalytic efficiency for both molecules (**A**). A similar lack of diastereoselectivity is observed in the case of L-glyceraldehyde (**A**) and D-glyceraldehyde-3-phosphate (**B**). While the enzyme has similar activity with D- and L-glyceraldehyde in the condensation direction, cleavage of phosphorylated substrates is at least two orders of magnitude more efficient than in the case of their non-phosphorylated substrates (Buchanan *et al.*, 1999; Lamble *et al.*, 2003; Lamble, 2004; Lamble *et al.*, 2005).

SsKDGA was shown to exist as a single homotetrameric species and was confirmed as being a type I aldolase on the basis of borohydride-induced inactivation and its insensitivity to the presence of EDTA and Zn^{2+}. It was also identified as a member of

the NAL subfamily by sequence alignments and was found to share the highest sequence identity with *H. influenzae* DHDPS (30%), as well as 28% identity with HiNAL and EcDHDPS, and 26% with EcNAL (Figure 1.2.22; Buchanan *et al.*, 1999).

A pyruvate-dependent aldolase, SsKDGA was initially believed to catalyse the reversible aldol cleavage of 2-keto-3-deoxy-D-gluconate (D-KDG) to pyruvate and D-glyceraldehyde. Extensive characterisation of the enzyme has, however, revealed a high degree of promiscuity with respect to its substrates. In addition to D-KDG, the enzyme shows comparable catalytic efficiency towards the diastereomer at C_4, D-KDGal (Lamble *et al.*, 2003) and has even greater activity with their phosphorylated counterparts (Lamble *et al.*, 2005). Moreover, in the condensation direction the enzyme has similar catalytic activity for D- and L-glyceraldehyde (Buchanan *et al.*, 1999) and condensation of either with pyruvate gives equal proportions of the two C_4 diastereomers (Figure 1.2.21; Lamble, 2004).

Initial structural characterisation of SsKDGA was carried out as part of a study on enzyme thermoadaptation (Walden, 2001). A number of crystallisation conditions were found for SsKDGA, giving rise to different crystal forms (Hendry *et al.*, 2000), however, attempts made at molecular replacement using the NAL and DHDPS structures as phasing models were unsuccessful (Walden, 2001). The structure was determined from SeMet crystals ($P2_12_12_1$: a=131.5 Å, b=132.5 Å, c=84.1 Å) grown in 0.1 M HEPES, pH 6.0, 8% propan-2-ol, 13% PEG 4,000. MAD phasing was used for structure solution and the model refined at 2.5 Å to an R-factor of 19.7% and a Free R-factor of 24.1% with good stereochemistry, 99.2% of residues occupying core regions on the Ramachandran plot (Walden, 2001).

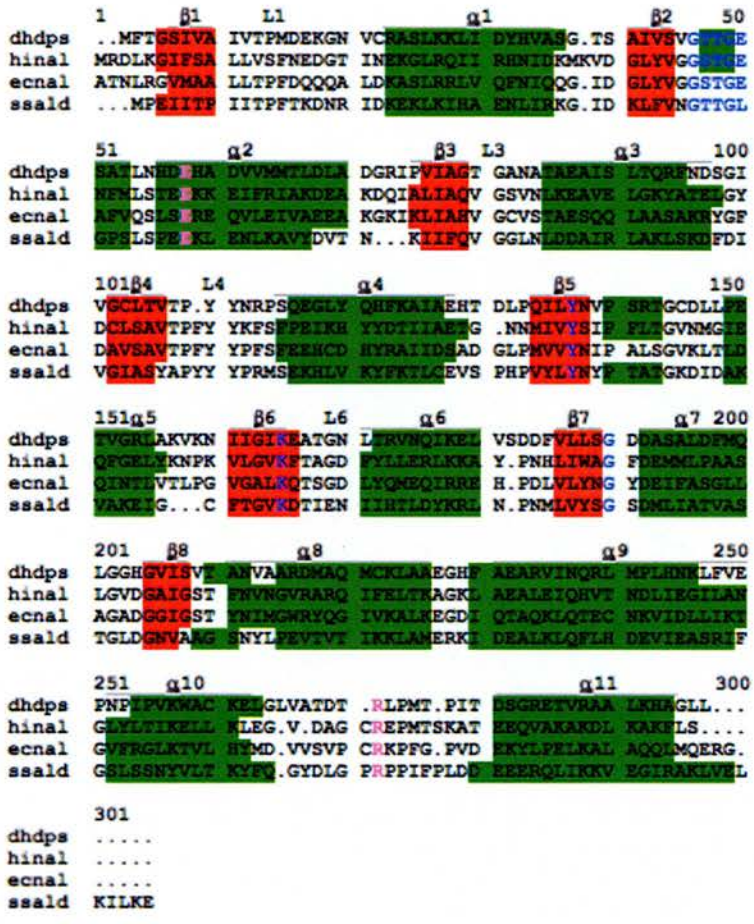


Figure 1.2.22 – The NAL subfamily. Alignment of SsKDGA (ssald), EcNAL (ecnal), HiNAL (hinal) and EcDHDPS (dhdps). β-strands are coloured in red, α-helices are in green, catalytic residues are in blue and the two residues of the conserved salt-bridge are in magenta. Reproduced from Walden, 2001.

The MAD structure was subsequently used as a search model to solve a 2.15 Å dataset obtained from a second crystal form ($P2_12_12$: a=135.1 Å, b=135.9 Å, c=188.7 Å). The crystals had been grown in 0.1 M citrate, pH 4.0, 1.6 M ammonium sulfate and predicted to consist of two tetramers in the asymmetric unit related by pseudo-centering (0, 0.04, 0.5; Hendry *et al.*, 2000). Following solution, however, the structure could not be refined below an R-factor of 25% and R-free of 28% and was characterised by distorted regions that presented difficulties during manual rebuilding. Consequently the structure was abandoned and all further analyses were carried out on the $P2_12_12_1$ structure.

As observed from the MAD structure, SsKDGA adopts the expected $(\beta/\alpha)_8$-barrel folds, including the three C-terminal α-helices observed in the structures of NAL and

DHPDS (Figure 1.2.23). Four monomers were found in the asymmetric unit, arranged as a homotetramer ABCD with a volume of 162,000 $Å^3$ (Figure 1.2.24). It was observed that the AD interface (equiv. to BC) buries 1,640 $Å^2$ compared to 1,100 $Å^2$ at the AC interface (equiv. to BD) and is therefore more extensive. Hence the SsKDGA tetramer can also be thought of as a dimer of dimers. The SsKDGA structure was found to be highly homologous with those of NAL and DHPDS, their monomer β-strands being highly superimposable and their tetramer rmsd values ranging from 2.1 to 2.3 Å. The Schiff base forming lysine (Lys-155) was ordered and clearly visible in each of the four monomers and in the same conformation as in the NAL and DHDPS structures. Moreover, the conserved intrasubunit salt-bridge between Glu-54 and Arg-261, responsible for tethering the C-terminal α-helices to the barrel, was observed. An intrasubunit disulfide formed by Cys-120 and Cys-150 was also identified in SsKDGA, but does not occur in the other two enzymes (Walden, 2001).

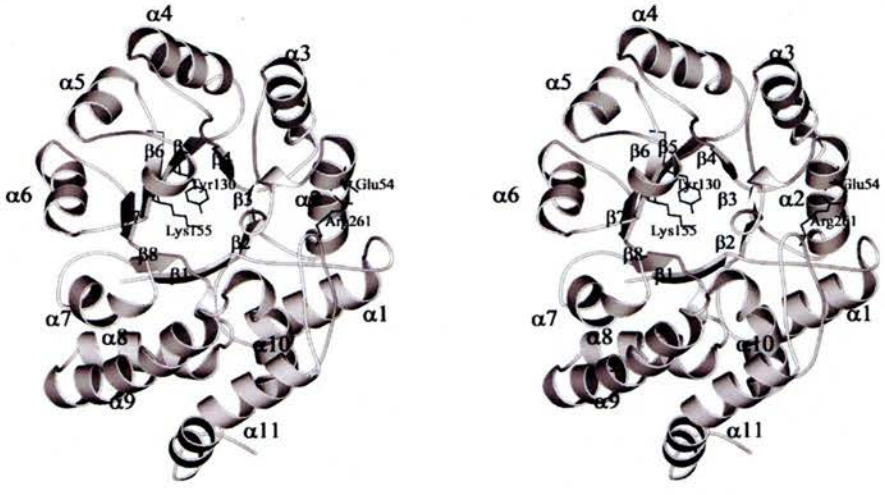


Figure 1.2.23 – The $(\beta/\alpha)_8$-barrel fold of SsKDGA. Stereo diagram of an SsKDGA monomer with secondary structural elements labelled. The enzyme adopts an $(\beta/\alpha)_8$-barrel fold with three additional α-helices at the C-terminal, tethered to the barrel by a salt bridge (E54-R261). The active site is situated in a pocket at the C-terminal mouth of the barrel and forms around the Schiff base forming lysine (K155). Reproduced from Walden, 2001.

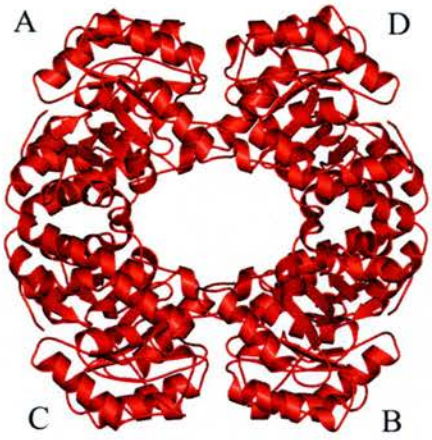


Figure 1.2.24 – The tetramer structure of SsKDGA. The SsKDGA is formed by the association of two close dimers. Consistent with this description, the AD inteface (equiv. to BC) is more extensive than the AC interface (equiv. to BD). Reproduced from Walden, 2001).

Project Aims

The focus of this project is the detailed characterisation of the mechanisms of substrate binding and catalysis in the 2-keto-3-deoxygluconate aldolase of *Sulfolobus solfataricus* (SsKDGA), using the technique of X-ray crystallography. Its aims, to achieve an understanding of the structural basis for this aldolase's promiscuous activity and permit a rational design approach to engineering desired specificities into the enzyme.

SECTION 2 - RESULTS

Part 1

Extending the resolution of the *S. solfataricus* KDGA structure

2.1.1 Cloning the KDGA gene from *S. solfataricus*

Although KDG aldolase activity has been assayed in a number of organisms, the only enzyme to have been biochemically characterised is from *S. solfataricus*. KDGA was purified from cell extract of the strain P1 (DSM 1616) and the N-terminal amino acid sequence was used to probe a genomic DNA library in lambda phage (Connaris *et al.*, 1998) for the gene sequence. The identified gene (SSO224174) was found to encode a 293 amino acid protein with M_r of 32,977 (Figure 2.1.1; Buchanan *et al.*, 1999). The complete genome of *S. solfataricus* strain P2 (DSM 1617) has since been determined using a cosmid/λ library cloning/sequencing strategy (Sensen *et al.*, 1998; She *et al.*, 2001) and an open reading frame (ORF) has been identified that corresponds to KDGA (SSO3197).

SSO3197 shares an identical sequence with SSO224174, with the exception of the former's N-terminus, which in the annotated genome includes an additional 14 amino acids (Figure 2.1.1). In *S. solfataricus* P1 the corresponding 42 base pair nucleotide sequence is a non-coding region located upstream of the KDGA gene's start codon and within it has been identified a putative Shine-Dalgarno ribosome binding sequence, consistent with the gene's position within an operon and its expression as a polycistronic transcript (Figure 2.1.1; Tolstrup *et al.*, 2000). Moreover,

characterisation of the native gene product from P1 and the 100% sequence identity observed at the level of DNA in this region of the two genomes would suggest that the additional 14 amino acids reported for the KDGA of *S. solfataricus* P2 represents a miss-annotation rather than a real difference between the enzyme in the two strains.

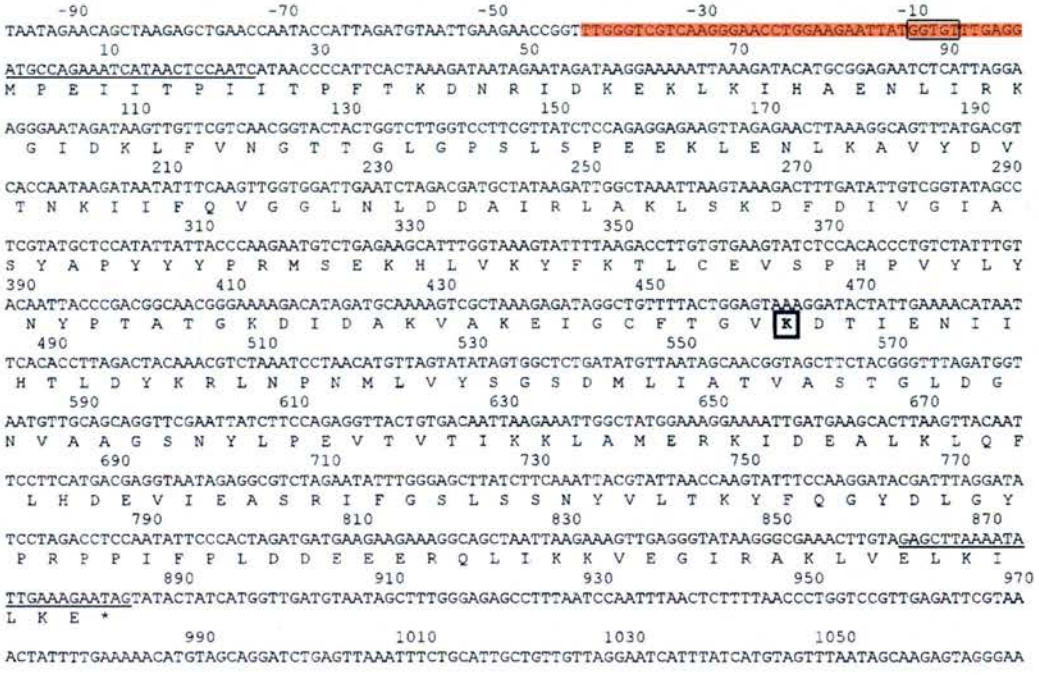


Figure 2.1.1 – The KDGA gene from *S. solfataricus* and its flanking regions. The nucleotide sequence encoding KDGA from *S. solfataricus* P1 (SSO224174) is numbered from 1-885 (including start and stop codons), with the corresponding amino acids shown below. Boxed and in bold type, the active site lysine residue (K155). Underlined; the 5' and 3' regions of the gene corresponding to the forward and reverse (reverse complement sequence) primer sequences used in amplifying the gene from the *S solfataricus* P2 genomic DNA. Highlighted in red; the 42 bases upstream of the *S. solfataricus* P1 KDGA start codon, which include a putative Shine-Dalgarno sequence (boxed). In the P2 genome, this nucleotide sequence is annotated as part of the coding region of the KDGA gene, encoding an additional 14 amino acids (MGRQGNLEELWCLR; TTG start codon) at the N-terminus. The DNA sequence shown above is identical in the two strains. Adapted from Buchanan *et al.*, 1999.

Initial work during this project (described below) focused on cloning the SsKDGA gene from P2 genomic DNA. Due to differences between the published and

amplified sequences, however, this gene was abandoned and all subsequent work was carried out using the *S. solfataricus* P1 gene for KDGA.

PCR amplification of the SsKDGA gene from genomic DNA:

The KDGA gene was amplified by PCR from *S. solfataricus* P2 genomic DNA, kindly provided by Professor M. White at the University of St Andrews. Forward and reverse primers corresponding to the 5' and complementary 3' ends of the P1 SsKDGA gene, respectively, were designed by Dr M. Ryan (Figure 2.1.1). Amplification was carried out in 50 µl reactions using *Taq* polymerase (Promega) and the products were subsequently analysed by agarose gel electrophoresis. A single band was observed in the case of certain reactions, corresponding to the expected 885 base pair (bp) product (Figure 2.1.2).

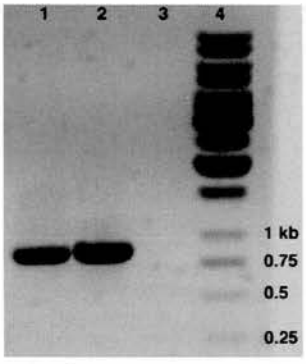


Figure 2.1.2 – PCR amplification of KDGA from P2 genomic DNA. 50 µl PCR reactions (116 ng genomic DNA, 100 pmol primers, 200 µM dNTP's and 2 mM $MgCl_2$) were set up with (lanes **1** and **2**) and without (lane **3**) 2.5 units *Taq* pol. Thermocycling protocols were carried out with (lane **2**) or without (lane **1**) an initial heat step (96 °C, 5 min → 82 °C, 2 min) prior to addition of *Taq*. 20-30 µl of reactions were subsequently run along side a 1 kb DNA ladder (lane **4**), on a 0.8% agarose gel containing 150 µM ethidium bromide. Single product bands observed in lanes **1** and **2** are consistent in size with to the 885 bp gene.

A simple thermocycling protocol [20x(96 °C, 10 s → 55 °C 10 s → 72 °C, 60 s) → 72 °C, 7 min → 4 °C ∞] gave efficient amplification of a single product in the presence of 2 mM $MgCl_2$ (Figure 2.1.2; lane 1), but no product was observed in the presence of higher or lower concentrations of Mg^{2+} (data not shown). Moreover, there was no appreciable effect on yield when an initial heat step (96 °C, 5 min → 82 °C, 2 min)

was included prior to addition of the polymerase (Figure 2.1.2; lane 2), while use of more and longer cycles [30x(96 °C, 75 s → 52 °C 90 s → 72 °C, 60 s) → 72 °C, 10 min → 4 °C, ∞] gave rise to secondary products (data not shown). Both of these modifications to the standard protocol were required when amplifying the KDGA gene from *T. acidophilum* genomic DNA (Lamble, personal communication) and may reflect that genome's higher G-C% content (46%) relative to *S. solfataricus* (36%).

Cloning SsKDGA into the pGEMT vector:

The product band (Figure 2.1.2; lane 2) was excised from the gel and purified by the protocol given in the QIAquick gel extraction kit (Qiagen). The eluted PCR product was then ligated directly into the pGEM-T Easy vector (Promega), which confers ampicillin (Amp) resistance (no prior digestions with restriction enzymes were necessary). 2 h ligation reactions were carried out at room temperature (RT) with rapid ligation buffer (Promega) and then used to transform competent *E. coli* JM109 cells (Promega), according to the manufacturer's instructions.

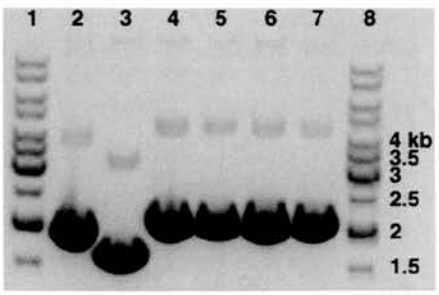


Figure 2.1.3 – Ligation of PCR amplified KDGA into the pGEM-T Easy vector. 0.8% agarose gel of plasmid DNA isolated from JM109 cell clones transformed with various ligation products and grown at 37 °C on LB-agar plates supplemented with 100 μg/ml Amp, 0.5 mM IPTG and 40 μg/ml X-Gal. Lanes **1** and **8**: 1 kb DNA ladders. Lane **2**: plasmid DNA of white colony from cells transformed with positive control ligation product (3,018 bp vector + 542 bp control insert). Lane **3**: plasmid DNA from blue colony - negative control ligation (vector + no insert). Lanes **4-7**: white colonies from cells transformed with ligation mixtures of vector and the 885 bp PCR product. The migration of plasmid DNA is consistent with the presence, absence and type of insert. The three most intense bands visible in each of lanes 2-7 correspond most likely to the supercoiled, nicked and relaxed forms of the plasmid (in order of increasing size).

Successful ligations/transformations were identified by blue/white colony screening on LB(Luria-Bertani)-agar plates supplemented with Amp, isopropyl-β-D-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (X-Gal). Selected clones were grown overnight (O/N) at 37 °C in LB broth supplemented with 35 μg/ml Amp and the plasmid DNA purified using the Wizard Plus SV miniprep purification system (Promega).

Presence of insert was further tested by electrophoresis of the plasmid DNA (Figure 2.1.3) and confirmed as being the SsKDGA gene by DNA sequencing using the T7 and SP6 primers. The sequencing results identified three point mutations relative to the published gene sequence. Two of the mutations result in an amino acid change (L84S and L170P), while the third is silent. These mutations may well have been introduced during amplification, due to the lack of proofreading by *Taq*. Unfortunately however, attempts to amplify the gene with *Pfu* polymerase, which possesses 3'-5' exonuclease activity, were unsuccessful. Consequently, further work on the P2 gene was not carried out and a pET3a vector containing the KDGA gene from *S. solfataricus* P1 was made available by collaborators at the University of Bath.

2.1.2 Expression, purification and crystallisation of recombinant SsKDGA

For the initial characterisation described by Buchanan and coworkers (1999) the *S. solfataricus* P1 KDG aldolase gene was cloned into the pREC7 plasmid and expressed in *E. coli* JM109 cells. However, for the purposes of SeMet incorporation in the methionine auxotrophic *E. coli* strain B834 (DE3), the gene had to be transferred to a pET3a expression plasmid (Kydd, 1999; Walden, 2001). The

construct (SsKDGA-pET3a) was generated by insertion of the P1 KDGA gene into the plasmid's multiple cloning site (MCS) using the restriction enzymes *Nde*I and *Bam*HI.

Both native SsKDGA and the SeMet derivative were over-expressed and purified for crystallographic studies using very similar protocols. Briefly, cells were grown at 37 °C to an OD_{600} of 0.5-0.6 and following induction, expression of SsKDGA was carried out for a further 21-24 h under the same conditions. After harvesting, cells were lysed by a combination of lysozyme treatment and sonication in 20 mM Tris/HCl pH 8.5, containing 1 mM PMSF, 1 mM EDTA and 0.1% (w/v) Triton X100. The protein was purified to homogeneity by bulk precipitation at 78 °C for 30 min, followed by anion-exchange chromatography using a 0-2 M NaCl gradient in 20 mM Tris/HCl pH 8.5 to elute the bound protein (Buchanan *et al.*, 1999; Walden, 2001).

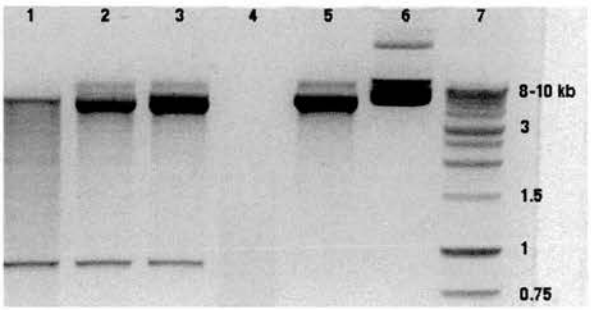


Figure 2.1.4 – Restriction analysis of the SsKDGA-pET3a vector. SsKDGA-pET3a plasmid DNA purified from transformed BL21(DE3) clones, was cut with *Nde*I and *Bam*HI (NEB) at 37 °C, according to the manufacture's instructions. Reactions were terminated by running on a 1.5% agarose gel with 150 μM ethidium bromide. A single fragment consistent in size with the SsKDGA gene was observed following double digestion (lanes **1-3).** This fragment was, however, not observed after an *Nde*I single digest (lane **5)** or in the case of plasmid DNA incubated in the absence of restriction enzymes (lane **6**). Lane **7**: 1 kb DNA ladder. The poor resolution at higher molecular weights does not permit comparison of the migration properties of the 4.6 kb vector bands.

Initial crystallisation trials of native SsKDGA were carried out using the hanging-drop diffusion method and protein at a concentration of 12 mg/ml, in 20 mM Tris pH 8.5, 0.2 M NaCl. A number of crystallisation conditions and several crystal forms

were identified (Hendry *et al.*, 2000), including orthorhombic crystals that grew in 0.1 M citrate pH 4.0, 1.6 M ammonium sulfate and diffracted to 2.15 Å on a synchrotron source. Screening around the same conditions did not, however, yield crystals of the SeMet derivative, which had to be subjected to additional trials. Following lead optimisation crystals that diffracted to ~3.5 Å in-house were obtained in 0.1 M HEPES pH 6, 8% propan-2-ol, 13% PEG 4,000, using SsKDGA at 12 mg/ml.

Transforming *E. coli* BL21(DE3):

JM109 cells containing the SsKDGA-pET3α expression vector were kindly provided by Professor M. Danson at the University of Bath and the plasmid was purified as described above. *E. coli* cells of the expression strain BL21(DE3) were made competent, transformed with the purified vector and plated onto LB-agar plates supplemented with Amp. Clones containing the vector formed colonies O/N at 37 °C.

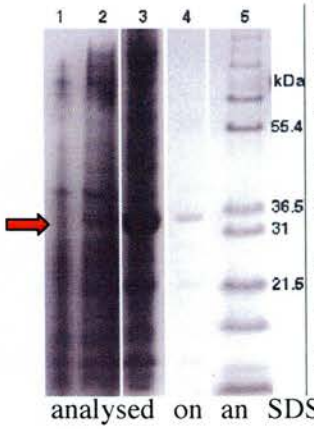


Figure 2.1.5 – Expression of SsKDGA. 10 ml starter cultures were grown O/N in LB broth containing 35 μg/ml Amp, at 37 °C in a shaking incubator. The next day 500 ml LB-Amp medium were pre-warmed and then inoculated with 5 ml of the O/N culture. Cells were grown in 2 L baffled flasks to an OD_{600} of 0.5-0.6, induced with 1 mM IPTG and grown for a further 21 h at 180 rpm. Cells were pelleted by centrifugation at 11,000 g and 4 °C, for 20 min. Pellets were stored at -20 °C. Whole cell samples of the culture taken at various stages were analysed on an SDS PAGE gel (NuPAGE; Invitrogen). A band corresponding to the monomeric M_r of SsKDGA (33 kDa) accumulated to ~30-40% of the total cellular protein during the course of induction (red arrow). Lane **1**: pre-induction. Lanes **2** and **3**: 3 and 21 h post-induction, respectively. Lane **4**: supernatant. Lane **5**: Mark12 marker (Invitrogen).

In order to confirm the integrity of the KDGA gene, plasmid DNA was once again purified from selected clones and subjected to digestion reactions with the *Nde*I and

*Bam*HI enzymes (NEB), which had originally been used to clone the gene into the vector's MCS (Kydd, 1999; Walden, 2001). Reaction products were analysed by agarose gel electrophoresis and a fragment consistent in size with the KDGA gene was observed (Figure 2.1.4). The insert's identity was confirmed by DNA sequencing, using the T7 forward and reverse primers.

Expression and purification:

The native recombinant aldolase was expressed and purified by a similar protocol to those described previously (Buchanan *et al.*, 1999; Walden, 2001; Lamble, 2004), with modifications and additions where necessary. Cell cultures were grown from a

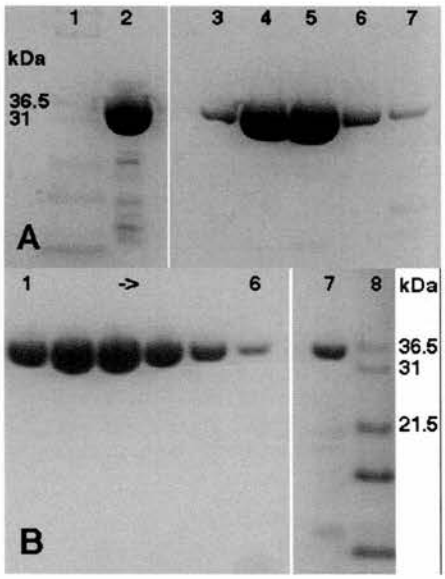


Figure 2.1.6 – Purification of recombinant SsKDGA. The cell pellet was resuspended at 0.2 g/ml on ice, in 20 mM Tris/HCl pH 8.5 containing 1 mM PMSF and 2 Complete (EDTA+) protease inhibitor tablets (Roche). Incubation with 100 μg/ml lysozyme in a water bath at 37 °C, with occasional mixing, was followed by treatment with 10 μg/ml DNaseI at RT for 15 min, in the presence of 1 mM $MgCl_2$. Complete lysis was achieved by 5x30 s bursts of sonication at 10 μm on a Soniprep 150 (MSE). The cell debris was pelleted at 11,000 g, 4 °C, for 30 min. After a 30 min incubation in a water bath at 78-80 °C, the lysate was spun for a further 20 min, to pellet the denatured protein. The supernatant was syringed through a 0.2 μm filter (Millipore) loaded onto a HiTrap Q-Sepharose HP column (Amersham) and the column washed with several volumes of 20 mM Tris/HCl pH 8.5. Bound protein was eluted using a 0-2 M NaCl gradient. Fractions containing the aldolase were pooled, concentrated and run on a HiLoad Superdex 200 16/60 prep-grade gel filtration column (Amersham) in 20 mM Tris/HCl pH 8.5, 200 mM NaCl. Samples ware taken during purification for SDS PAGE analysis. (**A**) Anion exchange. Lane **1**: molecular marker. Lane **2**: post-heat treatment lysate. Lanes **3-7**: aldolase elution. (**B**) Gel filtration. Lanes **1-6**: eluted aldolase. Lane **7**: pooled aldolase fractions from anion exchange. Lane **8**: molecular marker.

selected BL21(DE3) clone in LB medium, inside a shaking incubator at 37 °C. SsKDGA expression was induced with 1 mM IPTG and cells harvested after 21 h (Figure 2.1.5). Cell pellets stored at -20 °C, were resuspended on ice and treated with lysozyme, followed by DNaseI. Complete lysis was achieved by sonication and the cell debris removed by centrifugation (Figure 2.1.6).

An initial bulk purification step was carried out, involving incubation in a water-bath at 78-80 °C, followed by centrifugation to remove the precipitated protein. The supernatant (Figure 2.1.6**A**, lane 2) was further purified by anion exchange chromatography. Although a high degree of homogeneity was achieved in this step low molecular weight contaminants were observed (Figure 2.1.6**A**, lanes 3-7; **B**, lane 7). Size exclusion chromatography was therefore employed to achieve single band purity (Figure 2.1.6**B**, lanes 1-6).

Protein identity and integrity during purification was monitored by SDS-PAGE, MALDI-TOF mass spectrometry (MS) of protein tryptic digests and whole protein ESI MS. Also, the gel filtration column was calibrated according to the manufacturer's instructions, permitting confirmation of the enzyme's tetrameric assembly in solution (results not shown). Furthermore, the purified enzyme was shown to be active by the discontinuous TBA assay (results not shown), as described previously (Buchanan *et al.*, 1999; Lamble, 2004). Briefly, the method involved incubation at 70 °C of enzyme, pyruvate and glyceraldehyde, followed by enzyme precipitation using trichloroacetic acid and periodate cleavage of the enzyme catalysed condensation product [KDG(al)] to give formylpyruvate. The latter was then reacted with 2-thiobarbituric acid (TBA) to give a chromophore that absorbs light strongly at 549 nm.

Crystallisation:

SsKDGA was concentrated to 12 mg/ml and salt removed by buffer exchange with 20 mM Tris/HCl pH 8.5, using a centrifugal filter device with a molecular weight cut-off of 10,000 (Millipore). Crystallisation experiments were carried out at 21 °C in 4 μl drops using the sitting drop vapour diffusion method. Crystals of native SsKDGA appeared in both of the conditions mentioned above (0.1 M citrate pH 4.0, 1.6 ammonium sulfate; 0.1 M HEPES, pH 6.0, 8% propan-2-ol, 13% PEG 4,000), but were small and diffracted poorly (Figure 2.1.7**a**, **c**). Screening was carried out around the published pH and precipitant concentrations, while the protein concentration (15, 10, 7.5, 5, 3.5 mg/ml) was also varied. pH 6 crystals grew under many of these conditions, but an overall improvement in size and diffraction quality (typically ~2.1 Å in-house) was observed with protein concentrations between 10 - 7.5 mg/ml and in 13% PEG 4,000 (Figure 2.1.7**b**). The best pH 4 crystals, diffracting to about 2.5 Å in-house, grew with 7.5 mg/ml SsKDGA in 0.1 M citrate pH 4.3, 1.6 M ammonium sulfate (Figure 2.1.7**d**). Crystallisation at 4 °C gave only poorly diffracting needle like crystals in the pH 6 conditions (≥ 3.2 Å) and was not considered further.

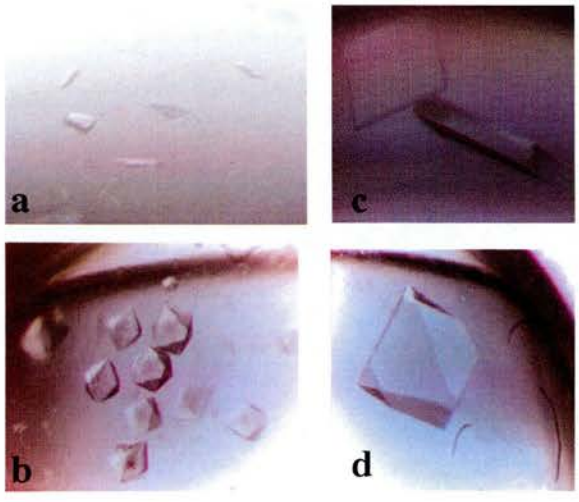


Figure 2.1.7 – Optimising KDGA crystals. Recombinant aldolase crystals were grown at 21 °C in 4 μl drops, using the sitting drop vapour diffusion method. The previously described conditions gave rise to small (long edge, 0.1-0.2 mm), poorly diffracting crystals that grew within 1-2 days: **a.** 12 mg/ml protein with 0.1 M HEPES pH 6.0, 13% PEG 4k, 8% propan-2-ol. **b.** 12 mg/ml protein with 0.1 M citrate pH 4.0, 1.6 M ammonium sulfate. Screening around these precipitant, concentration and pH values gave rise to larger crystals (longest edge 0.5-0.8 mm) of improved diffraction quality (2.5 – 2 Å in-house) that grew within 1-2 weeks. Examples of optimised crystals: **c.** 7.5 mg/ml protein with 0.1 M HEPES pH 6.3, 13% PEG 4,000, 6% propan-2-ol. **d.** 7.5 mg/ml protein with 0.1 M citrate pH 4.3, 1.6 M ammonium sulfate.

2.1.3 The native structure of SsKDGA at pH 6.0

The structure of SsKDGA has been previously solved by MAD phasing (Walden, 2001). The dataset was collected at the ESRF from a single SeMet crystal ($P2_12_12_1$: a=83.5 Å, b=131.1 Å, c=132.4 Å) that diffracted to 2.5 Å after being equilibrated against mother liquor containing 25% glycerol and flash-frozen in a stream of nitrogen at 100 K. A homotetramer ABCD was observed in the asymmetric unit, arranged as a dimer of the dimers AD and BC. The model was refined in *CNS* (Brunger *et al.*, 1998) to an R-factor of 19.7% and a Free R-factor of 24.1% with good stereochemistry, 89% of residues occupying core regions on the Ramachandran plot (Walden, 2001). Despite the excellent quality of the model, it was desirable that the resolution of the apo-structure be increased before embarking on the analysis of structures of the enzyme in complex with substrates.

Data collection and processing:

A pH 6 crystal of the native enzyme was cryoprotected by stepwise equilibration against mother liquor containing 10%, followed by 20% (v/v) glycerol and then flash frozen in a nitrogen stream at 100 K. The crystal diffracted to 2 Å on an in-house generator and an R-Axis IV++ image plate detector (MSC/Rigaku) was used to collect 180 images, with an oscillation angle of 0.5° per image and 10 min exposures. An additional dataset was subsequently obtained from a crystal that diffracted to 1.7 Å on ID14-EH2, at the ESRF in Grenoble. In this case an oscillation angle of 0.5° per image was used to collect 120° of data on a Q4 CCD detector (ADSC). The 1.7 Å data were obtained in a high-resolution pass using 5 s exposures, while a subsequent 2

Å resolution pass was made with 1 s exposures, in order to re-record previously overloaded spots.

	Native In-house	Native ID14-2 ESRF	SeMet (Peak) ID14-4 ESRF
Wavelength (Å)	1.54178	0.934	0.979
Resolution limits (Å)*	32 - 2.0	42 - 1.7	30 - 2.5
	(2.11 – 2.00)	(1.79 – 1.70)	(2.59 – 2.50)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 83.58	a = 83.63	a = 83.50
	b = 131.55	b = 130.41	b = 131.13
	c = 132.56	c = 132.46	c = 132.44
No. observations	287,808	1,103,606	268,882
No. unique reflections	98,295	158,423	52,011
Mosaicity	0.38	0.34	0.45
Completeness (%)*	99.1(94.5)	99.5 (98.9)	97.7 (89.1)
R_{merge} # (%)*	4.9 (15.6)	6.3 (16.8)	6.8 (22.9)
$< I/\sigma(I) >$*	14.5 (5.7)	22.2 (7.1)	12.0 (4.6)
Multiplicity*	3.0 (2.2)	7.0 (4.9)	~3.5
Wilson B (Å^2)	18.6	17.0	-

* Values in parentheses refer to the highest resolution shell.
# $R_{\text{merge}} = \Sigma_{hkl} \Sigma_i |I_{hkl,\,i} - \langle I_{hkl} \rangle| \,/\, \Sigma_{hkl} \langle I_{hkl} \rangle$

Table 2.1.1 – Data collection and processing statistics for native SsKDGA. Summary of statistics for the in-house and synchrotron datasets of native SsKDGA. Values for one of the three MAD datasets collected from a SeMet derivative crystal and used by Walden (2001) to solve the structure of SsKDGA are presented in the last column for comparison. In the case of the native synchrotron dataset, integrated intensities for the low- and high-resolution passes were merged using *SORTMTZ* (CCP4, 1994) before processing in *SCALA* to give the values shown.

In order to achieve maximum completeness and minimum overlap of spots, suitable parameters had to be selected prior to each data collection, based on an initial calculation of unit cell dimensions, mosaic spread and crystal symmetry. The determination and refinement of cell parameters, generation of reflection lists and

integration of spot intensities were carried out in *MOSFLM* (Leslie, 1992), while *SCALA* (Evans, 1993) was used for the subsequent scaling and merging of multiple observations, and *TRUNCATE* (French and Wilson, 1978; CCP4, 1994) for the generation of structure factor amplitudes.

Although processing was initially carried out in P222, both datasets were reprocessed in $P2_12_12_1$ once the space group was confirmed by examination of systematic absences in the axial reflections (*h00, 0k0* and *00l* reflections only when $h,k,l = 2n$). The statistics for the two datasets can be seen in Table 2.1.1, alongside those of the 'peak' data from the SeMet crystal used to solve the aldolase structure (Walden, 2001).

The data processing results confirmed that the native enzyme crystallises in the same space group and with near identical cell dimensions as its SeMet derivative. Therefore, the native crystals could also be expected to accommodate one tetramer per asymmetric unit, giving an estimated solvent fraction of ~0.55 (Matthews coefficient = 2.7 $Å^3$/Da; Matthews, 1968). Indeed, the high degree of isomorphism between crystals permitted direct refinement of the MAD derived model against the 2 Å native data. In the case of the 1.7 Å dataset, however, the crystal displayed a significant enough deviation along its b axis (0.9% relative to the in-house crystal) to allow changes in the crystal packing that would hinder refinement. For this reason, molecular replacement (MR) was initially carried out, in order to identify the position/orientation of protein in the asymmetric unit.

Molecular replacement:

The SsKDGA structure, its resolution extended to 2 Å, was used to generate both a monomeric and tetrameric search model with all solvent, ligands and alternate

conformations removed. Molecular replacement was carried out using the CCP4i implementation of *AMoRe* (version 5.0; Navaza, 1994; CCP4, 1994), with separate routines used for the calculation of cross-rotation and translation functions. As expected, a clear solution was found when using the tetrameric starting model, data in

α	β	γ	X	Y	Z	CC_F	Rf	CC_I
93.55	47.72	36.53	0.0893	0.4761	0.1490	76.5	29.2	75.2
93.70	*47.70*	*36.51*	*0.0888*	*0.4755*	*0.1495*	*81.0*	*25.8*	*81.9*

Table 2.1.2 – Molecular replacement using a tetramer search model. Native synchrotron data were used within the resolution range of 10-4 Å and the cross-rotation function search sphere for the Patterson correlation was set to 30 Å. The *AMoRe* rotation and translation searches were carried out sequentially and gave a clear solution, shown in the second row. Rigid body refinement was then applied to the solution to give the values shown in the bottom row, in italics. Abbreviations: **α β γ**, rotation solution in Eulerian angles; **X Y Z**, fractional unit cell coordinates for translation solution; **CC_F**, correlation coefficient between observed and calculated structure factor amplitudes; **Rf**, R-factor between observed and calculated amplitudes; **CC_I**, correlation coef. for intensities.

	α	β	γ	X	Y	Z	CC_F	Rf	CC_I
1	120.70	35.14	357.80	0.3853	0.1246	0.4664	20.7	53.0	21.0
2	174.78	83.19	69.88	0.2990	0.4688	0.1024	36.6	48.0	37.2
3	71.00	56.60	218.72	-0.0585	-0.3702	-0.1948	52.4	42.2	51.9
4	13.34	82.35	331.60	-0.1174	0.4400	0.1588	69.4	33.7	66.7
	120.11 *174.70* *71.59* *12.17*	*35.34* *82.43* *56.65* *81.91*	*357.73* *69.82* *218.94* *333.05*	*0.3860* *0.2994* *-0.0597* *-0.1178*	*0.1241* *0.4687* *-0.3697* *0.4398*	*0.4670* *0.1017* *-0.1952* *0.1585*	*79.4*	*27.9*	*80.4*

Table 2.1.3 – Molecular replacement with monomer search model. Rotation function solutions were obtained for the SsKDGA monomer, using data in the range of 15-4 Å and a search sphere of 25 Å. Four sequential iterations of the translation function were then carried out, giving the solutions shown above (**1-4**). Applying rigid body refinement to all four solutions, gave the values shown in the last row. Abbreviations used as above (Table 2.1.2).

the range of 10–4 Å (Table 2.1.2) and a rotation search sphere of 30 Å. In the case of the monomeric model as well, four rotation/translation solutions were identified sequentially (Table 2.1.3), although only when the low-resolution limit was extended to 15 Å. MR output models were generated by applying the two sets of rotations/translation solutions to their respective search models, using the CCP4i implementation of *PDBSET* (version 5.0; CCP4, 1994).

The final refined solution from the four monomer searches gave marginally worse statistics that that of the tetramer search. This difference probably reflects errors in the positioning/orienting of individual monomers relative to each other, due to the resolution limits of the data used. Searching with a tetramer model is in effect even more 'crude', as it is assumed that the four monomers behave as a single rigid body. Such an assumption could be justified in this case, however, as there are no differences in the molecule or crystallisation conditions that would be expected to disrupt the intersubunit interactions.

α	β	γ	X	Y	Z	a	b	c
0.000	0.000	-0.033	-0.0004	0.0058	0.5011	-0.0323	0.7687	66.4289

Table 2.1.4 – Superposition of MR input and output models. The *AMoRe* output model from the tetramer search was superimposed onto the 2 Å native SsKDGA model using *LSQKAB* (Kabsch, 1976), by fitting Cα's for each of the 293 residues of chains A-D. The 0.5 shift along **Z** is due to the initial repositioning of the search model by *AMoRe* at a point of origin. Abbreviations: **α β γ**, rotation around the cell axes in Eulerian angles; **X Y Z**, translation vector in fractional unit cell coordinates; **a b c**, translation vector in Ångströms along the cell axes.

The MR results indicate that crystal packing is the new unit cell is not significantly altered. Superposition of the input and output models from the tetramer search reveals a small shift along the b axis of nearly 70% the difference in that cell length

between the two crystals (1.1 Å), but no change in orientation of the molecule (Table 2.1.4). In fact, the small change in the molecule's position was also identified by direct rigid body refinement (RBr) of the 2 Å native SsKDGA model against the synchrotron data, without molecular replacing first. After 20 cycles of RBr in *REFMAC5* (version 5.2.0005; Murshudov *et al.*, 1997) with a resolution range of 40-4 Å, the model had shifted by 0.83 Å along the b axis, accompanied by a drop in the R-factor (Rf) of 28.5% to 24.5% and of 29% to 26.5% in the Free R-factor (R-free).

The results clearly show that molecular replacement was not necessary in this situation. Nevertheless, the MR solution was selected as the starting model for refinement against the native synchrotron data. In general, the speed with which the process can be carried out in *AMoRe* makes its execution in situations where there is uncertainty over the crystal isomorphism a worthwhile step prior to refinement and one that can ultimately save time and effort.

Model refinement and validation:

With both native datasets refinement of the starting models was carried out using the rigid body (RBr) and restrained refinement (Rr) routines in *REFMAC5*, with simultaneous maximum likelihood minimisation of geometric parameters and B-factors. In the case of the in-house data, the MAD derived model was stripped of solvent molecules and subjected to three rounds of RBr, during which the resolution limit was extended from 5 Å to 3.5 Å. Several rounds of Rr were then carried using the full resolution range (32-2 Å). Each round of restrained refinement was followed by inspection of the generated Fourier difference maps ($2F_o$-F_c and F_o-F_c) and manual

adjustments of the model coordinates in *O* (Jones *et al.*, 1991), in order to improve the density fit.

Having improved the protein model, the solvent building routine in *ARP/wARP* (Lamzin and Wilson, 1993) was used to identify density peaks corresponding to water molecules and add their coordinates to the structure. The new atoms were inspected in *O* against the unbiased maps and necessary alterations made before refining the updated model. With most of the electron density accounted for, more ambiguous regions were re-examined and final adjustments made. Several peaks were identified corresponding to ligand molecules (see Chapter 2.1.4), which had to be manually incorporated into the model for refinement against the data. Coordinates and geometric restraints for these were extracted from the *REFMAC* monomer library using *LIBCHECK* (CCP4, 1994).

For the purposes of cross-validation, a random subset of approximately 10% of all reflections (the *test* set) had to be left out of the refinement process (Kleywegt and Brunger, 1996). These reflections are flagged and then used to compute the Free R-factor. To avoid model bias, it was necessary to retain the same *test* set that had been used during refinement of the MAD derived model. So, the R-free flag was extracted from the MAD dataset using *CAD* (CCP4, 1994) and applied to the in-house native data, while additional non-equivalent reflections were also flagged in order to ensure a 10% *test* set.

The MR output model was refined against the 1.7 Å synchrotron data in a similar way, its *test* set generated using the flagged reflections from the in-house dataset. Figure 2.1.8 demonstrates the overall quality of the electron density in the two structures at the end of refinement. Validation of the two higher resolution structures was carried out in *PROCHECK* (Laskowski *et al.*, 1993) and *SFCHECK* (Vaguine *et*

al., 1999) by assessing the stereochemical quality of the models and their agreement with the experimental data, respectively. In the case of the 1.7 Å model, additional checks were carried out using the *MOLPROBITY* server (Lovell *et al.*, 2003; http://molprobity.biochem.duke.edu) and the validation tasks in *COOT* (version 0.0.33; Emsley and Cowtan, 2004).

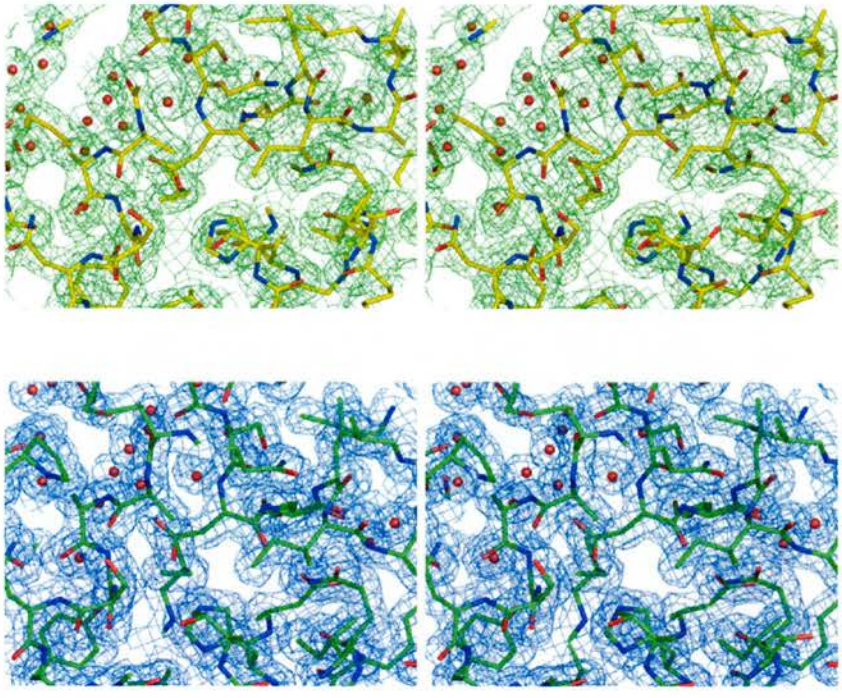


Figure 2.1.8 – Native SsKDGA. Stereoview showing a stick representation of the region around Glu-181 in the final 2 Å (**A**) and 1.7 Å (**B**) models. Superimposed, the $2F_o$-F_c maps for the two structures as a wire-frame contoured at 1 σ. Colour key: yellow or green sticks, carbon in the 2 Å and 1.7 Å models, respectively; blue sticks, nitrogen; red sticks or spheres, oxygen.

The final model and refinement statistics are presented in Table 2.1.5. They demonstrate only a small improvement in model quality at higher resolution, due largely to an increase in overall coordinate precision rather than accuracy. In fact, several stereochemical parameters (particularly the number of main chain bond length

	Native In-house	Native ID14-2 ESRF	MAD/SeMet ID14-4 ESRF
Resolution (Å)	2.0	1.7	2.5
Working set reflections	88,470	142,442	44,039
Test set reflections	9,825	15,754	4,952
Data completeness (%)	99.1	99.4	96.1
R-factor† (%)	15.3	15.0	19.7
R-free† (%)	20.2	18.7	24.1
•Number of atoms*:			
- protein	9,570 (19.1)	9,485 (15.5)	9,300 (29.7)
- water	901 (29.1)	1,557 (34.1)	324 (32.5)
- ligand	49 (33.4)	102 (30.3)	-
rmsd bonds (Å)	0.023	0.021	0.005
rmsd angles (°)	1.73	1.72	1.41
•*PROCHECK* results (%):			
- Ramachandran plot regions			
core	91.4	91.5	89.3
additionally allowed	8.2	8.1	9.9
generously allowed	0.0	0.0	0.3
disallowed	0.4	0.4	0.5
- M/c bond lengths within limits	99.4	98.9	100
- M/c bond angles within limits	98.0	98.4	98.3
- Planar groups within limits	99.8	99.5	99.3
•*SFCHECK* results:			
- R_stand(F)#	0.044	0.031	0.062
- DPI (Å)	0.161	0.085	0.542
- CC_F	0.941	0.962	0.895
- radial error in coordinates (Å) ‡	0.187	0.144	0.292
optical resolution (Å) §	1.5	1.4	1.8

* Values in parentheses - average B factor ($Å^2$) for all atoms

† $Rf = (\Sigma \mid |F_o| - |F_c| \mid)/(\Sigma |F_o|)$

The R-free is calculated in the same way, but using the *test* set reflections

$R\_stand(F) = \langle \sigma (F) \rangle / \langle F \rangle$

‡ The atomic error is computed from the Luzzati plot (R-factor vs. resolution)

§ Expected minimum distance between two resolved peaks in the electron density.

Table 2.1.5 – Refinement and model statistics for native SsKDGA structures. The table presents a summary of parameters indicating the quality of the final SsKDGA models after refinement against the native data. The respective values for the MAD derived structure are shown in the far right column for comparison. Average B-factors were calculated using *BAVERAGE* (CCP4, 1994). Selected results from the *PROCHECK* and *SFCHECK* analyses are presented. Abbreviations: **rmsd**, root mean square deviation; **M/c**, main chain; **R-stand(F)**, uncertainty in the average structure-factor amplitudes; **DPI**, diffraction component precision indicator - an estimate of the error in atomic coordinates; $\mathbf{CC_f}$, correlation coefficient between observed and calculated structure-factors.

outliers) appear to worsen at higher resolution, or do not vary consistently between the models. Such irregularities can be accounted for in terms of the tighter geometric restraints imposed in *CNS* compared to *REFMAC*, as well as the increased weighting of the experimental data relative to the geometric parameters when refining at higher resolution.

2.1.4 Key features of the apoenzyme models

The structure of the SeMet derivative of SsKDGA has been previously described as part of a study into the structural basis of protein thermostability (Walden, 2001). Each monomer adopts a $(\beta/\alpha)_8$ barrel fold decorated with three additional helices at the C-terminus. The enzyme's tetrameric assembly has a volume of 162,000 $Å^3$ and consists of two distinct subunit interfaces; AD and AC (equivalent to BC and BD, respectively). The AD interface has a greater hydrophobic character and is more extensive, covering a surface area of 1,660 $Å^2$, compared with the 1,060 $Å^2$ of accessible surface at the AC interface. Similar observations have also been made for NAL and DHDPS, the two other structurally characterised members of the subfamily (Izard *et al.*, 1994; Mirwaldt *et al.*, 1995; Blickling *et al.*, 1997a), and the subunit arrangement has been described as that of a dimer of dimers.

The MAD derived model consists of the four protein monomers and 324 water molecules. Every residue of the protein was observed in the density from Pro-2 to Glu-294 (the N-terminal methionine having been cleaved on production of the mature protein). An intrasubunit disulfide (Cys-150/Cys-120) and two salt bridges (Glu-54/Arg-261; Arg-283/Glu-294) were seen in each of the monomers, while two

intersubunit ion pair networks were also highlighted as important for the enzyme's thermostability:

AC or BD interface	AD or BC interface
Glu-234$_A$/Arg-169$_C$/Asp-230$_A$	Glu-271$_D$/Lys110$_A$/Asp-268$_D$
Glu-234$_C$/Arg-169$_A$/Asp-230$_C$	Glu-271$_A$/Lys110$_D$/Asp-268$_A$

Overall structure:

The SeMet and native models were superimposed using the CCP4 implementation of *LSQKAB* (Kabsch, 1976; CCP4, 1994). The small rmsd values calculated (Table 2.1.6) indicate that the overall structure of the two forms of the enzyme is the same and the difference in Cα rmsd's between the SeMet and native models appears to be due to deviations at only a small number of residue positions. The regions around Gly-74 and Pro-262 (in loops β3-α3 and α10-α11, respectively) in particular had to

	1.7 Å Native	2 Å Native	2.5 Å SeMet
2.5 Å SeMet	0.25 (0.68)	0.23 (0.67)	*0.25 (0.74)*
2 Å Native	0.14 (0.49)	*0.21 (0.65)*	-
1.7 Å Native	*0.20 (0.62)*	-	-

Table 2.1.6 – Superposition of the SsKDGA models. Root mean square deviation (rmsd) of Cα's and all atoms (in brackets) of SsKDGA in the three apoenzyme models (units in Å). Superpositions were carried out between residues 2-293 in equivalent subunits using *LSQKAB*. Glu-294 was excluded as it displays various conformations and greater disorder in the native models. Consistent with the remodelling carried out, the rmsd's between the 2 Å and 1.7 Å models are smaller in magnitude than their deviations from the SeMet structure. Also shown are the rmsd's between separate subunits of the same model (in italics), generated by superimposing chain A onto B and C onto D. These values suggest that there are 'real' differences between equivalent subunits in the tetramer, possibly arising from their distinct interactions with symmetry related molecules.

be remodelled during refinement of the SeMet structure against the 2 Å native data and display very different main chain torsional angles in the final 2 Å model. These and other changes made to the initial structure were conserved during subsequent refinement against the 1.7 Å data and, with the exception of the C-terminal glutamate, no further adjustments had to be made to main chain positions in producing the higher resolution native model.

Remodelling of the residues Pro-262 and Pro-263 resulted in main chain torsion angles in all monomers consistent with a *cis*-peptide conformation between the two residues (2 Å model, $\langle\omega\rangle$ = 10°; 1.7 Å model, $\langle\omega\rangle$ = 12°), in place of the original *trans*-peptide (2.5 Å model, $\langle|\omega|\rangle$ = 179°). In their new conformation the residues were found to occupy more favourable regions of the Ramachandran plot, as determined by *PROCHECK* (data not shown). Moreover, the B-factors for the two

	2.5 Å SeMet-A	2 Å Native-A	1.7 Å Native-A	2.5 Å SeMet-D	2 Å Native-D	1.7 Å Native-D
Pro-260	33.1/1.1	19.4/1	15.2/1	34/1.15	15.4/0.8	14.2/0.9
Arg-261	31.5/1.1	16.4/0.9	13.2/0.85	40/1.35	15.6/0.8	12.6/0.8
Pro-262	40.3/1.4	17/0.9	16.0/1	45.9/1.55	16.2/0.85	16.0/1
Pro-263	36.9/1.2	17/0.9	15.3/1	37.4/1.3	15.2/0.8	14.5/0.9
Ile-264	33.6/1.1	14.5/0.8	14.7/0.95	25.5/0.9	14.2/0.7	12.9/0.8
Phe-265	25.8/0.9	16.4/0.9	13.9/0.9	23.5/0.8	14.9/0.8	13.5/0.9

Table 2.1.7 – Remodelling loop α10-α11. Comparison of B-factors for P260-F265 in chains A and D of the three KDGA models. The values on the left and right in each cell are average and relative B-factors for all residue atoms, respectively. Average B-factors were obtained using *BAVERAGE* (CCP4, 1994), however, due to differences in resolution and the restraints used, meaningful comparison between structures is difficult. Dividing by the average B-factor for all protein atoms in the relevant model gives a relative value, which should be more informative (see Table 2.1.5). The values for Pro-262 and Pro-263 are highlighted in red. The two residues appear somewhat more disordered in the MAD versus the native structures; a possible indication of incorrect fitting to the density.

residues showed a modest improvement (Table 2.1.7); lower values were observed, as well as better agreement with flanking residues. In addition to the consistent observations in subsequent KDGA structures, these results suggest that the lower resolution maps permitted incorrect model building of the main chain at this position and that *cis*- is the true configuration.

Superposition of the three models revealed large deviations in side chain positions, although the most significant occur in surface residues located outside the subunit interfaces. On the other hand, the conformations of residues involved in intersubunit and symmetry related interactions (determined using *CONTACT*; CCP4, 1994) are highly conserved, with only a few exceptions. Deviations within the protein's interior are mainly due to side chain 'flips' of a few valines, leucines, asparagines and histidines.

Some of the similarities and differences between the three structures can be illustrated by comparison of the ion pairs mentioned previously. So, while the $E234_x/R169_y/D230_x$ and $E54_x/R261_x$ salt bridges (where x and y are different subunits) were clearly seen in all monomers of the three models, the presence of $E271_x/K110_y/D268_x$ and $R283_x/E294_x$ was more variable. The former was observed at three interfaces in the SeMet structure and at only two in either of the native models, despite clear density for the participating residues. Closer examination revealed that the side chain of Lys-110 adopts a number of well ordered conformations in the structures, only some of which bring its Nζ atom within 4 Å of the other residues' carboxylates.

Although present in all monomers of the original structure, the $R283_x/E294_x$ salt bridge was observed in only one subunit of each of the native models. In these structures the participating residues adopt a number of conformations, which in most

cases do not permit interaction of the two ionic groups. Consistent with the loss of this anchoring interaction Glu-294 is significantly more disordered in the native models, as is Lys-293 to a lesser extent. Arg-283, on the other hand, remains well ordered. The instability of this ionic interaction, like that of the $E271_x/K110_y/D268_x$ network, is not surprising given the high solvent exposure of the residues involved. Table 2.1.8 lists relative B-factor values for the salt bridge residues and those adjacent to them.

Chains A & B	2.5 Å SeMet-A	2 Å Native-A	1.7 Å Native-A	2.5 Å SeMet-B	2 Å Native-B	1.7 Å Native-B
Ile-282	1.4/1.4	1.3/1.4	1.4/1.4	0.9/0.9	0.9/0.9	0.9/1
Arg-283	1.4/1.9	1.5/2.2	1.5/2.2	1/1.4	0.95/1.1	0.9/1.6
Leu-292	1.7/1.5	1.9/1.7	2/1.7	1.3/1.2	1.2/1	1.2/1
Lys-293	2/2.15	2.5/2.7	2.7/3.2	1.6/1.8	1.8/2.15	1.8/2.3
Glu-294	2.25/2.4	2.9/3.2	3.2/3.4	1.9/2	2.5/2.8	2.4/3.1

Chains C & D	2.5 Å SeMet-C	2 Å Native-C	1.7 Å Native-C	2.5 Å SeMet-D	2 Å Native-D	1.7 Å Native-D
Ile-282	1.35/1.4	1.3/1.2	1.4/1.4	1.3/1.3	1.2/1.2	1.2/1.3
Arg-283	1.4/1.5	1.3/1.35	1.4/2.4	1.4/1.8	1.2/1.5	1.3/2.2
Leu-292	1.6/1.5	1.8/1.7	1.8/1.7	1.7/1.6	1.9/1.8	2.1/2
Lys-293	2/2.1	2.4/2.7	2.6/2.9	2.05/2.2	2.6/2.9	2.8/3
Glu-294	2.3/2.4	2.8/3	3.1/3.2	2.3/2.4	3.05/3.3	3.3/3.6

Table 2.1.8 – Relative B-factor analysis for the SeMet and native KDGA models. Comparison of disorder between the three aldolase models for residues I282, R283 and L292-E294. Average B-factors (in $Å^2$) for these residues were calculated from each model using *BAVERAGE* and divided by the average value for all protein atoms (see Table 2.1.5) to standardise for comparison. The numbers on the left and right in each cell are the relative main chain and side chain B-factors, respectively. The top and bottom tables contain values for chains A-B and C-D, respectively. Deviations in relative B-factor with respect to the SeMet structure of 0.4 or more are highlighted in red.

Occupancy refinement:

A number of side chains were identified in the native structures that had more than one ordered position. Although observation of multiple conformers is generally associated with atomic resolution (1.2 Å, or better; Sheldrick, 1990), in the case of several residues the data strongly supported interpretation of alternative conformations in the electron density. The two most consistently observed cases (also seen in subsequent structures) involve Ser-189 and Met-215.

CHAIN A	2.5 Å SeMet-A	2 Å Native-A	1.7 Å Native-A
Ala-188	28.8/26	18.9/17	16.4/14.5
Ser-189	32.7/32.5	20.8/24 (17.9/17.4)	15.4/21.8 (14.7/15.7)
Thr-190	32.7/34	17.2/15.3	14.4/13.2
Ala-214	34.8/34.5	25.2/24.5	19.1/17.6
Met-215	36.8/36.4	26.5/34.4 (26.2/27.2)	22.6/28.7 (20.8/21.2)
Glu-216	43.9/49.7	29.4/33.5	23.9/31.4

CHAIN D	2.5 Å SeMet-D	2 Å Native-D	1.7 Å Native-D
Ala-188	26.4/25.9	16.4/15.3	13.2/12.8
Ser-189	25.1/30.2	15.1/19.2 (14.3/13.5)	11.7/15.5 (10.4/10.7)
Thr-190	23.6/23.2	14.5/13.2	10.9/10.9
Ala-214	40.1/38.4	21.8/21.7	16.1/15.5
Met-215	41.7/44.3	25.3/34.5 (23.3/24.8)	22.8/30 (20.2/22)
Glu-216	42.2/49.6	25.3/34.9	22.1/28.8

Table 2.1.9 – Alternative conformers. B-factor analysis for residues A188-T190 and A214-E216 in chains A and D of the three KDGA models. The values on the left and right in each cell represent averaged main chain and side chain B-factors (Å^2), respectively (calculated using *BAVERAGE*). For Ser-189 and Met-215, which display two conformations in the native structures, B-factors outside brackets are for the original conformation at full occupancy. Inside the brackets are the averaged values for the two conformations at half occupancy.

These residues appeared well ordered during initial refinement of the MAD derived model against the 2 Å native data, however, both the $2F_o$-F_c and F_o-F_c difference

maps at the equivalent coordinates in each monomer included additional unexplained density not consistent with unaccounted for solvent molecules. Both residues were also characterised by high side chain B-factor values. The residues were refined with two sets of coordinates at different fractional occupancies and best agreement with the electron density maps was observed with each conformation set to half occupancy. Moreover, more suitable B-factors were obtained after refinement (Table 2.1.8), with values for atoms of alternative conformers being in most cases very similar. The results were consistent across the four monomers and similar observations made during subsequent refinement at 1.7 Å were dealt with in the same way.

The two conformations (*A* and *B*) of Ser-189 and Met-215 are conserved across monomers and models. *A* in each case corresponds to the original conformation observed in the MAD derived model and *B*, the additional position seen only in the native structures. Ser-189 is located on helix α-7 (S180-T190) at the AC (or BD) interface. In its original conformation, the side chain forms a hydrogen bond with its equivalent group on the adjacent monomer. The hydroxyl also forms an intrasubunit interaction with the main chain carbonyl of Thr-186. In its *B* conformation, on the other hand, the side chain is no longer involved in either of these interactions. Instead it is within hydrogen bonding distance of the Ala-185 carbonyl and an ordered water molecule.

Hydrogen bonding of a side chain hydroxyl to a main chain carbonyl in an $i \rightarrow (i\text{-}3)$ or $i \rightarrow (i\text{-}4)$ pattern (where i represents the residue number of the participating side chain) is commonly observed at the C-termini of helices (Bordo and Argos, 1994; Rossmann and Arnold, 2001). In this case, however, the Ala-185 interaction is probably weaker due to the relative non-planarity of the hydroxyl proton and carbonyl oxygen lone pairs. Moreover, due to the relative orientation of its hydrogen bonding partners in

conformation *A* the hydroxyl of Ser-189 cannot serve as proton donor to more than one acceptor (i.e. in a three-way hydrogen bond). Consequently, the intersubunit interaction made by the residues' side chains can only be satisfied if one of the two participating groups breaks its intrasubunit hydrogen bond with Thr-186.

The conditional nature of the hydrogen bonding in *A* and poorer angle of approach between interacting groups in *B* probably reduce the side chain's stability in either conformation, which could account for the observation of both orientations. An explanation, however, was less forthcoming in the case of Met-215. Located at the end of helix α8 (A198-E216), the residue is surface exposed and engages only in intrasubunit van der Waals and hydrophobic contacts. Although Met-215 is located close to Ser-189 (Cα215-Cα189 = 6.7 Å) the residues do not interact.

Disulfides:

SsKDGA contains two cysteines per monomer, one at the C-terminal of helix α4 (Cys-120) and the other in loop α5-β6 (Cys-150). The two residues are juxtaposed on the surface of the protein, although their side chains point down into the centre of the enzyme so that the thiol groups are exposed to a relatively hydrophobic cleft and are largely buried. In the native models the Sγ atoms of Cys-120 and Cys-150 have an average solvent accessible area of 0.2 and 1.4 $Å^2$, respectively, calculated using *AREAIMOL* (probe radius 1.4 Å; Lee and Richards, 1971; CCP4, 1994). The equivalent values for the MAD structure are 0.8 $Å^2$ (C120-Sγ) and 1.6 $Å^2$ (C150-Sγ) if subunit B in not included.

The side chain of Cys-150 in monomer B of the SeMet model has an unusually large solvent accessible area (Sγ of 8.5 $Å^2$) and its thiol group is rotated by ~88° around $\chi 1$

relative to the 'native' conformation. This difference in orientation results in a torsion angle outlier for the disulfide bond ($\chi 3$ = -91.4°) not observed in any other monomer across the three structures ($\chi 3$ between 95° and -85°), indicating that the side chain orientation in this case was probably modelled incorrectly.

The low solvent exposure of the cysteine thiol groups in KDGA may increase the stability of the disulfides. This is consistent with their presence in the MAD structure (Walden, 2001) despite the addition of 5 mM dithiothreitol (DTT) to all buffers during lysis and purification of the SeMet enzyme (used to prevent selenium oxidation). Moreover, the electron density observed around the two cysteines in both the native structures is also consistent with the presence of a covalent bond between the residue side chains. Negative density peaks at -3 σ or higher were also seen in the 1.7 Å F_o-F_c maps, however, suggesting that the disulfides in that structure might not be completely intact.

Refinement was carried out using different fractional occupancies and, with a value of 0.7 for the Sγ atoms, the negative density peaks were reduced to σ levels below -2. This was also accompanied by an improvement in the B-factors for these cysteine thiols with respect to the equivalent atoms in the 2 Å model (data not shown). The results indicate that partial reduction (~30%) of the disulfides may indeed have occurred in the higher resolution native structure, although an alternative conformation was not observed for the free thiol groups. The difference in apparent stability of the disulfide bridges between this and the MAD structure could not be accounted for.

The active site:

The location and composition of the SsKDGA active site have so far not been described. It has been possible, however, to make detailed predictions from structural

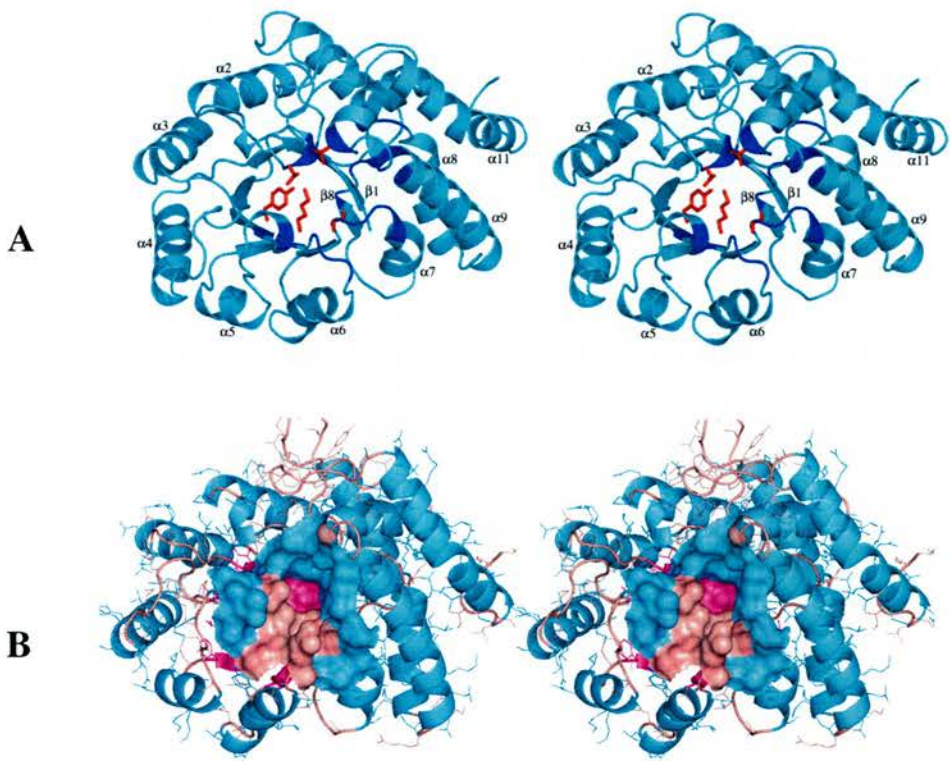


Figure 2.1.9 – The active site of SsKDGA. Stereoviews of the KDGA monomer as a cartoon representation. The monomer is oriented with the C-terminal side of the barrel facing out of the page. Contributions to the active site pocket are made by residues from strands β1 and β8; helices α5, α7 α8, α9 and α10; and loops β5-α5 and β6-α6. **A**. The relevant residue positions are mapped onto the secondary structure in blue. Previously predicted active site residues (Lys-155, Tyr-130, Thr-43, Thr-44 and Gly-179 - Asp-181 not shown) are highlighted as red sticks. **B**. Surface representation of the active site pocket. Colouring is with respect to secondary structure (helices, cyan; strands, magenta; loops, salmon). The identity of residues lining the pocket was confirmed using the *CASTp* server (Binkowski *et al.*, 2003), which uses a 1.4 Å probe to identify solvent accessible cavities and analyse their dimensions. The active site was calculated as having a molecular (or Connolly) surface area (Richards, 1977) of ~400 $Å^2$, equivalent to a solvent accessible (Richards) area of ~250 $Å^2$ (Lee and Richards, 1971). The molecular volume was estimated at ~615 $Å^3$ (See also Table 2.1.11).

and mechanistic information available on other aldolases, particularly NAL and DHDPS (see Sections 1.2.3 and 1.2.4). Six key active site residues have been proposed for KDGA on the basis of sequence alignments with NAL subfamily members (Barbosa *et al.*, 2000). Lys-155 has been identified as the Schiff base forming lysine, Tyr-130 as the proposed catalytic tyrosine and Thr-43/Thr-44 as the central residues of the GxxG motif. Gly-179 and Asp-181 have also been predicted to play a role in substrate binding.

Lys-155 is located on strand β6 of each monomer, with its side chain pointing into the centre of the TIM-barrel. Together with Tyr-130 (strand β5), Gly-179 (loop β7-α7), T-43 and T-44 (loop β2-α2), Lys-155 forms the base of a deep pocket on the C-terminal side of the barrel. Asp-181 (helix α7) also lines the active site, although the residue is located more than 9 Å from its base. Additional contributions from several secondary structural elements of the same monomer (Figure 2.1.9**A**) result in a pocket with an estimated volume of 600 $Å^3$, calculated using the *CASTp* server (http://sts.bioengr.uic.edu/castp/; Binkowski *et al.*, 2003; Figure 2.1.9**B**).

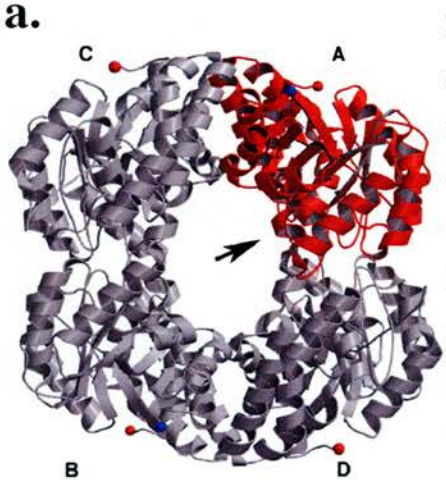
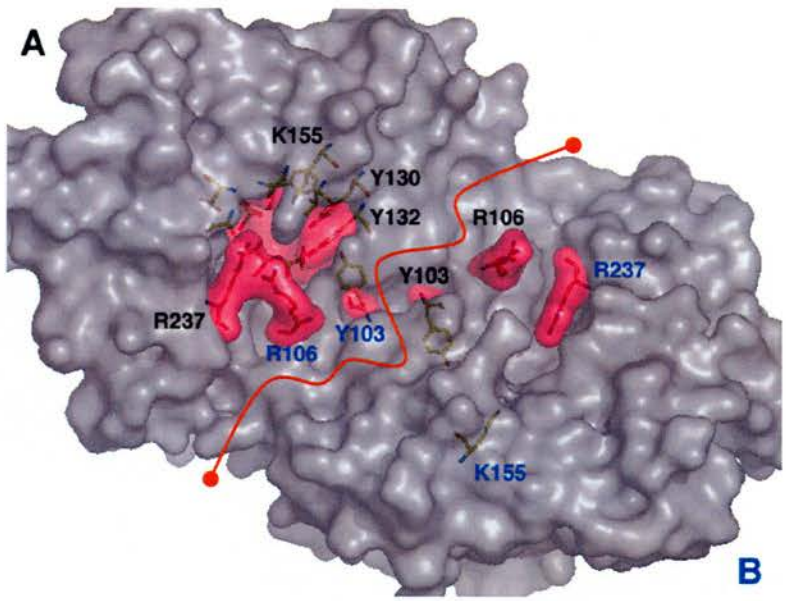


Figure 2.1.10 – Tetrameric assembly. The active site of each KDGA monomer is located within a large cavity accessible from the tetramer's central channel. **a.** Side-view of the enzyme. The four active site entrances are visible when standing at the centre of the tetramer looking out (black arrow). **b.** Semi-transparent surface representation of a close dimer (AD), as seen from the position of the arrow in **a**. The core of each active site is composed of monomer residues surrounding the catalytic lysine, K155. The adjacent 'close' subunit contributes additional residues at the periphery of the active site, creating a larger enclosed cavity. Residues of monomer A are labelled in black, while those of monomer B are in blue. An imaginary boundary between the subunits is drawn in red.

b.



Upon formation of the close dimer (AD or BC), access on one side of the active site (the side equivalent to helix α3) is blocked by residues of the adjacent subunit (loop β4-α4 and helix α4). This results in an even greater volume being enclosed, estimated at over 1,100 $Å^3$, with an internal molecular surface of ~780 $Å^2$. This cavity is accessible via a relatively small opening that faces the tetramer's central channel (Figure 2.1.10). The entrance is formed by residues Pro-133, Thr-134, Ile-158, Glu-159, Ser-180, Met-182, Arg-237, Gly-240 of a given monomer, as well as Pro-105 and Arg-106 from loop β4-α4 of the adjacent subunit.

Lys-155 is located at the base of the active site pocket. Its Nζ atom forms close contacts (≤ 4 Å) with the side chains of Val-196 and Pro-7, as well as a potential hydrogen bond with the hydroxyl of Tyr-130. The side chain amine also interacts strongly with two ordered water molecules, W2 and W3 (Figure 2.1.11) and more weakly with a third, W1. These form part of an extensive water structure within the active site cavity that is well conserved between monomers and models. Consequently, the catalytic lysine is able to make water mediated interactions with several other residues, including Thr-43, Thr-44, Tyr-130 and Thr-157.

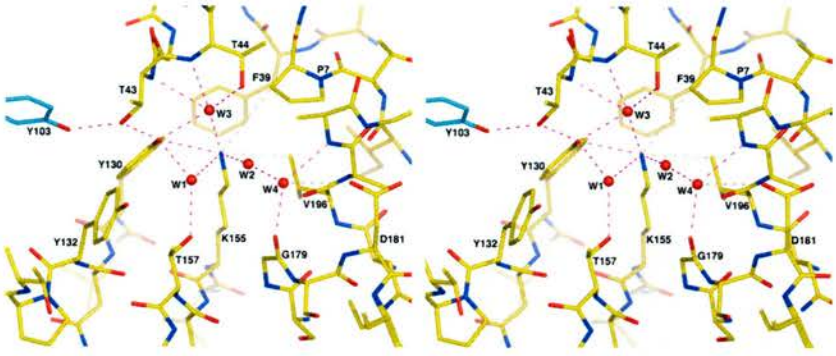


Figure 2.1.11 – Active site architecture. Stereoview of key residues and interactions in the SsKDGA active site. Residues are shown as sticks (**carbon**, yellow; **nitrogen**, blue; **oxygen**, red) and water molecules as spheres. Hydrogen bond interactions are shown as magenta dashes and close contacts (≤ 4 Å) as wheat coloured dashes. All potential hydrogen bonds of Lys-155 are shown. Tyr-103 from the close dimer partner is drawn with cyan carbons. [Average distances of interactions: **K155-Y130**, 3.1 Å; **K155-W1**, 3.2 Å; **K155-W2**, 2.8 Å; **K155-W3**, 2.8 Å; **W1-T157**, 2.9 Å; **W1-Y130**, 2.8 Å; **Y130-T43**, 2.6 Å; **T43-Y103**, 2.7 Å; **W2-W4**, 2.7 Å; **W3-Y130**, 2.8 Å; **W3-T43**, 3.1 Å; **W3-T44(N)**, 3 Å; **W3-T44(Oγ)**, 3 Å; **W4-G179**, 2.7 Å; **W4-A198**, 2.8 Å.]

Residues within the active site display highly conserved conformations across monomers of the two native models. The only exceptions are Tyr-130 and Phe-39, which display conformational flexibility, as indicated by variations in their $\chi 1$ $\chi 2$ torsion angles (Table 2.1.10**A**). Moreover, the two residues display elevated side chain B-factor values with respect to flanking residues (Table 2.1.10**B**), despite appearing well ordered in all monomers. The flexibility of Phe-39 and Tyr-130 is best illustrated by comparing with Tyr-132. The latter is also located in the active site and is more solvent accessible, yet it displays a highly conserved side chain conformation and low atomic B-factors.

The active site architecture displayed by KDGA is highly conserved in other members of the NAL subfamily (Figure 2.1.12). Superpositions reveal a similar shape to the

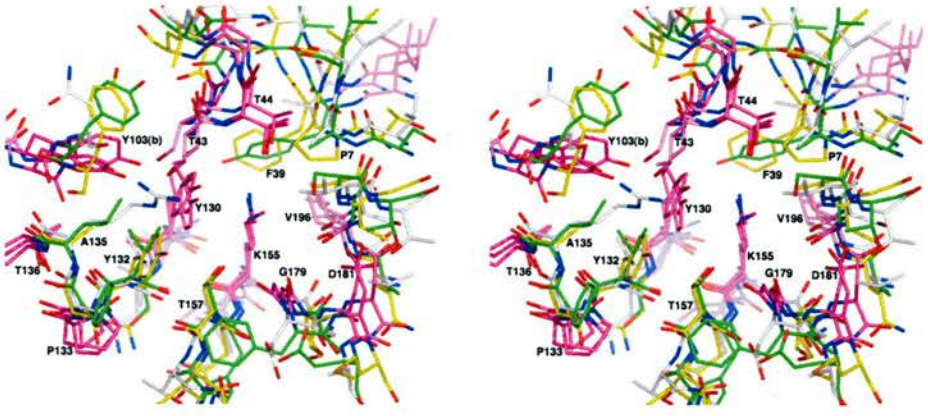
pocket, with key groups occupying equivalent positions. SsKDGA shares the highest

A	Phe-39 χ1		Phe-39 χ2		Tyr-130 χ1		Tyr-130 χ2	
	2 Å	1.7 Å	2 Å	1.7 Å	2 Å	1.7 Å	2 Å	1.7 Å
A	-156°	-168°	114°	106°	-175°	-177°	-33°	-26°
B	-155°	-172°	114°	109°	-173°	-177°	-42°	0°
C	-160°	-164°	110°	106°	-171°	-174°	-49°	-67°
D	-150°	-160°	117°	112°	-170°	-175°	-58°	-69°

B	Leu-38	Phe-39	Val-40	Leu-129	Tyr-130	Asn-131	Tyr-132
1.7 Å	10.9/10.5	9.5/18.3	9.1/11.7	9/9.9	8.8/15.2	9/9.8	9.1/10.6
2.0 Å	14.2/14.3	13.1/18.8	12.1/12.6	13.3/11.6	13.2/19.5	13.6/12.5	12.9/13.7

Table 2.1.10 – Active site side chain deviations. ***Table A.*** Values for the χ1 and χ2 torsion angles of Phe-39 and Ty-130 in the native SsKDGA structures. These residues display the only significant active site conformational differences across the monomers of the two models. The largest difference is observed between the Tyr-130 χ2 angles in the 1.7 Å model (B-D, 68°). Additionally, the χ1 torsion of Phe-39 deviates by as much as 21° between the two structures (monomers B and D of the 1.7 Å and 2 Å models, respectively). ***Table B.*** B-factor analysis for Phe-39, Tyr-130 and flanking residues. The average main chain and side chain B-factors (Å^2) are given on the left and right in each cell, respectively. Values for each residue were averaged across the four monomers of the respective model.

degree of similarity with HiNAL, as judged by the rms deviations of the strictly conserved Thr-44, Tyr-130, Lys-155 and Gly-179 (Figure 2.1.12**A**). In addition to these four, the active site residues Thr-43, Pro-133, Asp-181, Val-196, as well as Tyr-103 from an adjacent subunit, are also identical or subject to conservative mutations between subfamily members (Thr-43 to serine in NAL; Val-196 to isoleucine in NAL and DHDPS).



E. coli NAL	*H. influenzae* NAL	*E. coli* DHDPS	*T. maritima* DHDPS	*M. tuberculosis* DHDPS
0.43/0.62	0.24/0.54	0.57/0.75	0.55/0.79	0.52/0.96

Figure 2.1.12 – The NAL subfamily. A. Superposition of SsKDGA with each of the NAL subfamily enzymes, EcNAL (PDBid: 1nal), HiNAL (1f5z), EcDHDPS (1yxc), TmDHDPS (1o5k) and MtDHDPS (1xxx). The models were superimposed with *LSQKAB*, using the four strictly conserved residues Thr-44, Tyr-130, Lys-155 and Gly-179 (KDGA numbering). The values on the left and right in each cell are the rms deviations (Å) for Cα's and all atoms of the four residues, respectively. **B.** Stereoview of the superimposed active sites of SsKDGA, EcDHDPS and HiNAL. Colour key: **nitrogen**, blue; **oxygen**, red; **sulfur**, orange; **carbon**, yellow (KDGA specific), green (NAL specific), or white (DHDPS specific). Additionally, magenta carbons indicate strictly conserved residues and violet carbons conservative mutations between the three enzymes.

The only large variation in residue functionality between the active sites is the arginine observed in all DHDPS enzymes (Arg-138 in *E. coli*) and replaced by an alanine in KDGA (Ala-135), or a leucine in NAL. DHDPS also displays the greatest deviation in terms of the surrounding cavity's overall size (Table 2.1.11). In both the monomeric and oligomeric state, the *E. coli* and *T. maritima* enzymes have significantly smaller active site cavities than KDGA. *E coli* and *H. influenzae* NAL, on the other hand, do not differ significantly from KDGA.

The residues Tyr-130$_x$, Thr-43$_x$ and Tyr-103$_y$ (where *x* and *y* are different subunits) form a hydrogen-bonding network in KDGA that has also been observed in the other

NAL subfamily members (Blickling *et al.*, 1997b; Lawrence *et al.*, 1997) and in

		SsKDGA Native	EcNAL (1NAL)	HiNAL (1F5Z)	EcDHDPS (1YXC)	TmDHDPS (1O5K)
Monomer	s.a. ($Å^2$)	405	370	430	315	335
	V. ($Å^3$)	615	545	655	340	315
Dimer	s.a. ($Å^2$)	780	645	680	425	430
	V. ($Å^3$)	1,130	1,020	1,120	505	465

Table 2.1.11 – The active site cavity. The surface area (s.a.) and volume (V.) of the active site cavities in five NAL subfamily enzymes. Calculations were carried out using the *CASTp* server (http://sts.bioengr.uic.edu/castp/; Binkowski *et al.*, 2003). The values shown correspond to the protein cavities' molecular surface, i.e. that which is 'seen' by the surface of a probe with a 1.4 Å radius (Richards, 1977). Calculations were carried out on isolated monomers, as well as close dimers, and results averaged across all chains in a given model.

SsKDGA Native	EcNAL (1NAL)	HiNAL (1F5Z)	HiNAL (1F6P)	EcDHDPS (1YXC)	EcDHDPS (1DHP)	TmDHDPS (1O5K)	MtDHDPS (1xxx)
3.3/0	15.8/0.2	7.5/0	7.2/0.1	0/14.1	0.1/11.8	0/10.2	0/4.8
3.3 (0)	16 (0)	7.5 (0)	7.3 (0)	14.1 (3.7)	11.9 (5.6)	10.2 (5.1)	4.8 (4.3)

Table 2.1.12 – Solvent accessibility analysis. Solvent accessibility of Tyr-103 in SsKDGA and of equivalent residues in EcNAL (Y110), HiNAL (Y110), EcDHDPS (Y107), TmDHDPS (Y106) and MtDHDPS (Y117). PDBid's of the models used in calculations are shown in the top row, in brackets. Solvent accessibilities were calculated with a 1.4 Å probe using *AREAIMOL*. Values on the left and right in each cell of the second row are the main chain and side chain solvent accessible areas ($Å^2$), respectively. The third row shows the total solvent accessibility and, in brackets, the accessibility of the side chain hydroxyl oxygen. Values shown have been averaged across the subunits in a given model.

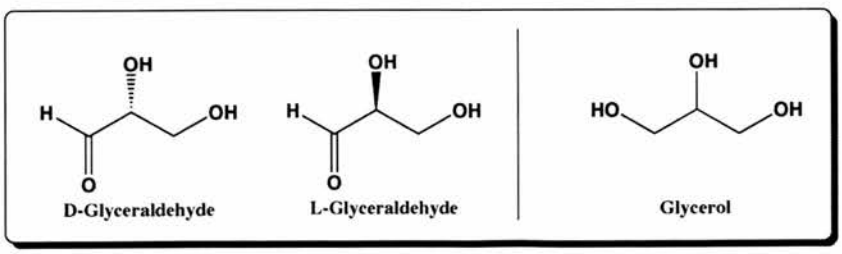
which the intersubunit tyrosine consistently occupies a disallowed region of the Ramachandran plot. In KDGA, however, the hydrogen-bonding network extends from Tyr-130 to Thr-157 via the ordered water W1 (Figure 2.1.11). While Thr-157 is conserved in NAL, a bridging water molecule is not present in any of the available

structures. In all DHDPS enzymes, on the other hand, the residue is replaced by an alanine.

The residues of the hydrogen bond network have been assigned an essential role in DHDPS on the basis of structural and kinetic studies of the *E. coli* enzyme (Blickling *et al.*, 1997b; Dobson *et al.*, 2004; Dobson *et al.*, 2005). In particular, this triad ($Y133_x$-$T44_x$-$T107_y$) have been proposed to facilitate proton shuffling during catalysis by providing a bridge between the active site and bulk solvent. Central to this hypothesis is the intersubunit tyrosine, which is solvent accessible in DHDPS. This is not the case in KDGA and NAL, however, where the presence of additional residues (phenylalanine, tyrosine or proline) in that region of the AD(BC) interface result in solvent exclusion and side chain burial (Table, 2.1.12). Consequently, it is unlikely that the proton shuttle proposed for DHDPS also functions in the latter two enzymes.

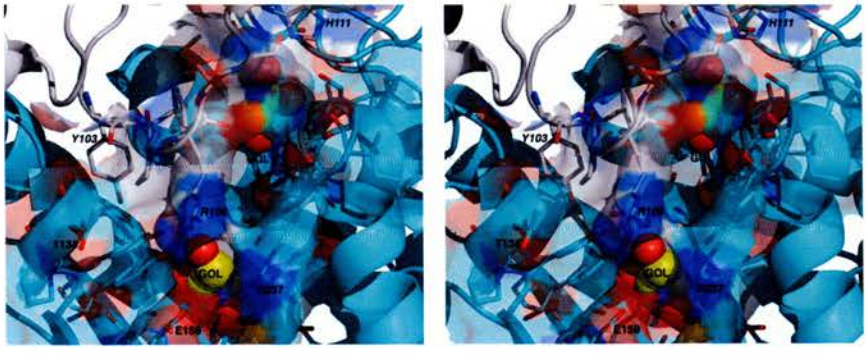
Bound ligands:

During refinement of the two native SsKDGA structures, a number of glycerol molecules were identified in the electron density. This in itself is not surprising given the high concentration of this molecule in the final cryoprotection solution: 20% (v/v), or ~2 M. Nevertheless, their interactions with the protein are worth some consideration due to the similarity in molecular structure between glycerol and the enzyme's natural substrate, glyceraldehyde:

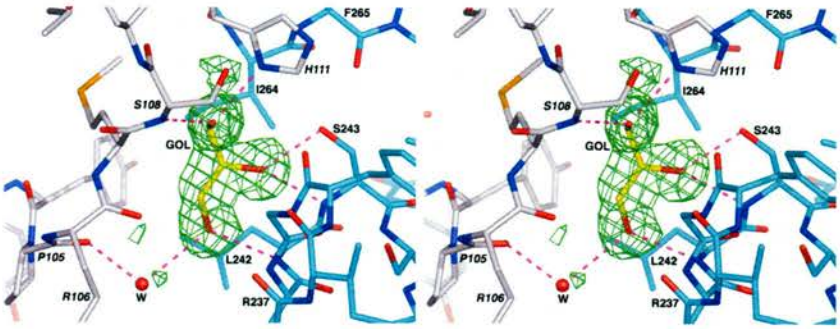


This achiral analogue of glyceraldehyde was consistently observed across the subunits of the 1.7 Å structure, at two locations within the active site cavity (Figure 2.1.13**A**). One of these binding sites is in close proximity to the catalytic lysine (Figure 2.1.13**C**, **D**), while the other is more remote (Figure 2.1.13**B**). Although, glycerol molecules were observed at the latter site in the 2 Å structure, none were bound near the base of the pocket. Moreover, several molecules were found on the protein's surface in both structures, however, these will not be considered further.

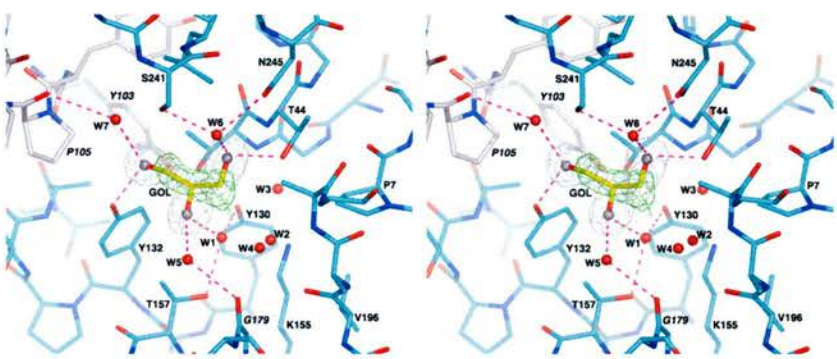
Figure 2.1.13 – Glycerol binding in SsKDGA. Stereoviews of the glycerol binding sites in SsKDGA. Map interpretation in the two native structures revealed glycerol molecules (GOL) bound at two distinct locations within the active site cavity (**A**). One binding site is located in a pocket near the entrance to the cavity (**B**) and the other is near the base of the active site. Moreover, different positions and orientations were observed at the latter site (**C**, **D**).



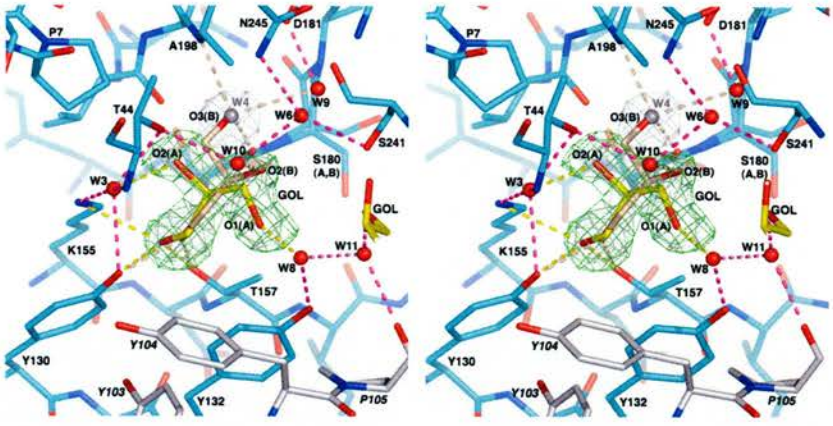
A. Cartoon representation of the close KDGA dimer, with key residues shown as sticks. Residues of the 'second' subunit are labelled in italics. A semi-transparent surface view of the active site cavity has been overlayed. The glycerol molecules are shown in space-fill mode (GOL). Element colour key: **nitrogen**, blue; **oxygen**, red; **sulfur**, orange; **carbon**, yellow (glycerol), cyan (monomer contributing the catalytic lysine to active site), or grey (adjacent monomer, contributing additional residues to active site). This key also applies to Figures 2.1.13**B**, **C** and **D**.



B. Protein-glycerol interactions at the remote binding site. The raw F_o-F_c density interpreted as glycerol is superimposed as a wire-frame contoured at 2.5 σ. Hydrogen bonds are shown as magenta dashes. [Average distances of interactions: **GOL-*H111***, 2.8 Å; **GOL-*S108***, 3 Å; **GOL-S243(Oγ)**, 2.9 Å; **GOL-S243(N)**, 3.1 Å; **GOL-L242**, 3 Å; **GOL-W**, 3.1 Å; **W-*P105***, 2.7 Å].



C. Protein-glycerol interactions for one of the positions observed near the base of the active site. A glycerol molecule was found at this position in two subunits of the 1.7 Å structure. Interpretation was based on the continuous F_o-F_c density (green mesh; 2.5 σ) observed after refinement of the model with three water molecules incorporated at this site by *ARP/warp* (silver spheres). Also shown, the $2F_o$-F_c map density observed for the three water molecules contoured at 1 σ (white mesh). Numbering of water molecules is consistent with Figure 2.1.11. [Average distances of interactions: **GOL-T44**, 2.7 Å; **GOL-W6**, 2.7 Å; **GOL-W1**, 2.9 Å; **GOL-W5**, 2.8 Å; **GOL-Y132**, 2.5 Å; **GOL-W7**, 2.7 Å; **W6-S241**, 2.8 Å; **W6-N245**, 3.1 Å; **W5-G179**, 3 Å; **W7-*P105***, 3.5 Å (close contact).]



D. In two subunits the glycerol molecules were observed within hydrogen bonding distance of the catalytic lysine. Three different orientations were seen, two of which represent alternative conformations of the same molecule. The F_o-F_c density interpreted as the first glycerol conformer (*A*; yellow carbons) is shown as a green wire-frame contoured at 2.5 σ. The observed density, however, could not be accounted for fully by a single set of glycerol coordinates. Best agreement with the Fourier difference maps was achieved by refining the model with two alternative glycerol conformations (*A* and *B*) at partial occupancies. The second conformer (*B*; 'wheat' carbons) extends into an area of electron density previously interpreted as a water molecule (W4 - silver sphere) and B-factors after refinement of this glycerol indicated that W4 is in fact also partially present. The $2F_o$-F_c density corresponding to W4 at full occupancy is shown as a white wire-frame contoured at 1 σ. Potential hydrogen bonds made by the *A* and *B* glycerol conformers are drawn as yellow and wheat coloured dashes, respectively. Hydrogen bonding interactions between other atoms are in magenta. The W7-W8 interaction is drawn in violet to highlight the greater distance (3.5 Å). W11 is equivalent to W in Figure 2.1.13**B** and is shown interacting with the glycerol molecule bound at the remote site. [Interaction distances: **GOL(*A*)-K155**, 3.1 Å; **GOL(*A*)-Y130**, 2.6 Å; **GOL(*A*)-T157**, 2.3 Å; **GOL(*A*)-W3**, 3 Å; **GOL(*A*)-W8**, 2.9 Å; **W8-Y132**, 2.6 Å; **GOL(*B*)-Y130**, 2.7 Å; **GOL(*B*)-T157**, 2.8 Å; **GOL(*B*)-W10**, 2.8 Å; **W10-T157**, 2.8 Å; **W10-W6**, 2.8 Å; **GOL(*B*)-G179**, 2.7 Å; **GOL(*B*)-A198**, 2.8 Å; **GOL(*B*)-W9**, 3.1 Å; **W9-D181**, 2.7 Å.]

The remote site is a pocket located near the entrance of the active site cavity and is formed by residues 240-243, 263 and 264 of a given monomer (the one contributing the catalytic lysine to the active site), as well as residues 104-108 and 111 of the adjacent 'close' subunit. The difference Fourier maps clearly indicated the presence of a glycerol molecule in this pocket in all monomers of the two native structures,

with the exception of one at 1.7 Å for which the electron density was ambiguous. In the refined model the molecule adopts the same conformation in all subunits and makes equivalent interactions with the surrounding residues (Figure 2.1.13**B**).

Binding near the base of the active site appears to be more variable and two alternative positions were identified in the 1.7 Å structure, each one occupied in two of the four subunits. The glycerol molecule's conformation at the first of these positions is highly conserved between the monomers (Figure 2.1.13**C**) and is stabilised by hydrogen bonding to Thr-44 and Tyr-132, as well as water mediated interactions with Tyr-130, Thr-157, Gly-179, Ser-241 and Asn-245. In contrast, three different orientations were observed at the second position, roughly corresponding to rotations within the same plane around the molecule's centre (carbon C-2).

Despite the inconsistency of their conformations, the glycerol molecules at this second position are characterised by their closer proximity to the catalytic lysine (3-3.5 Å from K155 Nζ) and stabilisation via direct and water mediated interactions with a similar set of residues. So, one of the glycerol hydroxyl groups in each of the three orientations is within hydrogen bonding distance of Tyr-130 and Thr-157. Water mediated interactions are made with Thr-44 and Tyr-132 in two orientations, while additional stabilising groups include Lys-155, Gly-179, Ala-198 and Val-196. The two subunits in which this binding is observed also display the most significant deviations in the $\chi 2$ torsion angle of Tyr-130 (Table 2.1.10).

Two of the identified orientations near Lys-155 belong to a single glycerol molecule that was observed in two alternative conformations with fractional occupancies of 0.4 [GOL(*A*)] and 0.6 [GOL(*B*)], respectively (Figure 2.1.13**D**). This interpretation gave the best results in terms of both electron density fit and average B-factors (18 $Å^2$ for *A* and *B*). A single outlier with a very low B-factor (7 $Å^2$), the O3 atom of conformer *B*,

suggested the additional presence of a water molecule at partial occupancy in that area of density. Indeed, further refinement with a water (W4) included in the model at 0.4 occupancy gave reasonable B-factors for both atoms [O3(*B*), 11 Å^2; W4, 13 Å^2].

The three orientations closest to Lys-155 are probably not relevant to glyceraldehyde binding during catalysis, as indicated by the steric clashes observed upon superposition of the native KDGA structure with pyruvate Schiff base complexes of NAL and DHDPS (e.g. 1fdz and 1o5k). However, this does not exclude the possibility that the substrate can be stabilised in an alternative orientation involving interactions with many of the same residues. The significance of glycerol binding at either of the other two sites (Figure 2.1.13**B**, **C**) is also unclear at this stage. While binding at these locations might be unproductive, it could also form part of the trajectory for glyceraldehyde entry and release from the active site cavity. Moroever, the observed glycerol binding sites and residue interactions could be relevant to pyruvate, which has a similar size - if not structure - as glyceraldehyde.

Summary

The resolution of the MAD derived SsKDGA structure has been extended to 2 Å and 1.7 Å using isomorphous crystals of the native enzyme. With the exception of some minor adjustments and corrections in a few main chain and side chain positions, the native and SeMet derivative structures are identical. On the other hand, the higher resolution models do display greater coordinate precision and additional information. Several alternative side chain conformations have been observed, as well as a more extensive solvent structure.

The enzyme's proposed catalytic residues have been found lining a solvent accessible pocket with a conserved architecture and similar composition to the active sites of other NAL subfamily members. This pocket is located within a large cavity formed by residues of two monomers, which constitute a close dimer pair in the enzyme's tetrameric structure. Analysis of the interactions made by glycerol molecules found ordered within the active site cavity has served to identify a number of residues that might be involved in substrate binding and/or stabilisation during catalysis. It has also highlighted the potential importance of water-mediated interactions in this process.

Part 2

Substrate recognition and catalysis by *S. solfataricus* KDGA

2.2.1 The structural basis of substrate promiscuity

Crystal structures of proteins in complex with ligands, substrates, or their analogues, have made and continue to make an important contribution to functional studies. In particular, such structures have proved invaluable in investigations of the mechanisms of substrate recognition, specificity and catalysis in both type I and type II aldolases (see Chapters 1.2.2, 1.2.3, 1.2.4). In the case of the Schiff base forming aldolases, a large number of complexes have been reported over the last decade (Table 2.2.1) involving natural substrates, as well as substrate analogues and inhibitors.

A number of strategies have been employed in trapping substrates (cocrystallisation, soaking, borohydride reduction, mutagenesis) and have yielded both non-covalent and covalent intermediate complexes. With the exception of FBPA and DERA, however, success has been restricted to the donor substrate involved in the condensation reaction, which cannot turnover in the absence of an electrophilic acceptor. Therefore, an understanding of aldol product binding and stabilisation in members of the NAL subfamily has relied on substrate analogues (Blickling *et al.*, 1997b; Barbosa *et al.*, 2000).

Sugar metabolism in *S. solfataricus* has long been a subject of considerable interest, owing to its unique features (see Chapter 1.1.2). Moreover, detailed investigations have in recent years demonstrated a promiscuity in the organism's Entner-Doudoroff

Enzyme	PDBid	Reference	Crystal Condition	Substrate Concentration	Complex Formed	Trapping Method
O.cuniculus (muscle) FBPA	1ado	Blom *et al.*, 1997	42% $(NH_4)_2SO_4$	-	DHAP non-covalent complex	-
H.sapiens (muscle) FBPA	4ald	Dalby *et al.*, 1999	Imidazole (pH 6.0), 70% $(NH_4)_2SO_4$	D-FBP (100 mM)	D-FBP Michaelis complex	Soak (several days)
O.cuniculus (muscle) FBPA	1j4e	Choi *et al.*, 2001	0.1 M Tris-Hcl (pH 7.4), 28% PEG 6k	D-FBP (480 μM)	DHAP Schiff base analogue	Borohydride reduction
T.tenax FBPA	1ok4	Lorentzen *et al.*, 2003	9% (w/v) PEG 4k, 0.1 M Na^+ acetate (pH 5.0), 1-8% (v/v) glycerol	DHAP (100 mM)	DHAP Schiff base intermediate	Cocrystallisation
T.tenax FBPA	1w8s	Lorentzen *et al.*, 2005	9% (w/v) PEG 4k, 0.1 M Na^+ acetate (pH 5.0), 1-8% (v/v) glycerol	D-FBP (100 mM)	β-D-FBP Michaelis complex	Mutant enzyme, soak (2 min)
T.tenax FBPA	1w8r	Lorentzen *et al.*, 2005	9% (w/v) PEG 4k, 0.1 M Na^+ acetate (pH 5.0), 1-8% (v/v) glycerol	D-FBP (100 mM)	D-FBP carbinol-amine intermediate	Mutant enzyme, soak (2 min)
O.cuniculus (muscle) FBPA	1zai	St-Jean et al, 2006	0.1 M HEPES (pH 7.5), 17.5% PEG 4k	D-FBP (10 mM)	D-FBP Schiff base	Soak (3 min)
E.coli DERA	1jcl	Heine *et al.*, 2001	4 °C, 13-18% MPEG 5k, 0.1 M cacodylate (pH5.5), 15-20% glycerol	D-DRP (10.6 mM)	D-DRP carbinol-amine	Short soak
E.coli DERA	1jcj	Heine *et al.*, 2001	4 °C, 13-18% MPEG 5k, 0.1 M cacodylate (pH5.5), 15-20% glycerol	D-DRP (10.6 mM)	D-DRP Schiff base	Mutant enzyme, short soak
T.thermophilus DERA	1ub3	Lokanath *et al.*, 2004	21 °C, 44% MPD, 0.05 $MgCl_2$, 0.1 M Tris-HCl (pH 7.9)	D-DRP (100 mM)	D-DRP carbinol-amine	Soak (8 min)
E.coli KDPGA	1eua	Allard *et al.*, 2001	0.1 M acetate (pH4.6), 18.5% PEG 3.5k, 0.2 M $(NH_4)_2SO_4$	Pyruvate (10 mM)	Pyruvate carbinol-amine	Soak (15 min)
T.maritima KDPGA	1wa3	Fullerton *et al.*, 2006	22 °C, 75 mM Na^+ acetate (pH 4.6), 0.1 M $(NH_4)_2SO_4$, 27% PEG 4k	-	Pyruvate Schiff base	-

Enzyme	PDBid	Reference	Crystal Condition	Substrate Concentration	Complex Formed	Trapping Method
E.coli NAL	1fdz	Lawrence *et al.*, 1997	75 mM Na_3PO_4 (pH 6.9), 53% $(NH_4)_2SO_4$	Neu5Ac (25 mM)	Pyruvate Schiff base analogue	Borohydride reduction
E.coli DHDPS	-	Blickling *et al.*, 1997 (Biochem.)	2.3 M K_3PO_4 (pH10), 50 mM hydroxyproline, β-octyl glucoside	Pyruvate (50 mM)	Pyruvate Schiff base	Soak (1 day)
E.coli DHDPS	-	Blickling *et al.*, 1997 (Biochem.)	2 M K^+ citrate (pH 5.7), β-octyl glucoside	Pyruvate (40 mM), SAS (120 mM), L-lysine (80 mM)	Pyruvate-SAS aldol product Schiff base	Soak (4 days)
N.sylvestris DHDPS	-	Blickling *et al.*, 1997 (JMB)	0.1 M Na_3PO_4 (pH6.3), 1.8 M $(NH_4)_2SO_4$, β-mercaptoethanol	Pyruvate (10 mM), L-lysine (10 mM)	Pyruvate Covalent	Cocrystallisation
N.sylvestris DHDPS	-	Blickling *et al.*, 1997 (JMB)	0.1 M Tris-HCl (pH 8), 1.3 M $(NH_4)_2SO_4$, 10 mM $MgSO_4$	Pyruvate (10 mM)	Pyruvate Covalent	Cocrystallisation
T.maritima DHDPS	1o5k (2003)	-	20 °C, 0.1 M HEPES (pH 7.5), 28% PEG 400, 0.2 M $CaCl_2$ dihydrate	-	Pyruvate Schiff base	-

Table 2.2.1 – Aldolase substrate complexes. Published/deposited enzyme-substrate complex crystal structures of fructose-1,6-bisphosphate aldolase (FBPA), 2-deoxyribose-5-phosphate aldolase (DERA), 2-keto-3-deoxy-6-phosphogluconate aldolase (KDPGA), N-acetylneuraminate lyase (NAL) and dihydrodipicolinate synthase (DHDPS). Only complexes with at least one natural substrate have been included in the list. A larger number of structures exist involving complexes with substrate analogues and/or inhibitors (**NAL**: 1f7b; 1f73; 1f74; 1fdy; 1hl2. **DHDPS**: 1yxd; (Blickling, S., Renner, C. *et al.*, 1997 – not deposited). **FBPA**: 1zaj; 1zal). Dashes indicate one of the following: the structure has not been deposited; the findings have not been published; the information was not made available. Abbreviations: **FBP**, fructose-1,6-bisphosphate; **DHAP**, dihydroxyacetone phosphate; **DRP**, 2-deoxyribose-5-phosphate; **SAS**, succinate β-semialdehyde – a substrate analogue.

variant pathway that enables the catabolism of both D-glucose and D-galactose via a single set of broad-specificity enzymes (Lamble *et al.*, 2003; Lamble *et al.*, 2004; Lamble *et al.*, 2005). Particularly remarkable was found to be the degree of substrate promiscuity displayed by the organism's KDG aldolase (see Chapter 1.2.4; Figure

1.2.21), prompting additional interest in its biotechnological potential. Consequently, docking studies were carried out with the natural substrates pyruvate, D-KDG and D-KDGal, in order to gain a better understanding of the enzyme's mechanism of substrate binding and catalysis.

Trapping covalent intermediates:

Fundamental to the structural studies of enzyme-substrate complexes undertaken throughout this project was the availability of *mg* quantities of high purity substrate molecules. Where such compounds were not commercially available, as was primarily the case, procedures for their synthesis and isolation had to be developed. This challenging work was carried out by Dr Henry Lamble at the University of Bath, as part of a collaboration between Professor Garry Taylor, Professor Michael Danson and Dr David Hough.

The synthesis of D-KDG and D-KDGal has been previously described (Lamble *et al.*, 2003; Lamble, 2004). Briefly, D-glyceraldehyde was synthesised by periodate cleavage of D-mannitol-1,2-5,6-di-*O*-acetonide (Hirth and Walther, 1985) followed by treatment with H_2SO_4. The product was shown to be free from contaminants by HPLC and ^1H NMR spectroscopy and its enantiomeric purity was assessed by polarimetry. Pyruvate and D-glyceraldehyde were subsequently incubated at 50 °C for 9 h in the presence of SsKDGA, yielding approximately equal quantities of D-KDG and D-KDGal. The two diastereomeric products were separated by Dowex 1X8-formate anion exchange chromatography and analysed by HPLC, polarimetry and NMR spectroscopy. Pyruvate was obtained commercially as a sodium salt at 99% purity (Sigma-Aldrich).

The crystallographic complexes described below were trapped by soaking the substrates into preformed crystals. In all cases the crystals were transferred into crystallisation buffer (Buffer A: 0.1 M HEPES (pH 6), 13% PEG 4k, 8% propan-2-ol) containing a given substrate concentration. They were then cryoprotected and flash frozen by placing in liquid nitrogen or in a nitrogen stream. Cryoprotection was achieved by rapid stepwise equilibration against buffer A containing the substrate and 10% (v/v), followed by 20% (v/v) glycerol.

Substrate	K_m (mM)	V_{max} (units.mg^{-1})	k_{cat} (s^{-1})	k_{cat} / K_m ($s^{-1}.mM^{-1}$)
Pyruvate	1.0 (± 0.1)	15.7 (± 0.3)	8.7	8.7
D-Glyceraldehyde	3.9 (± 0.3)	18.0 (± 1.0)	9.9	2.5
D-KDG	25.7 (± 1.2)	51.4 (± 2.5)	28.2 (± 1.4)	1.1 (± 0.08)
D-KDGal	9.9 (± 0.4)	12.3 (± 0.4)	6.8 (± 0.2)	0.7 (± 0.04)

Table 2.2.2 – Kinetic parameters for SsKDGA. Kinetic analysis in the condensation direction was carried out at 70 °C using a modified version of the TBA assay. Aldol cleavage was monitored at 60 °C using a coupled assay with the L-lactic acid dehydrogenase from *B. stearothermophilus*. Adapted from (Buchanan *et al.*, 1999; Lamble *et al.*, 2003; Lamble, personal communication).

The kinetics of the SsKDGA catalysed reaction with each of the substrates have previously been characterised (Table 2.2.2) by assaying in both the condensation (discontinuous TBA assay), as well as in the cleavage direction (continuous coupled assay). Consequently, efforts were made to carry out soaking experiments with saturating substrate concentrations. This was not always possible and in practice the substrate concentration used in the soaking experiments corresponded to the maximum that could be tolerated by the crystals.

Pyruvate soaks were carried out at room temperature, typically for 15 to 30 min, although crystals could be incubated for several hours in the presence of the compound without visible deterioration. In order to successfully trap complexes of

D-KDG and D-KDGal, on the other hand, it was necessary to decrease the soaking temperature to 4 °C, thus exploiting the temperature-dependence of the enzyme's activity to sufficiently decrease turnover (Figure 2.2.1). Even at this temperature, however, where the enzyme should retain less than 1% of its maximum activity, the soaking time had to be limited to ~1 min.

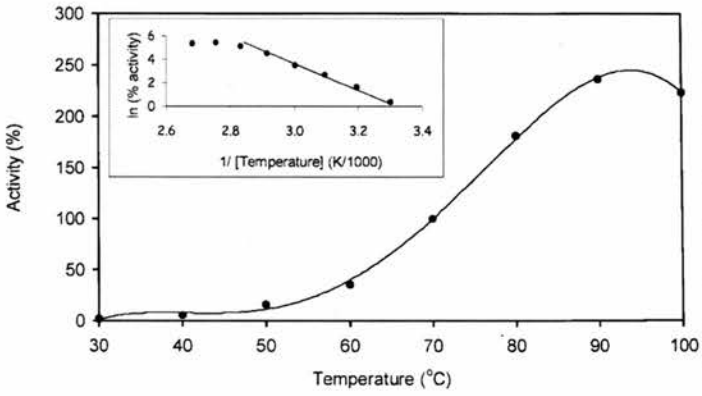


Figure 2.2.1 – Temperature-dependence of KDG aldolase activity. Recombinant KDGA is active over a wide range of temperatures. Activity doubles approximately with every 10 °C rise in temperature from 30 to 80 °C. Activity was measured at temperatures from 20 to 100 °C using the TBA assay and is expressed as a percentage of the activity at 70 °C, the routine assay temperature. An Arrhenius plot of the data is shown inset. Reproduced from Buchanan *et al.*, 1999.

Datasets for each complex were collected on beamline ID14-EH1, at the ESRF in Grenoble. Processing and refinement was carried out in a similar manner to that described previously for the native structures (see Chapter 2.1.3) and outlined in Appendix 1. Table 2.2.3 presents a summary of the processing and refinement statistics for the three complexes. No significant changes were observed in the overall structure of the complexes relative to the apoenzyme (Table 2.2.4). Moreover, with the exception of Phe-39 and Tyr-130, the active site residue conformations were found to be identical and the glycerol molecule identified at the remote site in the apo structure (see Chapter 2.1.4; Figure 2.1.13**B**) was also seen in each of the complexes.

Table 2.2.3 - Data collection and refinement statistics. Summary of statistics for the refined models of SsKDGA in complex with pyruvate, D-KDG and D-KDGal, as well as from collection/processing of the respective synchrotron data. Abbreviations: **rmsd**, root mean square deviation; **R-stand(F)**, uncertainty in the average structure-factor amplitudes; **DPI**, diffraction component precision indicator; $\mathbf{CC_f}$, correlation coefficient between observed and calculated structure-factors.

	Pyruvate	**D-KDG**	**D-KDGal**
	ID14-1 ESRF	**ID14-1 ESRF**	**ID14-1 ESRF**
Wavelength (Å)	0.934	0.934	0.934
Resolution limits (Å)*	40 - 1.7 (1.79 - 1.7)	40 - 2.1 (2.21 – 2.1)	40 - 2.1 (2.21 – 2.1)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å)	a = 83.7 b = 131.0 c = 132.3	a = 83.8 b = 131.3 c = 132.8	a = 83.4 b = 131.7 c = 132.9
No. observations	513,553	494,560	672,636
No. unique reflections	157,939	83,235	85,990
Mosaicity	0.5	0.6	0.6
Completeness (%)*	98.9 (99.7)	97.0 (98.3)	99.9 (100)
R_{merge}‡ (%)*	7.1 (19.7)	8.0 (31.1)	8.1 (21.4)
$< I/\sigma(I) >$*	10.6 (4.4)	17.5 (5.9)	21.6 (9.2)
Multiplicity*	3.3 (3.4)	6 (5.9)	7.8 (7.5)
Wilson B (Å^2)	18.7	22.0	20.3
Refinement			
No. of working / test set reflections	142,220/15,718	74,888/8,347	77,251/8,678
Data Completeness (%)	98.8	96.8	99.9
No. of protein / water / ligand atoms	9,542/878/80	9,438/812/62	9,389/819/112
<B> for protein / water / ligand atoms (Å^2)	19.0/ 31.5/24.4	22.2/33.0/29.2	20.3/31.4/26.2
R-factor / R-free (%)†	17.6 / 21.1	15.1 / 21.2	14.8/20.2
rmsd bond lengths (Å) / bond angles (°)	.015 / 1.8	.023 / 2.3	.022 / 2.3
•*PROCHECK* results (%):			
- Ramachandran plot regions			
core	91.6	91.4	92.3
additionally allowed	8	8.2	7.3
generously allowed	0	0	0
disallowed	0.4	0.4	0.4
- M/c bond lengths within limits	99.9	98.8	99.1
- M/c bond angles within limits	98.8	98.4	98.3
- Planar groups within limits	100	99	99.5
•*SFCHECK* results:			
- R_stand(F)#	0.05	0.041	0.031
- DPI (Å)	0.093	0.159	0.147
- CC_F	0.946	0.942	0.941
- radial error in coordinates (Å)	0.173	0.191	0.188
optical resolution (Å)	1.4	1.6	1.6

*, values in parentheses refer to the highest resolution shell. #, R_stand(F) = $\langle\sigma(F)\rangle/\langle F\rangle$

‡, $R_{merge} = \Sigma_{hkl}\,\Sigma_i |I_{hkl,\,i} - \langle I_{hkl}\rangle| \,/\, \Sigma_{hkl}\langle I_{hkl}\rangle$ †, Rf and R-free = $(\Sigma\,||F_o| - |F_c||)/(\Sigma\,|F_o|)$

The difference Fourier maps for each of the complex structures revealed that the substrates were bound in all four active sites of the asymmetric unit. Moreover, continuous electron density between K155(Nζ) and the C2 position of the substrates in each case, as well as the planar trigonal geometry around these atoms, identified the complexes as covalent Schiff base intermediates. Consequently, coordinates for the relevant substrate atoms were incorporated into the models and covalent links defined before refining against the data.

During refinement of the updated models, the B-factors for several of the substrate/ligand atoms unexpectedly dropped to ~2 Å^2, the minimum allowed value as set by default in the *REFMAC* run scripts (2 – 1,000 Å^2). Use of more reasonable limits (5, or 7 – 500 Å^2) resulted in the B-factors for the relevant atoms refining to their expected values (higher than the lower limit and in agreement with adjacent atoms), with insignificant changes to their coordinates. Moreover, the effect on B-factors and coordinates of protein and water atoms was negligible.

	2 Å apo	Pyruvate Complex	D-KDG Complex	D-KDGal Complex
2 Å apo	*0.21 (0.65)*	-	-	-
Pyruvate Complex	0.16 (0.51)	*0.20 (0.62)*	-	-
D-KDG Complex	0.17 (0.43)	-	*0.20 (0.66)*	-
D-KDGal Complex	0.14 (0.40)	-	-	*0.19 (0.57)*

Table 2.2.4 – Superposition of the apo and substrate complex structures of KDGA. Root mean square deviation (rmsd) of Cα's and all atoms (in brackets) between refined models of the apoenzyme and the three substrate complexes. Superpositions were carried out between residues 2-293 in equivalent subunits using *LSQKAB*. Also shown are the rmsd's between separate subunits of the same model (in italics), generated by superimposing chain A onto B and C onto D.

Coordinates and geometric parameters exist in the *REFMAC* monomer library for small molecules like glycerol and pyruvic acid. This, however, is not the case for D-

KDG, D-KDGal, or any of the Schiff base intermediate forms. Therefore, full geometric descriptions had to be created. Coordinates for an iminium between each substrate and an analogue of the lysine side chain (1-aminopropane) were generated in the CCP4 implementation of *LIBCHECK* (Figure 2.2.2). The coordinates were subsequently modified and subjected to energy minimisation in the molecular design and analysis package, *SYBYL* (version 6; Tripos). Values for chirality, bond lengths, angles, torsional and planar restrains corresponding to the substrate atoms were then output to a geometry file for use in *REFMAC5*. While the PYR(C2)-Lys155(Nζ) bond type (double) and distance (1.3 Å) were defined, bond angles and dihedrals between substrate and lysine atoms were not included in the geometry file.

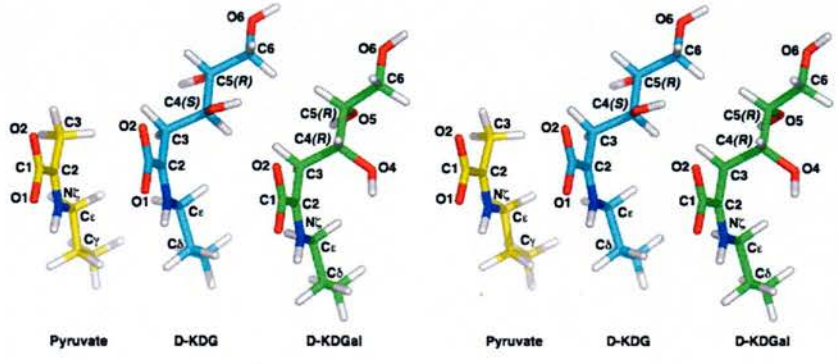


Figure 2.2.2 – Schiff base intermediates. Coordinates were generated for the Schiff base (iminium) between each of the three substrates and 1-aminopropane (mimicking the side chain of lysine). The carboxylate was defined as unprotonated (a delocalised electron system). The coordinates output from *SYBYL* following energy minimisaton are shown as stick representations. Colour key: **carbon**, yellow (pyruvate Schiff base), cyan (D-KDG Schiff base), or green (D-KDGal Schiff base); **oxygen**, red; **nitrogen**, blue; **hydrogen**, white. [Average bond distances/angles: **C1-O1**, 1.26 Å; **C1-O2**, 1.26 Å; **C1-C2**, 1.48 Å; **C2-Nζ**, 1.29 Å; **C2-C3**, 1.52 Å; **C1-C2-Nζ**, 115°; **C1-C2-C3**, 118°; **C3-C2- Nζ**, 127°; **O1-C1-C2-C3**, 180°; **O2-C1-C2-Nζ**, 180°].

The pyruvate Schiff base intermediate complex:

In the case of the pyruvate complex, best results were obtained by soaking a KDGA crystal in a 50 mM substrate solution for 30 minutes, before flash freezing in a nitrogen stream at 100 K. The difference Fourier maps clearly identified pyruvate covalently bound to Lys-155 as a Schiff base and substrate coordinates corresponding to the iminium intermediate form refined well in the density (Figure 2.2.3). The bond distances and angles for the refined substrate are almost identical with the computed values. Moreover, the geometry of the pyruvate Schiff base is in good agreement with other aldolase structures (eg 1o5k, 1wa3), the deviations most likely reflecting different choices of geometric parameters.

Of the existing experimental models of an enzyme-substrate Schiff base complex, by far the most reliable is the 1.1 Å structure of D-DRP bound to an active site mutant of *E. coli* DERA (PDBid: 1jcj). The ultra-high resolution of the data in this case permitted the substrate atoms to be refined without restraints (Heine *et al.*, 2001). The resulting geometric parameters for the DRP Schiff base [C1-Nζ, 1.29 Å; C1-C2, 1.52 Å; C2-C1-Nζ, 125°] are consistent with the values observed here for equivalent atoms of pyruvate in the SsKDGA complex [DRP(C1) ≈ PYR(C2); DRP(C2) ≈ PYR(C3)].

Residue interactions and solvent structure in the active site of the pyruvate complex are highly conserved with respect to the apo enzyme. The hydrogen bonding network described previously ($T157_x$-W1-$Y130_x$-$T43_x$-$Y103_y$; Figure 2.2.3**A**) is observed, as are the ordered waters molecules W4–W7 (Figure 2.2.4). Moreover, a glycerol molecule is bound in two subunits at a previously identified position near the base of the active site pocket (Figure 2.2.4; see also Figure 2.1.13**C**). In addition to its

previously observed direct and water-mediated interactions with the enzyme, the molecule's C1 is positioned 4 Å from the pyruvate methyl group.

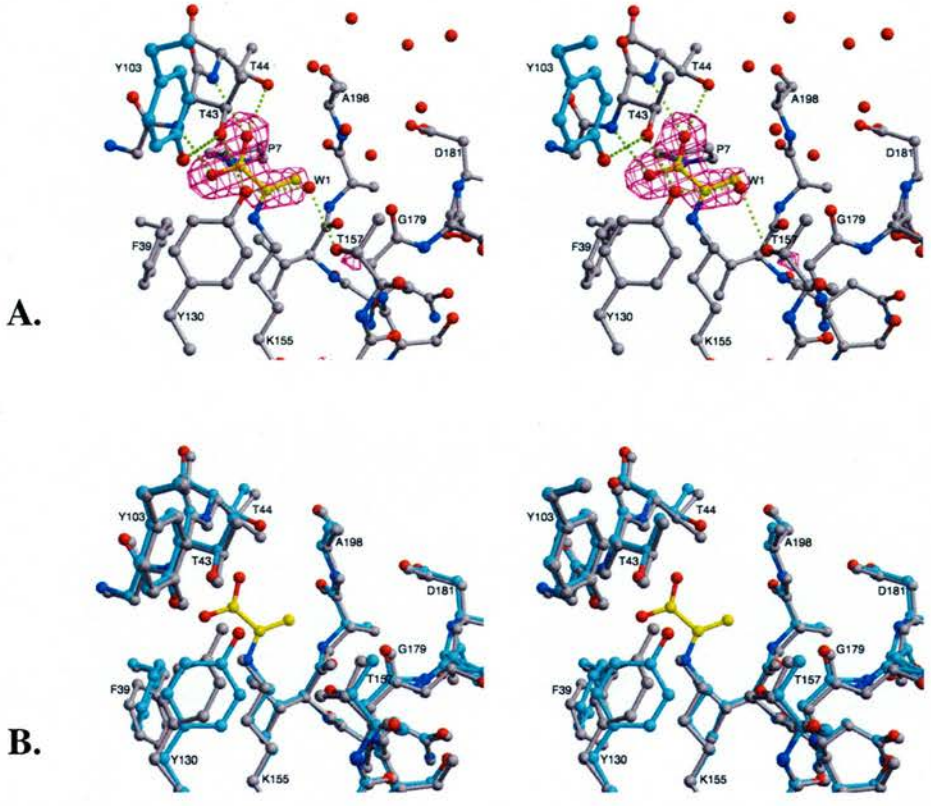


Figure 2.2.3 – The pyruvate Schiff base complex. Stereoviews of the SsKDGA active site as a stick representation. Residues correspond to any subunits *x* and *y* in an *xy* close dimer. **A.** Refined model of the enzyme in complex with pyruvate in the Schiff base intermediate form. Overlayed, the F_o-F_c map density for pyruvate contoured at 2.5 σ. Hydrogen bonds are shown as dashes. **B.** Superposition of the native and pyruvate Schiff base structures of SsKDGA.

Colour key: **nitrogen**, blue; **oxygen**, red; **carbon**, grey (**A.** monomer *x*; **B.** native apo-structure), cyan (**A.** monomer *y*; **B.** pyruvate Schiff base complex), or yellow (pyruvate). [Average interaction distances: **PYR(O2)-T44(Oγ)**, 2.6 Å; **PYR(O2)-T44(N)**, 2.8 Å; **PYR(O1)-T43(N)**, 2.8 Å; **Y130(Oη)-W1**, 2.6 Å; **Y130(Oη)-T43(Oγ)**, 2.6 Å; **T43(Oγ)-Y103$_y$(Oη)**, 2.7 Å; **T157(Oγ)-W1**, 2.8 Å; **PYR(C3)-W1**, 3.2 Å; **PYR(O1)-Y130(Oη)**, 3.2 Å (close contact)]. [Average pyruvate bond distances/angles: **C1-C2**, 1.5 Å; **C2-C3**, 1.5 Å; **C2-Nζ**, 1.3 Å; **C1-C2-C3**, 120°; **C1-C2-Nζ**, 111°; **Nζ-C2-C3**, 127°; **O1-C1-C2- C3**, 173°; **O2-C1-C2- Nζ**, -174°].

The substrate occupies equivalent positions in each of the monomers (rmsd, 0.06 Å) and makes the same interactions with the enzyme in each case (Figure 2.2.3**A**). Its carboxylate group is within hydrogen bonding distance of the backbone amides of Thr-43 and Thr-44, as well as the Oγ of Thr-44. Moreover, the pyruvate's O1 and C3 atoms are within interacting distance (3.2 Å) of Y130(Oη) and the water molecule W1, respectively. While the relative orientations of the tyrosyl's hydroxyl and the substrate's carboxylate oxygen are not consistent with a hydrogen bond, the methyl and ordered water could represent an example of C-H···O hydrogen bonding (Derewenda *et al.*, 1995; Wahl and Sundaralingam, 1997).

It is worth noting at this stage the criteria used in identifying potential hydrogen bond (HB) interactions in this and subsequent SsKDGA substrate complexes. Because the energy of hydrogen bonds falls off slowly compared to other electrostatic interactions (Hagler *et al.*, 1974) it is not possible to choose an exact cutoff distance. Rather both distances and angles must be considered together, the latter being important because of the directionality of HB's. Additional parameters that should be considered where possible are the hybridisation states of the atoms involved and the degree to which any approach lies in the plane of the lone pair(s) (Rossmann and Arnold, 2001).

In an ideal HB the donor heavy atom (*D*), hydrogen atom, acceptor lone pair and acceptor heavy atom (*A*) should all lie in a straight line (Legon and Millen, 1987). In general, however, a HB can be inferred where the angle at the acceptor heavy atom (*a*-***A***···*D*) is greater than 90° (*a* is an atom covalently bonded to the acceptor) and the donor-acceptor distance (*d*-***D***···***A***-*a*) is less than 3.5 Å (Baker and Hubbard, 1984; McDonald and Thornton, 1994). During this project more conservative limits were applied during analysis of interatomic contacts and a HB between two atoms in a

given model was treated as significant only if characterised in most cases by a distance of ≤ 3.2 Å and an angle of ≥ 100°.

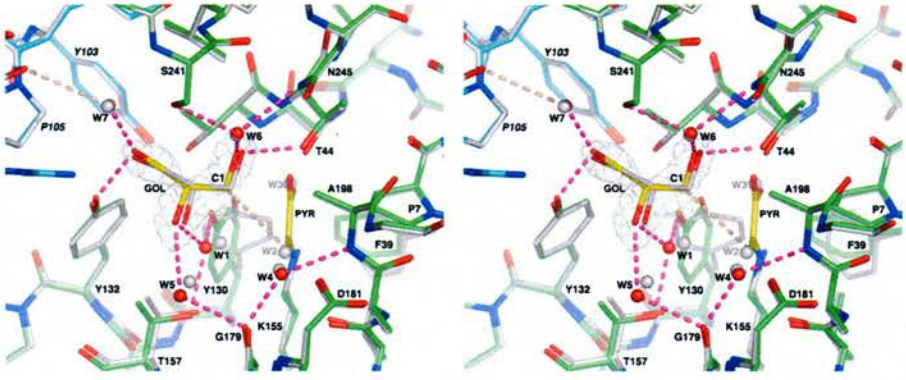


Figure 2.2.4 – Glycerol binding in the active site. Stereoview of the superimposed native and pyruvate complex structures of SsKDGA, displaying the active site as observed for two subunits of their asymmetric units. In addition to the pyruvate adduct (PYR) seen only in the Schiff base complex structure, a glycerol molecule (GOL) was found occupying an equivalent position in both. With the exception of two molecules (W3 and W2) that have been displaced by pyruvate, the two models share a highly conserved water structure in the active site region. The glycerol is stabilised by a number of conserved hydrogen bonds with ordered water molecules and active site residues, and is oriented so that one of its terminal carbons is making a close contact (~4 Å) with the methyl group of the pyruvate intermediate. Overlayed, the F_o-F_c map density for the glycerol molecule shown as a white wire-frame and contoured at 2 σ.

Colour key: **carbons**, yellow (Schiff base complex structure, glycerol and pyruvate), grey (native apo-structure), green (Schiff base complex structure, monomer x), or cyan (Schiff base complex structure, monomer *y*); **water molecules**, red spheres (pyruvate Schiff base complex), or grey spheres (native structure); **hydrogen bonds**, magenta dashes; **close contacts** (≤ 4 Å), wheat dashes. [Average interaction distances: **GOL-T44**, 2.8 Å; **GOL-W6**, 2.7 Å; **GOL-W1**, 2.8 Å; **GOL-W5**, 3.2 Å; **GOL-Y132**, 2.6 Å; **GOL-W7**, 2.7 Å; **W1-T157**, 2.8 Å; **W1-Y130**, 2.6 Å; **W5-G179**, 2.9 Å; **W6-S241**, 2.8 Å; **W6-N245**, 3.1 Å; **W7-*P105*$_y$**, 3.3 Å; **GOL-PYR**, 4.0 Å].

The D-KDG Schiff base complex:

A protein crystal was soaked in buffer A containing 30 mM D-KDG for ~1 min at 4 °C, resulting in clear density for the substrate Schiff base intermediate (Figure 2.2.5**A**). D-KDG adopts a similar conformation across the four subunits of the a.s.u. in the refined model (rmsd, 0.34 Å; maximum deviation, 0.77 Å), with the C5 and C6 positions being the most variable. The keto moiety displays the same configuration to that observed for the pyruvate Schiff base, its carboxylate group making identical interactions with the enzyme as those previously described (Figure 2.2.5**B**).

In the D-KDG complex the ordered water W1 has been displaced by the substrate's C4 hydroxyl (O4), which is positioned within hydrogen bonding distance of T157(Oγ). Moreover, O4 forms a very short interaction with the hydroxyl of Tyr-130 (2.3 Å to 2.6 Å; mean, 2.4 Å). The conformation of substrate atoms around the cleaved bond (C3-C4) in this structure of SsKDGA is in general agreement with observations made for other aldolases, including the equivalent complexes of RAMA (St-Jean *et al.*, 2005) and EcDERA (Heine *et al.*, 2001), as well as substrate analogue Schiff base complexes of EcDHDPS (Blickling *et al.*, 1997b) and HiNAL (Barbosa *et al.*, 2000).

The C5 and C6 hydroxyls also make stabilising interactions with the enzyme. O5 interacts with the conserved water W4, which in turn is within hydrogen bonding distance of the carbonyl oxygen of Gly-179 and the amide of Ala-198. O6 on the other hand forms a direct HB with Tyr-132, as well as water-mediated interactions with Tyr-132 and Thr-44 via the ordered waters W8 and W10, respectively. Despite conformational differences at the C5 and C6 positions, each of these interactions was observed in at least three of the four subunits.

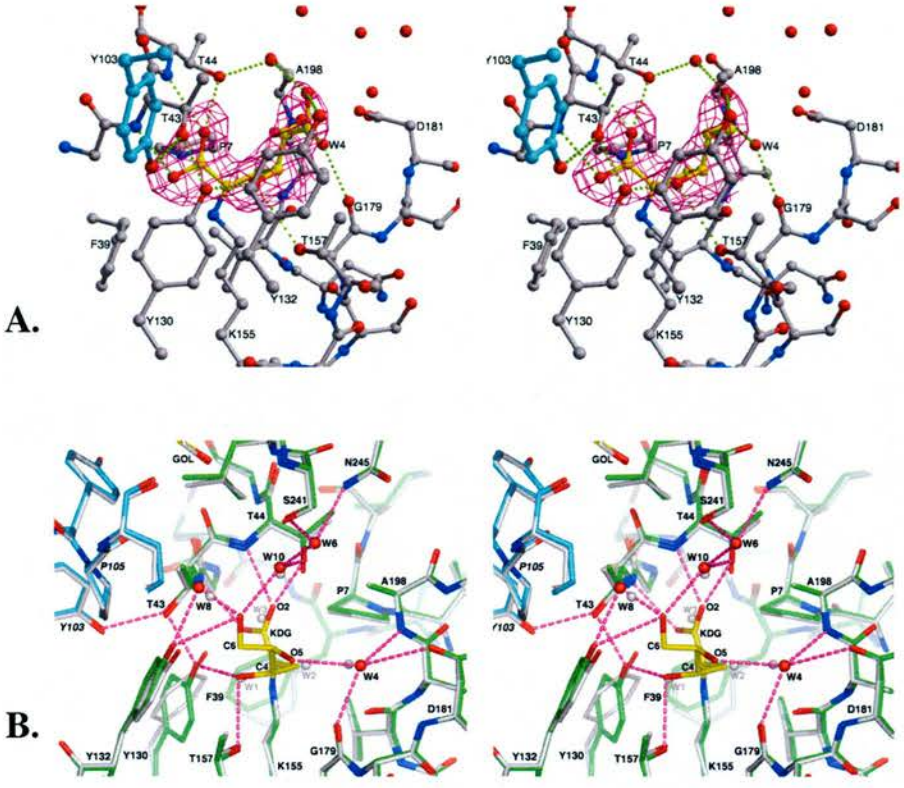


Figure 2.2.5 – The Schiff base complex of SsKDGA with D-KDG. Stereoviews of the enzyme active site as a stick representation. Residues correspond to any subunits *x* and *y*, in an *xy* close dimer. Hydrogen bonds are shown as dashes. **A.** Refined model of the enzyme in complex with D-KDG in the Schiff base intermediate form. Overlayed, the F_o-F_c map density for the substrate contoured at 2.5 σ. **B.** Superposition of the native and D-KDG complex structures of SsKDGA. With the exception of W1, W2 and W3, the solvent structure in the active site of the D-KDG complex is conserved with respect to the apo enzyme.

Colour key: **carbon**, grey (**A.** monomer *x*; **B.** native apo-structure), green, (**B.** D-KDG complex monomer *x*), cyan (D-KDG complex monomer *y* in **A** and **B**), or yellow (D-KDG in **A** and **B**); **water molecules**, red spheres (D-KDG complex), or grey spheres (native structure). [Average interaction distances: **KDG(O2)-T44(Oγ)**, 2.6 Å; **KDG(O2)-T44(N)**, 2.8 Å; **KDG(O1)-T43(N)**, 2.8 Å; **KDG(O4)-Y130**, 2.4 Å; **KDG(O4)-T157**, 3.1 Å; **KDG(O5)-W4**, 2.9 Å; **KDG(O6)-Y132**, 3.1 Å; **KDG(O6)-W8**, 2.5 Å; **KDG(O6)-W10**, 3.1 Å; **KDG(O1)-Y130**, 3.2 Å (close contact); **W4-G179**, 2.8 Å; **W4-A198**, 2.8 Å; **W4-D181**, 3.3 Å; **W8-Y132**, 2.9 Å; **W10-T44(Oγ)**, 2.8 Å; **W10-W6**, 2.6 Å; **W6-S241**, 2.8 Å; **W6-N245**, 3 Å]. [Average D-KDG bond distances/angles: **C1-C2**, 1.5 Å; **C2-C3**, 1.5 Å; **C2-Nζ**, 1.3 Å; **C1-C2-C3**, 116°; **C1-C2-Nζ**, 113°; **C3-C2-Nζ**, 129°].

The D-KDGal Schiff base complex:

The crystallographic complex was obtained using a 25 mM D-KDGal soaking solution, the unbiased Fourier difference maps clearly showing that substrate was present in all four monomers as a Schiff base intermediate (Figure 2.2.6**A**). In addition to continuous density corresponding to the C3-C4 bond of D-KDGal, however, the raw maps also revealed density for the pyruvate methyl. Moreover, coordinates for the D-KDGal adduct at full occupancy did not refine well, the results suggesting partial cleavage of the C3-C4 bond.

A number of models were tested including Schiff base and carbinolamine intermediates of the aldol or its cleavage products, and occupancies were also refined. The best interpretation of the electron density was achieved with a combination of substrate and product Schiff bases at occupancies of 20% (D-KDGal) and 80% (pyruvate and D-glyceraldehyde), respectively. In this final model the two keto moieties adopt equivalent conformations, which are also consistent with the previously described pyruvate and D-KDG complexes. On the other hand, the aldehyde moieties display alternative configurations and significant differences relative to the D-KDG complex, particularly at the C5 and C6 positions.

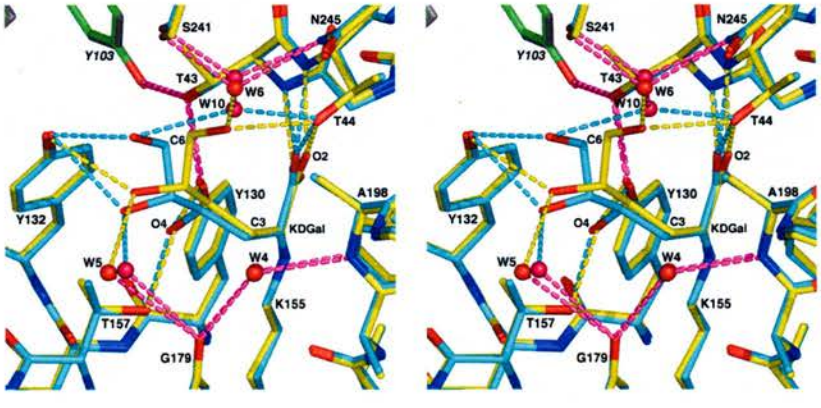
Two alternative conformations of D-KDGal (*X* and *Y*) were observed in the complex structure (Figure 2.2.6**B**), each in two of the subunits of the a.s.u. In all cases the C4 hydroxyl was found making similar interactions as those observed for D-KDG, although there are differences in the relative orientations of atoms owing to the opposite chirality of C4 in the two diastereomers. Consequently, in the D-KDGal complex O4 is positioned more equidistantly from T157($O\gamma$) (mean distance, 2.6 Å) and Y130($O\eta$) (mean distance, 2.5 Å).

Figure 2.2.6 – Stereoviews of the D-KDGal Schiff base complex.

A. Refined model of the covalent complex between Lys-155 (grey carbons) and D-KDGal (yellow carbons), or pyruvate plus free D-glyceraldehyde (orange carbons). Overlayed, the F_o-F_c map density for the substrates contoured at 3 σ (cyan) and 2 σ (magenta).

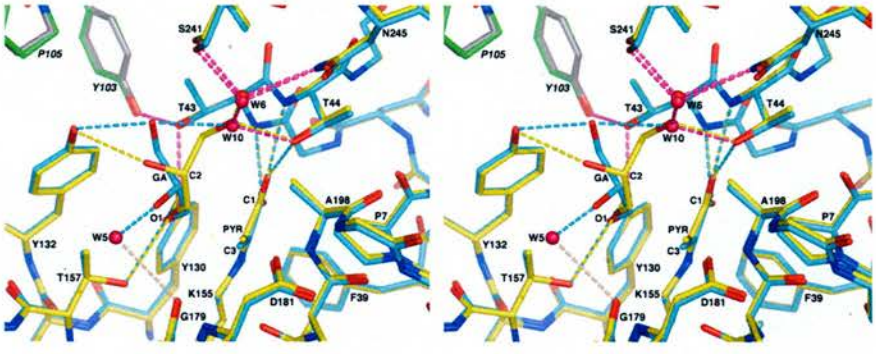
B. Superposition of two active sites in the complex structure showing the distinct D-KDGal conformations *X* (yellow) and *Y* (cyan).

Colour key: **carbon**, yellow (active site containing conformer *X*), cyan (active site containing conformer *Y*), grey or green (Y103 from the adjacent subunit); **water molecules**, red (active site *X*), or hotpink (active site *Y*); **hydrogen bonds**, yellow (made by D-KDGal in *X*), cyan (made by D-KDGal in *Y*), magenta (made between non-substrate atoms). [Average interaction distances: **KDGal(O2)-T44(Oγ)**, 2.6 Å; **KDGal(O2)-T44(N)**, 2.7 Å; **KDGal(O1)-T43(N)**, 2.7 Å; **KDGal(O1)-Y130**, 3 Å (close contact); **KDGal(O4)-Y130**, 2.5 Å; **KDGal(O4)-T157**, 2.6 Å; **KDGal(O5)-Y132**, 3.2 Å; **KDGal(O5)-W5**, 3.2 Å; **KDGal(O6)-T44(Oγ)**, 3.1 Å; **KDGal(O6)-W6**, 3.1 Å; **KDGal(O6)-W10**, 2.7 Å; **W5-G179**, 3.3 Å (close contact); **W10-T44(Oγ)**, 2.6 Å; **W10-W6**, 2.6 Å; **W6-S241**, 2.8 Å; **W6-N245**, 3 Å]. [Average D-KDGal bond distances/angles: **C1-C2**, 1.5 Å; **C2-C3**, 1.6 Å; **C2-Nζ**, 1.3 Å; **C1-C2-C3**; 114°; **C1-C2-Nζ**, 114°; **C3-C2-Nζ**, 125°].



C. Active site superposition showing the pyruvate Schiff base intermediate (PYR) and distinct D-glyceraldehyde (GA) conformations (*X'* and *Y'*) in the final model.

Colour key: **carbon**, yellow (active site containing conformer *X'*), cyan (active site containing conformer *Y'*), grey or green (Y103 and P105 from the adjacent subunit); **water molecules**, red (active site *X'*), or hotpink (active site *Y'*); **hydrogen bonds**, yellow (made by GA and PYR in *X'*), cyan (made by GA and PYR in *Y'*), magenta (made between non-substrate atoms). [Average interaction distances: **PYR(O2)-T44(Oγ)**, 2.5 Å; **PYR(O2)-T44(N)**, 2.8 Å; **PYR(O1)-T43(N)**, 2.8 Å; **PYR(O1)-Y130**, 3.1 Å (close contact); **GA(O1)-Y130**, 2.7 Å; **GA(O1)-T157**, 3 Å; **GA(O2)-Y132**, 3.3 Å; **GA(O2)-W5**, 2.5 Å; **GA(O3)-Y132**, 3.1 Å; **GA(O3)-T44(Oγ)**, 2.8 Å; **GA(O3)-W6**, 2.6 Å; **GA(O3)-W10**, 2.8 Å]. [Average pyruvate bond distances/angles: **C1-C2**, 1.5 Å; **C2-C3**, 1.5 Å; **C2-Nζ**, 1.3 Å; **C1-C2-C3**; 119°; **C1-C2-Nζ**, 114°; **C3-C2-Nζ**, 127°].



The C5 hydroxyl of D-KDGal occupies a similar position in *X* and *Y* and consistently interacts with Y132(Oη). The orientation of the C6 hydroxyl, however, differs significantly in the two conformations. In *X*, O6 is within hydrogen bonding distance of T44(Oγ) and forms a water-mediated interaction with S241(Oγ) and N245(Nδ) via W6. In *Y*, on the other hand, the hydroxyl interacts with T44 via W10. Furthermore, two additional interactions were observed in subunit B. The substrate's conformation in this monomer positions O6 within hydrogen bonding distance of Y132(Oη), while W5 mediates an interaction between O5 and G179(O).

D-glyceraldehyde (GA) also displays the two distinct configurations (*X'* and *Y'*), although their positions are shifted relative to the equivalent atoms of D-KDGal (Figure 2.2.6**C**). O1 is also within hydrogen bonding distance of Tyr-130 and Thr-

157, but with average distances of 2.7 Å and 3 Å, respectively. In its alternative conformations O2 is able to variably interact with Y132(Oη) (in *X'*) and make water mediated interactions with Ala-198 and Gly-179 (in *Y'*). Finally, O3 forms similar interactions to those observed for KDGal(O6) in the equivalent orientations.

Active site conformational changes:

The only significant conformational change observed in the active site upon binding of the three substrates involves the side chains of Phe-39 and Tyr-130, which also display the only active site variability between subunits in the apo structures (Figure 2.2.3**B**; Figure 2.2.5**B**; see also Table 2.1.10). The two residues adopt the same positions in the three complexes, resulting from a large rotation around their $\chi 2$ angles with respect to the free enzyme, as well as a smaller change in $\chi 1$ (Table 2.2.5). Furthermore, the side chains display smaller deviations between subunits in their substrate-bound relative to their free-enzyme conformations, as well as lower B-factors (particularly in the case of Tyr-130), indicating reduced flexibility.

	2 Å apo	1.7 Å apo A,B/C,D	Pyruvate Complex	D-KDG Complex	D-KDGal Complex
Phe-39 $\chi 1$	-155° ±4°	-170° ±3°/-162° ±2°	-175° ±2°	-175° ±2°	-175° ±1°
Phe-39 $\chi 2$	114° ±3°	108° ±2°/109° ±5°	77° ±5°	74° ±4°	76° ±4°
Tyr-130 $\chi 1$	-172° ±2°	-177° ±0.4°/-175° ±0.4°	177° ±1°	178° ±3°	178° ±1°
Tyr-130 $\chi 2$	-46° ±11°	-14° ±18°/-68° ±1°	12° ±2°	15° ±1°	12° ±2°

Table 2.2.5 – Active site conformational changes. The $\chi 1$ and $\chi 2$ torsion angles for Phe-39 and Tyr-130 in the apo and substrate complex structures of SsKDGA. In each cell the mean value and standard deviation of a given torsion angle are presented on the left and right, respectively. In the case of the 1.7 Å apo structure, the pairs of monomers correspond to the two different glycerol binding positions observed near Tyr-130.

The altered conformation of Tyr-130 observed in the substrate complexes is characterised by an increase in the distance of its hydroxyl from i) the amine of Lys-155 (3.1 Å → 3.5 Å) and ii) the carboxylate O1 atom of substrates (2.3 Å → 3 Å). Moreover, in all complexes the aromatic ring of Tyr-130 is rotated so that its plane is parallel to that of the Schiff base, while Pro-7 is positioned at ~ 3.5 Å on the opposite side of the substrate. Therefore, the conformational change in Tyr-130 appears to aid in accommodating and properly orienting the approaching α-keto acid moieties within the binding pocket. The residue's apparent flexibility in the apoenzyme structure is consistent with this, as it reduces the barrier to substrate binding.

In the case of Phe-39, on the other hand, the shift in position of the aromatic ring results in a significant decrease in its distance from K155(Nζ) (mean, 3.9 Å) and the carboxylate O1 atom of substrates [F39(Cδ2), 3.5 Å; F39(Cε2), 3.4 Å]. Moreover, the plane of the aromatic ring is re-oriented so that it becomes approximately parallel to the long axis of the lysine side chain. Active site superpositions show that in the complex structures the side chain of Phe-39 can be accommodated in either its free-enzyme or substrate-bound conformation, indicating that the latter is preferentially stabilised. The Cε2-O1 interaction may be particularly significant in this respect.

In each of the complex structures the distances and angles between F39(Cε2) and the substrate O1 atom are consistent with the formation of a C-H···O hydrogen bond {Derewenda, 1995). Moreover, the relative geometry of the two heavy atoms would indicate that the aryl proton is directed towards the free lone pair of O1 the other lone pair being able to form a HB with the amide of Thr-43). Therefore, in addition to helping stabilise the conformation of the residue's aromatic ring, this interaction may provide the donor atom needed by the substrate carboxylate in order to satisfy its full HB potential.

Interestingly, the movements displayed by the active site residues of SsKDGA have not been observed in either NAL or DHDPS. The strictly conserved tyrosine of both these subfamily members adopts the same orientation in apo and complex structures as that of Tyr-130 in the KDGA complexes. Moreover, a second tyrosine in NAL that occupies the position of Phe-39 is also observed in a single conformation, equivalent to that of the phenylalanine in its substrate-bound state. The reduced flexibility of these residues in NAL and DHDPS appears to be due to additional steric restraints not present in SsKDGA, which maintain the side chains in a suitable conformation for Schiff base formation.

A model for promiscuous substrate binding:

The complex structures described above identify the interactions involved in stabilisation of substrate intermediates in the active site of SsKDGA, confirming several predictions made from structural studies of other NAL subfamily members (Blickling *et al.*, 1997b; Blickling *et al.*, 1997a; Lawrence *et al.*, 1997; Barbosa *et al.*, 2000). Lys-155 is conclusively identified as the Schiff base forming residue, while Thr-43 and Thr-44 (second and third residues of the -GXXG- motif) stabilise the α-keto acid carboxylate of the substrates. Moreover, Tyr-130 is in close contact with the carboxylate O1 atom in all three complexes and interacts with the C4 hydroxyl of both D-KDG and D-KDGal, consistent with its proposed catalytic role.

On the other hand, the experimental evidence is not in agreement with the predictions made for Gly-179 and Asp-181 from the substrate-analogue complexes of EcDHDPS (Blickling *et al.*, 1997b) and HiNAL (Barbosa *et al.*, 2000). The main chain carbonyl oxygen of Gly-179 is not involved in stabilising the C4 hydroxyls of D-KDG(al), but

rather forms water mediated interactions with the C5 hydroxyls. Moreover, Asp-181 does not appear to play any significant role in substrate binding, as it only interacts weakly with W4 in the D-KDG complex (mean, 3.3 Å). The remaining stabilising interactions made by SsKDGA have not been previously predicted, however, they do closely resemble those made by glycerol in the apoenzyme and pyruvate complex structures.

The glycerol molecules found near the base of the active site pocket in these structures do not adopt orientations equivalent to the aldehyde moieties of either D-KDG or D-KDGal. Nevertheless, a conserved set of residue interactions are implicated in binding both the ligand and substrates. In addition to the previously mentioned Tyr-130 and Gly-179, these include direct interactions with T157(Oγ); direct or water mediated interactions with T44(Oγ) and Y132(Oη); and water mediated interactions with G179(O), A198(N), S241(Oγ) and N245(Nδ).

Re-examining the stabilising interactions made by EcDHDPS and HiNAL in light of the above observations reveals previously unrecognised similarities with SsKDGA. In addition to its role in stabilising the α-keto acid carobxylate, T48(Oγ) of HiNAL (equivalent to KDGA-T44) also makes a water-mediated interaction with the N-acetyl groups of the substrate analogues (Neu5Ac2ol, 4d-Neu5Ac, 4oxo-Neu5Ac). Moreover, while Ala-198 is replaced by Ser-207 in HiNAL, this residue's amide group is also involved in stabilising interactions. Moreover, in the EcDHDPS complex with the aldol product of pyruvate and SAS, the carboxylate of SAS interacts with the side chain of Asn-248, which is positioned near that of KDGA-N245 (replaced by Val-252 in EcDHDPS).

There are also, however, unique features to substrate binding by SsKDGA. Water mediated interactions are far less common in the HiNAL complex structures and absent in EcDHDPS. Moroever, Tyr-132 and Ser-241 are not conserved in either of the other two subfamily members, nor are there interactions made by the residues that replace them. Finally, interactions equivalent to those made by Thr-157 in SsKDGA are not observed in HiNAL, although the residue is conserved (Thr-166). In the apo structures of both enzymes the residue adopts the same position, but undergoes a conformational change in the HiNAL complexes, orienting its hydroxyl away from the substrate analogues. Although this conformational change may be an artefact of the experiments, it could also be relevant to the binding of natural substrates in this enzyme.

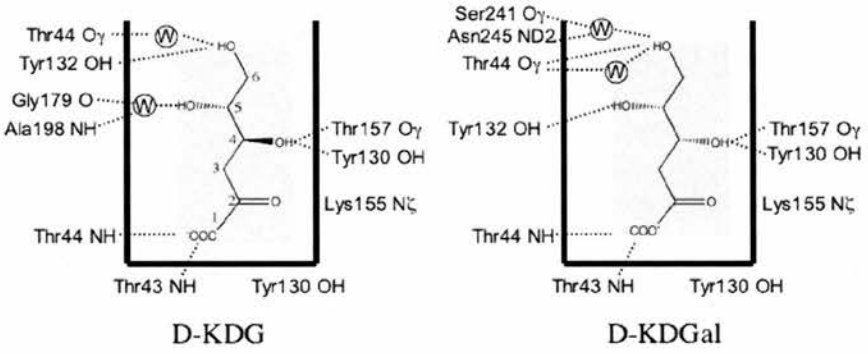


Figure 2.2.7 – Promiscuous substrate binding in SsKDGA. A summary of significant interactions between the enzyme and its two diastereomeric substrates, D-KDG and D-KDGal, represented in each case as a consensus between subunits. HB distances are indicated by dashes and ordered water molecules by circled W's.

The complex structures of SsKDGA allow the formulation of a model for substrate binding by this enzyme that is quite different from that proposed by Barbosa *et al.* (2000), particularly with respect to the identity of the 'secondary group' of residues. As illustrated in Figure 2.2.7, a single set of residues are involved in stabilising different substrate groups, or the same groups oriented differently. In this way the

active site provides alternative hydrogen bonding opportunities for substrates, allowing stereoisomers to bind in distinct conformations.

Ordered water molecules appear to play a key role in this process by permitting a greater economy of protein functionalities and eliminating the need for additional residue flexibility. Moreover, by alternating direct and water-mediated interactions with residues, the enzyme retains a similar affinity towards different substrates. On this basis, the shorter HB with Thr-157 and direct interaction with Tyr-132 may explain, at least in part, the lower K_m of D-KDGal (10 mM) relative to D-KDG (26 mM). Given that similar soaking concentrations and times were used for both substrates, the lower K_m of D-KDGal is also consistent with the diastereomer's observed cleavage.

Insights to the catalytic mechanism:

A comparison of the D-KDG and D-KDGal complexes shows that the two Schiff base intermediates are bound in different orientations with respect to their C5 and C6 positions, but retain a similar conformation around the cleaved bond (C3-C4). Moreover, the overall arrangement of substrate and protein atoms at this position appears consistent with equivalent complexes of other the type I aldolases (Figure 2.2.8). These include the D-FBP Schiff base complex of RAMA (St-Jean *et al.*, 2005; 1zai), the D-DRP complex with the K201L mutant of EcDERA (Heine *et al.*, 2001; 1jcj), the complex of EcDHDPS with the aldol product between pyruvate and SAS (Blickling *et al.*, 1997b), and one of two conformations (that of the hydrate) in the 4oxo-Neu5Ac complex with HiNAL (Barbosa *et al.*, 2000; 1f7b).

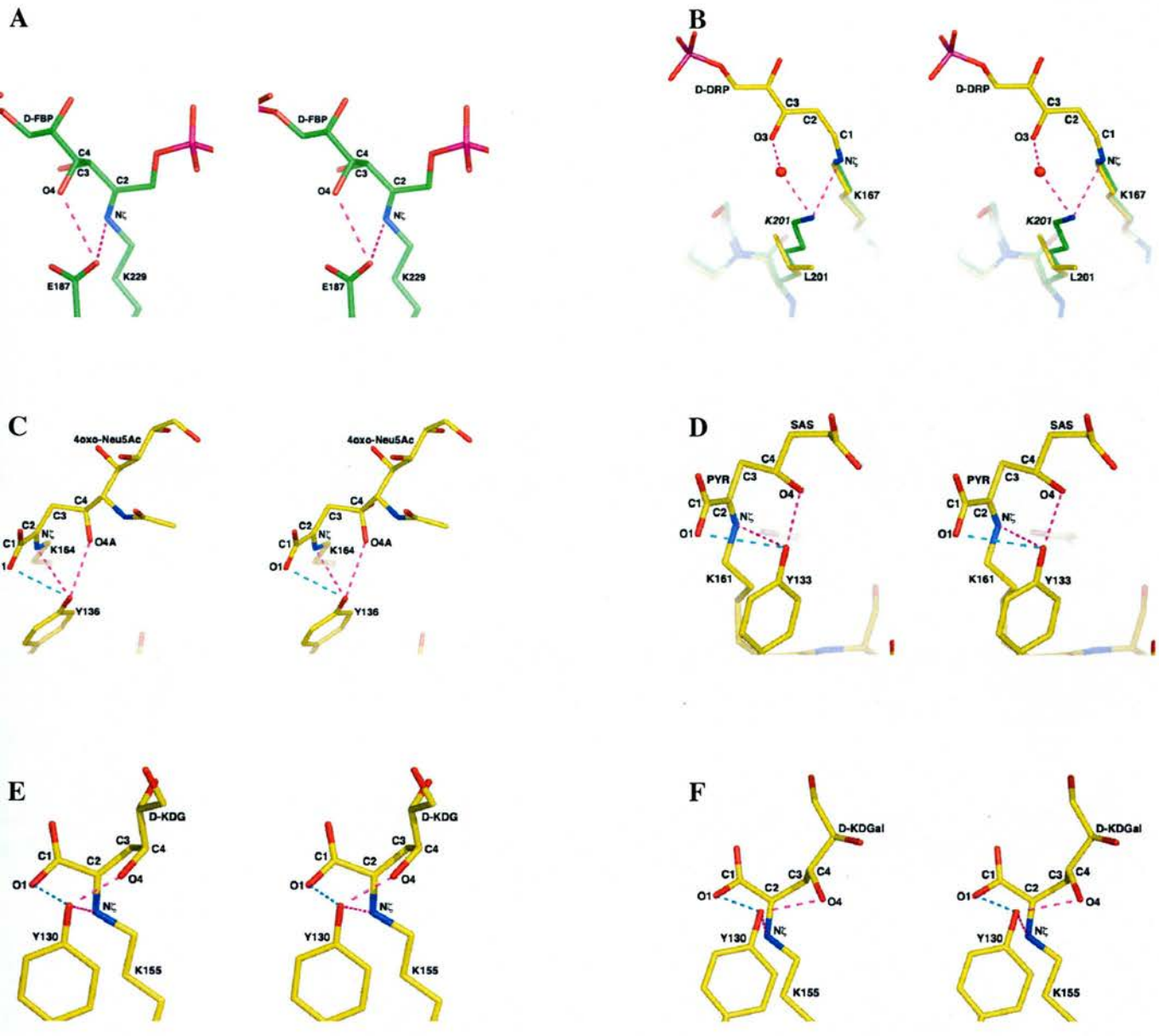


Figure 2.2.8 – Stereoviews of Type I aldolase Schiff base complexes. A. D-FBP bound to RAMA (St-Jean *et al.*, 2005). Atoms FBP(O2), -(O3) and –(O4), together with K229(Nζ) and E187(Oε) outline a chair-like configuration. **B.** D-DRP bound to the K201L mutant of EcDERA (Heine *et al.*, 2001). Superimposed in green carbons, the wild type apoenzyme structure (Heine *et al.*, 2004). The conserved water involved in proton relay is also shown. Joining DRP(C1), (C2), (C3) and (O3), with K201(Nζ), W and K167(Nζ), also results in a chair-like conformation [assuming a straight line between O3, W and K201(Nζ)]. **C.** HiNAL in complex with the hydrate form of its substrate analogue 4oxo-Neu5Ac. Atoms C2, C3, C4, O4, Y136(Oη) and K164(Nζ) possess a half-chair geometry. Atoms O1, C1, C2, C3, C4, O4 and Y136(Oη) also give a half-chair, assuming a straight line between C1, C2 and C3. **D.** The complex of EcDHDPS with the aldol product of pyruvate and the substrate analogue SAS (Blickling *et al.*, 1997b). The two sets of atoms equivalent to those in **E.** outline a chair-like conformation. **E** & **F.** SsKDGA in complex with D-KDG and D-KDGal, respectively. The highlighted atoms [C2, C3, C4, O4, Y130(Oη) and K155(Nζ); or O1, C2, C3, C4, O4 and Y130(Oη)] possess a chair-like or half-chair geometry in the case of D-KDGal and D-KDG, respectively.

While these atoms are positioned on opposites sides of the Schiff base in RAMA and EcDERA, compared to the NAL subfamily members, their relative geometry is highly conserved, with a proposed catalytic residue occupying an equivalent position in each of the structures. Moreover, in the case of RAMA the authors noted the near chair like configuration adopted by the six atoms; K229(Nζ), E187(Oε) and the substrate's C2, C3, C4 and O4. In making their observations they drew comparisons with the pericyclic transition state of nonenzymatic aldol condensations involving preformed enolates (Figure 2.2.9; Dubois and Fort, 1972; Heathcock *et al.*, 1980; Zhang and Houk, 2005).

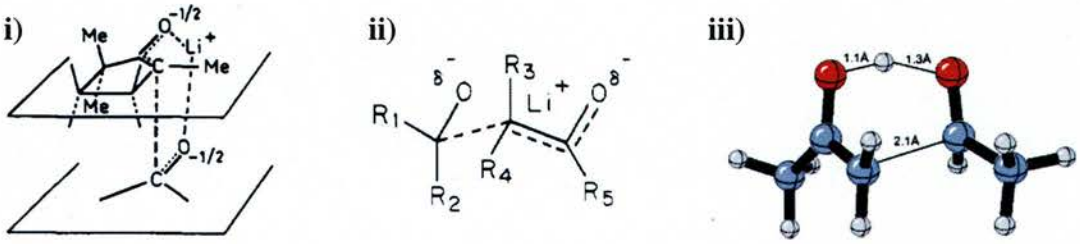


Figure 2.2.9 – Transition states of non-enzymatic aldol condensations. **i)** cyclic TS proposed for metal ion catalysed aldol condensations in weakly polar solvents, or in the presence of small cation catalysts (e.g. Li^+). Reproduced from Dubois *et al.* (1972). **ii)** Six-centre TS proposed for aldol condensations of preformed lithium enolates with aldehydes. Reproduced from Heathcock *et al.* (1980). **iii)** Geometry of the TS in the aqueous aldol reaction between acetaldehyde and acetone enol, where enolisation is catalysed by pure water. Reproduced from Zhang *et al.* (2005).

A chair-like configuration is also formed by the equivalent substrate/protein atoms in the DERA complex, assuming a straight line between DRP(O3), the ordered water and K201(Nζ) of the superimposed wild type enzyme. A similar arrangement is observed in the EcDHDPS complex and that of SsKDGA with D-KDGal. On the other hand, in the two remaining NAL subfamily complexes shown in Figure 2.2.8, the geometry resembles that of a half-chair. Nevertheless, there is a clear consistency

that would suggest the observed substrate/analogue conformations are relevant to catalysis.

The position of Tyr-130 in the SsKDGA complex structures and its interactions with the C4 hydroxyls of D-KDG(al) are consistent with a role in catalysis, as the residue's hydroxyl is able to carry out the proton abstraction that initiates C-C bond cleavage. Moreover, the Y130(Oη)-O4 distances in these structures are very short (D-KDG, 2.4 Å; D-KDGal, 2.5 Å) and could be indicative of low barrier hydrogen bonds, facilitating proton transfer. However, they may also be artefacts of refinement, as is most likely the case for three additional close interactions observed in the complexes, which would be difficult to rationalise as LBHB's (e.g. D-KDG(O6)-W8, 2.3 - 2.8 Å).

In the pyruvate complex, however, the residue's hydroxyl is not able to interact directly with the substrate methyl (~ 4.4 Å), but rather the interaction is mediated by W1. This bridging water could permit Tyr-130 to also carry out the proton abstraction necessary for enamine formation in the condensation reaction. Despite similar large Tyr(Oη)-PYR(C3) distances in equivalent complexes of EcNAL (1fdz), TmDHDPS (1o5k) and EcDHDPS (Blickling *et al.*, 1997b), however, an equivalent ordered water has only been observed in the latter structure.

The inconsistent observation of a bridging water molecule in the enzymes of the NAL subfamily calls into question its relevance to the mechanism of pyruvate activation. While absence of the ordered water molecule in the EcNAL complex may be associated with the trapping method (borohydride reduction), or the low resolution/narrow range used in structure refinement (2.6 – 6.0 Å), it cannot be accounted for in the case of TmDHDPS. Nevertheless, the involvement of explicit

water molecules in proton shuffling has been widely reported in other type I aldolases (Heine *et al.*, 2001; Heine *et al.*, 2004; Fullerton *et al.*, 2006) and is therefore at least a plausible mechanism in SsKDGA, too.

SsKDGA Pyruvate	SsKDGA D-KDG	SsKDGA D-KDGal	EcNAL (1FDY)	EcNAL (1FDZ)	HiNAL (1F73)	HiNAL (1F74)	HiNAL (1F7B)	
3.7/0	3.8/0	3.6/0	14.4/0.2	13.7/0.1	36.7/10.3	38.6/31.1	7/0	43/33.4
3.7 (0)	3.8 (0)	3.6 (0)	14.6 (0)	13.8 (0)	47 (3.4)	69.7 (9)	7 (0)	76.4 (8.4)

Table 2.2.6 – Solvent accessibility analysis. Solvent accessibility of Tyr-103 in SsKDGA and of equivalent residues in EcNAL (Y110) and HiNAL (Y110). PDBid's of the models used in calculations are shown in the top row, in brackets. Solvent accessibilities were calculated with a 1.4 Å probe using *AREAIMOL*. Values on the left and right in each cell of the second row are the main chain and side chain solvent accessible areas ($Å^2$), respectively. The third row shows the total solvent accessibility and, in brackets, the accessibility of the side chain hydroxyl oxygen. Values shown have been averaged across the subunits in a given model. In the case of the *H. influenzae* NAL complex with 4oxo-Neu5Ac (1f7b), the values for the two subunits (A & C) in the a.s.u are shown separately. In this structure the substrate analogue was found bound in different forms in the two active sites. Moreover, in monomer C the residues 138-145 are disordered, resulting in a similar solvent accessibility for Tyr-110 as that observed in the equivalent subunit of the apo structure (1f6k; see Table 2.1.12). The residues of subunit A, on the other hand, were ordered and in an equivalent conformation to that seen in 1f6k. In the Neu5Ac2ol complex (1f73), residues 140-144 appear disordered throughout the a.s.u., in contrast to the apo structure of the same crystal form (1f6p). In the 4d-Neu5Ac complex (1f74), the residue region is ordered in both subunits of the a.s.u., but has undergone a large conformational change with respect to monomer A of its equivalent apo structure (1f6k).

The ability of the NAL subfamily's strictly conserved tyrosine to function in proton transfer has been linked to either its participation in a catalytic triad (Tyr_x-Thr/Ser_x-Tyr_y; Blickling *et al.*, 1997b), or its mediating role in a substrate assisted mechanism involving the α-keto acid carboxylate group (Smith *et al.*, 1999; Barbosa *et al.*, 2000). Although the residues comprising the catalytic triad are present in all three subfamily

members, the intersubunit tyrosine displays the necessary solvent accessibility only in DHDPS. Analysis of the apo structures of KDGA and NAL has shown that the residue's side chain is buried in these enzymes, an observation which is inconsistent with its proposed role in shuttling protons to and from bulk solvent (see Chapter 2.1.4).

The KDGA complex structures display the same solvent accessibilities as the apoenzyme (Table 2.2.6), confirming that there are no conformational changes upon substrate binding that would expose the Tyr-103 hydroxyl group. Moreover, the EcNAL complex structures (1fdz; 1fdy) reveal that Tyr-110 also remains buried (Table 2.2.6). The situation is, however, very different in the case of the *H. influenzae* enzyme (1f7b; 1f73; 1f74). In the analogue complexes of HiNAL there is disordering or large conformational changes in residues 136-147, which result in Tyr-110 appearing highly exposed. While it is difficult to draw conclusions from the HiNAL observations regarding their relevance to this enzyme's normal activity, the structural evidence clearly indicates that the residues of the catalytic triad cannot operate as a proton shuttle in KDGA and EcNAL.

A substrate-assisted mechanism for NAL was initially proposed by Smith *et al.* (1999). It accounted for the paucity of functional groups within the active site and was supported by modelling studies of the enzyme-substrate Schiff base intermediate, the results of which showed the C4 hydroxyl of Neu5Ac forming an intramolecular hydrogen bond with the α-keto acid carboxylate group. However, the experimental evidence for this and other subfamily members that has emerged since then has identified no such interaction (Blickling *et al.*, 1997b; Barbosa *et al.*, 2000).

A role for the conserved tyrosine (Tyr-136 in HiNAL) in mediating the activity of the substrate carboxylate was subsequently put forward by Barbosa *et al.* (2000) following the observation of mutual hydrogen bonding with both the α-keto acid carboxylate and C4 hydroxyl in the 4oxo-Neu5Ac hydrate complex of HiNAL (1f7b, subunit A). Although the arrangement of atoms in the substrate complexes of SsKDGA are in general agreement with this hypothesis, they do highlight certain inconsistencies. Specifically, the angles of interaction between the carboxylate O1 atom of substrates and the hydroxyl of the conserved tyrosine, as well as the expected orientation of protons and electron lone pairs, do not indicate efficient hydrogen bonding.

SsKDGA PYR	SsKDGA D-KDG	SsKDGA D-KDGal	EcDHDPS PYR	EcDHDPS PYR-SAS	TmDHDPS PYR (1o5k)	EcNAL PYR (1fdz)	EcNAL HPYR (1fdy)	HiNAL 4oxo-Neu5Ac (1f7b)
90 ± 1	93 ± 2	92 ± 2	86 ± 8	83 ± 1	78 ± 1	86 ± 3	86 ± 2	97 ± 15
82 ± 1	85 ± 4	86 ± 15	83 ± 5	82 ± 6	64 ± 12	94 ± 9	94 ± 10	93 ± 0.1

Table 2.2.7 – Hydrogen bond interaction angles. A list of angles between the interacting atoms of the substrate carboxylate group and strictly conserved tyrosine in Schiff base complexes of NAL subfamily members. The angle at the hydroxyl of the conserved tyrosine (O1····**Oη**-Cζ) is shown in the second row. The angle at the carboxylate O1 atom of a given substrate (Oη····**O1**-C1) is shown in the third row. The mean values and standard deviations across subunits of a given model are shown in each cell. Abbreviations: **PYR**, pyruvate; **PYR-SAS**, the aldol product of pyruvate and SAS; **HPYR**, hydroxypyruvate. [Average interaction distances: **EcDHDPS:PYR** PYR(O1)-Y133(Oη), 3.6 Å ± 0.2 Å; **EcDHDPS:PYR-SAS**, PYR(O1)-Y133(Oη), 3.4 Å ± 0.1 Å; **1O5K**, PYR(O1)-Y132(Oη), 4.4 Å ± 0.4 Å; **1FDZ**, PYR(O1)-Y137(Oη), 3.1 Å ± 0.2 Å; **1FDY**, HPYR(O1)-Y137(Oη), 3.2 Å ± 0.2 Å; **1F7B**, 4oxo-Neu5Ac(O1)-Y136(Oη), 3.2 Å ± 0.01 Å].

Accepting either the substrate carboxylate or Tyr-130 as hydrogen donor in the SsKDGA complexes results in interaction angles that are either near or below the acceptable limit for a HB (~90°; Table 2.2.7). In the NAL complexes, on the other hand, the angles are favourable in some subunits, but not all, as indicated by the larger

standard deviations. Nevertheless, given the interatomic distances and assuming a certain degree of flexibility in the carboxylate's orientation (alluded to in the crystal structures), HB formation with the conserved tyrosine may be possible in NAL and KDGA. In contrast, both distances and angles are incompatible with such an interaction in the DHDPS complexes analysed (Table 2.2.7).

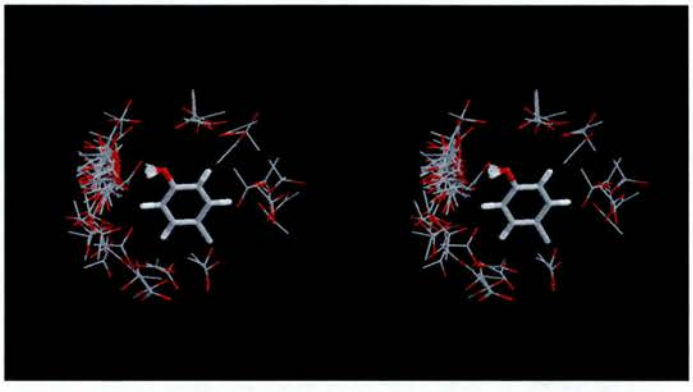
A rigorous assessment of the potential for hydrogen bond formation requires consideration of the relative positioning and orientations of not only the donor/acceptor heavy atoms, but also the electron lone pairs and hydrogen atoms on these groups. This information, however, is not available directly from the protein structures and chemical 'logic' must be applied in order to provide some insights. Another more precise and statistically relevant approach, involving the analysis of equivalent non-bonded interactions in small-molecule crystal structures, is made possible through *IsoStar*, a knowledge-based library (Bruno *et al.*, 1997) provided by the EPSRC's Chemical Database Service (Fletcher *et al.*, 1996).

Using *IsoStar* to generate scatterplots of intermolecular contacts, it was possible to examine the distributions and interactions of carboxylic acid and 4-hydroxyphenyl, as both hydrogen bond donors and acceptors, across hundreds of structures deposited within the Cambridge Structural Database (Allen *et al.*, 1991). The experimental data show a significant preference for interactions forming along the planes of the two groups (Figure 2.2.10). This is consistent with the planar trigonal geometry expected for their lone pairs of electrons, as well as the reduced rotational freedom of their covalently bonded hydrogens.

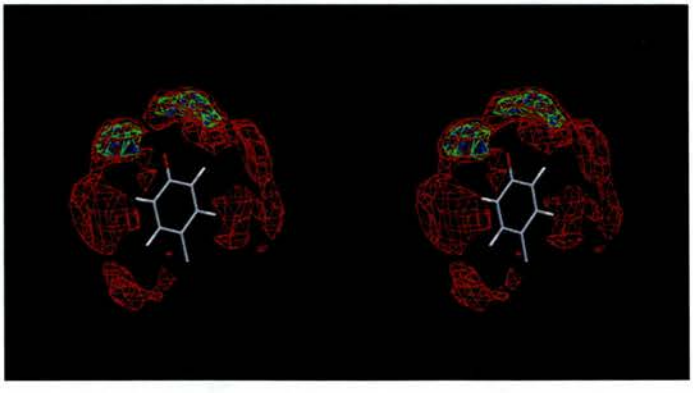
In view of the small molecule data, the probability of hydrogen bond formation between the conserved tyrosines of NAL or KDGA and the carboxylate groups of their substrates in the Schiff base intermediate form is even smaller than if judged

solely on the basis of heavy atom distances and angles. Nevertheless, the observed proximity between the two groups may still be significant. While a route for proton transfer is not apparent from the KDGA structures, more suitable conformations may

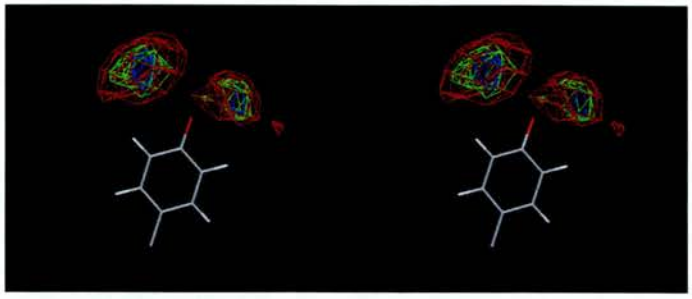
Figure 2.2.10 – *Scatterplots of non-bonded interactions.* Distributions of intermolecular contacts compiled by *IsoStar*. Non-bonded contacts are included up to a maximum distance of 0.5 Å greater than the sum of the van der Waals radii of the atoms involved. Colour key: **carbon**, green; **oxygen**, red; **hydrogen**, white.



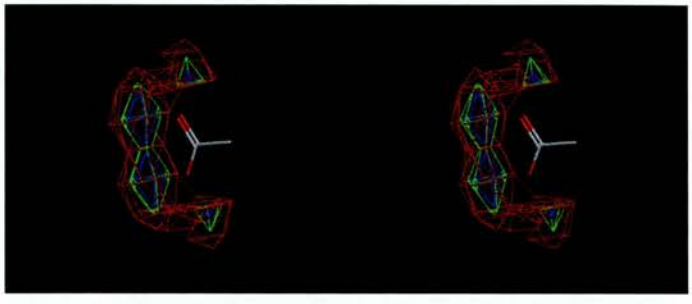
A. Distribution of carboxylate groups around 4-hydroxyphenyl in 57 CSD entries.



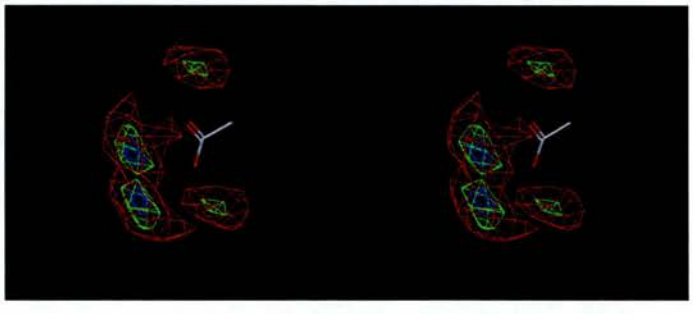
B. Distribution of the oxygen atoms of OH groups around 4-hydroxyphenyl in 610 CSD entries, shown as a contoured density surface. The likelihood that the oxygen of an OH group is found at a given position in space is separated into three levels, corresponding to different colours: **blue** defines the regions where the density of oxygen atoms is three times as high as would be expected by chance; **green**, the regions where the density of contacts is two times higher than the random expectation; **red**, the frequency of contacts is equal to that expected by chance alone.



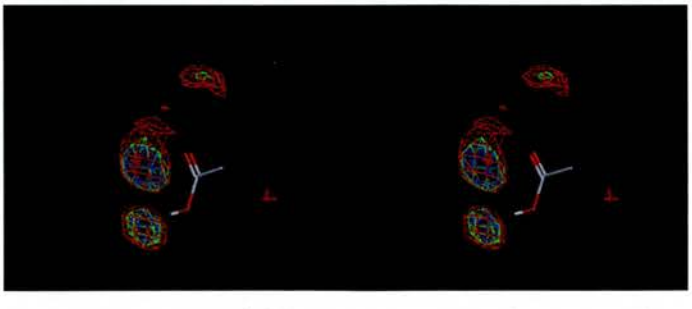
C. Contoured density surface showing the hydrogen atoms of OH groups around 4-hydroxyphenyl. The distribution was derived from the same scatterplot as that used in **B**.



D. Distribution of the hydrogen atoms of phenol OH groups around carboxylate anions in 256 CSD entries.



D. Contoured density surface of the oxygen atoms of phenol OH groups, distributed around carboxylate anions. Generated from the same scatterplot as that used in **D**.



E. Distribution of the oxygen atoms of OH groups around cis-carboxylic acids in 2,490 CSD entries.

be adopted at other stages in the catalytic cycle. Moreover, whether such interactions occur or not, the chemical properties of Tyr-130 are likely to be influenced by the proximity of the carboxylate and Schiff base forming lysine, permitting the tyrosyl hydroxyl to function as both general acid and general base.

2.2.2 Recognition of phosphorylated substrates

S. solfataricus has been shown to be capable of catabolising D-glucose and D-galactose via two alternative branches of its ED-variant pathway (see Chapter 1.1.2 & 1.2.4; Figures 1.1.2 & 1.1.3). In addition to the non-phosphorylative branch that proceeds with no consumption of ATP at the level of six-carbon substrates, there is a second, part-phosphorylative branch, in which phosphorylation occurs after dehydration. Furthermore, it has been established that SsKDGA is responsible for the aldol cleavage step in both branches (Ahmed *et al.*, 2005; Lamble *et al.*, 2005). However, determination of the kinetic parameters for catalysis of D-KD(P)G(al) (Table 2.2.8) has revealed that while SsKDGA recognises both non-phosphorylated and phosphorylated substrates, it displays far greater activity with the latter.

Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat} / K_m ($s^{-1}.mM^{-1}$)
D-KDG †	25.7 (± 1.2)	28.2 (± 1.4)	1.1 (± 0.08)
D-KDGal †	9.9 (± 0.4)	6.8 (± 0.2)	0.7 (± 0.04)
D-KDPG ‡	0.1 (± 0.01)	61.9 (± 0.4)	643
D-KDPGal ‡	0.17 (± 0.01)	34.8 (± 0.2)	207

Table 2.2.8 – Kinetic parameters for SsKDGA. Aldol cleavage was monitored at 60 °C using a coupled assay with the L-lactic acid dehydrogenase from *B. stearothermophilus*.
† Reported in Lamble *et al.*, 2003. ‡ Reported in Lamble *et al.*, 2005.

In an effort to better understand the remarkable promiscuity of SsKDGA and following initial success with D-KDG and D-KDGal, attempts were made to trap complexes of their phosphorylated counterparts in the enzyme's active site. Soaking

Table 2.2.9 - Data collection and refinement statistics. Summary of statistics for the complexes of SsKDGA with D-KDPG and D-KDPGal. Processing and refinement were carried out as described previously and outlined in Appendix I. Abbreviations: **rmsd**, root mean square deviation; **R-stand(F)**, uncertainty in the average structure-factor amplitudes; **DPI**, diffraction component precision indicator; $\mathbf{CC_f}$, correlation coefficient between observed and calculated structure-factors.

	D-KDPG In-house	D-KDPGal In-house	D-KDPGal ID14-2 ESRF
Wavelength (Å)	1.54178	1.54178	0.934
Resolution limits (Å)*	30 – 2 (2.11 – 2.0)	30 – 2.3 (2.42 – 2.3)	41 - 1.9 (2 - 1.9)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å)	a = 83.6 b = 131.8 c = 132.6	a = 83.7 b = 132.4 c = 132.5	a = 83.7 b = 131.7 c = 131.9
No. observations / No. of unique reflections	359,139 / 87,340	335,729 / 65,762	424,840 / 114,572
Mosaicity	0.45	0.8	0.7
Completeness (%)*	88.9 (89.5)	99.6 (98.8)	99.6 (98.8)
R_{merge}‡ (%)*	6.1 (28.5)	10.8 (27.6)	8.1 (28.2)
$< I/\sigma(I) >$*	19.3 (4.3)	13.6 (4.7)	13.0 (3.5)
Multiplicity*	4.1 (3.3)	5.1 (4.4)	3.7 (3.4)
Wilson B (Å^2)	18.9	23.6	20.6
Refinement			
No. of working / test set reflections	78,536 / 8,804	59,169 / 6,582	103,068 / 11,501
Data Completeness (%)	87.8	99.5	99.5
No. of protein / water / ligand atoms	9,473 / 958 / 78	9,300 / 588 / 76	9,356 / 1001 / 68
<B> for protein / water / ligand atoms (Å^2)	16.9 / 28.1 / 28.9	17.1 / 23.3 / 28.4	20.1 / 32.7 / 31.9
R-factor / R-free (%)†	15.8 / 20.8	15.7 / 21.9	16.4 / 21.2
rmsd bond lengths (Å) / bond angles (°)	.012 / 1.9	.019 / 1.7	.020 / 1.7
•*PROCHECK* results (%):			
- Ramachandran plot regions			
core / additionally. allowed / generously. allowed / dissallowed	91.2 / 8.4 / 0 / 0.4	90.9 / 8.7 / 0 / 0.4	91.2 / 8.4 / 0 / 0.4
- Main chain bond lengths / bond angles within limits	100 / 99.4	99.5 / 98	99.4 / 98.6
- Planar groups within limits	100	100	99.5
•*SFCHECK* results:			
- R_stand(F)# / DPI (Å) / CC_F	0.047 / 0.15 / 0.944	0.057 / 0.20 / 0.926	0.053 / 0.12 / 0.908
- radial error in coordinates (Å)	0.19	0.22	0.191
optical resolution (Å)	1.5	1.7	1.5

*, values in parentheses refer to the highest resolution shell. #, R_stand(F) = $\langle\sigma(F)\rangle/\langle F\rangle$

‡, $R_{\text{merge}} = \Sigma_{hkl}\Sigma_i|I_{hkl,i} - \langle I_{hkl}\rangle| / \Sigma_{hkl}\langle I_{hkl}\rangle$ †, Rf and R-free = $(\Sigma\,||F_o| - |F_c||)/(\Sigma\,|F_o|)$

experiments were carried out at 4 °C as described previously, using saturating concentrations of the two substrates. D-KDPG was a kind gift from Professor Eric Toone at Duke University. D-KDPGal was synthesised by Dr Henry Lamble at the University of Bath, as part of the collaborative effort between the research groups of Prof. Taylor, Prof. Danson and Dr Hough. Complexes were obtained with both compounds and the statistics for the respective structures are shown in Table 2.2.9.

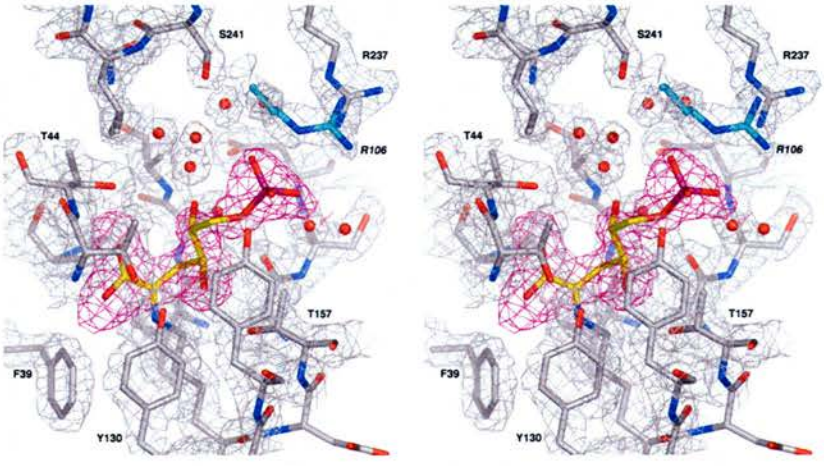
The D-KDPG complex:

A crystal of SsKDGA was soaked in 30 mM D-KDPG for 1 min at 4 °C before flash freezing in liquid nitrogen. Data refined to 2 Å with good statistics (Table 2.2.9) and the difference Fourier maps revealed substrate bound in each of the four subunit active sites of the a.s.u. (Figure 2.2.11**A**). The electron density closely resembles that previously seen for D-KDG, with an additional peak beyond C6 corresponding to the phosphate group. Continuous density extending from K155(Nζ) and a planar trigonal geometry about the C2 position of the substrate are consistent with the covalent Schiff base intermediate form observed in all complexes so far.

Coordinates for the substrate atoms of the D-KDPG Schiff base were incorporated into the model and refined as described previously, using geometric restraints for the (*4S,5R*)-configured intermediate derived from *SYBYL* (Figure 2.2.11**A**, **B**). D-KDPG refined well in the electron density, adopting the same conformation across subunits (rmsd, 0.13 Å). Moreover, superposition of the final model with that of the D-KDG complex revealed that the equivalent atoms of the two substrates occupy similar positions (Figure 2.2.11**C**), with O6 displaying the only significant deviation. The enzyme interactions made by the two substrates are also conserved.

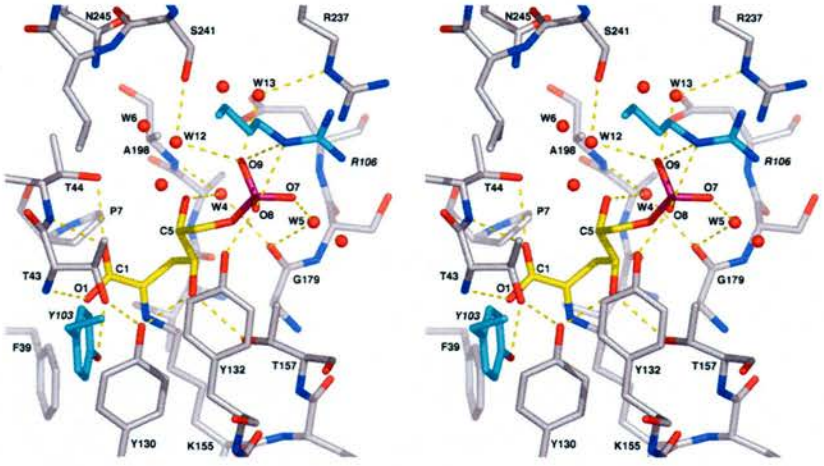
In the refined model the carboxylate group of D-KDPG is positioned within hydrogen bonding distance of the backbone amides of Thr-43 and Thr-44, as well as T44(Oγ). Moreover, the carboxylate O1 atom also interacts with Y130(Oη) (3.3 Å) and F39(Cε) (3.4 Å). The substrate's C4 hydroxyl is within hydrogen bonding distance of Y130(Oη) and T157(Oγ), while the C5 hydroxyl forms water-mediated interactions with the carbonyl oxygen of Gly-179 and amide of Ala-198, via W4. Unlike the C6 hydroxyl oxygen of the non-phosphorylated substrate, however, O6 in D-KDPG makes no contact with the protein or solvent and all additional stabilising interactions take place via the phosphate oxygens (Figure 2.2.11**B**).

Figure 2.2.11 - D-KDPG Schiff base complex. Stereoviews of the SsKDGA active site as stick representations. Residues correspond to any subunits *x* and *y* in an *xy* close dimer (where the catalytic lysine is contributed by *x*).

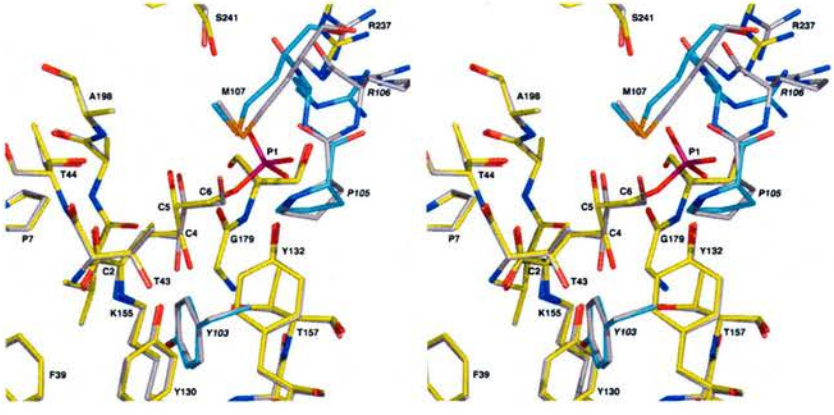


A. Refined model of the enzyme in complex with D-KDPG in the Schiff base intermediate form. Overlayed, the unbiased F_o-F_c map density for D-KDPG contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.5 σ (white).

Colour key: **carbon**, grey (subunit *x*), cyan (subunit *y*) and yellow (D-KDPG); **oxygen**, red; **nitrogen**, blue; **phosphorus**, purple; **water molecules**, red spheres. [Average D-KDPG bond distances/angles: **C1-C2**, 1.5 Å; **C2-C3**, 1.5 Å; **C2-Nζ**, 1.3 Å; **C1-C2-C3**, 120°; **C1-C2-Nζ**, 117°; **Nζ-C2-C3**, 123°; **O1-C1-C2- C3**, -175°; **O2-C1-C2- Nζ**, -175°].



B. Refined model of KDGA in complex with D-KDPG, showing key stabilising interactions as dashes. Atoms are coloured as in **A**. [Average interaction distances: **KDPG(O2)-T44(Oγ)**, 2.6 Å; **KDPG(O2)-T44(N)**, 2.8 Å; **KDPG(O1)-T43(N)**, 2.8 Å; **KDPG(O1)-F39**, 3.4 Å (close contact); **KDPG(O1)-Y130**, 3.3 Å (close contact); **KDPG(O4)-Y130**, 2.5 Å; **KDPG(O4)-T157**, 2.8 Å; **KDPG(O5)-W4**, 3 Å; **KDPG(O5)-T44(Oγ)**, 3.4 Å (close contact); **KDPG(O5)-W6**, 3.4 Å (close contact); **KDPG(O7)-W5**, 2.5 Å; **KDPG(O7)-*R106(Nε)***, 3.8 Å (close contact); **KDPG(O7)-*R106(Nη)***, 3.4 Å (close contact); **KDPG(O7)-R237(Nη)**, 3.6 Å (close contact); **KDPG(O8)-Y132(Oη)**, 2.4 Å; **KDPG(O8)-*R106(Nε)***, 3 Å; **KDPG(O8)-*R106(Nη)***, 3.6 Å (close contact); **KDPG(O9)-*R106(Nε)***, 3.1 Å; **KDPG(O9)-W12**, 2.6 Å; **KDPG(O9)-W13**, 2.7 Å; **T43(Oγ)-*Y103***, 2.6 Å; **T43(Oγ)-Y130**, 2.6 Å; **W4-G179**, 2.7 Å; **W4-A198**, 2.9 Å; **W5-G179**, 3 Å; **W6-N245**, 2.9 Å; **W12-S241**, 2.7 Å; **W13-R237(Nε)**, 2.9 Å.]



C. Superposition of the D-KDG and D-KDPG complexes, showing the similarity in their bound conformations. The different conformations of R237 and $R106_y$ in the two complexes are also shown. Colour scheme: **carbon**, yellow (D-KDPG complex, subunit *x*), cyan (D-KDPG complex, subunit *y*), grey (D-KDG complex, *x* and *y*); **oxygen**, red; **nitrogen**, blue; **phosphorus**, purple; **sulfur**, orange.

The substrate's phosphate group is stabilised by several active site residues, including Arg-106 from the adjacent subunit in the close dimer. The phosphate oxygen O7 interacts with G179(O) via W5, as well as forming close contacts with the guanidinium groups of Arg-106 and Arg-237 [R106(Nε), 3.8 Å; R106(Nη), 3.4 Å; R237(Nη), 3.6 Å). O8 also interacts with Arg-106 [R106(Nε), 3 Å; R106(Nη), 3.6 Å) and is positioned within hydrogen bonding distance of Y132(Oη). Finally, O9 forms water-mediated interactions with S241(Oγ) and R237(Nε), as well as a direct interaction with R106(Nε) (3.1 Å).

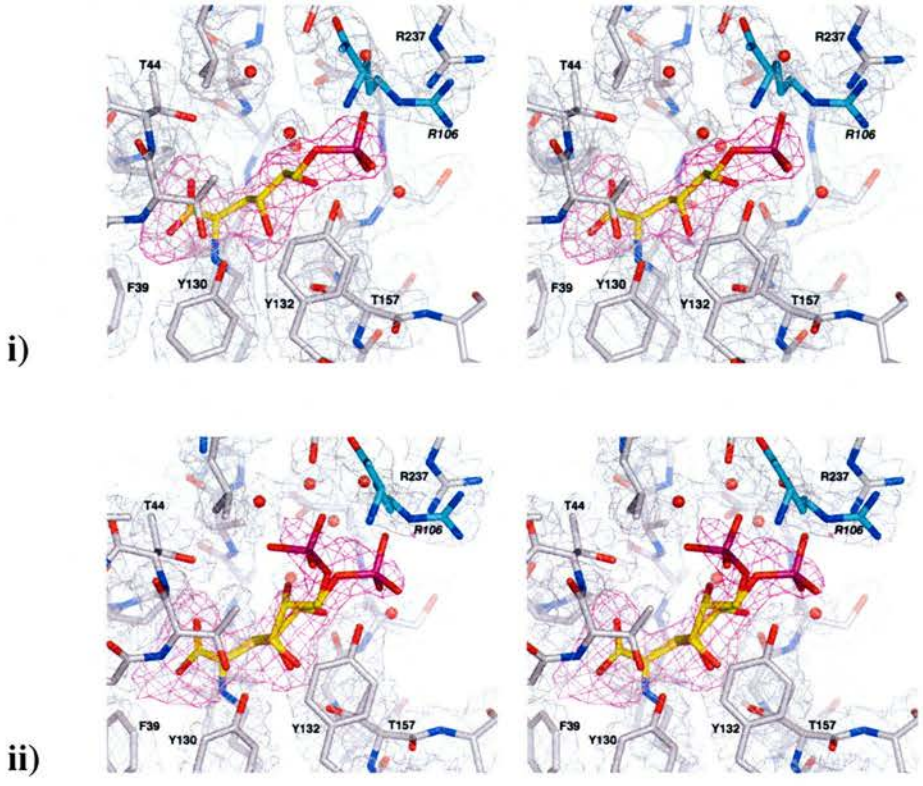
In addition to the potential LBHB between the substrates' C4 hydroxyl and Tyr-130 (mean, 2.5 Å) that has been previously discussed, two other short HB distances were observed in this complex (W5-O7, 2.5 Å; Y132-O8, 2.4 Å). Although both could be artefacts of refinement, the Y132 hydrogen bond may actually play an important role in stabilising the substrate. A similar interaction has also been observed in the 1.8 Å resolution Schiff base complex of RAMA with D-FBP (1zai), involving S271(Oγ) and the P1 oxyanion (2.45 Å). Even in the absence of such a LBHB in D-KDPG binding by SsKDGA, however, the phosphate is likely to be significantly stabilised through ionic interactions with the guanidinio groups of Arg-106 and Arg-237.

The D-KDPGal complex:

An initial complex was obtained using a 5 mM substrate soak and the 2.3 Å dataset collected on an in-house source (Table 2.2.9). The difference Fourier maps calculated during refinement revealed clear continuous density for the Schiff base intermediate of D-KDPGal bound in each active site of the tetramer (Figure 2.2.12**A**). Furthermore, in two of the subunits the electron density indicated that the phosphate

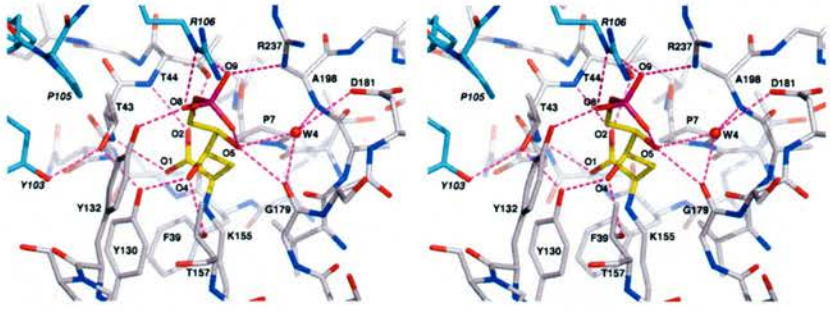
group was present in two alternative orientations (Figure 2.2.12**Aii**). Problems were encountered, however, in selecting a suitable substrate conformation as a starting point for refinement.

Figure 2.2.12 - The D-KDPGal Schiff base complex. Stereoviews of D-KDPGal bound in the active site of SsKDGA. Residues correspond to subunits *x* and *y* in an *xy* close dimer (where the catalytic lysine is contributed by *x*).

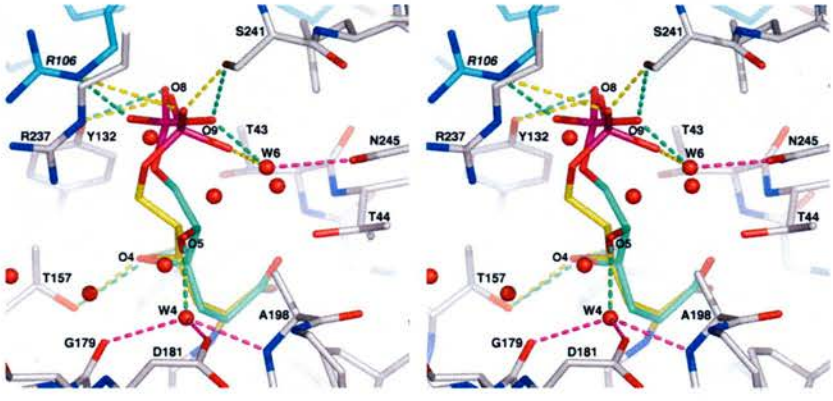


A. Refined model of the enzyme in complex with D-KDPGal in the Schiff base intermediate form. Overlayed, the unbiased F_o-F_c map density for D-KDPGal contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.0 σ (white). In two of the subunits (A and B) the substrate adopted only the *X* conformation (**i**), while in the other two (C and D) conformer *Y* was also observed (**ii**).

Colour key: **carbon**, grey (subunit *x*), cyan (subunit *y*) and yellow (D-KDPGal). [Average D-KDPGal bond distances/angles: **C1-C2**, 1.5 Å; **C2-C3**, 1.5 Å; **C2-Nζ**, 1.3 Å; **C1-C2-C3**, 118°; **C1-C2-Nζ**, 115°; **Nζ-C2-C3**, 122°; **O1-C1-C2- C3**, -162°; **O2-C1-C2- Nζ**, 174°].



B. Refined model of KDGA in complex with D-KDPGal conformer *X*, showing key stabilising interactions as dashes. Atoms are coloured as in **A**. [Average interaction distances: **KDPGal(O2)-T44(Oγ)**, 2.6 Å; **KDPGal(O2)-T44(N)**, 2.8 Å; **KDPGal(O1)-T43(N)**, 2.8 Å; **KDPGal(O1)-F39**, 3.4 Å (close contact); **KDPGal(O1)-Y130**, 3.2 Å (close contact); **KDPGal(O4)-Y130**, 2.7 Å; **KDPGal(O4)-T157**, 2.8 Å; **KDPGal(O5)-G179**, 2.7 Å; **KDPGal(O5)-W4**, 2.7 Å; **KDPGal(O5)-KDPGal(O7)**, 2.6 Å; **KDPGal(O6)-Y132**, 3.4 Å (close contact); **KDPGal(O8)-Y132**, 2.8 Å; **KDPGal(O8)-*R106(Nε)***, 3.1 Å; **KDPGal(O8)-*R106(Nη)***, 3.5 Å (close contact); **KDPGal(O9)-*R106(Nε)***, 2.9 Å; **KDPGal(O9)-*R106(Nη)***, 3.3 Å (close contact); **KDPGal(O9)-R237(Nη)**, 3.2 Å; **T43(Oγ)-*Y103***, 2.7 Å; **T43(Oγ)-Y130**, 2.5 Å; **W4-G179**, 2.8 Å; **W4-A198**, 2.9 Å; **W4-D181**, 3.2 Å].



C. Refined model of KDGA in complex with D-KDPGal conformer *Y* as seen in subunits C (green carbons) and D (yellow carbons). Key substrate interactions observed in each of the active sites are shown as green and yellow dashes, respectively. Interactions between non-substrate atoms are shown in magenta. The conformations of all residues in the two active sites were highly conserved and therefore only active site D is shown (grey carbons, chain D; cyan carbons, chain A). [Average interaction distances: **KDPGal(O5)-W4**, 2.6 Å; **KDPGal(O6)-Y132**, 3.3 Å (close contact); **KDPGal(O7)-*R106(Nε)***, 3 Å; **KDPGal(O8)-Y132**, 3 Å; **KDPGal(O9)-W6**, 2.5 Å; **KDPGal(O7/9)-S241**, 2.8 Å; **W6-N245**, 3.1 Å; **W6-T44**, 3.5 Å (close contact)].

Aided by the 'map-fitting' tools available in *COOT* (version 0.0.33) two configurations were identified for the substrate coordinates displaying the best overall consistency with the unbiased map density. The first (*X*) was observed in all four subunits (Figure 2.2.12**Ai**, **B**), while the second (*Y*) accounted for the additional phosphate position in two of the active sites (Figure 2.2.12**Aii**, **C**). Neither conformation, however, permitted the C5 hydroxyl to be properly accommodated, nor in some cases all three of the phosphate oxygens. Nevertheless, given the strong density (carboxylate and phosphate peaks visible at > 5.5 σ) and its continuity from C1-P1 (visible at > 4 σ) it was considered worthwhile to proceed.

The model was initially appended with only conformer *X* of the substrate and refined against the data using geometric restraints for the (*4R,5R*)-configured intermediate. The results revealed that the phosphate group of *X* is present at approximately half occupancy in all subunits and not only in the two where the maps indicate an additional orientation (C and D). As an alternative position for this group could not be identified in monomers A and B, its partial occupancy must reflect a degree of disorder. While such flexibility is surprising given the observations made in the D-KDPG complex, it is consistent with the alternative binding modes seen in the other two subunits.

The same process was also carried out separately for *Y* in subunits C and D, confirming that this conformer is also present at half occupancy. Both sets of coordinates were subsequently included in the model for a final round of refinement. Occupancies for the phosphate atoms of *X* were set to 0.5 in subunits A and B, while all atoms of *X* and *Y* were set to half occupancy in C and D. Refinement of *Y* alongside *X*, however, resulted in the C4 hydroxyl of the former adopting a position that was inconsistent with the unbiased maps. It was therefore necessary to constrain

this group by defining a single conformation for the substrate atoms from positions C1-C4 and alternative conformers only for positions C5-P1.

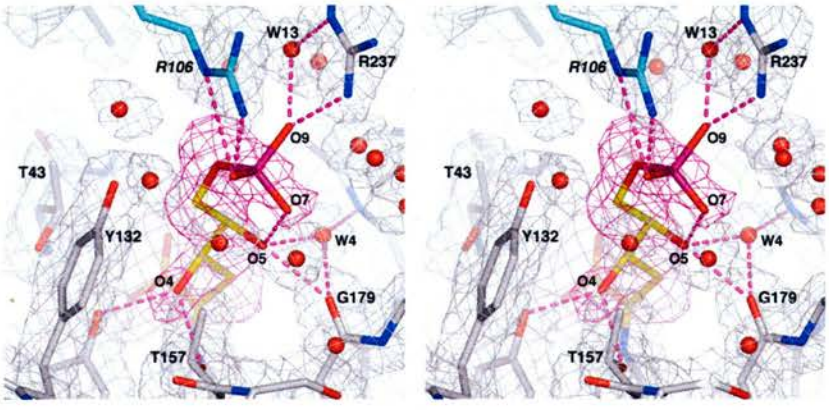
In the final model the substrate's *X* conformation is conserved across the four subunits, displaying only minor deviations (rmsd 0.28 Å), while *Y* is less consistent (rmsd, 0.68; Figure 2.2.12**C**). The α-keto acid moiety occupies its conserved position and engages in the same interactions that have been described previously. Moreover, the atoms around the C4 hydroxyl adopt the configuration observed for the non-phosphorylated diastereomer, although the Y130(Oη)-KDPGal(O4) distance is longer (~ 2.7 Å). Conformations *X* and *Y* of the aldehyde moiety's remaining groups, on the other hand, are distinct from those observed in both the D-KDGal and D-KDPG complexes.

Specifically, in the case of conformer *X* the C5 hydroxyl is within HB distance of Gly-179 (mean, 2.7 Å), as well as forming water-mediated interactions via W4 (Figure 2.2.12**B**). Furthermore, the plane defined by the three phosphate oxygens is oriented at 90° to the guanidinium group of Arg-106 and not parallel to it, as in the case of D-KDPG. In this conformation O9 is stabilised by both arginines, while O8 interacts with Tyr-132 and Arg-106. O7 only forms an intramolecular interaction with O5 (mean, 2.6 Å). In *Y*, on the other hand, the phosphate group points away from the guanidinio groups of the two arginines and the only interaction is between O7 and Arg-106 (Figure 2.2.12**C**). O8 maintains a HB with Tyr-132, while O9 is positioned within hydrogen bonding distance of Ser-241 and forms a water-mediated interaction with Asn-245. In this conformation O5 forms a HB only with W4.

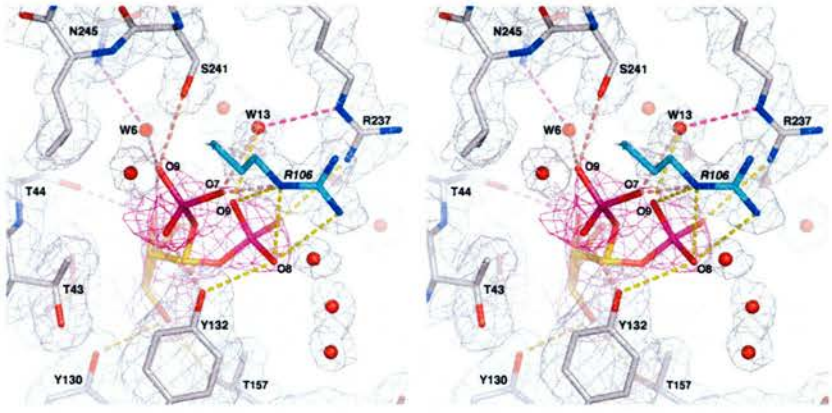
The phosphate group's greater proximity to Arg-106 and Arg-237 would be expected to confer greater stability to the substrate's *X* conformer relative to *Y*. Therefore, the

presence of the two configurations suggests that the former is in some way unfavourable. This could be due to repulsion rather than attraction between KDPGal(O5) and KDPGal(O7). Certainly the poorly defined density for the C5 hydroxyl observed in the unbiased maps does not appear consistent with the multiple interactions it is engaged in. In order to further investigate these features of D-KDPGal binding another complex was obtained using an 8 mM substrate soak and a 1.9 Å dataset was collected at the ESRF in Grenoble (Table 2.2.9).

Figure 2.2.13 - The 1.9 Å D-KDPGal complex. Stereoviews of D-KDPGal bound in the active site of SsKDGA. Residues correspond to subunits *x* and *y* in an *xy* close dimer (where the catalytic lysine is contributed by *x*).
Colour key: **carbon**, grey (subunit *x*), cyan (subunit *y*) and yellow (D-KDPGal); **oxygen**, red; **nitrogen**, blue; **phosphorus**, purple; **water molecules**, red spheres.



A. Active site B in the refined model of SsKDGA, showing D-KDPGal bound as conformer *X*. Key interactions are represented by dashes. Overlayed, the unbiased F_o-F_c map density for D-KDPGal contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.0 σ (white).
[Interaction distances: **KDPGal(O4)-Y130**, 2.7 Å; **KDPGal(O4)-T157**, 2.8 Å; **KDPGal(O5)-G179**, 2.5 Å; **KDPGal(O5)-W4**, 2.6 Å; **KDPGal(O7)-KDPGal(O5)**, 2.4 Å; **KDPGal(O8)-Y132**, 3.6 Å (close contact); **KDPGal(O8)-*R106(Nε)***, 3.2 Å; **KDPGal(O8)-*R106(Nη)***, 2.8 Å; **KDPGal(O9)-W13**, 2.5 Å; **KDPGal(O9)-*R106(Nε)***, 3.4 Å (close contact); **KDPGal(O9)-*R106(Nη)***, 3.6 Å (close contact); **KDPGal(O9)-R237**(Nη), 2.6 Å; **W4-G179**, 2.7 Å; **W4-A198**, 3 Å; **W13-R237**(Nε), 2.8 Å].



B. Active site C in the refined model of SsKDGA, showing D-KDPGal bound as a dual conformer. Overlayed, the unbiased F_o-F_c map density for D-KDPGal contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.0 σ (white). Interactions made by *Y* are shown as green dashes, while those made by the substrate's C4 hydroxyl and the phosphate group of conformer *X'* are represented by yellow dashes. Interactions between non-substrate atoms are shown in magenta.

[Interaction distances for *X'*: **KDPGal(O4)-Y130**, 2.9 Å; **KDPGal(O4)-T157**, 2.7 Å; **KDPGal(O5)-W4**, 2.8 Å; **KDPGal(O7)-R237(Nη)**, 3.4 Å (close contact); **KDPGal(O8)-Y132**, 3.2 Å; **KDPGal(O8)-*R106(Nε)***, 3.5 Å (close contact); **KDPGal(O8)-*R106(Nη)***, 3.4 Å (close contact); **KDPGal(O9)-W13**, 2.9 Å; **KDPGal(O9)-*R106(Nε)***, 3 Å; **W4-G179**, 2.6 Å; **W4-A198**, 2.9 Å; **W13-R237(Nε)**, 3 Å]. [Interaction distances for *Y*: **KDPGal(O5)-T44(Oγ)**, 3 Å; **KDPGal(O5)-W4**, 3.5 Å (close contact); **KDPGal(O7)-*R106(Nε)***, 2.6 Å; **KDPGal(O7)-W13**, 3.1 Å; **KDPGal(O8)-Y132**, 2.5 Å; **KDPGal(O9)-S241**, 3.1 Å; **KDPGal(O9)-W6**, 2.5 Å; **W6-N245**, 3 Å].

The Fourier difference maps calculated during refinement indicated that the substrate intermediate is also present in this structure. Despite the higher resolution and overall quality of the data, however, the D-KDPGal conformation could not be determined in one of the subunits, while in the other three there are inconsistencies in the configurations interpreted from the electron density. Nevertheless, the conformations determined in active sites B and D correspond, respectively, to *X* and *Y* of the lower resolution structure and are stabilised by the same enzyme interactions (Figure 2.2.13**A**). Also, the phosphate atoms in each case are again present at only half occupancy.

Despite some evidence in the unbiased maps there was not sufficient density in either B or D to define alternate substrate configurations and the only active site clearly displaying a dual D-KDPGal conformation is that of subunit C (Figure 2.2.13**B**). However, while one of the two conformers assigned in this case resembles *Y*, the other is not consistent with *X*. Rather the orientations observed for the C5 hydroxyl and phosphate group in this conformation (defined as *X'*) match those of D-KDPG. An occupancy of 0.5 was again determined for each of these conformations and best results during refinement were obtained by defining alternate atoms only from positions C5-P1.

The significance of these inconsistencies in substrate conformation and phosphate orientation across subunits of the two complex structures is unclear. The high salt concentration (> 50%) present as a contaminant in the D-KDPGal sample may have interfered with proper binding, although the electron density could also reflect a mixture of catalytically relevant modes. In any case it is unlikely that the results are due to the nature of the bound molecule. Both mass spectrometry and NMR (^{1}H and ^{13}C) have confirmed that D-KDPG(al) is the major organic species in the sample (results not shown), while a high diastereomeric purity was virtually ensured by the method of synthesis (Lamble *et al.*, 2005).

Briefly, D-KDGal was synthesised from D-glyceraldehyde and pyruvate using SsDKGA. The diastereomer was then separated from D-KDG by Dowex 1X8-formate anion exchange chromatorgraphy and its purity confirmed by HPLC, polarimetry and NMR spectroscopy (Lamble *et al.*, 2003). *S. solfataricus* KDG kinase (SsKDGK) was incubated at 50 °C for 20 h in the presence of D-KDGal and ATP, giving rise to a single product, D-KDPGal. This was subsequently separated from the other components in the reaction mixture by anion exchange

chromatography and selective precipitation in 80% (v/v) ethanol. Consequently, the crystallographic observations must reflect some feature of D-KDPGal binding.

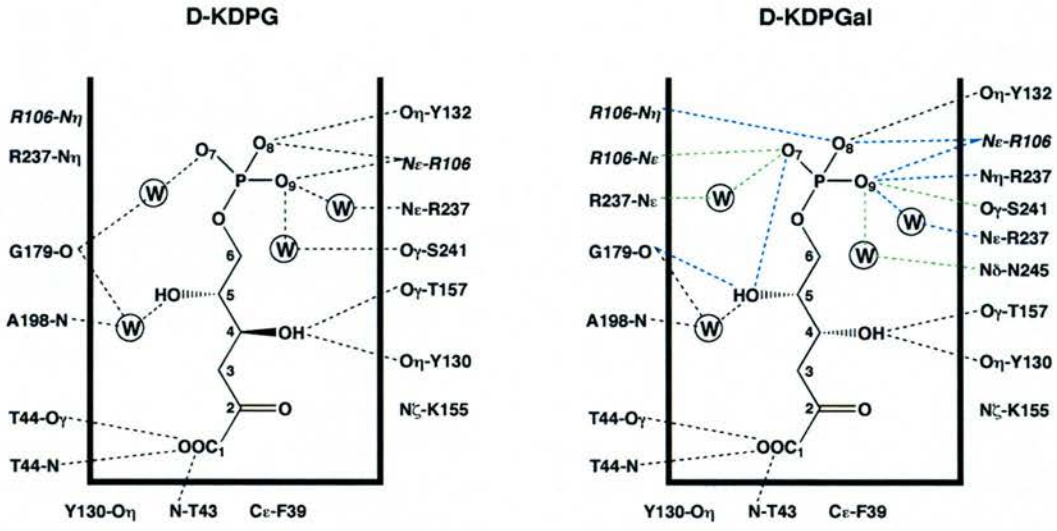


Figure 2.2.14 – Promiscuous binding of phosphorylated substrates. Schematic representation of key interactions involved in stabilising the Schiff base intermediate forms of the two diastereomeric substrates D-KDPG and D-KDPGal in the active site of SsKDGA. Hydrogen bond distances (2.4 – 3.2 Å) are represented by dashes. Protein groups participating in other non-bonded contacts of possible significance are also shown and are positioned nearest to the substrate group they interact with. The catalytic lysine is included for completeness. In the case of D-KDPGal only the most consistently observed interactions are presented. Specifically, HB distances conserved between substrate conformations are represented by black dashes. Those observed in at least 3 of the 6 active sites displaying conformers *X/X'* are shown in blue, while in green are potential hydrogen bonds seen in at least 2 of the 4 subunits in which *Y* was observed.

While a single and conclusive set of stabilising interactions could not be identified for D-KDPGal, the two complex structures implicate many of the same residues that are involved in stabilising the D-KDPG intermediate. In addition to Thr-43, Thr-44, Tyr-130 and Thr-157 involved in the strictly conserved interactions from positions C1-C4, the residues Arg-106, Tyr-132, Gly-179, Ala-198, Arg-237 and Ser-241 are important in stabilising the remaining groups of the aldehyde moiety in both diastereomers. Moreover, as in the case of the non-phosphorylated substrates, ordered water

molecules appear to play a significant role by bridging interactions with protein groups. The crystallographic observations for the two diastereomers are summarised schematically in Figure 2.2.14.

Active site conformational changes:

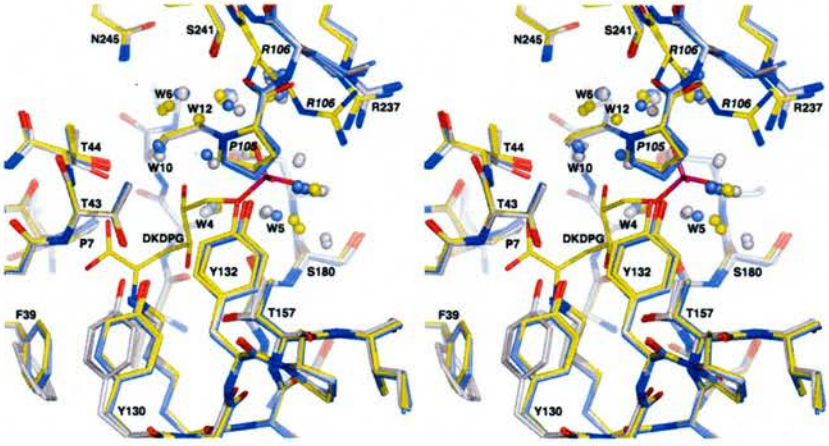
The active site cavity in the D-KDPG(al) complexes described above is highly conserved, with all residues occupying equivalent positions (Table 2.2.10; Figure 2.2.15). A number of differences are observed, however, within the catalytic site of these structures relative to the apoenzyme. Phe-39 and Tyr-130 are rotated about their $\chi 1$ and $\chi 2$ angles, adopting conformations identical to those observed in the pyruvate and D-KDG(al) complexes. Thr-44 also displays a shift, and although minor

	Native 1.7 Å	D-KDPGal 1.9 Å	D-KDPGal 2.3 Å	D-KDPG 2 Å
D-KDPG 2 Å	0.42 / 0.77	0.12 / 0.18	0.12 / 0.20	*0.12 / 0.34*
D-KDPGal 2.3 Å	0.40 / 0.73	0.13 / 0.17	*0.15 / 0.32*	-
D-KDPGal 1.9 Å	0.39 / 0.73	*0.13 / 0.32*	-	-
Native 1.7 Å	*0.14 / 0.43*	-	-	-

Table 2.2.10 – Active site architecture in the phosphorylated substrate complexes of SsKDGA. Comparison of the active sites in the refined models of the D-KDPG and two D-KDPGal complexes, as well as the apoenzyme. Superpositions were carried between active site residues (38-45, 103-111, 130-133, 155-158, 178-182, 236-245) in equivalent subunits of different models, as well as between separate subunits of the same model (in italics; chain A onto B and C onto D). Each cell shows the main chain and all atom rmsd values (Å) on the left and right, respectively.

it, too, appears to be important in binding the α-keto acid moiety's carboxylate group (Figure 2.2.15). The residue's small backbone shift and rotation around $\chi 1$, which is

seen in all of the substrate complexes, results in a T44(Oγ)-Carboxylate(O2) distance of 2.6 Å, instead of ≤ 2.3 Å.



Figirure 2.2.15 – Conformational changes in the active site of SsKDGA. Superposition of the two apoenzyme and five substrate complex structures of SsKDGA. Active site residues are presented as sticks. Selected ordered water molecules are shown as spheres and coloured the same as the carbon atoms of that model. The coordinates for D-KDPG have also been included as representative of the substrate intermediates. Active site conformational changes observed in the substrate complexes include: **i)** the consistent side chain movements of Tyr-130, Phe-39 and Thr-44, associated with binding of the α-keto acid moiety of all substrates; **ii)** the conformational changes in Arg-237 and Arg-106, which permit their guanidinium groups to interact with the D-KDPG(al) phosphates. The superposition also illustrates certain differences in the solvent structure: **iii)** W4 is shifted away from Lys-155 in the complexes, in order to accommodate the substrate groups; **iv)** W6 occupies a different position in the D-KDPG(al) structures, in which it no longer interacts with Ser-241, **v)** W10 is only conserved in the apoenzyme and non-phosphorylated substrate complexes.
Colour key: **carbon**, grey (2 Å and 1.7 Å apoenzyme), slate (pyruvate, D-KDG and D-KDGal complexes), yellow (D-KDPG and 1.9 Å D-KDPGal complexes); **oxygen**, red; **nitrogen**, blue; **phosphorus**, purple.

In addition, the D-KDPG(al) complexes provide further support for a small change in secondary structure in the -GXXG- motif upon binding of the α-keto acid moiety (Appendix 2). In these and previous substrate complexes residues T43-G45 are consistently identified as hydrogen bonded turns by the program *DSSP* (Kabsch and

Sander, 1983). In the two apoenzyme structures, on the other hand, the assignement varies between subunits and frequently the three residues are defined as a 3_{10}-helix. Although superpositions reveal a clear correlation between substrate binding and a change in backbone conformation around these residues, the change is minor. Consequently, it is unclear whether a transition in secondary structure can be considered significant.

Residue	APO 1.7 Å	PYR 1.7 Å	D-KDG 2.1 Å	D-KDGal 2.1 Å	D-KDPG 2 Å	D-KDPGal 2.3 Å	D-KDPGal 1.9 Å
2-294	13.7 / 17.3	17.5 / 20.5	21.5 / 24	18.5 / 22	15.8 / 18	16.2 / 17.9	18.7 / 21.7
A100	0.7 / 0.6	0.8 / 0.7	0.7 / 0.6	0.7 / 0.6	0.7 / 0.6	0.6 / 0.5	0.8 / 0.7
P101	0.8 / 0.7	0.8 / 0.7	0.7 / 0.7	0.8 / 0.7	0.8 / 0.7	0.7 / 0.6	0.8 / 0.7
Y102	0.7 / 1.0	0.8 / 0.9	0.7 / 0.8	0.7 / 0.9	0.8 / 1.0	0.7 / 0.9	0.8 / 1.0
Y103	0.7 / 0.5	0.7 / 0.6	0.7 / 0.6	0.7 / 0.6	0.7 / 0.6	0.7 / 0.6	0.7 / 0.6
Y104	0.7 / 0.6	0.8 / 0.6	0.7 / 0.6	0.8 / 0.5	0.8 / 0.6	0.8 / 0.6	0.8 / 0.6
P105	1.1 / 0.9	1.1 / 0.9	1.0 / 0.9	1.1 / 0.9	1.1 / 0.9	1.2 / 1.0	1.1 / 0.9
R106	1.6 / 2.1	1.4 / 1.9	1.3 / 1.8	1.5 / 2.1	1.4 / 1.4	1.6 / 1.6	1.4 / 1.5
M107	1.4 / 1.1	1.3 / 1.1	1.2 / 1.0	1.4 / 1.1	1.4 / 1.1	1.5 / 1.2	1.4 / 1.1
S108	1.5 / 1.2	1.3 / 1.1	1.2 / 1.1	1.3 / 1.1	1.3 / 1.1	1.3 / 1.2	1.3 / 1.1

Table 2.2.11 – Loop 4 in SsKDGA. Relative B-factors for residues of loop 4 (A100 – M107) in the apoenzyme and substrate complex structures of SsKDGA. Each cell shows the main chain and side chain values on the left and right, respectively. Relative values for each residue were calculated by dividing its average main chain (m.c.) and side chain (s.c.) B-factors by the respective average value for all residues (shown in the second row). Loop 4 appears well ordered in the seven structures, with only P105-S108 displaying m.c. and/or s.c. B-factors higher than the averages in a given model. Moreover, relative values are reasonably consistent across the structures. The only exception is the s.c. of Arg-106, which is characterised by significantly higher B-factors in the apoenzyme and non-phosphorylated substrate complexes, compared to the D-KDPG(al) structures.

In contrast to the above, two additional conformational changes observed in the phosphorylated substrate complexes do not appear in any of the other SsKDGA structures. The first is a small shift in Arg-237 (~ 0.5 Å), which nonetheless displays

clear directionality. In the case of the D-KDPG complex this movement results in the R237(Nη)-Phosphate(O7) distance consistently decreasing from above to below 4 Å (4.2 ± 0.16 Å → 3.6 ± 0.1 Å) and is a clear indication that while the two groups are not within hydrogen bonding distance they do interact electrostatically. This change in side chain position also results in closer interactions between R237(Nη) and the phosphate oxygen O9 of conformer *X* in the D-KDPGal complexes.

The second conformational change unique to the phosphorylated substrate complexes involves Arg-106. This residue is located at the entrance of the active site and in the apoenzyme, pyruvate and D-KDG(al) structures extends out into the tetramer's central channel. While the main chain (m.c.) adopts a consistent conformation in these structures, the side chain (s.c.) of Arg-106 is poorly ordered (Table 2.2.11). In the D-KDPG(al) complexes, on the other hand, the entire residue undergoes a significant shift (m.c. rmsd, 2.0 Å; s.c. rmsd, 3.4 Å) and its side chain is observed in a single, ordered configuration (Figure 2.2.15), in which the guanidinium group is positioned within interacting distance of the substrates' phosphate oxygens.

The significantly improved density and lower B-factors observed for the Arg-106 side chain in the D-KDPG structure are the result of the stabilising interactions formed between the phosphate and guanidinio groups (Table 2.2.11). On the other hand, the residue is characterised by poorer density and somewhat higher B-factors in the D-KDPGal complexes, which is consistent with the partial occupancy of the substrate's interacting conformer (*X*). No significant changes in average B-factors were observed in the main chain from P105-M107, despite significant shifts in atomic positions and a flip in the P105-R106 peptide bond (ψ_{105} = 160 ± 4° → ψ_{105} = -15 ± 6°; ϕ_{106} = 56° ± 6° → ϕ_{106} = -112° ± 4°).

The conformational change in Arg-106 has a significant effect on the shape and size of the active site cavity (Table 2.2.12). In the D-KDPG(al) structures the entire residue is shifted towards the interior of the active site, restricting the mouth opening (mean Δm.a., 40 Å^2; Figure 2.2.16**B**) and reducing the enclosed volume relative to the apoenzyme and non-phosphorylated substrate complexes (mean ΔV, 255 Å^2; Figure 2.2.16**A**). Moreover, the observed backbone shift results in the remote glycerol-binding pocket and its connecting channel becoming restricted (Figure 2.2.16**A**). *CASTp* (Binkowski *et al.*, 2003), in fact, treats this site as a separate cavity in most subunits of the D-KDPG(al) structures (Table 2.2.12).

	APOENZYME 1.7 Å	PYRUVATE 1.7 Å	D-KDG 2.1 Å	D-KDGal 2.1 Å	D-KDPG 2 Å	D-KDPGal 1.9 Å
s.a. (Å^2)	781 ± 13	832 ± 45	885 ± 88	857 ± 48	632 ± 44 (794 ± 47)	629 ± 56 (777 ± 51)
V. (Å^3)	1,134 ± 111	1,173 ± 73	1,243 ± 102	1,164 ± 74	827 ± 38 (958 ± 37)	822 ± 63 (947 ± 57)
m.a. (Å^2)	113 ± 30	110 ± 16	122 ± 24	99 ± 16	73 ± 10	68 ± 12

Table 2.2.12 – The active site cavity. Dimensions of the active site cavity in the apo and complex structures of SsKDGA. Row two presents the internal surface area (s.a.), row three the internal volume (V) and row four the surface area of the cavity's mouth opening (m.a.). The values shown correspond to the protein cavity's molecular surface, i.e. that which is 'seen' by the surface of a probe with a 1.4 Å radius (Richards, 1977). Measurements were carried out for 'close' subunit pairs (AD, BC) using the *CASTp* server and mean/standard deviation (s.d.) values calculated for the four subunits of a given model. The conformational changes observed in the D-KDPG(al) complexes result in restricted access to the remote glycerol binding site. Consequently, the program treated this pocket as a separate entity in almost all subunits of the two models. The 'Connolly' surface calculated by *PyMOL* (DeLano, 2002), on the other hand, consistently revealed a channel linking the remote binding site to the larger cavity (see Figure 2.2.16). To allow for this discrepancy mean and s.d. values were calculated for the active site cavity on its own, as well as for the cavity + the remote pocket. The two sets of values are shown in the relevant columns outside and inside brackets, respectively.

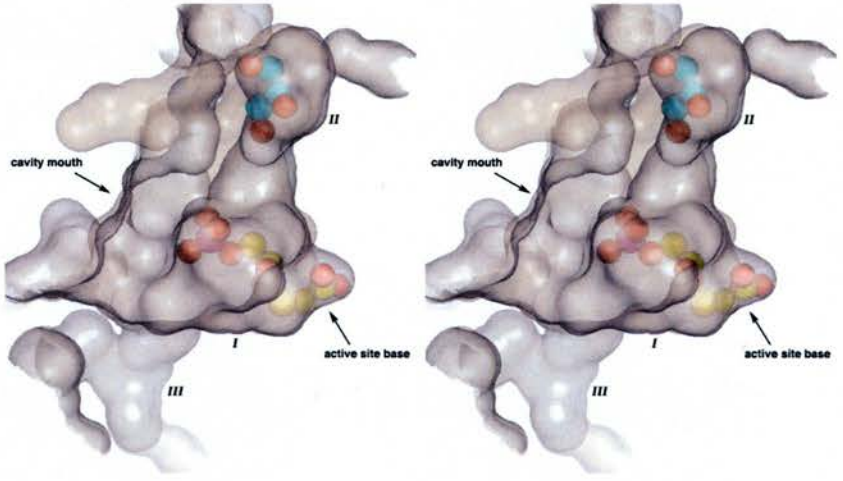
Given these changes to the active site cavity it is not surprising that glycerol was not found in the remote pocket in any of the D-KDPG(al) complexes, despite being consistently present in the other structures. Significant to the observations, however, is also the fact that the crystals were soaked with substrate before transferring to the cryoprotectant. Superpositions of the different structures reveal steric clashes that would make the backbone shift at Arg-106 unfavourable in the presence of glycerol at the remote site. Although it would be interesting to see what effect this might have on substrate binding, attempts made to test the hypothesis by changing the order in which the molecules were introduced to the protein were unsuccessful.

In addition to the clear constriction of the remote pocket in the D-KDPG(al) complexes, the active site's accessibility also appears to be significantly reduced. This can be demonstrated by considering the dimensions of the enzyme's substrates. Crystallographically derived estimates for the molecular volumes of pyruvic acid (100 $Å^3$; Refcode, PRUVAC) and glycerol (110 $Å^3$; GLCROL), as well as the pyranose form of glucose-6-phosphate (sodium salt, 250 $Å^3$; DACHEU) were retrieved from the CDS using *CONQUEST* (Bruno *et al.*, 2002). These volumes correspond to sphere diameters of 5.8 Å, 5.9 Å and 7.8 Å respectively and are consistent with free movement of substrates into the active site in the apoenzyme, as well as the non-phosphorylated substrate structures (circle diameter for cavity entrance, ~12 Å).

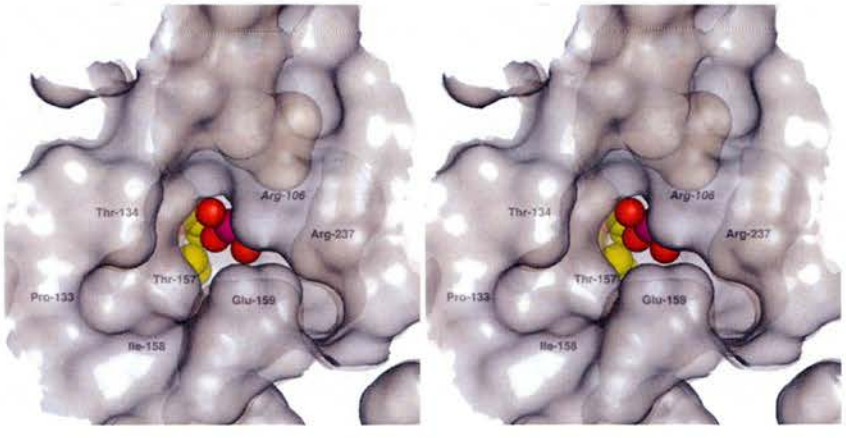
Despite the conformational change in Arg-106, the cavity's mouth in the D-KDPG(al) complexes (outlined in Figure 2.2.16**B** by residues T134, P133, I158, E159 R237 and R106) also retains a large surface area (circle diameter, 9.5 Å). However, Arg-106 causes a more significant narrowing near Thr-157, that further restricts access to the interior. While the cavity's length at this point is approximately 9 Å, it is only 4 Å at

its widest point (using van der Waals radii from Bondi, 1964). Given the size and

Figure 2.2.16 – Changes to the active site cavity. Semitransparent surface representation of the active site region of SsKDGA in the apoenzyme (wheat) and D-KDPG complex structures (grey). The Schiff base intermediate of D-KDPG is also shown in space-fill mode (yellow carbons), as is a glycerol molecule bound at the remote site (cyan carbons). Glycerol molecules have been consistently observed in this pocket in the apoenzyme, pyruvate, D-KDG and D-KDGal structures, but not in the phosphorylated substrate complexes.



A. Side view with key features highlighted. In addition to the main cavity (*I*), there is the remote glycerol binding site (*II*), the channel leading to which is restricted in the phosphorylated-substrate complexes, and a second inconsistently observed pocket (*III*).



B. View of the entrance to the active site cavity, showing the constriction that results from the conformational change of Arg-106.

irregular shape of the active site's entrance resulting from this constriction, as well as the intimate association of the phosphate group of bound substrates with Arg-106, it is unlikely that substrate/product exchange can occur and the active site in these complexes can therefore be thought of as being in a 'closed' state.

Main chain and side chain shifts have also been observed in the active site of RAMA (St-Jean *et al.*, 2005) and shown to promote binding of the P1 and P6 phosphate groups of D-FBP. Moreover, in the ultra-high resolution structures of EcDERA Heine, *et al.* (2004) have described a 'breathing' of the phosphate binding site, which they suggested is required for substrate capture and release. The observed motions involve both backbone and side chain groups and include partial unwinding of an α-helix N-terminus. The authors have shown that this loss of local secondary structure in the substrate complexes is compensated for by interactions formed with the bound phosphate group.

The exchange of stabilising interactions described for EcDERA presumably maintains a balance between the two conformations, facilitating the ready transition from one to the other during the course of the catalytic cycle. However, a similar mechanism could not be established in SsKDGA, despite crystallographic evidence that the conformational changes in Arg-106 are required for the binding and release of phosphorylated substrates and products, respectively. Instead, the 'closed' configuration observed in the presence of phosphorylated substrates appears to be significantly more favourable. In addition to the greater stability that the phosphate group interactions confer to the residue's side chain, the shift in the main chain at this position does not appear to affect its stability, despite a small change in secondary structure (Appendix 2).

Arg-106 is positioned close to the C-terminal end of loop 4 (A100-S108), which is involved in several intersubunit interactions at the AD(BC) interface, including the highly conserved $Y130_x$-$T43_x$-$Y103_y$ hydrogen bond network. In the apoenzyme, and non-phosphorylated substrate structures Ala-100, Pro-101, Tyr-102, Tyr-104, Met-107 and Ser-108 are assigned as irregular elements by *DSSP*. Moreover, three bends are identified (regions of high curvature, i.e. > 70°), centred on residues Tyr-103, Pro-105 and Arg-106. In the D-KDPG(al) complexes the secondary structure assignment in this region remains the same for all residues except Arg-106, which is defined as an irregular element.

Although the backbone conformational change observed in these complexes (ψ_{105} = 160 ± 4° → ψ_{105} = -15 ± 6°; ϕ_{106} = 56° ± 6° → ϕ_{106} = -112° ± 4°) results in a decrease in curvature around Arg-106, it is not accompanied by a loss of hydrogen bonding that might balance the increase in side chain interactions (as in the case of EcDERA (Heine *et al.*, 2004)). Indeed, in both conformations the main chain atoms of P105-M107 are involved in vary few consistent hydrogen bonds. The most significant change in the D-KDPG(al) structures is the exchange of a water mediated interaction between R106(O) and glycerol for one between the residue's carbonyl and S241(N). These observations do not provide a mechansism by which the apparent energetic barrier to phosphorylated product release might be overcome. Consequently, the possibility was considered that the conformation of Arg-106 observed in the apoenzyme and non-phosphorylated substrate structures is biased by the presence of glycerol at high concentrations and that another more stable configuration might otherwise be adopted. To test this hypothesis 2-methyl-2,4-pentanediol (MPD) was used as an alternative to glycerol. Despite being prone to rapid degradation in the presence of this molecule, protein crystals were successfully cryoprotected by rapid

transfer from 2 %, to 5% and then 10% MPD before flash freezing. A dataset was collected to 2 Å in-house and processed with good statistics (data not shown).

Refinement of the data against the apoenzyme model revealed no significant changes in the enzyme's overall structure. Also, consistent with its larger size relative to glycerol (MPD M_r, 118; glycerol M_r, 92), and the lower concentration used (MPD, 0.85 M; glycerol, 2.2 M) MPD was not observed in any of the active sites. Most importantly, however, an identical conformation was observed for the main chain of Arg-106 and its surrounding residues, despite the fact that the remote binding pocket was empty in all subunits. B-factor analysis further supports the previous conclusions that there is no significant difference in stability of the backbone of loop 4 between the 'open' and 'closed' conformations.

The only other plausible 'triggers' that can be put forward for the conformational change in Arg-106 necessary for product release, involve either entropic effects or charge repulsion between its side chain and that of the Arg-237. In its open configuration the side chain of Arg-106 is characterised by greater flexibility and solvation than when it is interacting with the substrate intermediates. Therefore, a shifting balance during the catalytic cycle between entropy loss, desolvation and the formation of electrostatic interactions between the guanidinium and phosphate groups may provide a driving force for conformational change.

The phosphate group's relative proximity to the guanidinio groups of Arg-106 and Arg-237 may also affect the stability of the closed conformation and may explain the poorer ordering of the Arg-106 side chain in the two D-KDPGal complexes. While the phosphate of D-KDPG and D-KDPGal conformer *X* can provide a counter-charge to both guanidinium groups, the same is not true of D-KDPGal conformer *Y*, which is only able to interact directly with Arg-106. Therefore a change in substrate

conformation during the catalytic cycle may result in greater repulsion between the two arginine side chains and trigger a transition of Arg-106 to the open configuration.

Phosphate binding in type I aldolases:

SsKDGA is the only identified NAL subfamily enzyme known to utilise phosphorylated substrates and consistent with its unique activity, appears to recruit a poorly conserved subset of active site residues in phosphate binding. Of the six identified in the D-KDPG(al) complexes (Figure 2.2.14) only Gly-179 is strictly maintained across the subfamily. Tyr-132, Arg-237 and Ser-241, on the other hand, are all mutated to unrelated residues (Appendix 3). Asn-245 is also not conserved in the amino acid sequences of subfamily members, although another asparagine side chain does occupy a close position in the tertiary structures of *E. coli* and *T. maritima* DHDPS.

Arg-106 is reasonably conserved in the NAL subfamily sequences, the only exception being EcNAL, in which the residue is mutated to a proline. In EcDHDPS an arginine is maintained at that position, while in TmDHDPS, MtDHDPS, NsDHDPS and HiNAL it is replaced by a lysine (Appendix 3). Moreover, in the three-dimensional structures these residues are also located on loop 4 near the active site entrance. The positions they occupy, however, are significantly different than both the open (Figure 2.2.17) and closed configurations of Arg-106 and do not appear to be consistent with an involvement in binding substrate intermediates.

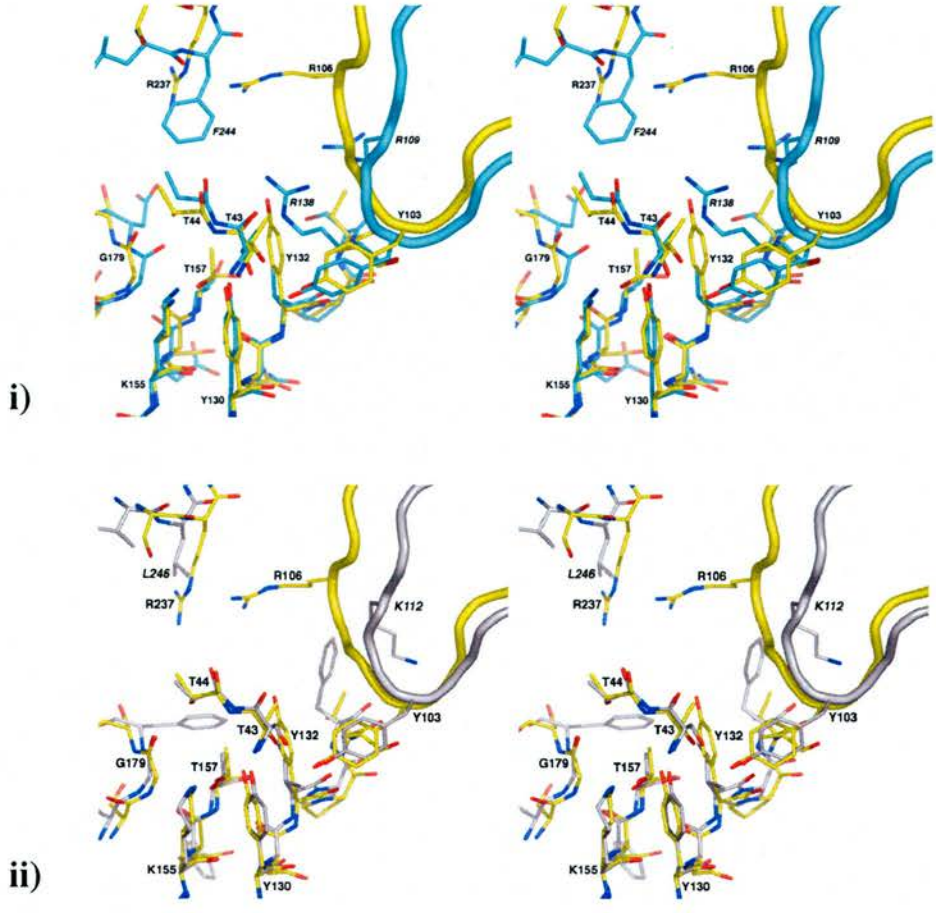


Figure 2.2.17 – Active site superpositions of NAL subfamily enzymes. Superposition of the active site region in the apoenzyme structures of SsKDGA, and **i)** EcDHDPS (PDBid, 1yxc), or **ii)** HiNAL (1f5z). The active site residues of a given subunit *x* (in an *xy* close dimer) are shown in stick representation. Loop 4 of the corresponding subunit *y* in each model is displayed as a ribbon, with key side chains represented by sticks. EcDHDPS and HiNAL residue labels are in italics. Colour key: **cabon**, yellow (SsKDGA), cyan (EcDHDPS), grey (HiNAL).

While Arg-106 extends out over the cavity mouth and can reach into the active site, its equivalents form one side of the entrance (Figure 2.2.17). Moreover, the positions observed for these residues are conserved across subunits and models, despite being characterised by high solvent accessibility and B-factors (Table 2.2.13). The only significant conformational change observed is a side chain shift in Lys-112 in the substrate analogue complexes of HiNAL (1f73, 1f74, 1f7b). This change, however, is not comparable to that of Arg-106 in the D-KDPG(al) structures and, in any case,

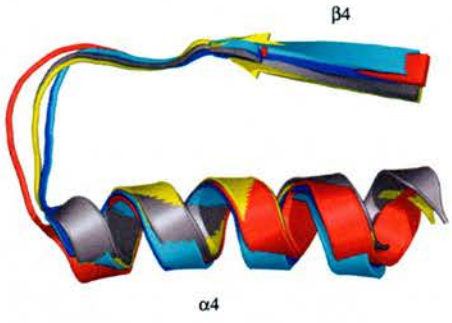
appears to be the result of unusual distortions and/or disorder in residues 136-147 that may not be relevant to normal enzyme function.

		HiNAL-APO (1F5Z)	HiNAL 4D-SIA (1F74)	HiNAL-APO (1F6K)	HiNAL-4OXO-SIA (1F7B)	EcDHDPS-APO (1YXC)	TmDHDPS-PYR (1O5K)	EcDHDPS-APO (1DHP)	NsDHDPS-APO (1XXX)
	i - 2	19.8/20.9	14.3/13.8	22.1/23.6	16.9/17.4	17.2/19.7	23.0/28.9	20.9/25.6	26.9/29.0
	i - 1	19.1/17.0	13.0/10.9	20.6/19.4	16.0/12.9	17.5/16.6	22.5/22.3	21.6/20.3	26.2/26.3
Bf (Å^2)	*i*	22.6/28.9	16.5/24.5	24.6/30.0	20.4/27.8	18.1/29.3	23.3/35.0	23.1/35.2	26.5/33.4
	i + 1	19.3/16.8	12.8/9.9	20.4/17.4	17.7/15.3	18.6/18.1	22.8/25.0	22.2/20.9	26.8/27.0
	i + 2	18.8/20.1	12.1/12.5	19.7/20.6	17.6/17.7	18.1/17.3	23.1/21.9	20.8/21.0	25.7/25.3
Rmsd (Å)	(*i*-4)-(*i*+4)	0.21/0.29	0.18/0.90	0.25/0.31	0.25/0.96	0.08/0.12	0.07/0.1	0.11/0.30	0.06/0.07
	(*i*-4)-(*i*+4)	0.47/0.77		0.30/0.69		0.22/0.37		0.31/0.62	

Table 2.2.13 – Loop 4 in NAL and DHDPS. Average B-factors for loop 4 residues (*i* – 2) to (*i* + 2) of DHDPS and NAL, where *i* corresponds to K112 in HiNAL, R109 in EcDHDPS, K108 in TmDHDPS and K119 in NsDHDPS. Main chain (m.c.) and side chain (s.c.) values are shown on the left and right in each cell, respectively. The last two rows contain the m.c./s.c. rmsd's for the region around *i*. Deviations between subunits in a given model are presented in row 7 and between equivalent subunits of different models, in row 8. Abbreviations: **APO**, apoenzyme; **4D-SIA**, 4d-Neu5Ac; **4OXO-SIA**, 4oxo-Neu5Ac; **PYR**, pyruvate.

Although also affected by tetrameric assembly, the difference in positioning of Arg-106 relative to its equivalent residues in DHDPS and NAL is clearly apparent at the level of secondary structure (Figure 2.2.18**A**). Superpositions and sequence alignments reveal that while strand β4 and helix α4 are well conserved, the intervening loop is longer in SsKDGA than in the other structurally characterised members of the subfamily (Figure 2.2.18**B**). This observation is interesting in view of the fact that hyperthermophilic enzymes often display shorter loops than their mesophilic counterparts (Russell and Taylor, 1995; Vieille and Zeikus, 2001; Yamagata *et al.*, 2001). In fact, all other loops in SsKDGA are either the same length or shorter than their equivalents in EcNAL , HiNAL and EcDHDPS (Walden, 2001).

A.



B.

```
                        β4      L4          α4
SsKDGA          ...KDFDIVGIASYAPYYYPRMSEKHLVKYFKTLCEVSP-HPVYLY... 130
HiNAL           ...TELGYDCLSAVTPFYY-KFSFPEIKHYYDTIIAETG-NNMIVY... 136
EcNAL           ...KRYGFDAVSAVTPFYY-PFSFEEHCDHYRAIIDSADGLPMVVY... 137
EcDHDPS         ...NDSGIVGCLTVTPYYN-RPSQEGLYQHFKAIAEHTD-LPQILY... 133
TmDHDPS         ...EKLGANGVLVVTPYYN-KPTQEGLYQHYKYISERTD-LGIVVY... 132
MtDHDPS         ...AAEGAHGLLVVTPYYS-KPPQRGLQAHFTAVADATE-LPMLLY... 143
                          ..    *:*. . :.  . .  :. : . .
```

Figure 2.2.18 – Loop 4 differences between NAL subfamily enzymes.

A. Superposition of the region from strand β4 to helix α4 in five NAL subfamily members, demonstrating the varitation in length of the intervening loop. Colour code: SsKDGA (red), TmDHDPS (cyan), EcDHDPS (blue), EcNAL (grey), HiNAL (yellow).

B. Segment of a multiple sequence alignment of six NAL subfamily members, showing the residues corresponding to β4 (turquoise) and α4 (green) in each enzyme. Arg-106 and its equivalent residues are highlighted in bold type. The degree of conservation relative SsKDGA is shown below each residue: (*), strictly conserved; (:), similar residue in all other enzymes; (.) similar or identical residue in at least two other enzymes. The alignment was carried out using *CLUSTALW* (version 1.82; (Chenna *et al.*, 2003)) and secondary structure was determined by *DSSP* (Kabsch and Sander, 1983). Abbreviations: **HiNAL**, *H. influenzae* NAL; **EcNAL**, *E. coli* NAL; **EcDHDPS**, *E. coli* DHDPS; **TmDHDPS**, *T. maritima* DHDPS; **MtDHDPS**, *M. tuberculosis*.

Loop 4 may be longer in SsKDGA by only one amino acid, however, this appears to have a significant effect on the position adopted by Arg-106 relative to the active site entrance and consequently, on the ability of this arginine to interact with bound substrates. Moreover, the presence of an additional residue within the loop is likely to provide greater backbone flexibility, facilitating the conformational changes that are

believed to be associated with substrate binding and release. In view of the evidence it would be reasonable to suggest that the increase in length of loop 4 represents a specific adaptation by SsKDGA towards improved binding of phosphorylated substrates.

A comparison was subsequently carried out between SsKDGA and four representatives of the type I aldolase family that are specific for phosphorylated substrates; *O. cuniculus* FBPA (RAMA), *T. tenax* FBPA, *E. coli* DERA and *T. maritima* KDPGA (see Chapter 1.2.3). Initial analysis of the residues involved in phosphate binding and their position within the secondary structure revealed that only Gly-179 and the Tyr-132 position are conserved in one or more of the other enzymes (Table 2.2.14). A glycine located on the loop that is preceded by strand β7 (loop 7) is involved in all cases, while strand β5 and the loop that follows it (loop 5) also make a contribution in RAMA and TtFBPA, respectively.

Although additional similarities are apparent in Table 2.2.14 between SsKDGA and the other aldolases they do not reflect true structural homology. β4 and loop 4 may also play a role in the two FBP aldolases, however, they do not represent contributions from an adjacent subunit, as in the case of the *Sulfolobus* enzyme. Moreover, the involvement of α9 in SsKDGA cannot be considered as being conserved in TtFBPA, RAMA and TmKDPGA. This is because in these enzymes there is a helix positioned before the first β-strand and consequently their ninth helix is equivalent to α8 in SsKDGA.

SsKDGA	TtFBPA	RAMA	EcDERA	TmKDPGA
	Glu-28 (α2′)	**Ser-35 (β1 - α2′)**		
	His-29 (α2′)	**Ser-38 (α2′)**		
	Tyr-114 (β4 - α5)	**Lys-107 (β3)**		
Arg-106 (β4 - α4)	**Arg-148 (β5 - α6)**	Asp-109 (β3 - β3′)	Gly-171 (β6 - α7)	
Tyr-132 (β5 - α5′)	**Lys-179 (β6)**	Lys-146 (β4)	Lys-172 (β6 - α7)	**Gly-157 (β7 - α8)**
Gly-179 (β7 - α7)	Arg-232 (α9)	Arg-148 (β4)	Gly-204 (β7 - α8)	**Gly-158 (β7 - α8)**
Arg-237 (α9)		Arg-42 (α2′)	**Gly-205 (β7 - α8)**	**Val-159 (β7 - α8)**
Ser-241 (α10)	**His-25 (α2′)**	Glu-189 (β5)	Val-206 (β7 - α8)	**Ser-179 (α9)**
Asn-245 (α10)	**Gly-204 (β7 - α8)**	**Ser-271 (β7 - α8)**	**Ser-238 (β8 - α9)**	
	Gly-231 (β8 - α9)	**Gly-272 (β7 - α8)**	Ser-239 (β8 - α9)	
	Arg-232 (α9)	**Gly-302 (β8 - α9)**		
	Asn-233 (α9)	**Arg-303 (α9)**		

Table 2.2.14 – Phosphate binding in type I aldolases. Residues involved in binding substrate phosphate groups in SsKDGA, TtFBPA, RAMA, EcDERA and TmKDPGA. Residues that interact directly with the substrate are shown in bold type, while all others are involved in water-mediated interactions. SsKDGA-R106 (highlighted in red) represents the only contribution from an adjacent subunit. In the case of the FBPA enzymes the residues in the top and bottom columns are involved in binding the P6 and P1 oxyanions, respectively (St-Jean *et al.*, 2005; Lorentzen *et al.*, 2005). The residues of TmKDPGA (highlighted in yellow) have been proposed on the basis of their observed interactions with a sulfate ion (Fullerton *et al.*, 2006). The positions of the residues in the secondary structure of each enzyme (*DSSP* assignment) are shown in brackets. Loops are identified by the flanking secondary structure, e.g. β7-α8 corresponds to the loop between strand 7 and helix 8.

Compared to the *Sulfolobus* aldolase, phosphate binding is significantly more conserved between RAMA, TtFBPA, EcDERA and TmKDPGA in terms of both the position and nature of the residues involved. While this may appear surprising given their low sequence identity (Table 2.2.15) the enzymes are in fact closely related in structure. Type I aldolases are all classified as part of the same structural family (*SCOPid*, 51570; Murzin *et al.*, 1995) and considered to share a common evolutionary ancestor (*CATHid*, 3.20.20.70; Orengo *et al.*, 1997). More importantly, however, phosphate binding within type I aldolases is an ancestral feature and typically involves a motif found around β7 that comprises primarily of backbone interactions

and includes at least one glycine (Copley and Bork, 2000; Nagano *et al.*, 2002; Heine *et al.*, 2004).

	TmDKDPG (1WA3)	EcDERA (1JCL)	TtFBPA (1ok4)	RAMA (1ZAI)	SsKDGA D-KDPG
SsKDGA	15.6 (0.75)	12.9 (0.88)	15.5 (0.54)	16.1 (0.75)	-
RAMA	17.9 (0.53)	16.4 (0.63)	19.8 (0.52)	-	-
TxFBPA	17.8 (0.71)	18.8 (0.81)	-	-	-
EcDERA	17.8 (0.25)	-	-	-	-

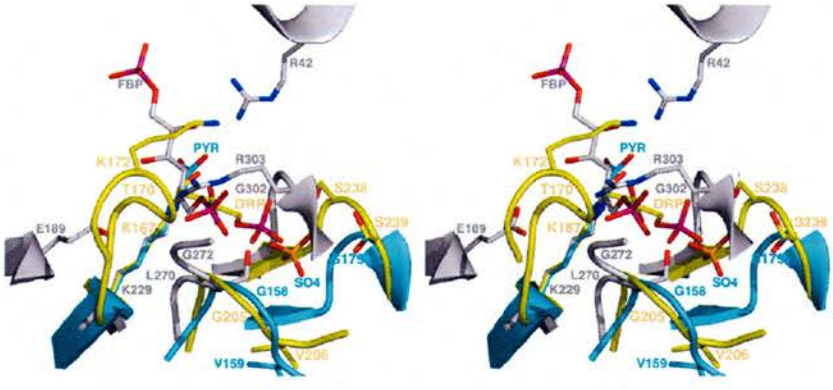
Table 2.2.15 – Comparison of type I aldolases. Global sequence identity between the type I aldolases SsKDGA, TmKDPGA, EcDERA, TtFBPA and RAMA. Also shown in brackets are the rmsd values (Å) between enzyme active sites, calculated using all atoms of two residues: **i)** the Schiff base forming lysine (SsKDGA-K155; TmKDPGA-K129; EcDERA-K167; TtFBPA-K177; RAMA-K229) and **ii)** a residue in the loop preceded by strand β7 (KDGA-G179; KDPGA-T156/-*G157*; DERA-A203/-*G204*; TtFBPA-S202/-*G203*; RAMA-L270/-*S271*). The smallest rmsd's were obtained by superimposing KDGA-G179 with each of the underlined residues. All other enzymes were superimposed using the second residue (in italics). Sequence alignments were carried out using *LALIGN* version 2.0 (Huang and Miller, 1991). Superposiitons were carried out in *LSQKAB*, using the enzyme models shown in brackets in the first column.

This phosphate-binding motif is conserved on loop 7 of SsKDGA as well as the other four aldolases. Moreover, its position relative to the catalytic centre also appears to be maintained between the five enzymes. This was demonstrated by superposition of the active sites using the Schiff base forming lysine and a suitable residue within loop 7 (Table 2.2.15). The resulting rmsd values are all comparable and do not appear to correlate with the sequence identity between the enzyme pairs. Consequently, it was possible to use the structurally conserved loop as a frame of reference in considering the positions of other phosphate binding residues (Figure 2.2.19).

In the superimposed models of RAMA, TtFBPA, EcDERA and TmKDPGA most of the phosphate-binding residues occupy similar positions relative to the Schiff base forming lysine (Figure 2.2.19**A**). Moreover, the substrate phosphate groups and the

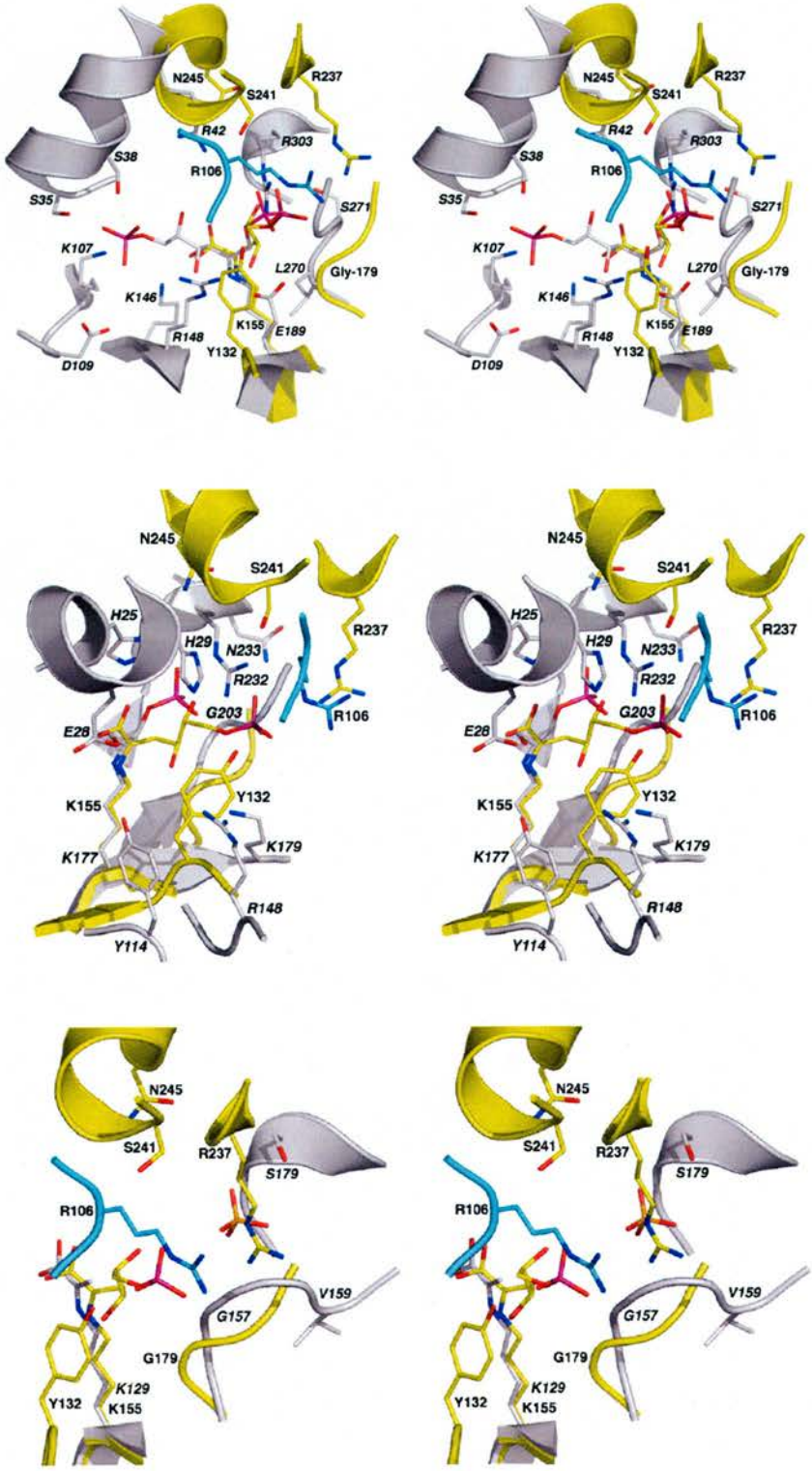
sulfate ion that were trapped in these structures superimpose well, all located between the backbones of loop 7 and loop 8 (or helix α9). In contrast, SsKDGA displays significant differences towards each of these aldolases (Figure 2.2.19**B**). With the exception of Tyr-132 (loop 5), which superimposes well with Glu-189 and Arg-148 of RAMA and TtFBPA, respectively, none of the enzyme's other residue positions appear to be structurally conserved. Even loop 7 has a distinct conformation and lacks the characteristic bulging associated with the phosphate-binding positions (Copley and Bork, 2000).

Figure 2.2.19 – Structural homology between the phosphate binding motifs of type I aldolases. Stereoviews of the superimposed active sites of *S. solfataricus* KDGA (D-KDPG complex), *T. tenax* FBPA (1ok4), *O. cuniculus* FBPA (RAMA; 1zai), *E. coli* DERA (1jcl) and *T. maritima* KDPGA (1wa3). In addition to the Schiff base forming lysine and bound substrate, all other residues shown as sticks are involved in phosphate binding (Table 2.2.14). Surrounding residues and all glycines are in cartoon representation.

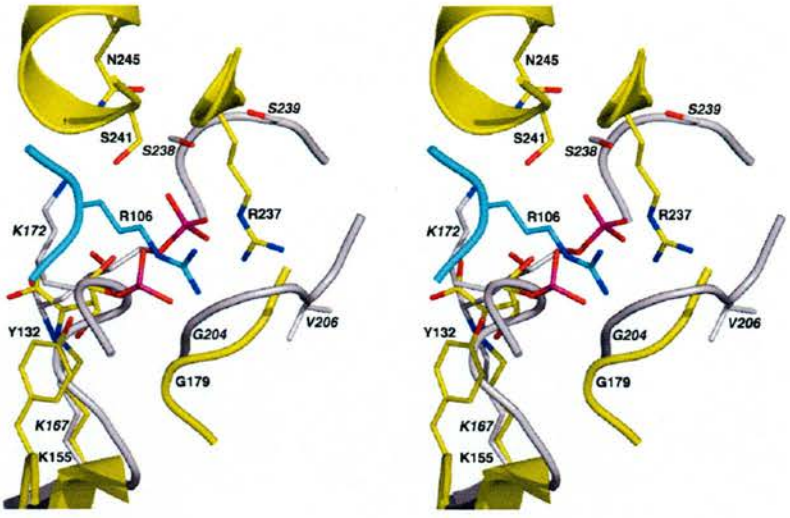


A. Comparison of the phosphate binding motifs of RAMA (grey carbons), EcDERA (yellow carbons) and TmDKDPG (cyan carbons). TmKDPGA and RAMA were superimposed onto EcDERA using only two residues: **i)** the Schiff base forming lysine (RAMA-K229; DERA-K167; KDPGA-K129) and **ii)** the first residue involved in phosphate binding on the loop preceded by strand β7 (RAMA-S271; DERA-G204; KDPGA-G157). The resulting rmsd's are shown in Table 2.2.15. In the case of RAMA only residues involved in binding the P1 phosphate group are shown. The residue labels match the colour of the carbon atoms in the respective model. Although omitted for clarity, TtFBPA also superimposed well.

B. Comparison of the phosphate binding motifs of SsKDGA and **i)** RAMA, **ii)** TtFBPA, **iii)** TmDKDPG and **iv)** EcDERA. SsKDGA residues are drawn with yellow carbons, except for the intersubunit contribution from loop 4 (P105-M107), which is in cyan. All residues of the other three enzymes have grey carbons. Labels corresponding to RAMA, TtFBPA, TmKDPGA and EcDERA are italicised.



iv)



The observations reveal a distinct mode of phosphate binding in SsKDGA compared to the other four aldolases. In this enzyme the substrate phosphate group is positioned on the opposite side of loop 7 and is stabilised primarily by side chain interactions with non-conserved secondary structure elements, including a unique contribution from an adjacent subunit. These represent fundamental differences with respect to the ancestral phosphate-binding function of aldolases (Copley and Bork, 2000; Nagano *et al.*, 2002) that appear even more significant when viewed alongside the fact that the relevant residues in SsKDGA are not conserved amongst its closest neighbours (NAL, DHDPS).

Copley and Bork (2000) have suggested on the basis of their phylogenetic studies that NAL and DHDPS may have arisen from an ancestor having the conserved phosphate-binding site and that this function was subsequently lost. In such a case the observations for SsKDGA may reflect the reappearance of a phosphate binding within the subfamily. The evidence presented above supports this hypothesis, although the study was too limited for any conclusions to be drawn. Consequently, while not within the scope of this project, a more extensive and rigorous comparison of

phosphate binding in SsKDGA relative to type I and type II aldolases, as well as other TIM-barrel enzymes, may be warranted.

2.2.3 The $P2_12_12$ crystal form

Crystallisation experiments with SsKDGA have identified two crystal forms that are characterised by diffraction properties (in-house, ≤ 2.5 Å) suitable for structural studies (see Chapter 2.1.2). In the soaking experiments described in previous chapters use of $P2_12_12_1$ crystals, grown at around pH 6.5, resulted in the consistent trapping of Schiff base complexes with natural substrates. However, SsKDGA also crystallises at acidic pH (~ 4.3) in space group $P2_12_12$, offering the possibility of comparing substrate binding in this enzyme under physiological and non-physiological conditions.

During initial crystallographic studies of SsKDGA a dataset of the $P2_12_12$ crystal form was solved by molecular replacement against the MAD derived model, revealing two tetramers per asymmetric unit related by an $X = 0$, $Y = 0.04$, $Z = 0.5$ translation (Walden, 2001; see Chapter 1.2.4). However, the structure was subsequently abandoned during refinement because of the persistence of high values for the Rf and R-free (25% and 28% respectively), as well as difficulties encountered in fitting certain regions of the model (particularly loop 4) into the electron density. Although the conformational changes and disorder observed were accounted for in terms of hydrogen bonding disruptions within the low pH environment, re-examination of the structure of SsKDGA in this crystal form, both in its apoenzyme and substrate-bound states, has provided a very different explanation.

Table 2.2.16 - Data collection and refinement statistics. Summary of statistics for the $P2_12_12$ apoenzyme and 'pyruvate soak' (PYR-1, -2) structures. Processing and refinement were carried out as described previously and outlined in Appendix 1. Abbreviations: **rmsd**, root mean square deviation; **R-stand(F)**, uncertainty in the average structure-factor amplitudes; **DPI**, diffraction component precision indicator; $\mathbf{CC_f}$, correlation coefficient between observed and calculated structure-factors.

	APO ID14-2 ESRF	PYR-1 ID14-2 ESRF	PYR-2 ID14-2 ESRF
Wavelength (Å)	0.934	0.934	0.934
Resolution limits (Å)*	55 – 2 (2.11 – 2.0)	57 – 2.0 (2.11 – 2.0)	57 – 2.3 (2.42 – 2.3)
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit cell dimensions (Å)	a = 135.3 b = 135.8 c = 189.1	a = 135.5 b = 136.3 c = 189.8	a = 135.6 b = 136.6 c = 189.8
No. observations / No. of unique reflections	806,285 / 221,537	872,484 / 234,710	681,186 / 135,835
Mosaicity	0.5	0.2	0.5
Completeness (%)*	95.0 (91.4)	99.5 (99.3)	90.7 (90.0)
R_{merge}‡ (%)*	5.4 (15.1)	5.7 (16.7)	8.6 (27.5)
$< I/\sigma(I) >$*	17.9 (6.6)	16.0 (6.1)	17.1 (4.6)
Multiplicity*	3.6 (3.1)	3.7 (3.7)	5.0 (5.0)
Wilson B (Å^2)	24.9	26.0	29.1
Refinement			
No. of working / test set reflections	210,432 / 11,104	222,941 / 11,768	122,280 / 13,553
Data Completeness (%)	94.6	99.4	90.3
No. of protein / water / ligand atoms	18,702 / 1,373 / 50	18,817 / - / 40	18,600 / 734 / 58
<B> for protein / water / ligand atoms (Å^2)	23.7 / 31.8 / 38.8	25.1 / - / 41.5	20.9 / 21.7 / 23.6
R-factor / R-free (%)†	25.5 / 31.5	22.6 / 26.2	19.7 / 26.0
rmsd bond lengths (Å) / bond angles (°)	.022 / 2.0	.025 / 2.0	.023 / 1.9
•*PROCHECK* results (%):			
- Ramachandran plot regions			
core / additionally. allowed / generously. allowed / disallowed	90.5 / 8.5 / 0.6 / 0.4	91.2 / 7.8 / 0.4 / 0.6	91.6 / 8.0 / 0 / 0.4
- Main chain bond lengths / bond angles within limits	97.4 / 95.6	96 / 96.4	98.3 / 97.2
- Planar groups within limits	99	98.5	99.8
•*SFCHECK* results:			
- R_stand(F)# / DPI (Å) / CC_F	0.034 / 0.19 / 0.858	0.041 / 0.15 / 0.925	0.051 / 0.23 / 0.894
- radial error in coordinates (Å)	0.32	0.27	0.33
optical resolution (Å)	1.6	1.6	1.7

*, values in parentheses refer to the highest resolution shell. #, R_stand(F) = $\langle\sigma(F)\rangle/\langle F\rangle$

‡, $R_{merge} = \sum_{hkl}\sum_i |I_{hkl,i} - \langle I_{hkl}\rangle| / \sum_{hkl}\langle I_{hkl}\rangle$ †, Rf and R-free = $(\Sigma\,||F_o| - |F_c||)/(\Sigma\,|F_o|)$

The apoenzyme structure at pH 4:

A pH 4 crystal of SsKDGA was cryoprotected by equilibration against mother liquor containing 20% (v/v) glycerol and then flash frozen in a nitrogen stream at 100 K.

The crystal diffracted to 2 Å on a synchrotron source (ESRF, Grenoble) and data were collected over 90° using half-degree oscillations. Processing was then carried out as described previously (see Chapter 2.1.3; Appendix 1) resulting in the statistics shown in Table 2.2.16 (*Apo* column). The crystal was found to be isomorphous with that studied by Hendry *et al.*, (2000) and Walden (2001), each of the cell dimensions deviating by only 0.2% or less. In addition, the observed distribution of axial reflections (*h00* and *00l*) was consistent with the expected space group ($P2_12_12$), while calculation of a native Patterson also confirmed the pseudo-centring relationship reported previously (data not shown).

α	β	γ	X	Y	Z	CC_F	Rf	CC_I
135.24	79.69	92.55	0.2431	0.2664	0.3745	22.1	52.7	24.1
135.24	79.69	92.55	0.2430	0.2281	-0.1253	51.8	48.1	47.5

Table 2.2.17 – Molecular replacement. A tetramer search model was generated from the $P2_12_12_1$ native apo structure. Data from the pH 4 crystal were used within the resolution range of 10 - 4 Å and the cross-rotation function search sphere for the Patterson correlation was set to 48 Å. The *AMoRe* rotation and translation searches were carried out sequentially, resulting in an initial solution shown in the second row. This solution was subsequently fixed and the translation function repeated, resulting in a second solution shown in the third row. Rigid body refinement of the two solutions together resulted in an R-factor of 44% and correlation coefficient of 52%. Abbreviations: **α β γ**, rotation solution in Eulerian angles; **X Y Z**, fractional unit cell coordinates for translation solution; **CC_F**, correlation coefficient between observed and calculated structure factor amplitudes; **Rf**, R-factor between observed and calculated amplitudes; **CC_I**, correlation coefficient for intensities.

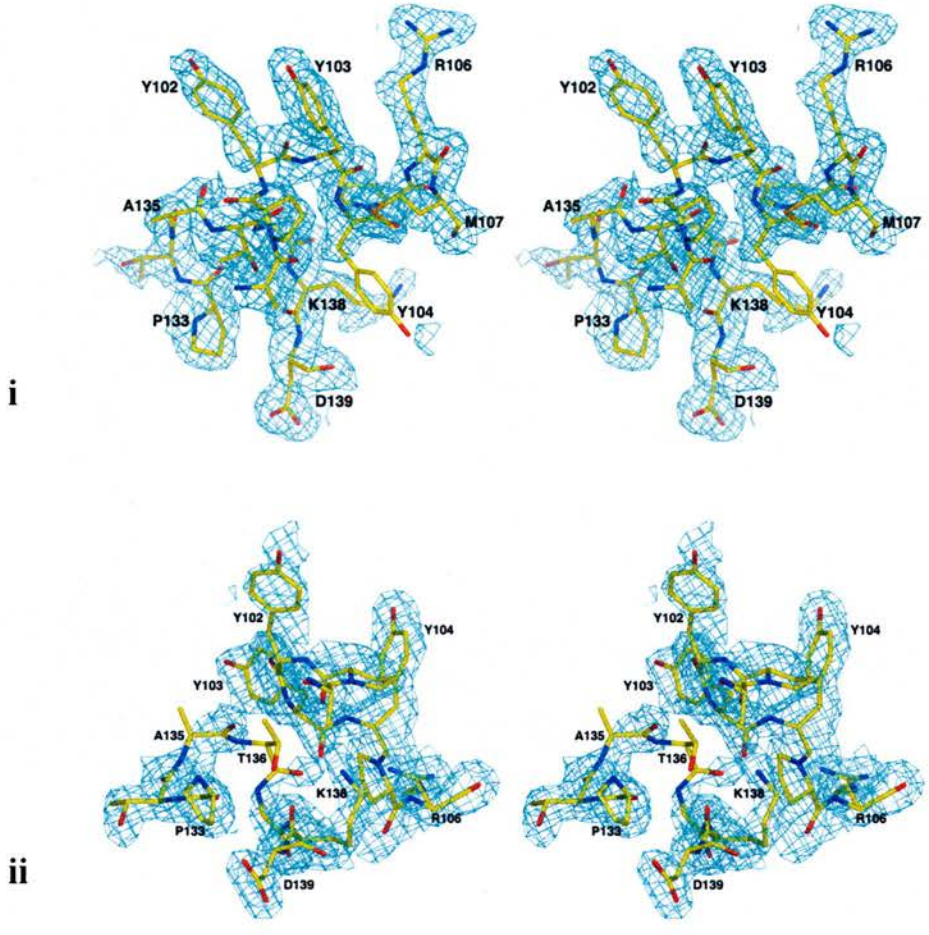
Molecular replacement was again carried out in *AMoRe* using a tetramer search model generated from the $P2_12_12_1$ native apo structure (see Chapter 2.1.3). Rotation and translation solutions were identified sequentially for each of the two tetramers in the asymmetric unit, the values indicating that the two molecules are related by the

expected 0, 0.04, 0.5 translation and are oriented in the same way (Table 2.2.17). Despite being characterised by poor statistics (Rf, 44%; CC_F, 52%) these results were consistent with those obtained by Walden (2001), indicating that the correct solution had been found. Consequently, the two sets of rotation/translation operations were applied to the search model to generate starting coordinates for refinement.

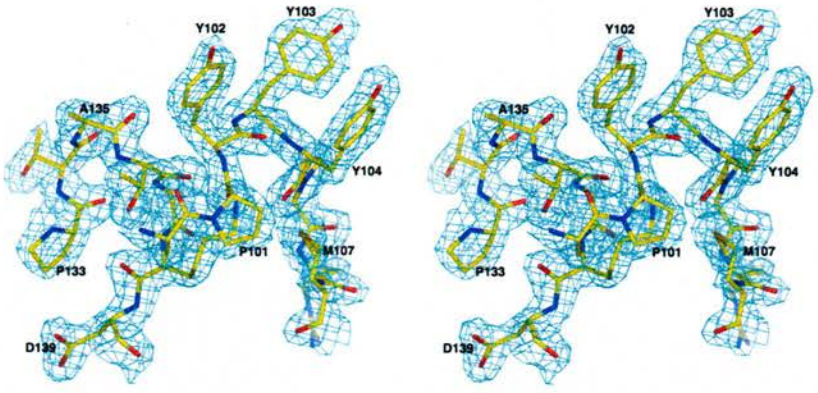
For the purposes of cross-validation a random list of reflections in the pH 4 crystal data were assigned to a *test* set using the program *FREERFLAG* (CCP4, 1994). The coordinates output from *AMoRe* were then refined against the data using *REFMAC5* as before (see Chapter 2.1.3; Appendix 1). Despite an Rf of 32% and Rfree of 36% in the early stages of restrained refinement, inspection of the maps revealed that most of the model was in agreement with the experimental data. However, significant errors were associated with loop 4, the region preceding helix α5 (helix α5′ and loop α5′-α5; Appendix 2) and points of contact between symmetry related molecules.

The model was subjected to iterative cycles of refinement and manual adjustments in *COOT*, which along with solvent building in *ARP/wARP* resulted in the Rf and Rfree dropping to 25.5% and 31.5%, respectively. Although these values are considered quite high, they do not indicate over-interpretation of the data. Moreover, other statistics for the final model are reasonably good and so is its overall fit to the electron density, although some problems could not be resolved. Consistent with previous results (Walden, 2001), two alternative conformations were identified for loop 4, in each of which most, but not all, of the residues could be fit into the maps. Also, some residues in the region between strand β5 and helix α5 (loop 5) were found to be disordered (Figure 2.2.20).

Figure 2.2.20 – A region of increased disorder in the $P2_12_12$ apoenzyme structure. Stereoviews of the residues in loop 4 and the region between strand β5 and helix α5 in the apo structure of SsKDGA at pH 4 (**A**). Also shown for comparison, the equivalent region as observed in the $P2_12_12_1$ structures (**B**). Colour key: **carbon**, yellow; **oxygen**, red; **nitrogen**, blue; **sulfur**, orange.



A. Representative views of P101-M107 and P133-D139 from two subunits in the $P2_12_12$ apoenzyme structure. Overlayed, the $2F_o$-F_c map density at the end of refinement (contoured at 1.0 σ). These residues display conformational differences and varying degrees of order across the eight subunits of the a.s.u. Moreover, residues P101-P105 of loop 4 can be modelled into the electron density in two distinct arrangements, each of which was used in one of the tetramers of the a.s.u. In the first conformation (shown in **i**) the side chain of Tyr-104 appears disordered in most of the subunits, while the other residues of the loop are in good agreement with the observed density. On, the other hand, in the second arrangement all residues are accommodated by the density, but are characterised by a poorer fit (**ii**).



B. Residues P101-M107 and P133-D139 in the SsKDGA apoenzyme structure at pH 6. Overlayed, the $2F_o$-F_c density for the residues contoured at 1.5 σ. With the exception of the Arg-106 side chain the region is well ordered in this crystal from, with all residues occupying highly conserved positions.

	Monomer	AC/BD	AD/BC	Tetramer
$P2_12_12_1$ - Apo	0.20 / 0.21	0.24 / 0.25	0.23 / 0.24	0.14 / 0.14
$P2_12_12$ - Apo	0.38 / 0.29	0.39 / 0.31	0.39 / 0.30	0.53 / 0.30
$P2_12_12_1$ *vs.* $P2_12_12$	0.76 / 0.47	0.89 / 0.65	1.10 / 0.91	1.11 / 0.93

Table 2.2.18 – Comparison of the crystal structures of SsKDGA at pH 4 and 6. Superpositions of the $P2_12_12_1$ and $P2_12_12$ apoenzyme models of SsKDGA involving one, two or four subunits. The values on the left and right in each cell correspond to the rmsd's (Å) for the Cα's of residues 2-293 and 2-100/141-293, respectively. Positions 101-140 include both regions that display the most significant deviations between the two crystal forms. They were therefore excluded from some calculations in order to assess their contribution to the overall rmsd values. In addition to the tetramer structures (column 5), individual subunits were also superimposed (column 2), as were equivalent 'loose' (column 3) and 'close' (column 4) dimers. For the $P2_12_12$ structure (row 2) tetramer rmsd's were calculated between the two molecules of the a.s.u. and the higher value for the superposition of residues 2-293 reflects the modelling of alternative conformations for loop 4 in each tetramer. In the case of the $P2_12_12_1$ structure, tetramer rmsd's were calculated between the two native models (2 Å and 1.7 Å).

In the $P2_12_12$ apoenzyme structure the residues of loops 4 and 5 are characterised by conformational variability across the a.s.u., which can be demonstrated by superposition of equivalent subunits in the final model either excluding the relevant regions (P101-I140; 0.29 Å) or using all residues (rmsd, 0.38 Å; Table 2.2.18).

Moreover, the two loops make a significant contribution to the overall difference in monomeric structure between the pH 4 and pH 6 crystal forms. This is indicated again by the rmsd's from subunit superpositions with (0.76 Å) and without (0.47 Å) the relevant regions included (Table 2.2.18).

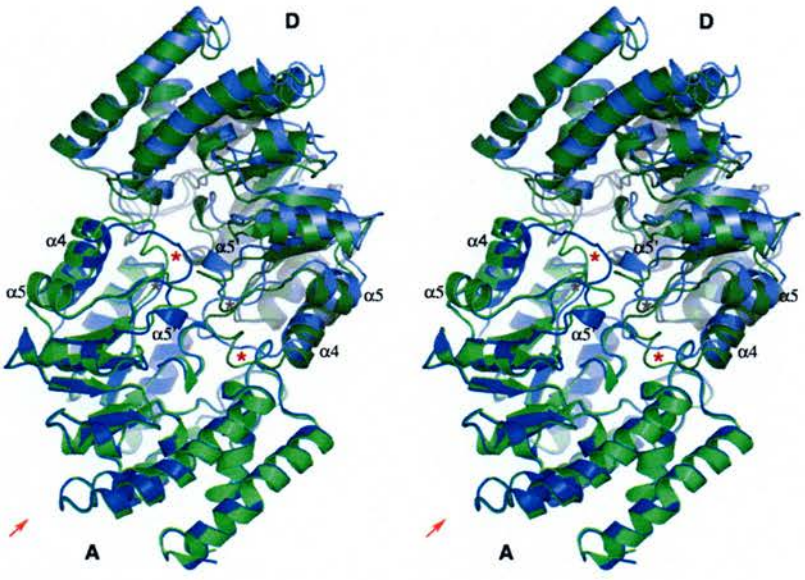
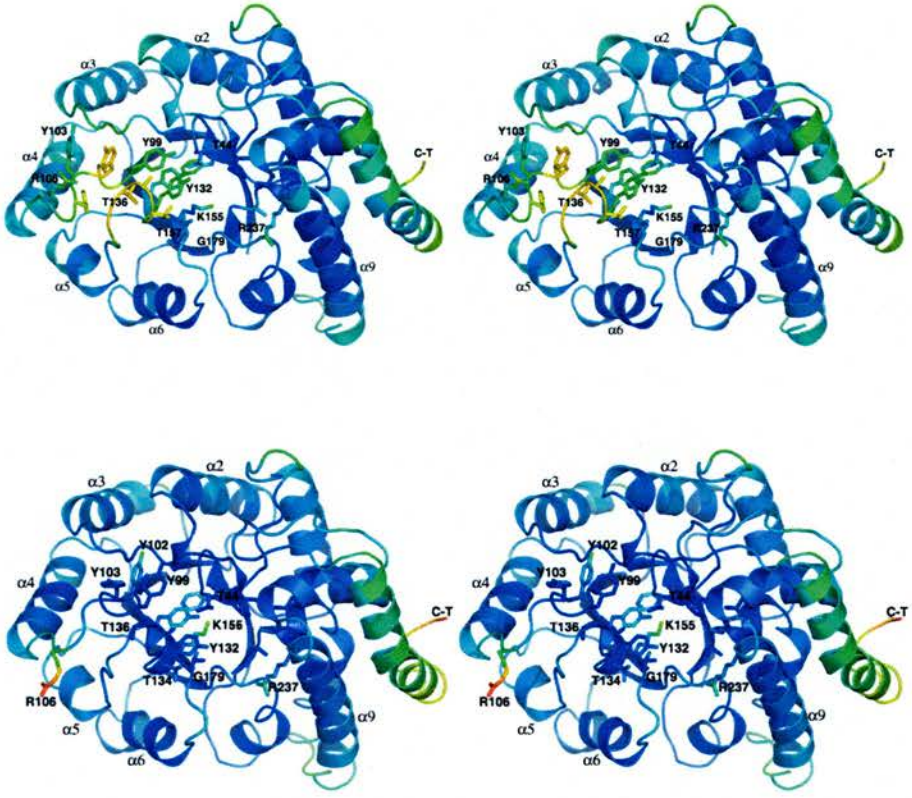


Figure 2.2.21 – The AD interface. Superposition of the $P2_12_12_1$ and $P2_12_12$ apoenzyme structures of SsKDGA using residues 2-100 and 141-293 of monomer A. Subunits A and D of two models are shown in cartoon representation. Loop 4 and the region between β5 and α5 in each subunit are indicated with a star (loop 4, red star; β5-α5, grey star). The location of the AC interface is shown by a red arrow. Colour key: **$P2_12_12_1$(A)**, blue; **$P2_12_12_1$(D)**, slate; **$P2_12_12$(A)**, light green; **$P2_12_12$(D)**, dark green. The superposition highlights three main points: **i)** the only significant changes in monomer structure between the two crystal forms involve helix α4, loop 4 and the region between β5 and α5. This includes loss of secondary structure at α5′; **ii)** these regions are located at the AD interface and affect the relative positioning of the two subunits; **iii)** the AC interface is not significantly affected.

Owing to their position at the AD(BC) interface loops 4 and 5 have an effect not only on the monomeric but also the oligomeric structure of the enzyme (Table 2.2.18; Figure 2.2.21). Superposition of equivalent 'close' dimers of the $P2_12_12_1$ and $P2_12_12$ apo structures (rather than individual subunits) reveals a significant increase in the

rmsd values (0.47 Å → 0.91 Å). In turn, the changes at the AD interface appear to affect the relative positioning of monomers forming a 'loose' dimer association, as they also display greater rmsd's (0.47 Å → 0.65 Å) despite the fact that superpositions do not indicate any large differences in monomeric structure at the AC(BD) interface (Figure 2.2.21).

Figure 2.2.22 – B-factor analysis. Cartoon representation of the SsKDGA monomer in the $P2_12_12$ (**A**) and $P2_12_12_1$ (**B**) apoenzyme structures. Key residue side chains are shown as sticks. All residues are coloured by B-factor; 10 Å^2 (blue) → 60 Å^2 (red).

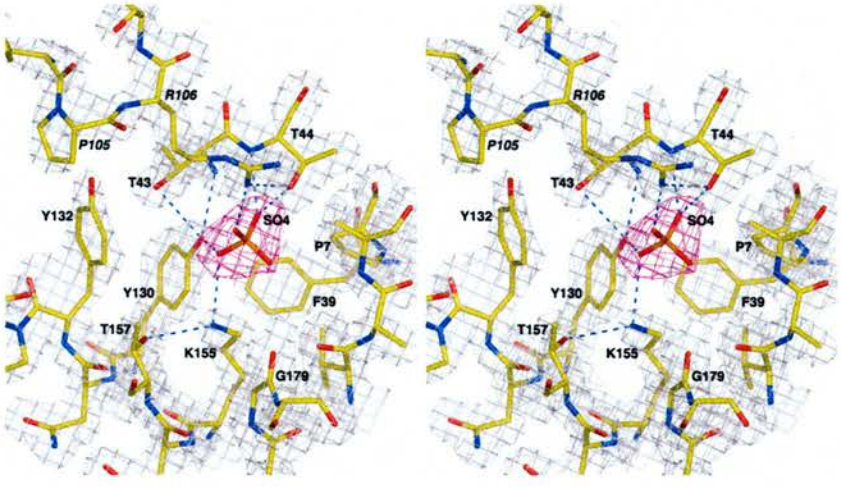


Consistent with their conformational variability, the residues in loop 4 and 5 are characterised by elevated B-factors relative to the equivalent regions in the $P2_12_12_1$ apoenzyme structure (Figure 2.2.22). This is clearly apparent even when considering the differences on an absolute scale, as the average B-factors for all protein atoms in the two structures are similar ($P2_12_12_1$, 19 Å^2; $P2_12_12$, 24 Å^2), and most regions of the

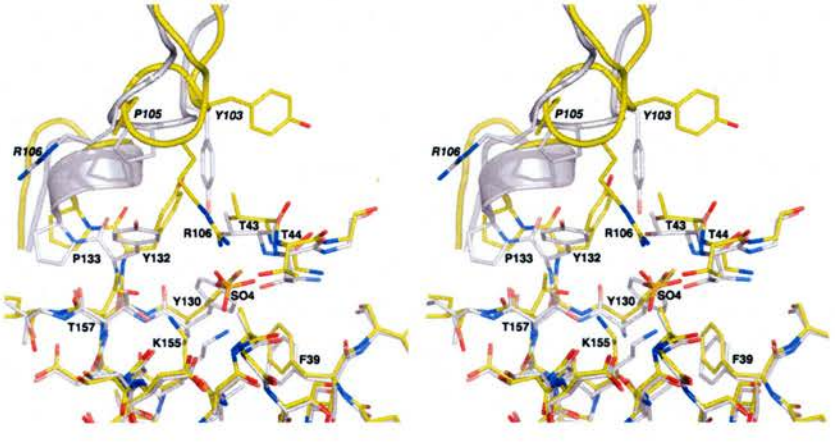
protein display comparable values. In fact, the two loops and the adjoining helices and strands are the only regions of the protein that display significantly greater disorder in the pH 4 structure. Consequently, the enzyme active site, with which these structural elements are intimately associated, was examined more closely.

Inspection of the active site in the $P2_12_12$ structure revealed a sulfate ion (a component of the crystallisation medium) bound at the base of the pocket in all subunits (Figure 2.2.23**A**). The ion occupies a conserved position, equivalent to that of the substrate's α-keto acid moiety, and is stabilised by the same set of residues (Thr-43, Thr-44, Tyr-130, Lys-155). In addition, however, the ion was also consistently observed forming an intersubunit interaction with Arg-106, the side chain of which was found protruding deep into the active site pocket. This represents a far greater shift in the residue than that observed in the D-KDPG(al) complexes, as illustrated by the fact that its guanidinium Nη atom is positioned 7 Å closer to the carbonyl carbon of Thr-43 ($P2_12_12$ = 4.5 Å ± 0.4 Å; $P2_12_12_1$ = 11.9 Å ± 0.2 Å). Superposition of the apoenzyme models for the two crystal forms reveals a number of main chain and side chain conformational changes within the active site, which appear to be directly or indirectly associated with the presence of Arg-106 and the sulfate ion (Figure 2.2.23**B**). Although this observation can also be extended to the larger conformational changes observed in loops 4 and 5, there is not sufficient evidence to determine a sequence of cause and effect. On the one hand, pH induced conformational shifts may have permitted binding of the sulfate ion and its interaction with Arg-106, while on the other, the ion itself may have induced all other changes. In any case, the active site remains solvent accessible in the pH 4 structure, offering at least the potential for substrate complex formation.

Figure 2.2.23 – The active site of SsKDGA. Stereoviews of the active site in the apoenzyme structures at pH 4 and pH 6. Residues correspond to any subunits *x* and *y* in an *xy* close dimer (where the catalytic lysine is contributed by *x*). Residues from subunit *y* are labelled in italics. Colour key: **carbon**, yellow ($P2_12_12$ structure), or grey ($P2_12_12_1$ structure); **oxygen**, red; **nitrogen**, blue; **sulfur**, orange.



A. The active site in the $P2_12_12$ apo structure. Overlayed, the unbiased F_o-F_c density for the bound sulfate ion contoured at 2.5 σ (cyan mesh) and the $2F_o$-F_c density for the protein contoured at 1.0 σ (white mesh). Key interactions between the sulfate ion and active site residues are shown as dashes. [Average interaction distances: **SO4-T44(Oγ)**, 2.8 Å; **SO4-T44(N)**, 2.9 Å; **SO4-T43(N)**, 3.2 Å; **SO4-Y130**, 2.4 Å; **SO4-K155**, 2.9 Å; **SO4-*R106***, 2.9 Å; **T157-K155**, 2.9 Å; **T43(Oγ)-Y130**, 2.8 Å; **T43(N)-Y130**, 3.0 Å].

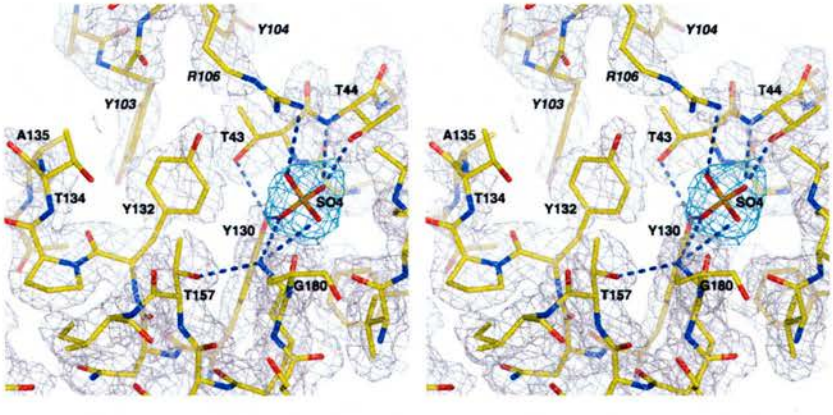


B. Superposition of the active site region of SsKDGA in the $P2_12_12$ and $P2_12_12_1$ apo structures.

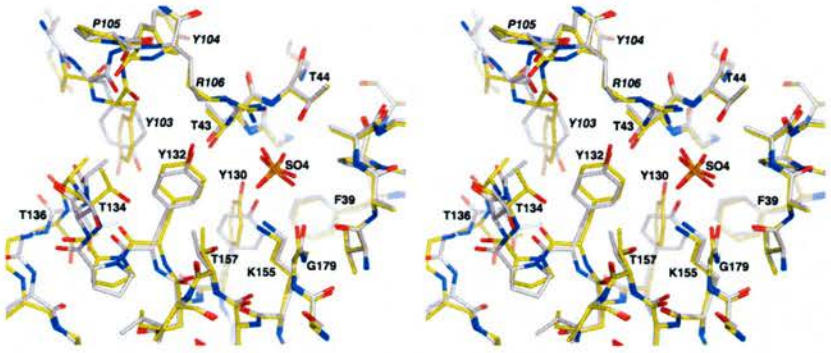
The pyruvate complex:

An initial soak was carried out at RT using a 50 mM solution of pyruvate in mother liquor. After 30 min the crystal was cryoprotected by equilibration against a substrate solution containing 20% (v/v) glycerol and then flash frozen. A 2 Å dataset was collected at the ESRF and processed as before with good statistics (Table 2.2.16, *PYR-1* column). Molecular replacement and subsequent refinement also proceeded as in the case of the pH 4 apo structure, with no significant deviations. Moreover, while the statistics for the MR solution were better (Rf, 41%; CC_F, 70%) and lower R/Free R factors were observed during refinement (23% and 26%, respectively, before solvent building), PYR-1 displays the same overall structure as the apoenzyme [rmsd in P2-K293 Cα's = 0.38 Å (monomer); 0.40 Å (tetramer)].

Figure 2.2.24 – The active site in the PYR-1 structure of SsKDGA. Stereoviews of the active site in PYR-1 and the apoenzyme at pH 4. Residues correspond to any subunits *x* and *y* in an *xy* close dimer (where the catalytic lysine is contributed by *x*). Residues from subunit *y* are labelled in italics. Colour key: **carbon**, yellow (PYR-1), or grey (apoenzyme); **oxygen**, red; **nitrogen**, blue; **sulfur**, orange.



A. The active site in PYR-1. Overlayed, the unbiased F_o-F_c density for the bound sulfate ion contoured at 2.5 σ (cyan mesh) and the $2F_o$-F_c density for the protein contoured at 1.0 σ (white mesh).



B. Superposition of PYR-1 and the $P2_12_12$ apo structure, showing the active site region as a stick representation.

In addition to loops 4 and 5 being characterised by the same features as described above, a sulfate ion was found in the active site of each subunit, bound in the same position and engaging in equivalent interactions as before (Figure 2.2.24). Moreover, all active site residues (including Arg-106) were seen in similar conformations as in the apoenzyme. In this structure, however, the sulfate and arginine were characterised by weaker density, while there was also some evidence in the maps for the side chain of Lys-155 adopting a conformation inconsistent with the presence of the ion. These observations were interpreted as evidence for slow exchange taking place between pyruvate and the sulfate, and prompted further soaking experiments.

A substrate complex was obtained by soaking a pH 4 crystal overnight in a 40 mM pyruvate solution, at RT. Cryoprotection was then carried out by stepwise equilibration against the substrate solution containing 10% (v/v), followed by 20% (v/v) glycerol, before flash freezing in liquid nitrogen. The 2.3 Å dataset was collected on ID14-2 at the ESRF, in Grenoble, and processed in $P2_12_12$, with good statistics (Table 2.2.16, *PYR-2* column). Moreover, subsequent molecular replacement against the $P2_12_12_1$ native apoenzyme tetramer resulted in an equivalent

solution to those observed for the other pH 4 crystals, although the statistics in this case were significantly improved (Table 2.2.19).

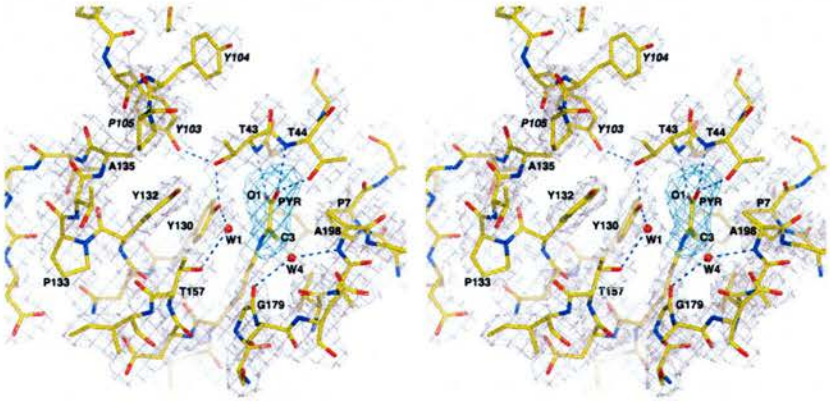
α	β	γ	X	Y	Z	CC_F	Rf	CC_I
135.43	79.52	92.55	0.2418	0.2653	0.3753	35.0	51.1	39.7
44.80	79.76	270.05	-0.2583	0.2650	0.1249	73.5	36.0	73.4

Table 2.2.19 – Molecular replacement. A tetramer search model was generated from the $P2_12_12_1$ native apo structure. Crystal data were used within the resolution range of 10 - 4 Å and the cross-rotation function search sphere for the Patterson correlation was set to 30 Å. The *AMoRe* rotation and translation searches were carried out sequentially, resulting in an initial solution shown in the second row. This solution was subsequently fixed and the translation function repeated, resulting in a second solution shown in the third row. Rigid body refinement of the two solutions together resulted in an R-factor of 31% and correlation coefficient of 79%. Abbreviations: **α β γ**, rotation solution in Eulerian angles; **X Y Z**, fractional unit cell coordinates for translation solution; **CC_F**, correlation coefficient between observed and calculated structure factor amplitudes; **Rf**, R-factor between observed and calculated amplitudes; **CC_I**, correlation coefficient for intensities.

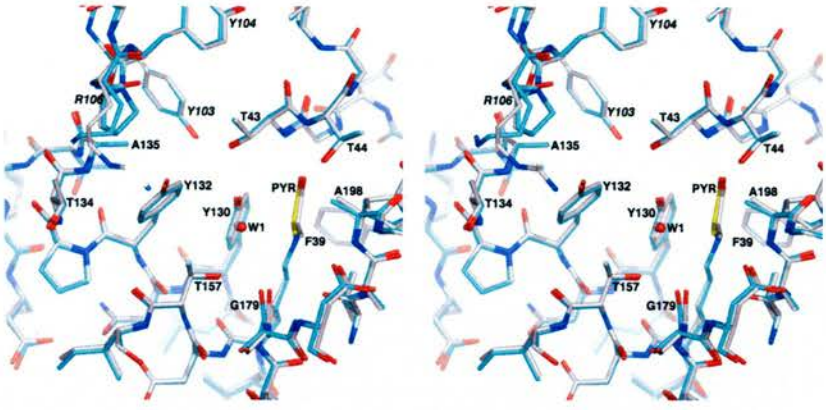
Refinement was carried out as described previously (Appendix 1) and resulted in an Rf of 20% and Rfree of 26%. Interestingly, significant corrections to the starting model did not have to be made in this case. Although some side chain adjustments were necessary, for example at points of contact between symmetry related molecules, the backbone structure remained unchanged. Consequently, the enzyme's overall monomeric and oligomeric structure in the final model is equivalent to that observed at pH 6 [rmsd in P2-K293 Cα's = 0.25 Å (monomer); 0.31 Å (tetramer)] and not that displayed by PYR-1 and the $P2_12_12$ apoenzyme [rmsd in P2-K293 Cα's = 0.75 Å (monomer); 1.1 Å (tetramer). The active site and the regions β4-α4 and β5-α5, in particular, are characterised by clear density and display the same architecture as in the $P2_12_12_1$ crystal form.

Although sulfate ions were found interacting with the protein at various locations within this structure, there was no evidence in any of the subunits for their presence within the active site pocket. Instead, at the position previously occupied by sulfate the Fourier maps revealed a Schiff base complex between Lys-155 and pyruvate (Figure 2.2.25**A**), indicating that complete exchange had taken place during the course of the soak, despite the anion's significantly higher concentration (x40). The substrate intermediate refined well within the electron density and in the final model occupies the same position and is involved in the same interactions as in the equivalent complex structure obtained for the $P2_12_12_1$ crystal form (Figure 2.2.20**B**).

Figure 2.2.25 – The pyruvate Schiff base complex at pH 4. Stick representations of the active site of SsKDGA. Residues correspond to any subunits *x* and *y* in an *xy* close dimer (where the catalytic lysine is contributed by *x*). Residues from subunit *y* are labelled in italics.



A. Refined model of the enzyme in complex with the Schiff base intermediate form of pyruvate at pH 4. Overlayed, the unbiased F_o-F_c density for the substrate contoured at 2.5 σ (cyan mesh) and the $2F_o$-F_c density for the protein contoured at 1.5 σ (white mesh). [Average interaction distances: **PYR(O2)-T44(Oγ)**, 2.5 Å; **PYR(O2)-T44(N)**, 2.8 Å; **PYR(O1)-T43(N)**, 2.9 Å; **PYR(O1)-Y130**, 3.2 Å (close contact); **PYR(O1)-F39**, 3.4 Å (close contact); **PYR(C3)-W1**, 3.0 Å; **Y130-T43(Oγ)**, 2.6 Å; **Y130-W1**, 2.6 Å; **W1-T157**, 2.6 Å; **T43(Oγ)-*Y103***, 2.8 Å; **A198-W4**, 2.9 Å; **G179-W4**, 2.8 Å]. [Average pyruvate bond distances/angles: **C1-C2**, 1.5 Å; **C2-C3**, 1.5 Å; **C2-Nζ**, 1.3 Å; **C1-C2-C3**, 119°; **C1-C2-Nζ**, 113°; **Nζ-C2-C3**, 124°; **O1-C1-C2- C3**, 176°; **O2-C1-C2- Nζ**, -178°].



B. Superposition of the Schiff base pyruvate complexes of SsKDGA at pH 6 (grey carbons) and pH 4 (protein, cyan carbons; pyruvate, yellow carbons).

The formation of this covalent complex requires that the side chain of Lys-155 remains deprotonated at pH 4. Consequently, the pKa of this residue's ε-amino group, which in solution would be 10.5 (Whitford, 2005), must be significantly lower within the active site environment. The two most common strategies employed by enzymes for increasing the nucleophilic character of a catalytic lysine at physiological pH involve charge destabilisation by either burial of the residue within a hydrophobic environment (Dao-pin *et al.*, 1991; Barbas *et al.*, 1997), or by electrostatic interaction with another basic residue (Westheimer, 1995; Maurady *et al.*, 2002; Heine *et al.*, 2004). Moreover, in the case of KDPGA, the structures of which (1eua, 1wa3) represent the only other observations in a type I aldolase of covalent intermediate formation at such a low pH, it has been suggested that both mechanisms may be in operation (Fullerton *et al.*, 2006).

Unlike most type I aldolases, SsKDGA does not possess a suitably positioned basic residue. The guanidinium groups of Arg-237 and Arg-106 are located more than 11 Å from K155(Nζ) in the apoenzyme and all substrate complex structures, and are therefore unlikely to influence it's pKa. Moreover, the only acidic residue within the

active site (Asp-181) is also too far away (> 8.5 Å) to permit initial activation of Lys-155, as has been proposed in the case of transaldolase B (Jia *et al.*, 1997), although a similar process could potentially occur via Tyr-130 and/or the α-keto acid carboxylate of the approaching substrate. However, on the basis of the available evidence it is most likely that the residue's unusual ionisation properties are the result of burial within a low dielectric environment (surrounding residues; P7, F39, Y130, T157, G179, V196), in which the ε-amino group retains a small solvent accessible area (~ 2.5 $Å^2$).

A similar mechanism has been demonstrated for the aldolase antibody 33F12 (Barbas *et al.*, 1997), the catalytic lysine of which (Lys-H93) is surrounded by hydrophobic residues (Ser-35, Val-H37, Tyr-H95, Ser-H100 Trp-H103, Phe-L98) and also displays a small solvent accessibility (Nζ, 5.8 $Å^2$). The side chain's pKa in this case was determined at 5.5, which would mean that at pH 4.5 only 9% of the lysine population remains deprotonated. Nevertheless, given the length of the soaking experiment carried out, the observation of a covalent intermediate at full occupancy in the SsKDGA structure is not inconsistent with a similarly low abundance of the reactive species.

Although nucleophilic attack clearly takes place at pH 4, the reaction may be sufficiently slow to permit trapping of a non-covalent (Michaelis) complex with pyruvate, if a suitable soaking time is used. Moreover, the observation of a carbinolamine might also be possible. Schiff base formation results from dehydration of this tetrahedral intermediate, which in turn requires proton donation by a general acid; a process that is also expected to be less efficient at lower pH (Clayden *et al.*, 2001). Consequently, providing the time required for exchange to take place between sulfate and pyruvate can be determined with some accuracy, the $P2_12_12$ crystal form

should prove useful in further studies of the SsKDGA mechanism of substrate binding and catalysis.

The pyruvate complex highlights an additional interesting feature. It clearly demonstrate that the distortions in the $P2_12_12$ apoenzyme structure are not due to pH, but rather are solely the result of a sulfate ion bound within the active site pocket. The extent of these changes (involving main chain and side chain shifts that affect the oligomeric assembly), as well as their reversibility, reveal a surprising degree plasticity in SsKDGA that has not been observed in other type I aldolases. Although tetrahedral anions have been found bound to the active sites in several structures, in almost all cases they have not been associated with significant conformational changes.

HiNAL represents the only notable exception, as extensive distortions/disordering in the region between strand β5 and helix α5 (residues 136-147) have been observed in the presence within the active site of both sulfate ions and substrate analogues (Barbosa *et al.*, 2000). Unlike SsKDGA, however, these distortions in HiNAL structure do not extend to either loop 4, or the tetrameric assembly. Given the fact that the region from β5 to α5 is well conserved between the two enzymes and makes similar contributions to the AD interface, it would be reasonable to suggest that differences in loop 4 are responsible for the more extensive changes observed in SsKDGA. Consequently, the plasticity displayed by this enzyme may be an indirect consequence of an increase in this loop's length and flexibility, required for more efficient phosphate binding. Mutagenesis studies focusing on this region could be useful in obtaining a better understanding of its physiological role and effect on the enzyme's stability.

2.2.4 Substrate binding at equilibrium

The mechanism of type I aldolases is characterised by complex energy profiles involving several intermediates, as illustrated in Figure 2.2.26. An apparent consequence of this is the difficulty of identifying which of the steps is rate limiting. Citing numerous studies Willson and co-workers (1995) concluded that in FBPA the rate is determined by product release in the condensation reaction and either product release or C-C bond cleavage, in the reverse direction. On the other hand, while discussing the synthetic utility of type I and type II aldolases, Machajewski and Wong (2000) more recently defined formation of the enolate as rate limiting in both families.

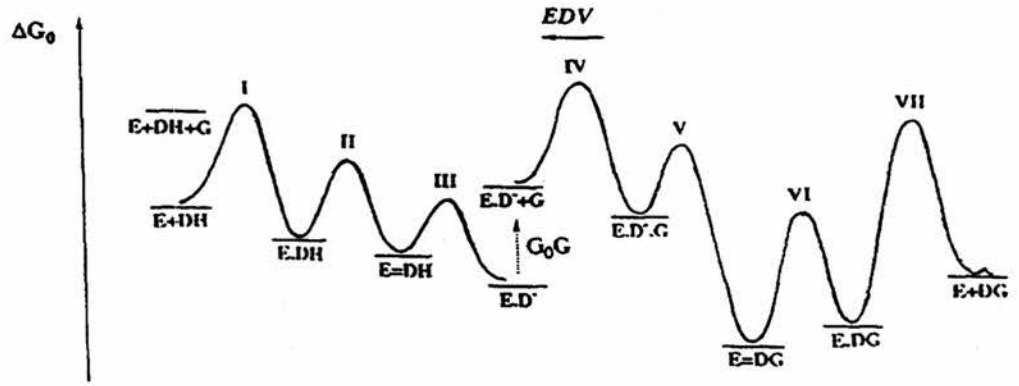


Figure 2.2.26 – Proposed energy level diagram for FBPA. The left part represents the relative energy levels of the complexes formed between FBPA and DHAP, while the right part those of FBP (noted DG). The gap between the two parts corresponds to the standard free enthalpy difference (G_oG) due to the contribution of GAP. The different complexes are indicated as follows: **E.DH**, non-covalent complex between DHAP and enzyme; **E=DH**, DHAP iminium intermediate; **ED⁻**, enamine; **ED⁻.G**, ternary complex of enamine and GAP; **E=DG**, FBP iminium; **E.DG**, non-covalent complex between FBP and enzyme; **E + DH**, enzyme + DHAP; **E + DG**, enzyme + FBP; **ED⁻ + G**, enzyme with DHAP enamine + GAP. Reproduced from (Gefflaut *et al.*, 1995).

The crystallographic evidence for type I aldolases indicates that the reaction profile varies between the enzymes, as Schiff base and carbinolamine intermediates (5 and 3 structures, respectively), as well as non-covalent complexes (2 structures) of natural substrates have been trapped (see Table 2.2.1). However, these results also display significant inconsistencies, as in two cases different types of complex have been observed within the same enzyme (FBPA: 4ald, 1zai), even under apparently similar conditions (KDPGA: 1eua, 1wa3), while in a third (DERA: 1jcl, 1ub3), very different experimental parameters have given rise to the same intermediate form.

In the case of SsKDGA, soaking experiments have resulted in the trapping of the iminium forms of D-KD(P)G(al) and either the iminium or enamine forms of pyruvate, the two intermediates being indistinguishable at the resolution they were observed. Although the results are relatively consistent and suggest that C-C bond cleavage is the limiting step in one direction (at least at 4 °C), either enolisation or C-C bond formation could determine the rate of the reverse reaction. Moreover, the prevalence of the Schiff base intermediate could be, at least in part, a result of the experimental conditions. Attempts were therefore made to obtain some independent confirmation of the crystallographic results using alternative biophysical techniques (primarily mass spectrometry), in order to detect complexes of SsKDGA with pyruvate and determine their identity.

Gas-phase observation of a pyruvate Schiff base intermediate:

Electrospray ionisation ESI is a gentle ionisation technique that permits the accurate determination of protein molecular weight (Fenn *et al.*, 1989) from a series of multiply charged gas-phase ions, while avoiding molecular fragmentation. It has found great utility in the detection and study of covalent (Stevenson *et al.*, 1990)

(Aplin *et al.*, 1990), as well as and non-covalent interactions (Loo, 1997; Jorgensen *et al.*, 1998), as it displays the capability of transferring even weakly bound complexes from solution to the vacuum. Moreover, the experimental evidence indicates that for many systems the binding characteristics observed by ESI-MS reflect to some extent the nature of the interactions in solution-phase.

With relevance to the current study, ESI-MS has been used to observe complexes between 3-dehydroquinate and dehydroquinase (Shneier *et al.*, 1991), as well as between pyruvate and DHDPS (Borthwick *et al.*, 1995), revealing an imine (Schiff base) in both cases. Detection of this reversible intermediate was achieved by flow injection analysis (FIA), in which the sample is injected into a carrier stream of solvent that propels it to the mass spectrometer. The advantages of this method are that diffusion is minimised, as is the length of exposure of the protein to the denaturing effects of methanol (a component of the carrier solvent). Consequently, the methodology was also applied to investigate the binding of pyruvate to SsKDGA.

Immediately prior to analysis the enzyme was dialysed from its native buffer (10 mM Tris/HCl, pH 8.5) into 10 mM ammonium acetate pH 7, while sodium pyruvate (Sigma) was dissolved in deionised water. SsKDGA and its substrate were mixed to the desired ratio and incubated at RT for 15 min. 20 µl samples were then injected into the mass spectrometer and data collected for five minutes in positive ionisation mode. Approximately 50 scans around the chromatogram peak were combined to give the raw spectrum, which was processed and deconvoluted, giving the data shown in Figure 2.2.27.

Initially, SsKDGA was analysed on its own (Figure 2.2.27-**1**) and gave rise to a spectrum with two main peaks, corresponding to the enzyme subunit (**a**) and a sodium

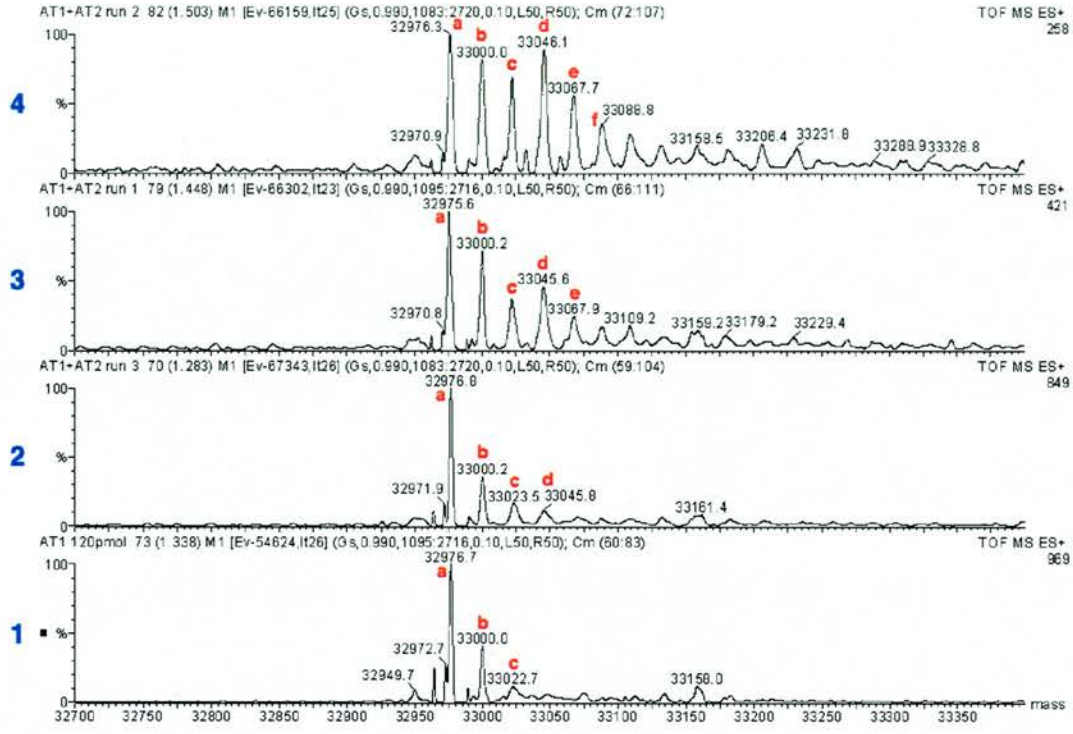


Figure 2.2.27 – Observation of a pyruvate Schiff base by ESI mass spectrometry. Samples were analysed by Flow Injection Analysis (FIA) on a Micromass LCT mass spectrometer equipped with a Waters 2795 HPLC. FIA carrier solvent conditions were 1:1 methanol water at 200 µl/min. Immediately prior to the experiment SsKDGA (in 20 mM Tris/HCl, pH 8.5) was dialysed into 10 mM ammonium acetate pH 7 and sodium pyruvate was dissolved in water. The protein and ligand were mixed to the desired ratio and 20 µl of sample was injected from the autosampler. Samples tested, were: **1.** 6 µM SsKDGA; **2.** 1 mM pyruvate in 1.8 µM SsKDGA; **3.** 10 mM pyruvate in 1.8 µM SsKDGA; **4.** 20 mM pyruvate in 1.8 µM SsKDGA. Mass spectrometry data was collected for five minutes over the range 100 – 3500 m/z, at 1 scan/sec. The instrument was operated in positive ionisation mode, with capillary at 3500 V, cone at 50 V, and RF lens at 500 Volts. Approximately 50 scans around the chromatogram peak were combined to give the raw spectrum, which was then processed using the Micromass Masslynx v4.0 software and deconvoluted using the MaxEnt algorithim. The mass spectrometer was calibrated prior to analysis with a solution of 10 µM horse heart myoglobin. The observed peaks are: **a.** SsKDGA (theoretical M_r, 32,977.2); **b.** SsKDGA + Na^+ (M_r, 23); **c.** SsKDGA + 2×Na^+; **d.** SsKDGA + pyruvate Schiff base (M_r, 71); **e.** SsKDGA + pyruvate Shiff base + Na^+; **f.** SsKDGA + pyruvate carbinolamine (M_r, 88) + Na^+.

adduct (**b**). A third signal arising from SsKDGA with two sodium ions bound (**c**), was also recorded, although at low intensity. Subsequent samples of SsKDGA were incubated in the presence of excess pyruvate before running on the mass spectrometer. Three separate concentrations of substrate were used, corresponding to ×1, ×10 and ×20 its K_m, and in the resulting spectra additional signals were observed (Figure 2.2.27-**2**, -**3**, -**4**). Peaks **d** and **e** were tentatively assigned as complexes of SsKDGA with a Schiff base intermediate of pyruvate (+ a sodium ion **e**), and **f**, as a carbinolamine or non-covalently bound pyruvate complex (+ one sodium ion).

Although the data did not permit these signals to be identified with certainty, they do provide some evidence for intact substrate complexes being transferred from the condensed- to the gas-phase. They also indicate that the Schiff base intermediate is the major species under these conditions and consequently, support the existing evidence for its prevalence at equilibrium. While discrimination between the iminium and enamine forms was again not possible in this experiment, additional inconclusive evidence obtained by NMR (data not shown) suggests that the two intermediates may exist in equilibrium. In any case, further experimentation and refinement of the existing protocols should help to provide more easily interpretable results.

Summary

Complexes have been obtained at physiological pH between SsKDGA and its natural substrates, pyruvate, D-KDG(al) and D-KDPG(al). In each case a Schiff base intermediate has been observed, in a configuration that appears to closely resemble the transition state. The conformational changes and interactions associated with the binding of each of these substrates have been characterised and have permitted a

rationalisation of the enzyme's unique promiscuity. Key groups have also been examined with respect to the catalytic mechanism, providing an assessment of their predicted roles, as well as additional insights. Further studies using crystals grown under different conditions have demonstrated SsKDGA's ability to form Schiff base intermediates with its substrates at acidic pH and have revealed a significant degree of plasticity inherent in the enzyme's structure. Substrate binding has also been investigated in the gas-phase, by electrospray ionisation mass spectrometry, providing some evidence for the preferred formation of Schiff base intermediates under conditions of equilibrium.

Part 3

Further investigations of SsKDGA

2.3.1 Altering specificity

Detailed knowledge of the active site environment, as well as the mechanisms of substrate binding and catalysis affords the opportunity of employing a rational design approach to altering enzyme specificity. Although directed evolution is generally considered as the preferred strategy for protein engineering (Fong *et al.*, 2000; Wymer *et al.*, 2001; Reetz, 2004; Hsu *et al.*, 2005; Gould and Tawfik, 2005), even where an understanding of the target is not limited, structure-based re-design has also proved successful. With respect to type I aldolases for example, site-directed mutagenesis studies have been effective in expanding the substrate specificities of DERA (Silvestri *et al.*, 2003) and NAL (Joerger *et al.*, 2003; Williams *et al.*, 2005).

In the case of SsKDGA, the detailed structural characterisation described in previous chapters identified key residues involved in substrate binding, as well as interactions that appear to play a role in preferentially stabilising either D-KDG or D-KDGal (see Figure 2.2.7). The proposed model has since formed the basis of a mutagenesis study being carried out at the University of Bath to try and induce greater selectivity for one or other of the two diastereomers in the condensation reaction between pyruvate and D-glyceraldehyde. Some success has been achieved so far, with mutants generated that display up to 80% selectivity for D-KDG (Table 2.3.1). However, increased specificity for D-KDGal has proved more elusive.

No.	Mutation	D-KDG (%)
w.t.	-	55
1	T44A	65
2	Y132E	73
3	T44A/S241A/N245L	75
4	T44A/Y132E	79
5	T44A/Y132E/S241A/N245L	79

Table 2.3.1 – Inducing stereoselectivity. SsKDGA mutants displaying improved selectivity for D-KDG relative to D-KDGal. Aldol condensation between pyruvate and D-glyceraldehyde catalysed by the wild-type enzyme results in the formation of both diastereomers in similar proportions. Consequently, structure-based re-design of SsKDGA has been carried out to try and induce greater stereospecificity in the enzyme reaction. 80% selectivity for D-KDG has been achieved by i) eliminating interactions involved in preferentially stabilising D-KDGal, and ii) introducing a new functionality to the active site, which is thought to lead to improved binding of D-glyceraldehyde in a conformation that favours *si*-face attack by the enamine. Reproduced from Lamble, personal communication.

Significantly greater success has been achieved by utilising a substrate-engineering strategy (Stampfer *et al.*, 2003; Silvestri *et al.*, 2003; Turner, 2003; Lamble, 2004; Lamble *et al.*, 2005a). It was reasoned that the use of a structurally-rigid analogue of glyceraldehyde, such as an acetonide derivative, might enable stereocontrol to be introduced to the SsKDGA reaction. Kinetic studies confirmed that both D- and L-glyceraldehyde acetonide are accepted as substrates by the enzyme (Table 2.3.2) and it was subsequently discovered that the SsKDGA catalysed condensation of pyruvate with D- or L-glyceraldehyde acetonide results in highly stereoselective formation of

Substrate	K_m (mM)	k_{cat} (min^{-1})	k_{cat} / K_m ($min^{-1}.mM^{-1}$)
D-Glyceraldehyde	3.9 (± 0.2)	594(± 33)	152(± 11)
L-Glyceraldehyde	7.1 (± 0.6)	594(± 63)	83 (± 11)
D-Glyceraldehyde acetonide	22.7 (± 2.4)	324 (± 18)	14 (± 1.7)
L-Glyceraldehyde acetonide	5.5 (± 0.4)	42 (± 1)	7 (± 0.5)

Table 2.3.2 – Kinetic parameters for SsKDGA. Analysis was carried out at 70 °C using a modified version of the TBA assay. Reproduced from Lamble, 2004.

D-KDG acetonide (>92% d.e.), or L-KDGal acetonide (>94% d.e.), respectively (Figure 2.3.1). Moreover, it was found that pure D-KDG or L-KDGal can then be obtained by a simple acid catalysed hydrolysis of the protecting group.

The results demonstrated that this reversible derivatisation of glyceraldehyde can induce stereocontrol in the SsKDGA reaction and indicated the possibility of the technique's application with other acceptor substrates (e.g. tetroses). Crystallographic experiments were subsequently undertaken in order to obtain complexes of SsKDGA

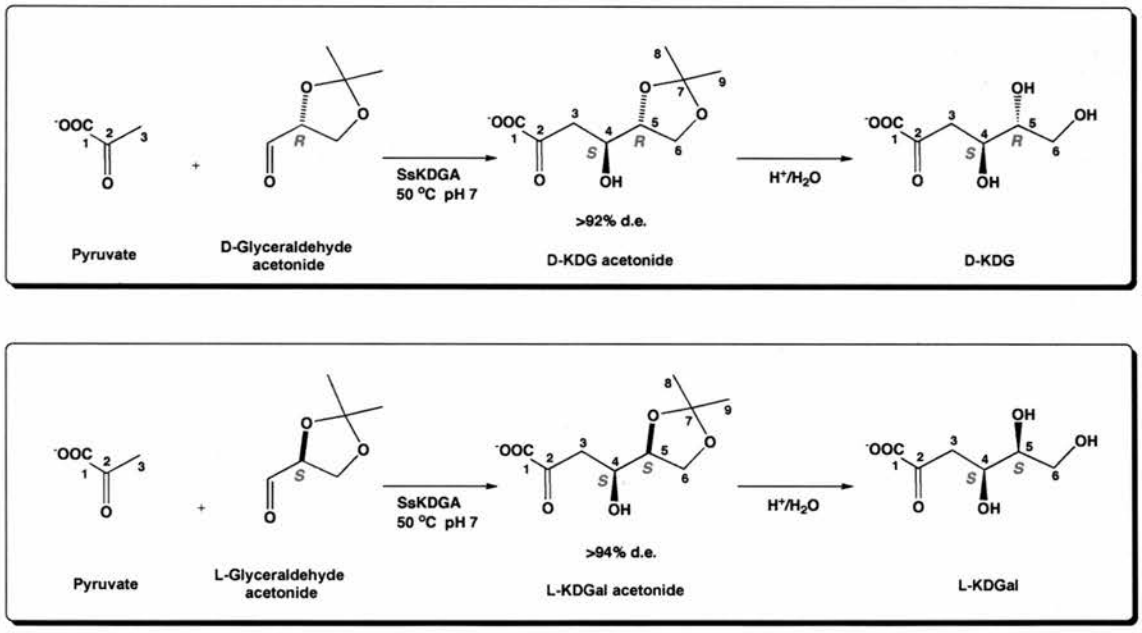


Figure 2.3.1 – Stereospecific biotransformations with SsKDGA. Synthesis of D-KDG and L-KDGal from acetonide derivatives. SsKDGA catalysed condensation of pyruvate with D-glyceraldehyde or L-glyceraldehyde acetonide results in highly stereoselective formation of D-KDG acetonide (>92% d.e.) or L-KDGal acetonide (>94% d.e.), respectively. Subsequent deprotection by acid-catalysed hydrolysis of the acetonide group gives rise to D-KDG and L-KDGal, at an equivalent diastereomeric excess (d.e.). Adapted from Lamble, 2004.

with each of the two products, L-KDGal and D-KDG acetonide, the results of which are described below. In addition to rationalising the structural basis for the observed stereoselectivity, it was considered that mutations might be identified that would

induce formation of the two opposite diastereomers, thus affording complete control over the stereochemical course of the SsKDGA catalysed condensation.

An additional aspect of the enzyme's selectivity that has been considered for rational re-design, is that of the donor species. Like other pyruvate aldolases SsKDGA displays strict specificity for its natural donor substrate (see Chapters 1.2.2 & 1.2.3)

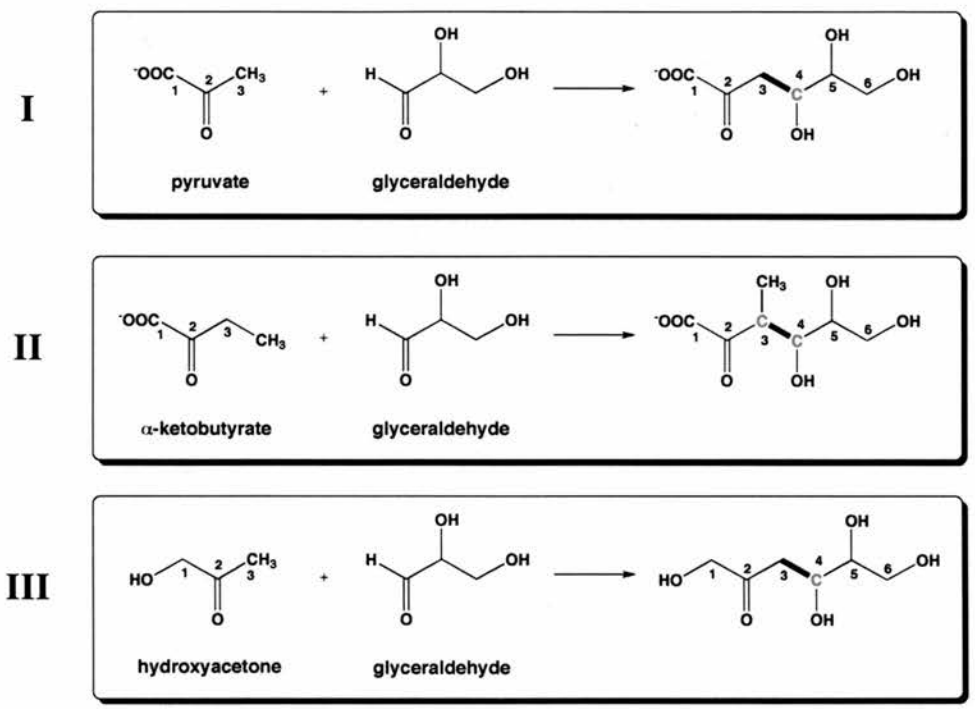


Figure 2.3.2 – Donor substrate specificity. Comparison of the products of the aldol condensations between glyceraldehyde, as acceptor, and either pyruvate (**I**), α-ketobutyrate (**II**), or hydroxyacetone (**III**) acting as donor substrates. The aldol condensations involving pyruvate or hydroxyacetone generate only one new chiral centre (red C), while the aldol product of α-ketobutyrate and glyceraldehyde (III) has two. Moreover, hydroxyacetone is unique in generating an uncharged aldol.

and no activity has been detected with any of the analogues tested so far (Buchanan *et al.*, 1999; Lamble, personal communication). Nevertheless, the potential for accepting alternative donor species does exist, as illustrated by the directed evolution of a EcKDPGA mutant displaying significantly greater activity with α-ketobutyrate

(Wymer *et al.*, 2001). Consequently, a structure guided mutagenesis study of SsKDGA has been initiated, with the pyruvate analogues α-ketobutyrate and hydroxyacetone selected as preliminary targets (Figure 2.3.2).

Complexes with acetonide substrates:

A complex with L-KDGal acetonide was obtained by soaking a protein crystal in a 50 mM substrate solution for 1 min at 4 °C, before flash freezing for data collection (Table 2.3.3). Inspection of the Fourier difference maps during refinement revealed a Schiff base of L-KDGal acetonide bound in each of the active sites of the asymmetric

Table 2.3.3 - Data collection and refinement statistics for acetonide complexes. Summary of statistics for the complexes of SsKDGA with the D-KDG and L-KDGal acetonide substrates. Processing and refinement were carried out as described previously and outlined in Appendix 1. Abbreviations: **rmsd**, root mean square deviation.

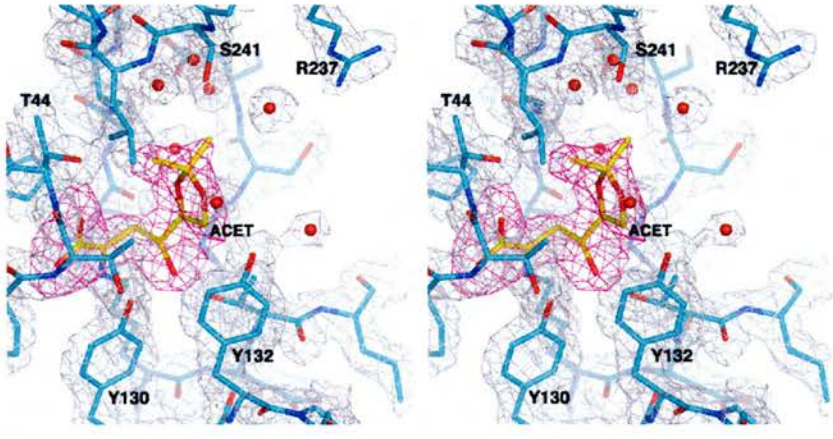
	D-KDG acetonide ID14-EH2	L-KDGal acetonide In-house
Wavelength (Å)	0.934	1.54178
Resolution limits (Å)*	42 – 1.8 (1.9 – 1.8)	30 – 2.1 (2.21 – 2.1)
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å)	*a* = 83.7 *b* = 131.5 *c* = 132.8	*a* = 83.9 *b* = 132.4 *c* = 132.7
No. observations / No. of unique reflections	628,802 / 135,648	349,065 / 84,311
Mosaicity	0.54	0.7
Completeness (%)*	99.9 (100)	97.6 (91.5)
R_{merge}‡ (%)*	5.7 (27.5)	9.1 (25.2)
$< I/\sigma(I) >$*	19.1 (4.9)	12.2 (4.1)
Multiplicity*	4.6 (4.3)	4.1 (3.2)
Wilson B (Å^2)	18.0	19.7
Refinement		
No. of working / test set reflections	122,056 / 13,592	75,806 / 8,503
Data Completeness (%)	99.8	97.2
No. of protein / water / ligand atoms	9,300 / 1,187 / 56	9,300 / 370 / 56
<B> for protein / water / ligand atoms (Å^2)	17.2 / 32.3 / 14.7	17.9 / 21.0 / 14.8
R-factor / R-free (%)†	16.9 / 20.6	17.6 / 22.8
rmsd bond lengths (Å) / bond angles (°)	.013 / 1.7	.018 / 2.1

*, values in parentheses refer to the highest resolution shell.
‡, $R_{merge} = \Sigma_{hkl} \Sigma_i |I_{hkl,i} - \langle I_{hkl} \rangle| / \Sigma_{hkl} \langle I_{hkl} \rangle$
†, Rf and R-free = $(\Sigma \, | |F_o| - |F_c| |)/(\Sigma \, |F_o|)$

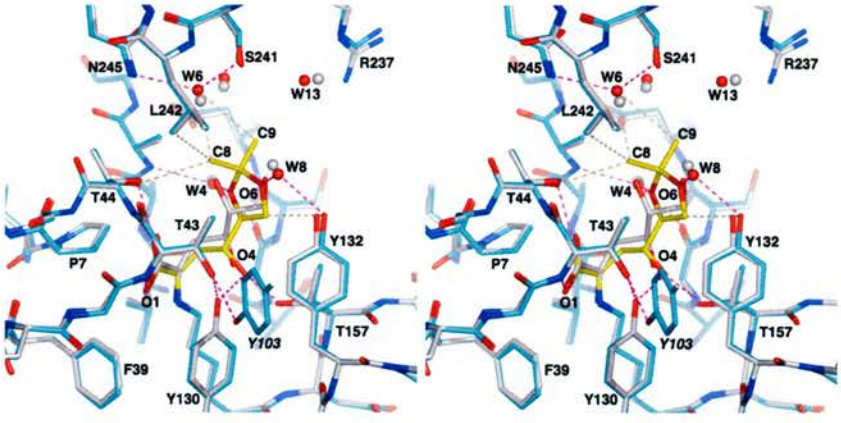
unit (Figure 2.3.3**A**). Coordinates for the substrate subsequently refined well within the density using geometric restraints for the (*4S,5S*)-configured intermediate, resulting in a highly consistent conformation across subunits (rmsd, 0.1 Å).

In the final model, the α-keto acid moiety and C4 hydroxyl occupy their conserved positions and engage in the same interactions as described previously (Figure 2.3.3**B**). In addition, one of the acetonide oxygens (O5) interacts with Ala-198 and Gly-179, via W4, while the other (O6) forms a water-mediated interaction with Tyr-132. The acetonide ring is also involved in several van der Waals contacts: C6 with Tyr-132 and Thr-157; C8 with Thr-44, Ala-198 and Leu-242. The acetonide methyl groups (C8 and C9) are also in close contact with several ordered water molecules, including W6 and W13, which in turn interact with protein side chains. These observations suggest that the substrate intermediate has very little conformational freedom within the active site, as is also indicated by its low average B-factor value (~ 15 $Å^2$).

Figure 2.3.3 – The L-KDGal acetonide complex.



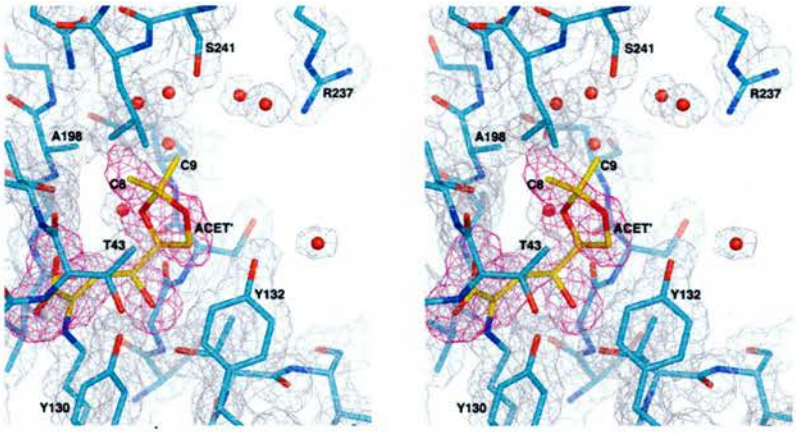
A. The refined model of SsKDGA in complex with the Schiff base intermediate of L-KDGal acetonide (ACET; yellow carbons). Overlayed the unbiased F_o-F_c map density for the substrate contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.5 σ (white). [Average pyruvate bond distances/angles: **C1-C2**, 1.5 Å; **C2-C3**, 1.5 Å; **C2-Nζ**, 1.3 Å; **C1-C2-C3**, 121°; **C1-C2-Nζ**, 108°; **Nζ-C2-C3**, 128°; **O1-C1-C2- C3**, -167°; **O2-C1-C2-Nζ**, 179°].



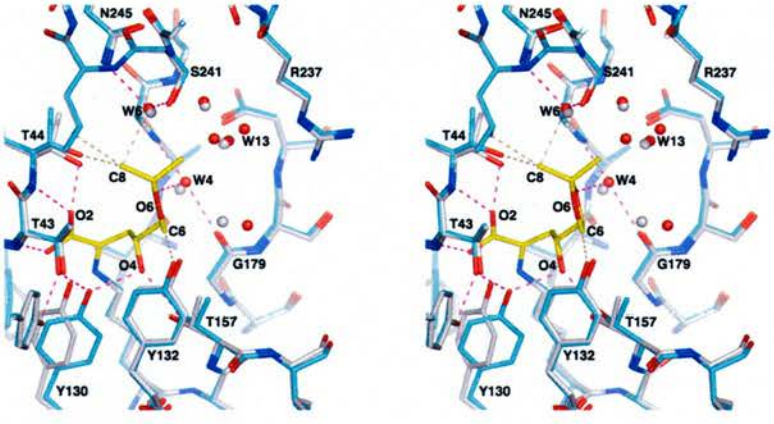
B. Superposition of the L-KDGal acetonide and D-KDG complexes. Key enzyme-substrate interactions in the acetonide structure are shown as dashes. Colour key: **carbon**, cyan (acetonide complex), grey (D-KDG complex), yellow (L-KDGal acetonide); **oxygen**, red; **nitrogen**, blue; **water molecules**, red spheres (acetonide complex), grey spheres (D-KDG complex); **HB distances**, magenta dashes; **close contacts (≤ 4 Å)**, wheat dashes.
[Average interaction distances: **ACET(O2)-T44(Oγ)**, 2.6 Å; **ACET(O2)-T44(N)**, 2.8 Å; **ACET(O1)-T43(N)**, 3.0 Å; **ACET(O1)-F39**, 3.4 Å (close contact); **ACET(O1)-Y130**, 3.4 Å (close contact); **ACET(O4)-Y130**, 2.5 Å; **ACET(O4)-T157(Oγ)**, 2.6 Å; **ACET(O5)-W4**, 2.8 Å; **ACET(O6)-W8**, 2.7 Å; **ACET(C6)-Y132**, 3.2 Å (close contact); **ACET(C6)-T157(Cγ)**, 3.8 Å (close contact); **ACET(O6)-Y132**, 3.5 Å (close contact); **ACET(C8)-T44(Oγ)**, 3.3 Å (close contact); **ACET(C8)-W6**, 3.1 Å (close contact); **ACET(C8)-A198(Cβ)**, 3.9 Å (close contact); **ACET(C8)-L242**, 3.9 Å (close contact); **ACET(C9)-W6**, 3.4 Å (close contact); **ACET(C9)-W13**, 3.3 Å (close contact); **Y130-T43**, 2.6 Å; ***Y103*-T43**, 2.7 Å; **W4-A198(N)**, 2.9 Å; **W4-G179**, 2.8 Å; **W8-Y132**, 2.7 Å; **W6-N245**, 3.0 Å; **W6-S241**, 2.9 Å; **W13-R237**, 2.9 Å].

Subsequently, a D-KDG acetonide complex was also obtained at 4 °C, using a 50 mM soaking concentration (Table 2.3.3) and refinement of the SsKDGA apoenzyme model against the 1.8 Å synchrotron data revealed a Schiff base intermediate of the substrate, bound in an almost identical conformation as L-KDGal acetonide (rmsd 0.2 Å; Figure 2.3.4**A**). In this case, however, the electron density indicated partial cleavage of the aldol, which was determined as being present at an occupancy of approximately 70%. Moreover, the electron density corresponding to the acetonide group was found to be incomplete.

Figure 2.3.4 – The D-KDG acetonide complex.



A. The refined model of SsKDGA in complex with the Schiff base intermediate of D-KDG acetonide (yellow carbons; ACET'). Overlayed the unbiased F_o-F_c map density for the substrate contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.5 σ (white). [Average pyruvate bond distances/angles: **C1-C2**, 1.5 Å; **C2-C3**, 1.5 Å; **C2-Nζ**, 1.3 Å; **C1-C2-C3**, 119°; **C1-C2-Nζ**, 109°; **Nζ-C2-C3**, 129°; **O1-C1-C2- C3**, -164°; **O2-C1-C2-Nζ**, 179°].



B. Superposition of the apoenzyme and D-KDG acetonide complex structures. Key enzyme-substrate interactions in the acetonide structure are shown as dashes. Colour key: **carbon**, cyan (acetonide complex), grey (apoenzyme), yellow (D-KDG acetonide); **oxygen**, red; **nitrogen**, blue; **water molecules**, red spheres (acetonide complex), grey spheres (apo structure); **HB distances**, magenta dashes; **close contacts (≤ 4 Å)**, wheat dashes.

[Average interaction distances: **ACET'(O2)-T44(Oγ)**, 2.7 Å; **ACET'(O2)-T44(N)**, 2.8 Å; **ACET'(O1)-T43(N)**, 2.8 Å; **ACET'(O1)-F39**, 3.4 Å (close contact); **ACET'(O1)-Y130**, 3.2 Å (close contact); **ACET'(O4)-Y130**, 2.6 Å; **ACET'(O4)-T157**, 2.7 Å; **ACET'(O5)-W4**, 2.9 Å; **ACET'(C6)-Y132**, 3.2 Å (close contact); **ACET'(O6)-Y132**, 3.4 Å (close contact); **ACET'(C8)-T44(Oγ)**, 3.0 Å (close contact); **ACET'(C8)-W6**, 2.9 Å (close contact); **ACET'(C8)-L242**, 3.5 Å (close contact)].

Despite clear density in the unbiased F_o-F_c maps for all other atoms of the acetonide group (visible at ≥ 4 σ), the C9 carbon was not observed in any of the subunits of the 1.8 Å structure, nor in the case of a lower resolution dataset obtained of the enzyme in complex with the same compound (data not shown). In contrast, the C9 methyl is clearly visible in the L-KDGal acetonide structure, although it is characterised by weaker density than C8. The discrepancy between the two complexes cannot be accounted for in terms of greater disorder. Moreover, both acetonide samples have been comprehensively characterised by NMR spectroscopy and their structures confirmed (Lamble, 2004). Loss of the methyl independently of the acetonide group either prior to, or during the soaking experiments is also highly unlikely.

Even in the absence of a plausible explanation for the incomplete substrate density, the overall evidence is consistent with D-KDG acetonide being the species that was trapped. Moreover, this structure displays no additional features and only serves to confirm the observations made in the L-KDGal acetonide complex. Both intermediates adopt the same rigid conformation (average B-factor, ~ 15 Å^2) and are involved in equivalent interactions with the enzyme (Figures 2.3.3**B** & 2.3.4**B**). These results are in agreement with the induction of stereoselectivity in the synthetic reaction. They indicate that the acetonide ring in both its D- and L-configurations can only bind in the active site in a single orientation. This in turn dictates the angle of approach of the acceptor, resulting in attack of the enamine onto a single face of the carbonyl.

Another unexpected feature of the two acetonide structures is highlighted by their superposition with the D-KDG Schiff base complex (Figure 2.3.3**B**). While demonstrating that all three substrates adopt a similar conformation with respect to their C5 and C6 positions (carbon and oxygen atoms), the superposition also reveals

an unexpected difference at C4. Although the three molecules are characterised by a (*4S*)-configuration, the two acetonides display a geometry around this carbon atom that resembles that of D-KD(P)Gal [(*4R*)-configuration]. Moreover, refinement using alternative geometric restraints resulted in the overall fit to the density becoming worse, indicating that the conformation is correct regardless of how it is assigned. Although the significance of this feature remains unclear, it is considered unlikely to significantly alter the overall understanding of how the acetonide compounds are bound to the active site of SsKDGA.

The crystallographic evidence indicates that the orientation of the acetonide ring is determined by its two methyl groups and their steric clashes with surrounding residues. Consequently, it has been possible to suggest mutations (e.g. Y132D and T157S) that should enable binding of the acetonide ring in its opposite orientation and may therefore result in a different face of the acceptor carbonyl being presented to the pyruvate enamine. If this is successful, a second round of mutations may be required to prevent binding of the acetonide in its current orientation (e.g. A198V). An alternative approach that may also be considered involves the elimination of a hydroxyl from position 157 (e.g. T157G), which may then permit greater flexibility in the orientation adopted by the acceptor carbonyl, without the need to alter the binding mode of the acetonide ring. It is first necessary, however, to establish the significance of the extended HB network describe previously ($Y103_y$-$T43_x$-$Y130_x$-W1-$T157_x$).

Novel donor substrates:

A complex with α-ketobutyrate (sodium salt, ≥ 99%; Fluka) was obtained by soaking a crystal of SsKDGA in a 50 mM solution at RT for 45 min. A dataset was then collected at the ESRF that processed to below 1.6 Å (Table 2.3.4, *KB* column). Inspection of the Fourier difference maps during refinement revealed a Schiff base intermediate of the substrate analogue bound in all four active sites and in the same position as that observed for the α-keto moiety in previous complexes (Figure 2.3.5).

Table 2.3.4 - Data collection and refinement statistics. Summary of statistics for the complexes of SsKDGA with hydroxyacetone (HYA) and α-ketobutyrate (KB). Processing and refinement were carried out as described previously and outlined in Appendix 1. Abbreviations: **rmsd**, root mean square deviation.

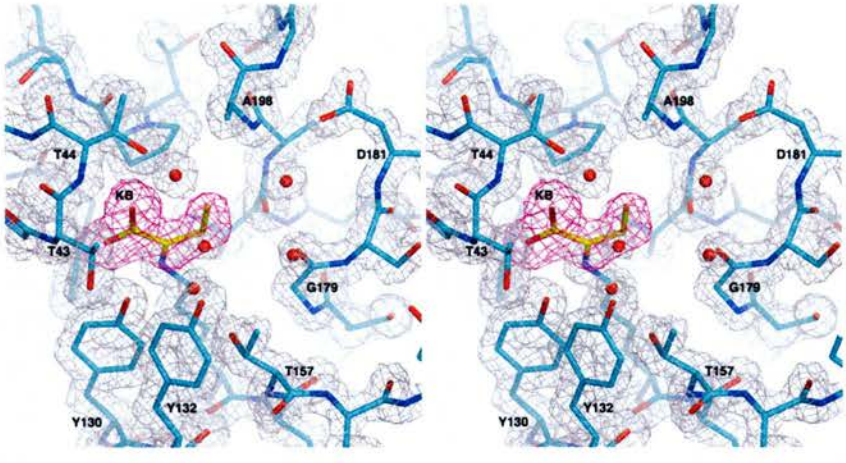
	HYA **ID14-EH1**	**KB** **ID14-EH1**
Wavelength (Å)	0.934	0.934
Resolution limits (Å)*	31 – 1.8 (1.9 – 1.8)	30 – 1.55 (1.63 – 1.55)
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å)	*a* = 83.7 *b* = 129.0 *c* = 132.4	*a* = 83.7 *b* = 131.5 *c* = 132.3
No. observations / No. of unique reflections	653,140 / 126,771	1,281,947 / 210,823
Mosaicity	0.26	0.37
Completeness (%)*	96.1 (96.8)	100 (100)
R_{merge}‡ (%)*	5.4 (26.0)	6.3 (29.5)
$< I/\sigma(I) >$*	18.7 (5.4)	18.5 (4.5)
Multiplicity*	5.2 (5.1)	6.1 (4.3)
Wilson B (Å^2)	21.0	16.2
Refinement		
No. of working / test set reflections	114,111 / 12,659	189,757 / 21,063
Data Completeness (%)	95.3	100
No. of protein / water / ligand atoms	9,300 / 737 / 16	9,300 / 1,120 / 24
<B> for protein / water / ligand atoms (Å^2)	21.0 / 34.3 / 22.8	16.2 / 31.0 / 17.2
R-factor / R-free (%)†	17.2 / 21.1	16.8 / 19.8
rmsd bond lengths (Å) / bond angles (°)	.018 / 1.9	.023 / 2.0

*, values in parentheses refer to the highest resolution shell.
‡, $R_{\text{merge}} = \Sigma_{hkl} \Sigma_i |I_{hkl,i} - \langle I_{hkl} \rangle| / \Sigma_{hkl} \langle I_{hkl} \rangle$
†, Rf and R-free = $(\Sigma \, | |F_o| - |F_c| |)/(\Sigma |F_o|)$

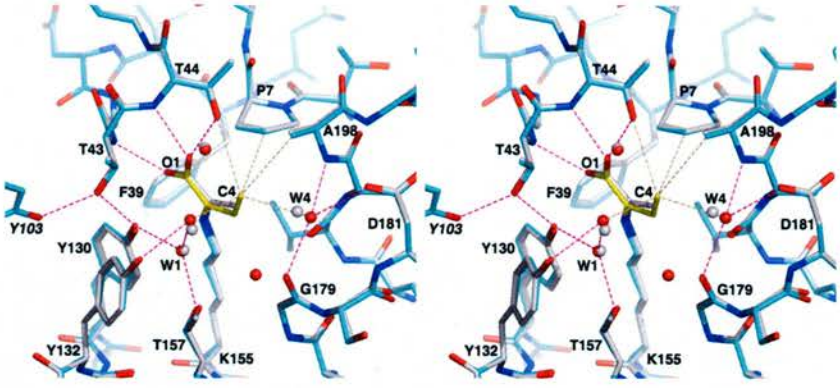
Moreover, continuous density was observed in each case extending to the substrate analogue's C4 methyl group, which was found in an eclipsed conformation, *cis-* to the O2 carboxylate oxygen.

In the final model atoms O1 to C3 of α-ketobutyrate adopt the same conformation as pyruvate and engage in identical interactions (Figure 2.3.5**B**). In addition, C4 forms van der Waals contacts with a number of residue and solvent atoms, including Pro-7, Thr-44, Ala-198, Gly-179, W1, W4 and W10, as well as with the KB(O2) atom. Despite these interactions, the methyl group does not display significant stability, as it is characterised by a higher average B-factor (26 $Å^2$) than the intermediate (17 $Å^2$) or protein (16 $Å^2$) as a whole. The group's increased disorder is consistent with its unfavourable eclipsed conformation relative to the carboxylate oxygens, in which it is likely to be forced by steric clashes with protein and solvent atoms.

Figure 2.3.5 – The SsKDGA complex with α-ketobutyrate.



A. The refined model of SsKDGA in complex with the Schiff base intermediate of α-ketobutyrate (yellow carbons; KB). Overlayed the F_o-F_c map density for the substrate contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.5 σ (white). [Average KB bond distances/angles: **C1-C2**, 1.5 Å; **C2-C3**, 1.5 Å; **C2-Nζ**, 1.3 Å; **C1-C2-C3**, 125°; **C1-C2-Nζ**, 108°; **Nζ-C2-C3**, 122°; **O1-C1-C2- C3**, 165°; **O2-C1-C2- Nζ**, -173°].



B. Superposition of the pyruvate (PYR) and α-ketobutyrate (KB) complexes. Key enzyme-substrate interactions in the KB structure are shown as dashes. Colour key: **carbon**, cyan (KB complex), grey (PYR complex), yellow (KB); **oxygen**, red; **nitrogen**, blue; **water molecules**, red spheres (KB complex), grey spheres (PYR complex); **HB distances**, magenta dashes; **close contacts (≤ 4 Å)**, wheat dashes.

[Average interaction distances: **KB(O2)-T44(Oγ)**, 2.6 Å; **KB(O2)-T44(N)**, 2.8 Å; **KB(O1)-T43(N)**, 2.9 Å; **KB(O1)-F39**, 3.4 Å (close contact); **KB(O1)-Y130**, 3.1 Å (close contact); **KB(C3)-W1**, 3.2 Å (close contact); **KB(C4)-W1**, 3.3 Å (close contact); **KB(C4)-KB(O2)**, 3.0 Å (close contact); **KB(C4)-W4**, 3.3 Å (close contact); **KB(C4)-T44(Oγ)**, 3.5 Å (close contact); **KB(C4)-A198(Cβ)**, 3.8 Å (close contact); **KB(C4)-P7**, 3.9 Å (close contact); **Y130-T43(Oγ)**, 2.5 Å; **T43(Oγ)-*Y103***, 2.7 Å; **Y130-W1**, 2.7 Å; **W1-T157**, 2.8 Å; **A198(N)-W4**, 2.8 Å; **G179-W4**, 2.8 Å; **D181-W4**, 2.8 Å].

The conformation adopted by the C4 methyl is unexpected, but demonstrates the substrate analogue's ability to exist as the activated enamine, in which formation of the double bond between C2 and C3 forces a planar geometry on the donor species. Although in the enamine form the methyl of α-ketobutyrate could theoretically adopt either a *cis-* or *trans-* configuration relative to the carboxylate O2 atom, the latter is prohibited within the active site of SsKDGA due to the positioning of Gly-179, and only the less energetically favourable *cis-* conformation is possible. Nevertheless, there is some evidence in this crystal structure for tautomerisation taking place.

Refinement of the coordinates for the α-ketobutyrate atoms of the Schiff base using geometric restraints for the imine intermediate, resulted in two distinct conformations in the final model. In two subunits the intermediate adopted a planar conformation,

while in the other two, a small but clear deviation from planarity was observed in the C3 and C4 atoms. The resulting rmsd of 0.12 Å is greater than that observed for all atoms of Lys-155 (0.07 Å) and given the high resolution and quality of the data, indicates a significant difference. It may therefore be the case that the electron density for α-ketobutyrate represents an average between the imine and enamine intermediate forms. Although this has yet to be carried out, at 1.55 Å this complex structure is just within the limits for refinement in SHELXL (Sheldrick and Schneider, 1997), which would permit an unrestrained determination of the K155(Nζ)-KB(C2) and other α-ketobutyrate bond lengths, as well as the visualisation of certain hydrogen atoms.

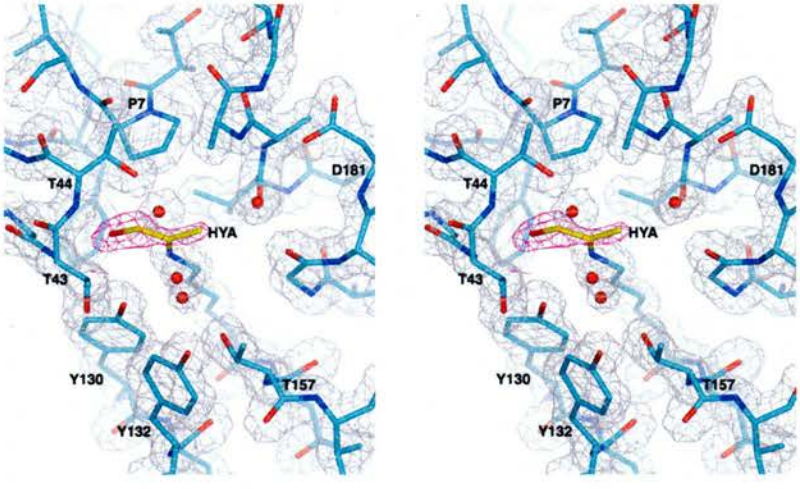
The crystallographic observations indicate that α-ketobutyrate should be able to act as the donor species in the SsKDGA catalysed aldol condensation. Therefore, the lack of observed activity is most likely due to steric clashes between the substrate analogue's additional methyl group and the acceptor carbonyl. This is supported by superpositions of the KB complex with the D-KDG and D-KDGal models. On the basis of these superpositions, mutations have been suggested that may suitably enlarge the active site pocket (e.g. Y132E, T157S), so as to permit simultaneous binding of α-ketobutyrate and glyceraldehyde. In fact, this may have indirectly been the effect of the EcKDPGA mutation, which shifted the catalytic lysine from strand β6 to β7 (Wymer *et al.*, 2001).

Although lower concentrations and shorter duration soaks were also attempted, a complex with hydroxyacetone (90%; Aldrich) was finally obtained by incubating an SsKDGA crystal in a 2 M substrate solution, at RT for 2 h. At that time the crystal was cryoprotected and placed in a nitrogen stream for data collection. In the structure

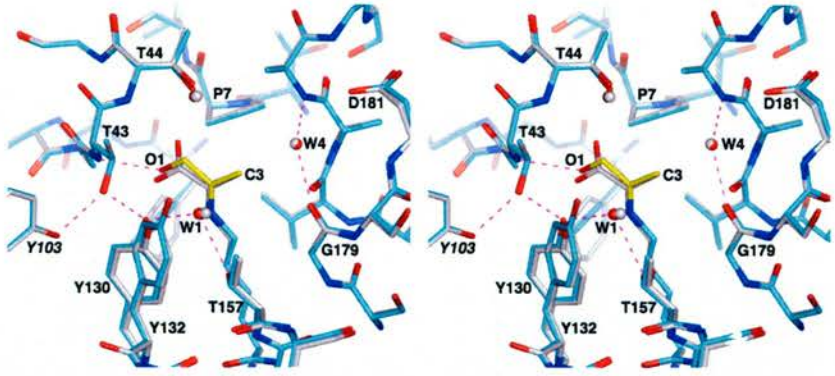
obtained (Table 2.3.4, *HYA* column), density was observed in each of the subunits extending from K155(Nζ), consistent with the Schiff base intermediate of the substrate analogue (Figure 2.3.6**A**). The coordinates for the hydroxyacetone intermediate were incorporated into the model and refined using appropriate restraints derived from *SYBYL*, with good results.

In the final model the substrate analogue occupies an equivalent position as the pyruvate Schiff base, with its C1 hydroxyl pointing in the same direction as the carboxylate O1 atom (Figure 2.3.6**B**). In this conformation, HYA(O1) is within HB distance of the amide of Thr-43 and the hydroxyl of Tyr-130, although the nature of the Y130(Oη)-HYA(O1) interaction is again not clear. The ordered water molecule W1 is also present in each of the subunits and forms potential hydrogen bonds with

Figure 2.3.6 – The SsKDGA complex with hydroxyacetone.



A. The refined model of SsKDGA in complex with the Schiff base intermediate of hydroxyacetone (yellow carbons; HYA). Overlayed the F_o-F_c map density for the substrate contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.5 σ (white). [Average pyruvate bond distances/angles: **C1-C2**, 1.5 Å; **C2-C3**, 1.5 Å; **C2-Nζ**, 1.3 Å; **C1-C2-C3**, 113°; **C1-C2-Nζ**, 127°; **Nζ-C2-C3**, 120°; **O1-C1-C2- C3**, 173°].



B. Superposition of the pyruvate (PYR) and hydroxyacetone (HYA) complexes. Key enzyme-substrate interactions in the HYA structure are shown as dashes. Colour key: **carbon**, cyan (HYA complex), grey (PYR complex), yellow (HYA); **oxygen**, red; **nitrogen**, blue; **water molecules**, red spheres (HYA complex), grey spheres (PYR complex); **HB distances**, magenta dashes; **close contacts (≤ 4 Å)**, wheat dashes.

[Average interaction distances: **HYA(C1)-T44(Oγ)**, 3.0 Å (close contact); **HYA(O1)-T43(N)**, 2.9 Å; **HYA(O1)-F39**, 3.7 Å (close contact); **HYA(O1)-Y130**, 3.0 Å (close contact); **HYA(C3)-W1**, 3.0 Å (close contact); ***Y103*-T43(Oγ)**, 2.7 Å; **Y130-T43(Oγ)**, 2.5 Å; **Y130-W1**, 2.7 Å; **T157-W1**, 2.9].

Tyr-130 and Thr-157, as well as interacting with the C3 methyl group of the intermediate. The only difference observed within the active site between the pyruvate and hydroxyacetone complexes is with respect to Phe-39, which in the latter case does not appear to interact with the intermediate and adopts a different orientation to that observed in the apoenzyme and substrate complex structures.

This complex reveals that hydroxyacetone is stabilised in the SsKDGA active site by fewer HB interactions, which is consistent with its apparently lower binding affinity. However, the structure does not identify differences in the binding mode of the substrate analogue that might account for its total inactivity in condensation reactions with glyceraldehyde. Instead, the evidence indicates that hydroxyacetone is unable to form the activated enamine. This, in turn, can only be explained in terms of the molecule's own structure, relative to that of pyruvate (i.e. the presence of a polar

group at C1 in place of an acidic one). In fact the observation of this Schiff base intermediate provides the most direct evidence yet for a substrate assisted mechanism operating in members of the NAL subfamily. Moreover, it demonstrates that the substrate carboxylate is involved only in general base catalysed proton abstraction and is not essential for the general acid activity that results in dehydration of the carbinolamine.

In view of these results, engineering activity for hydroxyacetone in SsKDGA would require eliminating the need for substrate assistance in catalysis. This could potentially be achieved by introducing an acidic residue into the active site, in a position suitable for it to interact with key substrate groups, effectively generating a mechanism similar to that of KDPGA or RAMA. Such an attempt has recently been reported for TtFBPA (Lorentzen *et al.*, 2005), in which the proposed catalytic tyrosine was substituted by a glutamic acid. However, the double mutant generated in this study was found to be inactive due to the formation of a salt bridge between the glutamate and the Schiff base forming lysine.

In the case of SsKDGA, an equivalent mutant has already been generated in the form of Y132E. This enzyme has been shown to retain high levels of activity and display increased selectivity for D-KDG. Although modelling suggests that Glu-132 is not in a position to interact with key substrate groups involved in C-C bond formation and cleavage, it is located near Tyr-130 and Thr-43. Therefore, it is possible that it could complement or replace the α-keto acid carboxylate, if indeed it carries out its function via the other residues, as has been proposed.

2.3.2 Identifying catalytic residues

Structural characterisation of the active site in SsKDGA has identified an extended hydrogen bonding network, consisting of Thr-43, Tyr-130, Thr-157, an ordered water molecule (W1) and Tyr-103 from an adjacent subunit (see Chapters 2.1.4). The network is conserved within the NAL subfamily and, on the basis of structural and mutagenesis studies, has been assigned a catalytic role in DHDPS as a proton shuttle (*E. coli* enzyme, $Y133_x$-$T44_x$-$T107_y$ triad; Blickling *et al.*, 1997b; Dobson *et al.*, 2004). Evidence presented in previous chapters (2.1.4; 2.2.1) indicates, however, that the equivalent residues of NAL and SsKDGA cannot fulfil the same function, due to differences in solvent accessibility.

Nevertheless, while not the same as that of DHDPS, the role of the HB network in SsKDGA may also be significant. This is supported by several factors, including: i) the participation of the proposed catalytic residue, Tyr-130; ii) the involvement of an ordered water (W1) that may assist in catalysis; iii) the fact that in the D-KD(P)G(al) complexes the position of the ordered water is occupied by a key substrate group (the C4 hydroxyl); and iv) the positioning of the intersubunit contribution, Tyr-103, in a disallowed region of the Ramachandran plot. In order to investigate this possibility, six mutants of the enzyme have been generated (Y130F, Y103F, Y103F/Y130F, T157V, T43V, T43V/Y130F).

Kinetic characterisation in the condensation reaction (Lamble, personal communication; Table 2.3.5) has revealed that the Tyr-130 mutation results in a considerable loss of activity (two orders of magnitude reduction in k_{cat}/K_m), primarily due to a decrease in V_{max}. The threonine to valine change at position 157 also appears to compromise catalytic efficiency (one order of magnitude), by affecting both K_m

and V_{max}. On the other hand, Y103F displays only a minor (2-fold) reduction in k_{cat}/K_m. Activity assays have also been carried out with T43V and T43V/Y130F (results not shown), and indicate a significant loss of activity relative to wild-type SsKDGA, although kinetic parameters have not been determined yet.

Mutants	Substrate	K_m (mM)	V_{max} (units.mg^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1}.mM^{-1}$)	Relative k_{cat}/K_m
Wild-type	Pyruvate	1.0 (± 0.1)	15.7 (± 0.3)	8.7	8.7	1
	D-Glyceraldehyde	3.9 (± 0.3)	18.0 (± 1.0)	9.9	2.5	1
Tyr103Phe	Pyruvate	1.33 (± 0.03)	9.92 (± 0.07)	5.45	4.10	0.47
	D-Glyceraldehyde	6.57 (± 0.18)	11.4 (± 0.4)	6.24	0.95	0.38
Tyr130Phe	Pyruvate	0.45 (± 0.02)	0.06 (± 0.01)	0.04	0.08	< 0.01
	D-Glyceraldehyde	9.61 (± 0.57)	0.06 (± 0.01)	0.03	(0.004)	< 0.01
Tyr103Phe/ Tyr130Phe	Pyruvate	0.38 (± 0.01)	0.09 (± 0.01)	0.05	0.14	0.02
	D-Glyceraldehyde	12.1 (± 0.4)	0.12 (± 0.01)	0.07	(0.005)	< 0.01
Thr157Val	Pyruvate	5.39 (± 0.1)	2.87 (± 0.03)	1.56	0.29	0.03
	D-Glyceraldehyde	14.2 (± 0.87)	3.22 (± 0.21)	1.77	0.12	0.05

Table 2.3.5 – Kinetic parameters for SsKDGA mutants. Kinetic analysis was carried out at 70 °C and pH 6.0 using a modified version of the TBA assay. 1 unit corresponds to the formation of 1 μmole D-KDG(al) per minute. Adapted from (Buchanan *et al.*, 1999; Lamble, personal communication).

These results do not only support an involvement of the HB network in SsKDGA catalysis, they also demonstrate its distinct nature from that proposed for DHDPS. For example, the role of the intersubunit contribution (Tyr-103), if indeed it has one in SsKDGA, appears to be less important than in EcDHDPS. This is indicated by the fact that the Y107F mutant of the latter displays a 10-fold decrease in V_{max}, and a 60-fold reduction in catalytic efficiency for L-ASA (Dobson *et al.*, 2004). Moreover, changing Tyr-133 to phenylalanine in DHDPS appears to have a significant effect

(two orders of magnitude) on both V_{max} and K_m, so that the overall catalytic efficiency is far lower than in the equivalent mutant of SsKDGA.

The kinetic data for the EcDHDPS mutants are complemented by apoenzyme structures for each (Dobson *et al.*, 2004). These have permitted comparisons to be made with the wild-type active site, as well as some rationalisation of the differences in activity between enzyme forms, particularly with respect to the Tyr-107. With the same aim in sight, crystallographic experiments were carried out with the SsKDGA mutants, The focus in this case, however, was on obtaining substrate complexes, which might afford greater insight to the effects on substrate binding and catalysis of each of the residue changes. The results of this study are presented below.

SsKDGA mutant structures:

All six enzymes expressed efficiently and were purified in the same way as the wild-type. Moreover, four of the mutants gave rise to crystals of good diffraction quality by screening around the central pH 6 condition (0.1 M HEPES, 8% propan-2-ol, 13% PEG 4k) as before. In the case of T43V and T43V/Y130F, crystals did grow after some optimisation of the conditions, but were small and did not diffract beyond 4 Å. Nevertheless, processing and refinement of this low resolution data did not reveal any large-scale changes to crystal packing or tetrameric structure. Therefore, it may be possible in the future to improve the size and quality of the crystals by further screening around the existing conditions. Along with a full kinetic characterisation, crystal structures of these two mutants may help to provide a more complete understanding of the hydrogen bonding network of SsKDGA.

Table 2.3.6 - Data collection and refinement statistics. Summary of statistics for the D-KDPG (KPG) and pyruvate (PYR) complexes of the SsKDGA mutants; Y130F, Y103F, Y103F/Y130F and T157V. Processing and refinement were carried out as described previously and outlined in Appendix 1. Abbreviations: **rmsd**, root mean square deviation.

	Y130F (KPG) ID14-EH3	**Y130F (PYR) ID14-EH3**	**Y103F (PYR) ID14-EH2**	**Y103F/Y130F (PYR) In-house**	**T157V (PYR) ID14-EH1**
Wavelength (Å)	0.931	0.931	0.934	1.54178	0.934
Resolution limits (Å)*	34 – 1.7 (1.79 – 1.7)	38 – 1.8 (1.9 – 1.8)	41 – 1.7 (1.79 – 1.7)	10 – 2 (2.11 – 2)	41 – 1.6 (1.69 – 1.6)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å)	a = 83.5 b = 131.3 c = 132.3	a = 83.6 b = 130.8 c = 132.3	a = 83.6 b = 130.9 c = 132.3	a = 83.5 b = 131.2 c = 132.2	a = 83.8 b = 131.1 c = 132.5
No. observations / No. of unique reflections	659,464 / 144,506	537,497 / 123,680	1,102,701 / 159,167	291,089 / 95,303	1,315,580 / 186,021
Mosaicity	0.29	0.54	0.45	0.5	0.4
Completeness (%)*	90.7 (93.7)	92.9 (95.5)	99.9 (100)	97 (90.6)	97.2 (98.7)
R_{merge}‡ (%)*	5.4 (19.7)	5.4 (21.8)	5.6 (17.6)	3.8 (7.4)	5.5 (28.5)
$< I/\sigma(I) >$*	16.4 (6.3)	16.7 (5.7)	25.8 (8.1)	25.6 (14.9)	22.6 (4.9)
Multiplicity*	4.6 (4.4)	4.3 (3.9)	6.9 (4.9)	3.1 (2.7)	7.1 (4.8)
Wilson B (Å^2)	17.6	18.9	17.6	16.4	18.3
Refinement					
No. of working / test set reflections	129,976 / 14,529	111,231 / 12,448	143,201 / 15,964	85,749 / 9,554	165,924 / 18,293
Data Completeness (%)	90.4	92.2	99.74	97.6	96.0
No. of protein / water / ligand atoms	9,300 / 1,366 / 120	9,300 / 1,183 / 20	9,300 / 1,428 / 20	9,300 / 934 / 20	9,300 / 1,710 / 20
<B> for protein / water / ligand atoms (Å^2)	16.6 / 32.6 / 15.3	17.8 / 32.6 / 13.7	16.3 / 33.5 / 11.6	12.5 / 25.7 / 9.2	17.9 / 37.9 / 12.8
R-factor / R-free (%)†	15.4 / 19.6	15.9 / 20.3	16.1 / 19.5	16.4 / 20.9	16.1 / 19.9
rmsd bond lengths (Å) / bond angles (°)	.021 / 2.3	.017 / 2.1	.014 / 2.0	.022 / 2.0	.019 / 2.0

*, values in parentheses refer to the highest resolution shell. ‡, $R_{\text{merge}} = \Sigma_{hkl}\Sigma_i |I_{hkl,\,i} - \langle I_{hkl} \rangle| / \Sigma_{hkl} \langle I_{hkl} \rangle$ †, Rf and R-free = $(\Sigma \mid |F_o| - |F_c| \mid)/(\Sigma |F_o|)$

In the case of the Y130F, Y103F, Y103F/Y130F and T157V mutants, high resolution complexes were obtained with bound pyruvate, as well as one of Y130F with D-KDPG (Table 2.3.6). In all five structures the substrate was observed as a Schiff base intermediate (Figure 2.3.7**A**-**E**), with the α-keto acid moiety occupying the same position and engaging in equivalent interactions as in the wild-type structure. This demonstrated that the mutations do not prevent formation of the covalent carbinolamine, or its subsequent dehydration. Moreover, none of the structures display differences in active site architecture that might affect activity (Figure 2.3.7**F**). The only residue variations involve small shifts in the side chains that participate in the HB network, including Thr-43, which although not mutated is affected by changes in its immediate environment.

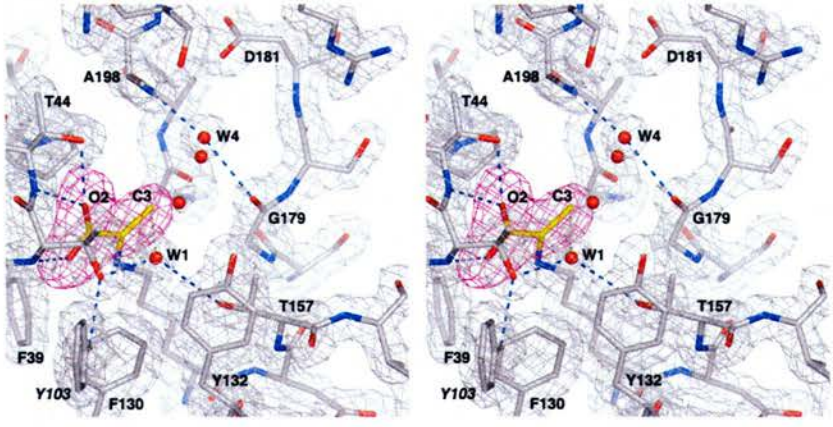
The ordered water, W1, also occupies different positions in each of the structures and displays varying degrees of order, in response to the availability of hydrogen bonding interactions. In the T157V mutant (Figure 2.3.7**D**) the molecule is shifted away from the valine side chain due to steric clashes. On the other hand, in Y130F (Figure 2.3.7**A**) it is positioned within hydrogen bonding distance of Thr-43 (2.9 Å) and makes a weak interaction with Thr-157 and the pyruvate carboxylate (3.8 Å and 3.6 Å, respectively), while in the double mutant (Figure 2.3.7**E**) W1 forms the HB with Thr-157 (3.2 Å) and the close contact with Thr-43 (3.8 Å), instead.

The changes in W1's position may provide clues to some of the observed kinetic parameters, such as the reduction in activity measured for the Thr157V mutant, which could partly be the result of steric repulsion of key substrate groups (e.g. the carbonyl of glyceraldehyde). Moreover, the residual activity retained by the Y130F mutants might be due to the presence of W1, which maintains an interaction with the pyruvate's C3 methyl while shifting enough in position to form a HB with Thr-43 and

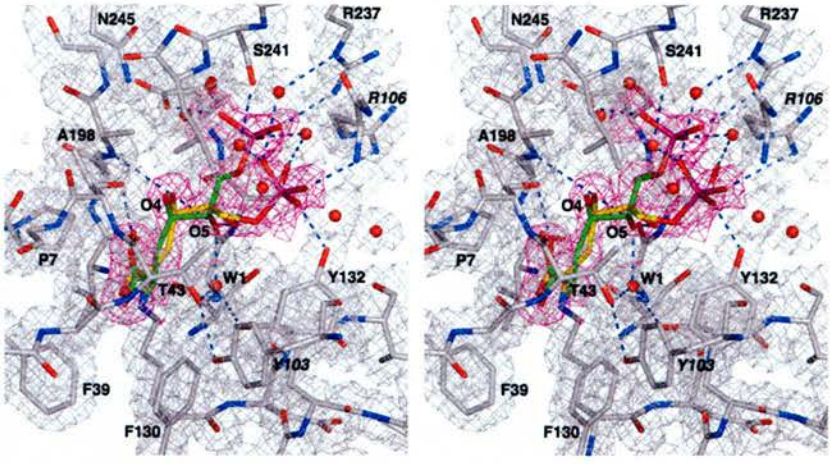
a close contact with the substrate's carboxylate group (3.6-3.8 Å). Also, the change in the water molecule's preferred pattern of interactions in the Y130F and double mutant, could be indicative of a change in the electrostatic nature of the Thr-43 hydroxyl in the absence of a hydrogen bond with Tyr-103. This in turn, could have some adverse effect on Tyr-130 during catalysis that was not apparent in the Y103F structure.

An additional, and potentially important, insight to catalysis was provided by the Y130F mutant complex with D-KDPG. In this structure the substrate was observed in two alternative conformations, corresponding to different positions of the phosphate and C5 hydroxyl groups. One of these phosphate positions, represents a previously

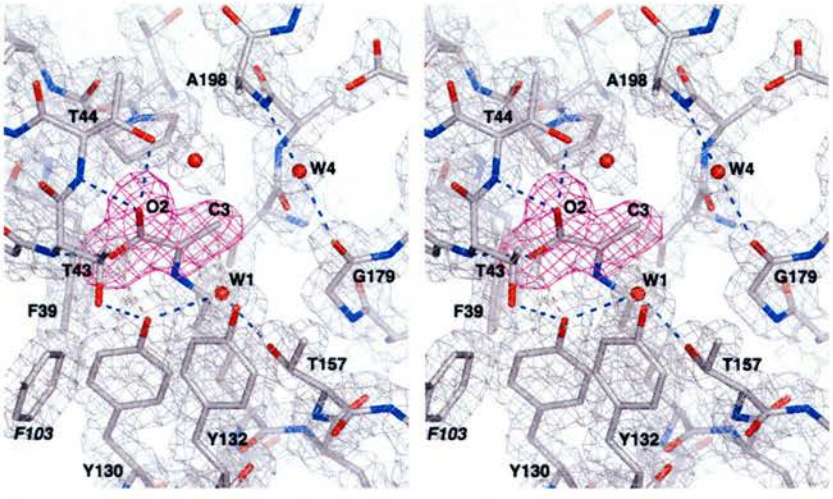
Figure 2.3.7 – SsKDGA mutant complexes.



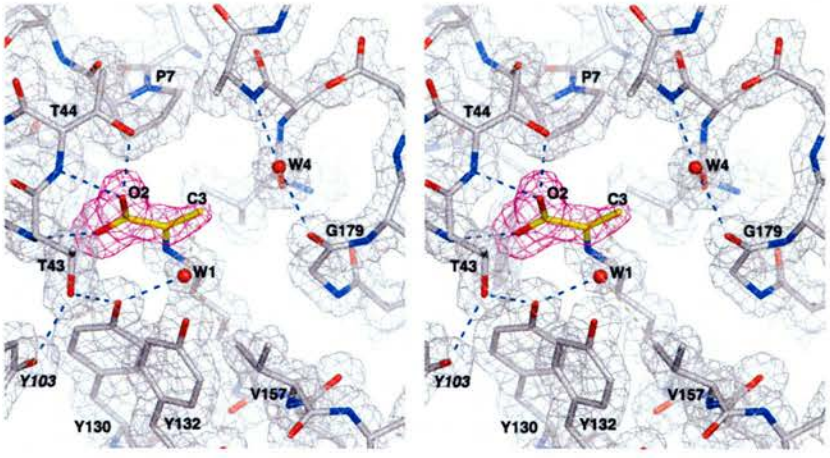
A. The refined model of SsKDGA mutant Y130F in complex with the Schiff base intermediate of pyruvate (PYR). Overlayed, the F_o-F_c map density for the substrate contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.5 σ (white). [Average interaction distances: **PYR(O2)-T44(Oγ)**, 2.6 Å; **PYR(O2)-T44(N)**, 2.8 Å; **PYR(O1)-T43(N)**, 2.8 Å; **PYR(O1)-F39**, 3.4 Å (close contact); **PYR(O2)-W1**, 3.6 Å (close contact); **PYR(C3)-W1**, 3.3 Å (close contact); **T43(Oγ)-W1**, 2.9 Å; **T157(Oγ)-W1**, 3.8 Å (close contact); **T43(Oγ)-*Y103***, 2.8 Å].



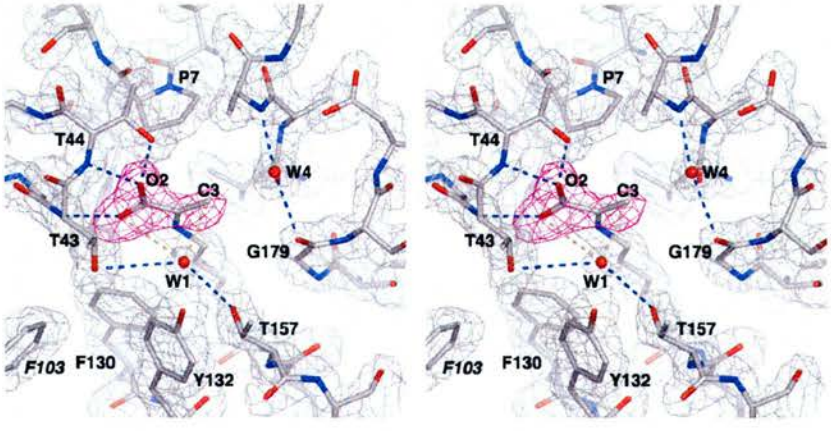
B. The refined model of SsKDGA mutant Y130F in complex with the Schiff base intermediate of D-KDPG (KPG). Overlayed, the F_o-F_c map density for the substrate contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.0 σ (white). [Average interaction distances: **KPG(O2)-T44(Oγ)**, 2.7 Å; **KPG(O2)-T44(N)**, 2.9 Å; **KPG(O1)-T43(N)**, 2.8 Å; **KPG(O1)-F39**, 3.6 Å (close contact); **KPG(O2)-W1**, 3.8 Å (close contact); **W1-T43(Oγ)**, 3.0 Å; **T157(Oγ)-W1**, 3.6 Å (close contact); **T43(Oγ)-*Y103***, 2.8 Å].



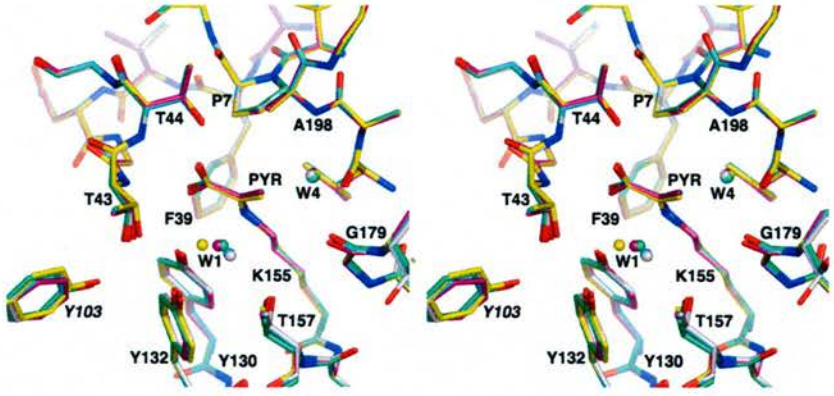
C. The refined model of SsKDGA mutant Y103F in complex with the Schiff base intermediate of pyruvate (PYR). Overlayed, the F_o-F_c map density for the substrate contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.5 σ (white). [Average interaction distances: **PYR(O2)-T44(Oγ)**, 2.6 Å; **PYR(O2)-T44(N)**, 2.8 Å; **PYR(O1)-T43(N)**, 2.8 Å; **PYR(O1)-F39**, 3.4 Å (close contact); **PYR(O1)-Y130**, 3.2 Å (close contact); **W1-Y130**, 2.7 Å; **PYR(C3)-W1**, 3.1 Å (close contact); **T43(Oγ)-Y130**, 2.7 Å; **T157(Oγ)-W1**, 2.8 Å].



D. The refined model of SsKDGA mutant T157V in complex with the Schiff base intermediate of pyruvate (PYR). Overlayed, the F_o-F_c map density for the substrate contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.5 σ (white). [Average interaction distances: **PYR(O2)-T44(Oγ)**, 2.7 Å; **PYR(O2)-T44(N)**, 2.8 Å; **PYR(O1)-T43(N)**, 2.8 Å; **PYR(O1)-F39**, 3.3 Å (close contact); **PYR(O1)-Y130**, 3.1 Å (close contact); **W1-Y130**, 2.7 Å; **PYR(C3)-W1**, 3.1 Å (close contact); **T43(Oγ)-Y130**, 2.6 Å; **T157(Cγ)-W1**, 3.3 Å (close contact); **T43(Oγ)-*Y103***, 2.7 Å].



E. The refined model of SsKDGA mutant Y103F/Y130F in complex with the Schiff base intermediate of pyruvate (PYR). Overlayed, the F_o-F_c map density for the substrate contoured at 2.5 σ (magenta) and the $2F_o$-F_c map density for the protein contoured at 1.5 σ (white). [Average interaction distances: **PYR(O2)-T44(Oγ)**, 2.6 Å; **PYR(O2)-T44(N)**, 2.8 Å; **PYR(O1)-T43(N)**, 2.8 Å; **PYR(O1)-F39**, 3.5 Å (close contact); **PYR(O2)-W1**, 3.8 Å (close contact); **PYR(C3)-W1**, 3.1 Å (close contact); **T43(Oγ)-W1**, 3.8 Å (close contact); **T157(Oγ)-W1**, 3.2 Å].



F. Active site superposition of the wild-type pyruvate (PYR) complex of SsKDGA (grey carbons), with the equivalent structures of the Y130F (yellow), Y103F (cyan), Y103F/Y130F (purple) and T157V (green) mutants. Residues are shown as sticks and labelled according to the wild-type enzyme. The ordered water molecules W1 and W4 are shown as spheres and coloured the same as the carbon atoms in the same model.

unknown coordination site that may be relevant to catalysis (e.g. at the carbinolamine stage), but which in any case provides yet another demonstration of the active site cavity's size and functionalisation. More importantly, however, the complex reveals the C4 hydroxyl bound in the position normally occupied by W4, where it is stabilised by Ala-198 and Gly-179.

From everything that is understood about the enzyme's catalytic mechanism this conformation is unlikely to represent a productive binding mode, capable of inducing C-C bond cleavage. Therefore, it appears that at least part of the catalytic role of Tyr-130 is to stabilise the C4 hydroxyl in a position suitable for proton abstraction. Moreover, preferential binding of D-Glyceraldehyde with its carbonyl group occupying the position of W4, could account for the significantly reduced activity recorded for the Y130F mutants in the condensation reaction. It is interesting that attempts to trap the D-KDG and L-KDGal acetonides in either Y130F or Y103F/Y130F have been unsuccessful, as these substrates lack the flexibility to bind in the manner observed here for D-KDPG. Consequently, determination of kinetic

parameters for these SsKDGA mutants with acetonide substrates could prove very instructive.

Summary

Rational design studies have been initiated in order to introduce altered specificities into the SsKDGA catalysed aldol condensation, including greater stereocontrol and activity with novel donor substrates. Crystal structures of the enzyme in complex with acetonide derivatives and pyruvate analogues have been obtained and characterised, providing additional insights to the enzyme's mechanisms of substrate binding and catalysis. The knowledge gained should help guide the process of selecting new targets for mutagenesis, so that initial successes (enzyme engineering, 80% selectivity for D-KDG; substrate engineering, > 90% specificity for D-KDG, or L-KDGal) can be built upon. In addition, mutants of SsKDGA shown by kinetic analysis to have reduced activity relative to the wild-type enzyme have been investigated structurally, in complex with natural substrates, in order to obtain a better understanding of their roles in catalysis.

SECTION 3 – DISCUSSION

SsKDGA has proved well suited to the crystallographic investigation of its mechanism. The preferred formation of covalent Schiff base complexes with natural substrates has enabled the study of intermediate stabilisation by the enzyme at an important step in the catalytic cycle. Moreover, the expected close approximation between the observed D-KD(P)G(al) intermediates and the TS leading to C-C bond cleavage, have permitted aspects of the catalytic process to be addressed, including the roles of particular residue and substrate groups, as well as the lack of stereoselectivity in the aldol reaction.

Central to the enzyme's promiscuity is its large and highly functionalised active site cavity, which is able to accommodate acceptor species of varying bulk and stereochemistry. A significant role has also been identified for water molecules forming the solvent structure in the active site. This ordered solvent, which can be considered as an integral part of the enzyme (Ringe and Petsko, 2003; Mattos, 2002), appears to bridge interactions between protein and ligand groups, thus helping to stabilise intermediates of alternative substrates in their preferred conformations.

Although the investigations described above have demonstrated and rationalised SsKDGA's ability to utilise substrates of different stereochemistries with comparable effieciency, they have also revealed that the enzyme has a significantly higher activity with phosphorylated relative to non-phosphorylated aldols, for which it also displays specific adaptations. This could raise question about the true extent of the enzyme's bifunctionality. It is worth noting, however, that SsKDGA's efficiency with non-phosphorylated substrates (k_{cat}/K_m, ~ 1 $s^{-1}mM^{-1}$) is far greater than that in other type I

aldolases that utilise phosphorylated substrates. EcKDPGA has been shown to display a catalytic efficiency of less than 0.02 $s^{-1}mM^{-1}$ for D-KDG (Fong et al., 2000), while in the case of EcDERA, the k_{cat}/K_m for D-2-deoxyribose (DR) has been measured at 2 $s^{1}M^{-1}$ (Silvestri et al., 2003). In contrast, all three enzymes catalyse the phosphorylated equivalents of these substrates with comparable efficiency (SsKDGA, 640 $s^{-1}mM^{-1}$; EcDERA, 110 $s^{-1}mM^{-1}$; EcKDPGA, 810 $s^{-1}mM^{-1}$).

It is also worth considering the fact that in solution D-KDPG(al) are present in measurable quantities as either an open chain keto or two furanose anomeric forms (Midelfort et al., 1977), while in the case of D-KDG(al) four different ring structures are detected by NMR, but not the keto (Plantierroyon et al., 1991a; Plantierroyon et al., 1991b). Therefore, a large contribution to the aldolase's affinity for phosphorylated substates, could be due to the greater availability of the active isomer, whichever one that may be.

The mechanistic studies carried out have revealed several key features associated with catalysis. Previous predictions for the involvement of Tyr-130 and the α-keto acid moiety's carboxylate have been confirmed. Moreover, it has been demonstrated that both groups are required for the efficient general base activity that leads to enamine formation, or C-C bond cleavage. The crystallographic complexes, however, have also highlighted an inconsistency between the relative orientations of the hydroxyl and carboxylate oxygen atoms and the proton transfer predicted to occur between them. Their roles may therefore require reinterpretation.

One possibility is that the positioning of the negatively charged carboxylate group in close proximity to Tyr-130 causes a change in the latter's ionisation properties. This

may then permit the residue to transiently bind a second hydrogen atom, leading to the water mediated proton abstraction from the C3 methyl, or from the C4 hydroxyl. However, the step that would follow this proton abstraction is less clear, as donation of the proton to the emerging enamine (in the cleavage direction), or aldol (in the condensation reaction) would first require exchange to take place between the acceptor carbonyl and a water molecule.

Another stage of the catalytic cycle that remains unclear is that of carbinolamine formation and breakdown. The structural evidence has revealed that this can take place in the absence of either the Tyr-130 or Thr-157 hydroxyls, as well as that of substrate's carboxylate group. Moreover, while a crystal structure of one of the T43V mutants in complex with a substrate has not been obtained yet, it is unlikely that this residue can influence the intermediate independently of the intervening tyrosine. Therefore, the substrates' transition from the Michaelis complex to the covalent iminium appears to only require the catalytic lysine and perhaps a water molecule.

Aldolases have emerged as a powerful tool in biocatalysis by virtue of their ability to catalyse stereoselective carbon-carbon bond formation (Silvestri et al., 2003). Furthermore, their broad acceptor substrate profile allows them to be used for the synthesis of a large number of hydroxylated natural products and non-natural carbohydrates that cannot be accessed by conventional synthetic techniques without complex protecting-group chemistry (Fessner, 1998). A number of aldolases have been developed for biocatalytic appliations. FBPA has been employed in the chemoenzymatic synthesis of a number of mycin antibiotics, thiosugars, azasugars and natural alkaloids (Fessner, 1998). EcNAL has been developed for the efficient

large-scale production of sialic acid, an intermediate of influenza drugs (von Itzstein, 1993). EcDERA has been employed in the production of key intermediates of the anticancer agents, epothilones (Machajewski and Wong, 2000) and of statin intermediates (Greenberg et al., 2004). L-threonine aldolase has been used in the synthesis of protent mycestericin immunosuppressants (Nishide et al., 2000).

SsKDGa also displays considerable potential for exploitation in biocatalysis: it has a broad specificity, is extremely thermostable and can be efficiently expressed in E. coli (Buchanan et al., 1999). Moreover, the 4-hydroxy-2-ketobutyrate product framework contains four different oxidation states on four contiguous carbons, potentially providing access to a range of different functional molecules (Shelton et al., 1996). It has also been demonstrated in principle that the enzyme's lack of stereocontrol, normally a problem in chiral synthesis, can be induced and that it may even be possible to tailor it to particular applications. Finally the enzyme's utility as a biocatalyst has been further demonstrated by its immobiliation and use for multi-gram syntheses inside a continuous column reactor (Lamble, 2004).

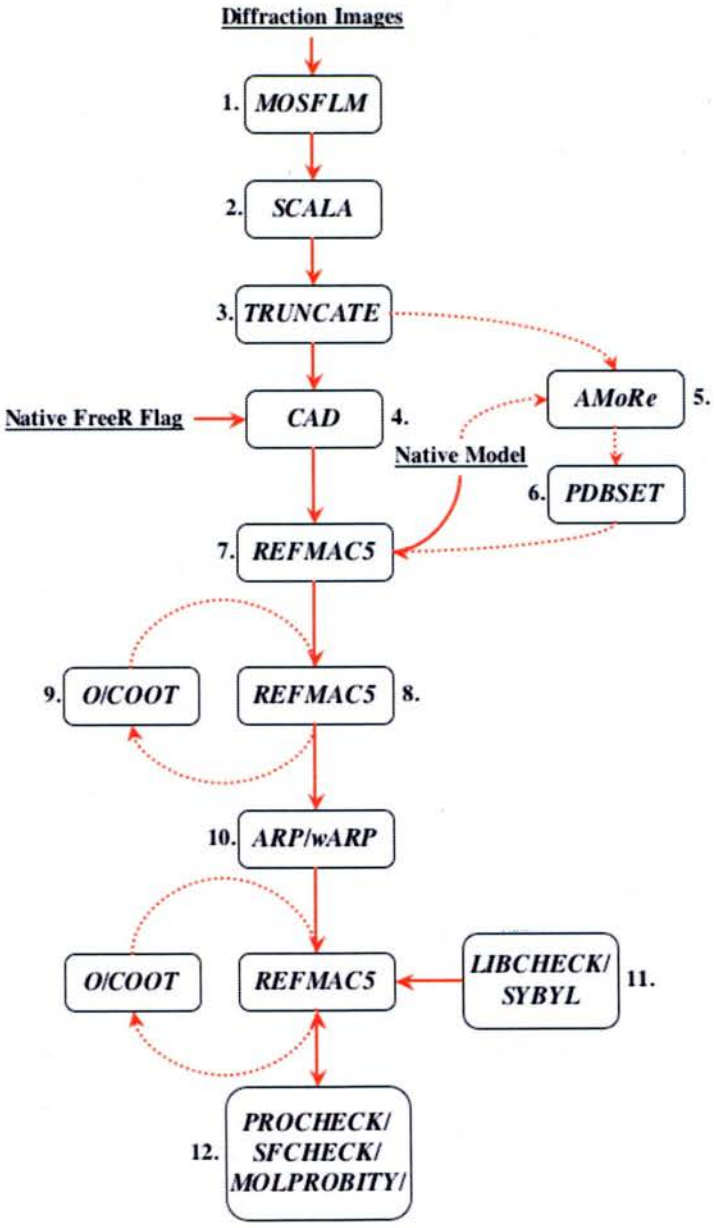
An interesting demonstration of SsKDGA's utility may be achieved by attempting to rationally engineer NAL activity into it. The similarity between the enzymes, particularly with respect to their active site architecture and likely mechanism suggests a high probability of success. Moreover, if sufficiently high activity can be achieved, SsKDGA's greater stability could make it a useful alternative to EcNAL in the synthesis of sialic acid (Ferrero et al., 1996).

The potential of the rational design approach in engineering *S. solfataricus* aldolase has yet to be fully proven and a directed evolution methodology may eventually need

to be implemented. This study has, however, afforded a detailed insight to the structural basis of the enzyme's mechanism of substrate binding and catalysis that will prove invaluable in future work.

APPENDIX 1

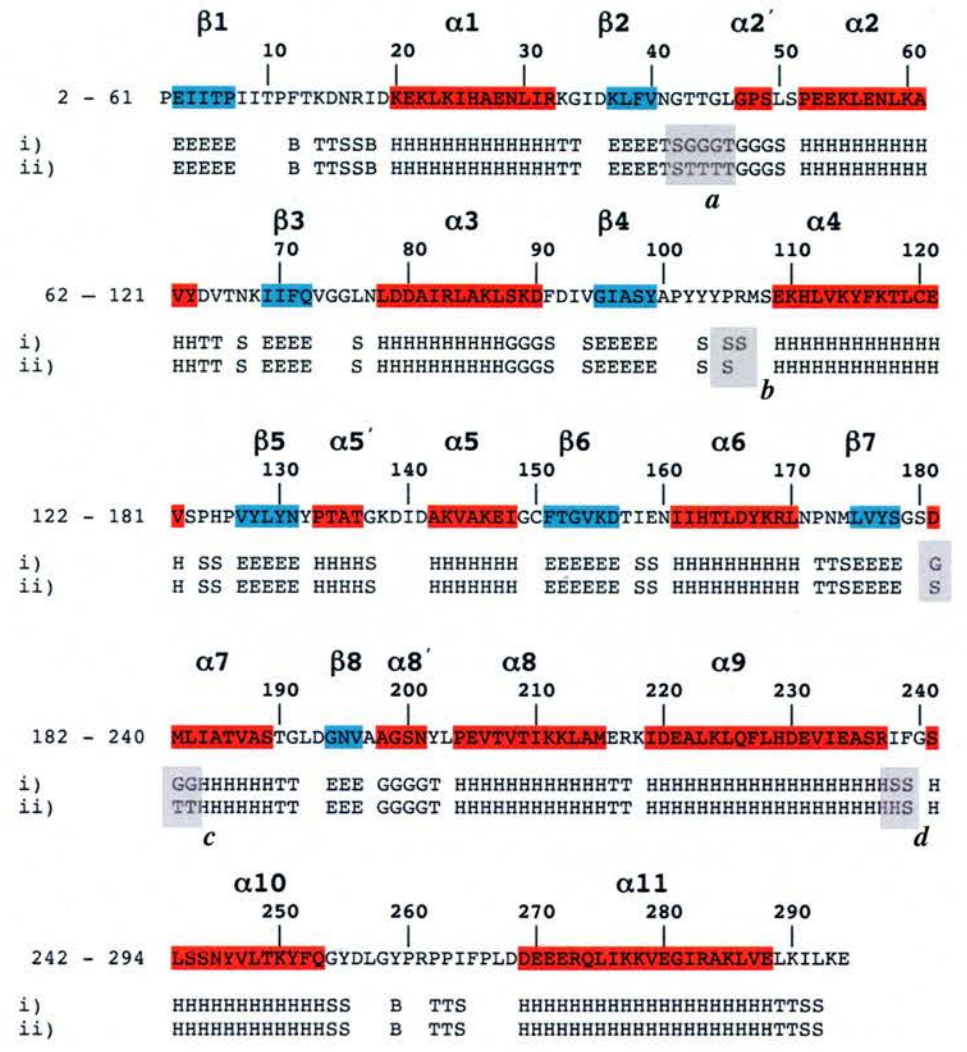
Data Processing and Refinement



Flow diagram outlining the general procedure used for analysing the crystallographic data obtained of the various complexes of SsKDGA and its natural substrates. Key: **1.** indexing and integration; **2.** scaling and merging; **3.** conversion of intensities to amplitudes; **4.** assigning reflections to a *test* set; **5.** molecular replacement; **6.** generation of a starting model from the MR solution; **7.** rigid body refinement; **8.** restrained refinement; **9.** manual corrections/adjustments to the model; **10.** automatic building of the solvent structure; **11.** generation of small molecule coordinates and geometric parameters; **12.** validation routines.

APPENDIX 2

S. solfataricus KDG Aldolase - Secondary Structure

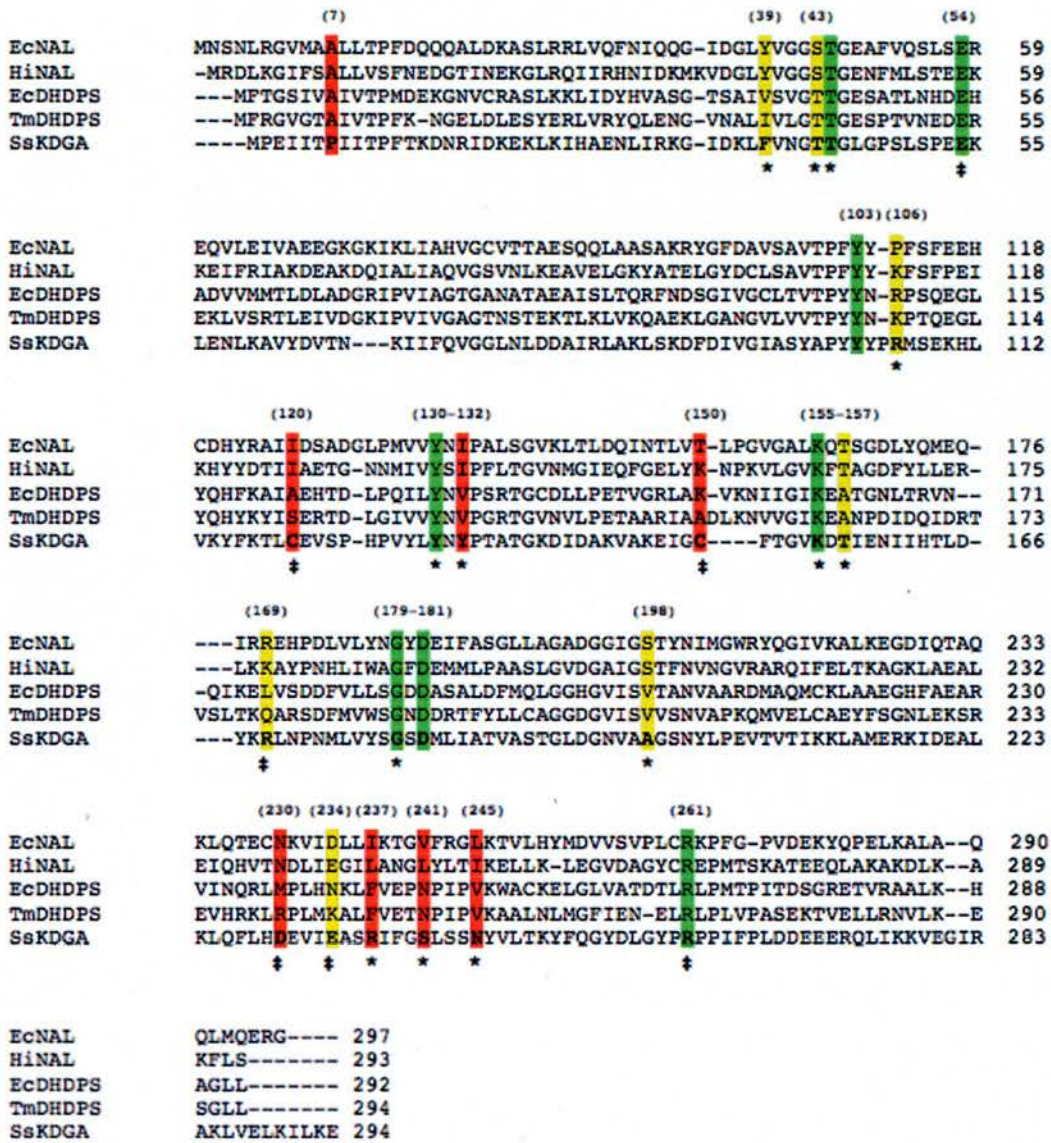


The amino acid sequence of SsKDGA with helices (α1 - α11) and sheets (β1 - β8) highlighted in red and tuquoise, respectively. Determination of secondary structure was carried out with *DSSP* (Kabsch and Sander, 1983), using the online server at http://bioweb.pasteur.fr/seqanal/interfaces/dssp.html. The assignment for individual residues is presented in rows **(i)** and **(ii)**, each displaying one of two alternatives consistently observed in the active site regions of different models (grey boxes): ***a.*** In all of the substrate complexes residues T43-G45 were consistently identified as turns, while in certain subunits of the two apoenzyme structures they were also defined as a 3_{10}-helix; ***b.*** Arg-106 was assigned as an irregular element in the D-KDPG(al) complexes and not as a bend, as in the other structures; ***c.*** Residues D181-L183 were consistently identified as a 3_{10}-helix in the apoenzyme and D-KDPG(al) structures, but in certain subunits of the non-phosphorylated substrate complexes were only defined as a bend and two turns; ***d.*** Ile-238 was inconsistently assigned in the different structures as either an isolated turn, or as part of helix α9.

DSSP code: B, residue in isolated β-bridge; **E**, extended strand, participating in a β-ladder; **G**, 3_{10}-helix; **H**, α-helix; **S**, bend; **T**, hydrogen bonded turn; **blank,** loop or irregular.

APPENDIX 3

Alignment of NAL Subfamily Members



Sequence alignment for five members of the NAL subfamily; *E. coli* NAL, *H. influenzae* NAL, *E. coli* DHDPS, *T. maritima* DHDPS and *S. solfataricus* KDGA. Key residues in SsKDGA are marked in bold and their sequence number shown in brackets. The degree of conservation within the subfamily relative to KDGA is highlighted in **green** (strict conservation), **yellow** (partial conservation) or **red** (no conservation). Moreover, those residues with a specific structural role in KDGA are identified by the symbol (‡). A known role in catalysis and/or substrate binding, on the other hand, is indicated by (*). The alignment was carried out using *CLUSTALW* version 1.82 (Chenna *et al.*, 2003).

REFERENCES

Abe, F., and Horikoshi, K. (2001). The biotechnological potential of piezophiles. Trends Biotechnol *19*, 102-108.

Agarwal, P. K. (2005). Role of protein dynamics in reaction rate enhancement by enzymes. J Am Chem Soc *127*, 15248-15256.

Agarwal, P. K. (2006). Enzymes: An integrated view of structure, dynamics and function. Microb Cell Fact *5*, 2.

Ahmed, H., Ettema, T. J., Tjaden, B., Geerling, A. C., van der Oost, J., and Siebers, B. (2005). The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. Biochem J *390*, 529-540.

Ahmed, H., Tjaden, B., Hensel, R., and Siebers, B. (2004). Embden-Meyerhof-Parnas and Entner-Doudoroff pathways in Thermoproteus tenax: metabolic parallelism or specific adaptation? Biochem Soc Trans *32*, 303-304.

Aisaka, K., Igarashi, A., Yamaguchi, K., and Uwajima, T. (1991). Purification, crystallization and characterization of N-acetylneuraminate lyase from Escherichia coli. Biochem J *276 (Pt 2)*, 541-546.

Albers, S. V., Elferink, M. G., Charlebois, R. L., Sensen, C. W., Driessen, A. J., and Konings, W. N. (1999). Glucose transport in the extremely thermoacidophilic Sulfolobus solfataricus involves a high-affinity membrane-integrated binding protein. J Bacteriol *181*, 4285-4291.

Allard, J., Grochulski, P., and Sygusch, J. (2001). Covalent intermadiate trapped in 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase structure at 1.95-A resolution. PNAS *98*, 3679-3684.

Allen, F. H., Davies, J. E., Galloy, J. J., Johnson, O., Kennard, O., Macrae, C. F., Mitchell, E. M., Mitchell, G. F., Smith, J. M., and Watson, D. G. (1991). The development of versions 3 and 4 of the Cambridge Structural Database system. J Chem Inf Comput Sci *31*, 187-204.

Andrews, P. R., Smith, G. D., and Young, I. G. (1973). Transition-state stabilization and enzymic catalysis. Kinetic and molecular orbital studies of the rearrangement of chorismate to prephenate. Biochemistry *12*, 3492-3498.

Aplin, R. T., Baldwin, J. E., Schofield, C. J., and Waley, S. G. (1990). Use of electrospray mass spectrometry to directly observe an acyl enzyme intermediate in beta-lactamase catalysis. FEBS Lett *277*, 212-214.

Auge, C., Bouxom, B., Cavaye, B., and Gautheron, C. (1989). Scope and Limitations of the Aldol Condensation Catalyzed by Immobilized Acylneuraminate Pyruvate Lyase. Tetrahedron Letters *30*, 2217-2220.

Auge, C., and Delest, V. (1993). Microbiological aldolisations - synthesis of 2-keto-3-deoxy-D-gluconate. Tetrahedron: Assymetry *4*, 1165-1168.

Auge, C., and Delest, V. (1995). The use of *Aspergillus terreus* extracts in the preparative synthesis of 2-keto-3-deoxy-ulosonic acids. Tetrahedron: Assymetry *6*, 863-866.

Auge, C., Gautheron, C., David, S., Malleron, A., Cavaye, B., and Bouxom, B. (1990). Sialyl Aldolase in Organic-Synthesis - from the Trout Egg Acid, 3-Deoxy-D-Glycero-D-Galacto-2-Nonulosonic Acid (Kdn), to Branched-Chain Higher Ketoses as Possible New Chirons. Tetrahedron *46*, 201-214.

Baker, E. N., and Hubbard, R. E. (1984). Hydrogen bonding in globular proteins. Prog Biophys Mol Biol *44*, 97-179.

Bakker, H. J., and Nienhuys, H. K. (2002). Delocalization of protons in liquid water. Science *297*, 587-590.

Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science *289*, 905-920.

Barbas, C. F., 3rd, Heine, A., Zhong, G., Hoffmann, T., Gramatikova, S., Bjornestedt, R., List, B., Anderson, J., Stura, E. A., Wilson, I. A., and Lerner, R. A. (1997). Immune versus natural selection: antibody aldolases with enzymic rates but broader scope. Science *278*, 2085-2092.

Barbas, C. F., 3rd, Wang, Y. F., and Wong, C. H. (1990). Deoxyribose-5-phosphate aldolase as a synthetic catalyst. J Am Chem Soc *112*, 2013-2014.

Barbosa, J. A., Smith, B. J., DeGori, R., Ooi, H. C., Marcuccio, S. M., Campi, E. M., Jackson, W. R., Brossmer, R., Sommer, M., and Lawrence, M. C. (2000). Active site modulation in the N-acetylneuraminate lyase sub-family as revealed by the structure of the inhibitor-complexed Haemophilus influenzae enzyme. J Mol Biol *303*, 405-421.

Barns, S. M., Delwiche, C. F., Palmer, J. D., Dawson, S. C., Hershberger, K. L., and Pace, N. R. (1996). Phylogenetic perspective on microbial life in hydrothermal ecosystems, past and present. Ciba Found Symp *202*, 24-32; discussion 32-29.

Baumann, W., Freidenreich, J., Weisshaar, G., Brossmer, R., and Friebolin, H. (1989). [Cleavage and synthesis of sialic acids with aldolase. NMR studies of stereochemistry, kinetics, and mechanisms]. Biol Chem Hoppe Seyler *370*, 141-149.

Bednarski, M., Simon, E., Bischofberger, N., Fessner, W. D., Kim, M., Lees, W., Salto, T., Waldmann, H., and Whitesides, G. M. (1989). Rabbit muscle aldolase as a catalyst in organic synthesis. Journal of American Chemical Society *111*, 627-635.

Benkovic, S. J., and Hammes-Schiffer, S. (2003). A perspective on enzyme catalysis. Science *301*, 1196-1202.

Berthiaume, L., Tolan, D. R., and Sygusch, J. (1993). Differential usage of the carboxyl-terminal region among aldolase isozymes. J Biol Chem *268*, 10826-10835.

Binkowski, T. A., Naghibzadeh, S., and Liang, J. (2003). CASTp: Computed Atlas of Surface Topography of proteins. Nucleic Acids Res *31*, 3352-3355.

Blickling, S., Beisel, H. G., Bozic, D., Knablein, J., Laber, B., and Huber, R. (1997a). Structure of dihydrodipicolinate synthase of Nicotiana sylvestris reveals novel quaternary structure. J Mol Biol *274*, 608-621.

Blickling, S., and Knablein, J. (1997). Feedback inhibition of dihydrodipicolinate synthase enzymes by L-lysine. Biol Chem *378*, 207-210.

Blickling, S., Renner, C., Laber, B., Pohlenz, H. D., Holak, T. A., and Huber, R. (1997b). Reaction mechanism of Escherichia coli dihydrodipicolinate synthase

investigated by X-ray crystallography and NMR spectroscopy. Biochemistry *36*, 24-33.

Blochl, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H. W., and Stetter, K. O. (1997). Pyrolobus fumarii, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. Extremophiles *1*, 14-21.

Blom, N., and Sygusch, J. (1997). Product binding and role of the C-terminal region in class I D-fructose 1,6-bisphosphate aldolase. Nat Struct Biol *4*, 36-39.

Blonski, C., De Moissac, D., Perie, J., and Sygusch, J. (1997). Inhibition of rabbit muscle aldolase by phosphorylated aromatic compounds. Biochem J *323 (Pt 1)*, 71-77.

Bondi, A. (1964). J Phys Chem *68*, 441.

Bordo, D., and Argos, P. (1994). The role of side-chain hydrogen bonds in the formation and stabilization of secondary structure in soluble proteins. J Mol Biol *243*, 504-519.

Borthwick, E. B., Connell, S. J., Tudor, D. W., Robins, D. J., Shneier, A., Abell, C., and Coggins, J. R. (1995). Escherichia coli dihydrodipicolinate synthase: characterization of the imine intermediate and the product of bromopyruvate treatment by electrospray mass spectrometry. Biochem J *305 (Pt 2)*, 521-524.

Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. (1972). Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch Mikrobiol *84*, 54-68.

Brown, J. R., and Doolittle, W. F. (1997). Archaea and the prokaryote-to-eukaryote transition. Microbiol Mol Biol Rev *61*, 456-502.

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., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54 (Pt 5)*, 905-921.

Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J., and Taylor, R. (2002). New software for searching the Cambridge

Structural Database and visualizing crystal structures. Acta Crystallogr B *58*, 389-397.

Bruno, I. J., Cole, J. C., Lommerse, J. P., Rowland, R. S., Taylor, R., and Verdonk, M. L. (1997). IsoStar: a library of information about nonbonded interactions. J Comput Aided Mol Des *11*, 525-537.

Buchanan, C. L., Connaris, H., Danson, M. J., Reeve, C. D., and Hough, D. W. (1999). An Extremely thermostable aldolase from Solfolobus sofataricus with specificity for non-phosphorylated substrates. Biochemistry Journal *343*, 563-570.

Budgen, N., and Danson, M. J. (1986). Metabolism of glucose via a modified entner-doudoroff pathway in the thermoacidophilic archaebacterium Thermoplasma acidophilum. FEBS Lett *196*, 207-210.

Cannon, W. R., and Benkovic, S. J. (1998). Solvation, reorganization energy, and biological catalysis. J Biol Chem *273*, 26257-26260.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

Chen, L., Brugger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H. P., and Garrett, R. A. (2005). The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota. J Bacteriol *187*, 4992-4999.

Chen, L., Dumas, D. P., and Wong, C.-H. (1992). Deoxyribose-5-phosphate aldolase as a catalyst in asymmetric aldol condensation. Journal of American Chemical Society *114*, 741-748.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Research *31*, 3497-3500.

Choi, K. H., Mazurkie, A. S., Morris, A. J., Utheza, D., Tolan, D. R., and Allen, K. N. (1999). Structure of a fructose-1,6-bis(phosphate) aldolase liganded to its natural substrate in a cleavage-defective mutant at 2.3 A. Biochemistry *38*, 12655-12664.

Choi, K. H., Shi, J., Hopkins, C. E., Tolan, D. R., and Allen, K. N. (2001). Snapshots of catalysis: the structure of fructose-1,6-(bis)phosphate aldolase covalently bound to the substrate dihydroxyacetone phosphate. Biochemistry *40*, 13868-13875.

Choi, K. H., and Tolan, D. R. (2004). Presteady-state kinetic evidence for a ring-opening activity in fructose-1,6-(bis)phosphate aldolase. J Am Chem Soc *126*, 3402-3403.

Ciaramella, M., Pisani, F. M., and Rossi, M. (2002). Molecular biology of extremophiles: recent progress on the hyperthermophilic archaeon Sulfolobus. Antonie Van Leeuwenhoek *81*, 85-97.

Clayden, J., Greeves, N., Warren, S. G., and Wothers, P. (2001). Organic chemistry, Repr. with corr. edn (Oxford: Oxford University Press).

Cleland, W. W., Frey, P. A., and Gerlt, J. A. (1998). The low barrier hydrogen bond in enzymatic catalysis. J Biol Chem *273*, 25529-25532.

Cleland, W. W., and Kreevoy, M. M. (1994). Low-barrier hydrogen bonds and enzymic catalysis. Science *264*, 1887-1890.

Clore, G. M., and Schwieters, C. D. (2006). Concordance of Residual Dipolar Couplings, Backbone Order Parameters and Crystallographic B-factors for a Small alpha/beta Protein: A Unified Picture of High Probability, Fast Atomic Motions in Proteins. J Mol Biol *355*, 879-886.

Comb, D. G., and Roseman, S. (1960). The sialic acids. I. The structure and enzymatic synthesis of N-acetylneuraminic acid. J Biol Chem *235*, 2529-2537.

Connaris, H., West, S. M., Hough, D. W., and Danson, M. J. (1998). Cloning and overexpression in Escherichia coli of the gene encoding citrate synthase from the hyperthermophilic Archaeon Sulfolobus solfataricus. Extremophiles *2*, 61-66.

Conway, T. (1992). The Entner-Doudoroff pathway: history, physiology and molecular biology. FEMS Microbiol Rev *9*, 1-27.

Cooper, R. A. (1978). Intermediary metabolism of monosaccharides by Bacteria, In Biochemistry of Carbohydrates II (International Review of Biochemistry: (Ed. Manners, D. J.)), pp. 37-73.

Copley, R. R., and Bork, P. (2000). Homology among (beta alpha)(8) barrels: Implications for the evolution of metabolic pathways. Journal of Molecular Biology *303*, 627-640.

Cotterill, I. C., Henderson, D. P., Shelton, M. C., and Toone, E. J. (1998). The synthetic utility of KDPGal aldolase. Journal of Molecular Catalysis B-Enzymatic *5*, 103-111.

Coulter, C. V., Gerrard, J. A., Kraunsoe, J. A. E., Moore, D. J., and Pratt, A. J. (1996). Tetrahedron *52*, 7127-7136.

Dalby, A., Dauter, Z., and Littlechild, J. A. (1999). Crystal structure of human muscle aldolase complexed with fructose 1,6-bisphosphate: mechanistic implications. Protein Sci *8*, 291-297.

Dalby, A. R., Tolan, D. R., and Littlechild, J. A. (2001). The structure of human liver fructose-1,6-bisphosphate aldolase. Acta Crystallogr D Biol Crystallogr *57*, 1526-1533.

Danson, M. J., Black, S. C., Woodland, D. L. and Wood, P. A. (1985). Citric acid cycle enzymes of the archaebacteria: citrate synthase and succinate thiokinase. FEBS Letters *179*, 120-124.

Danson, M. J. (1993). Central metabolism of the Archaea, In The biochemistry of Archaea (Archaebacteria) (Elsevier Biomedical Press: Amsterdam, The Netherlands), pp. 1-24.

Dao-pin, S., Anderson, D. E., Baase, W. A., Dahlquist, F. W., and Matthews, B. W. (1991). Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. Biochemistry *30*, 11521-11529.

DeLano, W. L. (2002). The PyMOL User's Manual (San Carlos, CA, USA: DeLano Scientific).

De Ley, J., and Doudoroff, M. (1957). The metabolism of D-galactose in Pseudomonas saccharophila. J Biol Chem *227*, 745-757.

De Rosa, M., Gambacorta, A., Nicolaus, B., Giardina, P., Poerio, E., and Buonocore, V. (1984). Glucose metabolism in the extreme thermoacidophilic archaebacterium Sulfolobus solfataricus. Biochem J *224*, 407-414.

Deijl, C. M., and Vliegenthart, J. F. (1983). Configuration of substrate and products of N-acetylneuraminate pyruvate-lyase from Clostridium perfringens. Biochem Biophys Res Commun *111*, 668-674.

Derewenda, Z. S., Lee, L., and Derewenda, U. (1995). The occurrence of C-H.O hydrogen bonds in proteins. J Mol Biol *252*, 248-262.

Dickerson, T. J., Lovell, T., Meijler, M. M., Noodleman, L., and Janda, K. D. (2004). Nornicotine aqueous aldol reactions: synthetic and theoretical investigations into the origins of catalysis. J Org Chem *69*, 6603-6609.

Dobson, R. C., Devenish, S. R., Turner, L. A., Clifford, V. R., Pearce, F. G., Jameson, G. B., and Gerrard, J. A. (2005a). Role of arginine 138 in the catalysis and regulation of Escherichia coli dihydrodipicolinate synthase. Biochemistry *44*, 13007-13013.

Dobson, R. C., Griffin, M. D., Jameson, G. B., and Gerrard, J. A. (2005b). The crystal structures of native and (S)-lysine-bound dihydrodipicolinate synthase from Escherichia coli with improved resolution show new features of biological significance. Acta Crystallogr D Biol Crystallogr *61*, 1116-1124.

Dobson, R. C., Valegard, K., and Gerrard, J. A. (2004). The crystal structure of three site-directed mutants of Escherichia coli dihydrodipicolinate synthase: further evidence for a catalytic triad. J Mol Biol *338*, 329-339.

Doudna, J. A., and Cech, T. R. (2002). The chemical repertoire of natural ribozymes. Nature *418*, 222-228.

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*, 458-466.

Dubois, J. E., and Fort, J. F. (1972). Dynamic Stereochemistry of Aldolization.21. Definition of Restoring Energy of a System of Reversible Competitive Reactions. Tetrahedron *28*, 1665-&.

Eaton, R. W. (1994). Organization and evolution of naphthalene catabolic pathways: sequence of the DNA encoding 2-hydroxychromene-2-carboxylate isomerase and trans-o-hydroxybenzylidenepyruvate hydratase-aldolase from the NAH7 plasmid. J Bacteriol *176*, 7757-7762.

Eisenmesser, E. Z., Bosco, D. A., Akke, M., and Kern, D. (2002). Enzyme dynamics during catalysis. Science *295*, 1520-1523.

Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., and Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. Nature *438*, 117-121.

Elferink, M. G., Albers, S. V., Konings, W. N., and Driessen, A. J. (2001). Sugar transport in Sulfolobus solfataricus is mediated by two families of binding protein-dependent ABC transporters. Mol Microbiol *39*, 1494-1503.

Elzainy, T. A., Hassan, M. M., and Allam, A. M. (1973). New pathway for nonphosphorylated degradation of gluconate by Aspergillus niger. J Bacteriol *114*, 457-459.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Entner, N., and Doudoroff, M. (1952). Glucose and gluconic acid oxidation of Pseudomonas saccharophila. J Biol Chem *196*, 853-862.

Evans, P. R. (1993). Data reduction, Paper presented at: Proceedings of the CCP4 study weekend. Data collection and processing (Warrington: Daresbury Laboratory).

Feingold, D. S., and Hoffee, P. A. (1972). Other Deoxy Sugar Aldolases, In The Enzymes, P. D. Boyer, ed. (New York; London: Academic Press), pp. 315-321.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science *246*, 64-71.

Ferrero, M. A., Reglero, A., Fernandez-Lopez, M., Ordas, R., and Rodriguez-Aparicio, L. B. (1996). N-acetyl-D-neuraminic acid lyase generates the sialic acid for colominic acid biosynthesis in Escherichia coli K1. Biochem J *317 (Pt 1)*, 157-165.

Fersht, A. (1985). Enzyme structure and mechanism, 2nd edn (New York: Freeman).

Fessner, W. D. (1998). Enzyme mediated C-C bond formation. Curr Opin Chem Biol *2*, 85-97.

Fessner, W. D., Schneider, A., Held, H., Sinerius, G., Walter, C., Hixon, M., and Schloss, J. V. (1996). The Mechanism of Class II, Metal-Dependent Aldolases. Angew Chem Int Ed Engl *35*, 2219-2221.

Fessner, W. D., and Walter, C. (1996). Top Curr Chem *184*, 97-194.

Fitz, W., Schwark, J. R., and Wong, C. H. (1995). Aldotetroses and C(3)-modified aldohexoses as substrates for N-acetylneuraminic acid aldolase: a model for the explanation of the normal and the inversed stereoselectivity. J Org Chem *60*, 3663-3670.

Fletcher, D. A., McMeeking, R. F., and Parkin, D. (1996). The United Kingdom Chemical Database Service. Journal of Chemical Information and Computer Sciences *36*, 746-749.

Fong, S., Machajewski, T. D., Mak, C. C., and Wong, C. (2000). Directed evolution of D-2-keto-3-deoxy-6-phosphogluconate aldolase to new variants for the efficient synthesis of D- and L-sugars. Chem Biol *7*, 873-883.

Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S., and Woese, C. R. (1977). Classification of methanogenic bacteria by 16S ribosomal RNA characterization. Proc Natl Acad Sci U S A *74*, 4537-4541.

Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., *et al.* (1980). The phylogeny of prokaryotes. Science *209*, 457-463.

French, G. S., and Wilson, K. S. (1978). Acta Crystallogr *A34*, 517.

Fullerton, S. W., Griffiths, J. S., Merkel, A. B., Cheriyan, M., Wymer, N. J., Hutchins, M. J., Fierke, C. A., Toone, E. J., and Naismith, J. H. (2006). Mechanism of the Class I KDPG aldolase. Bioorg Med Chem. *14*, 3002-10

Fusz, S., Eisenfuhr, A., Srivatsan, S. G., Heckel, A., and Famulok, M. (2005). A ribozyme for the aldol reaction. Chem Biol *12*, 941-950.

Gamblin, S. J., Cooper, B., Millar, J. R., Davies, G. J., Littlechild, J. A., and Watson, H. C. (1990). The crystal structure of human muscle aldolase at 3.0 A resolution. FEBS Lett *262*, 282-286.

Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D. G. (2004). How enzymes work: analysis by modern rate theory and computer simulations. Science *303*, 186-195.

Gefflaut, T., Blonski, C., Perie, J., and Willson, M. (1995). Class I aldolases: substrate specificity, mechanism, inhibitors and structural aspects. Prog Biophys Mol Biol *63*, 301-340.

Geissler, P. L., Dellago, C., Chandler, D., Hutter, J., and Parrinello, M. (2001). Autoionization in liquid water. Science *291*, 2121-2124.

Gerlt, J. A., and Gassman, P. G. (1993). Understanding the rates of certain enzyme-catalyzed reactions: proton abstraction from carbon acids, acyl-transfer reactions, and displacement reactions of phosphodiesters. Biochemistry *32*, 11943-11952.

Giardina, P., de Biasi, M. G., de Rosa, M., Gambacorta, A., and Buonocore, V. (1986). Glucose dehydrogenase from the thermoacidophilic archaebacterium Sulfolobus solfataricus. Biochem J *239*, 517-522.

Gijsen, H., and Wong, C.-H. (1994). Unprecedented Asymmetric Aldol Reactions with Three Aldehyde Substrates Catalysed by 2-Deoxyribose-5-phosphate Aldolase. Journal of American Chemical Society *116*, 8422-8423.

Gijsen, H. J., Qiao, L., Fitz, W., and Wong, C. H. (1996). Recent Advances in the Chemoenzymatic Synthesis of Carbohydrates and Carbohydrate Mimetics. Chem Rev *96*, 443-474.

Gould, S. M., and Tawfik, D. S. (2005). Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. Biochemistry *44*, 5444-5452.

Grazi, E., Meloche, H., Martinez, G., Wood, W. A., and Horecker, B. L. (1963). Evidence for Schiff base formation in enzymatic aldol condensations. Biochem Biophys Res Commun *10*, 4-10.

Grazi, E., Rowley, P. T., Cheng, T., Tchola, O., and Horecker, B. L. (1962). The mechanism of action of aldolases. III. Schiff base formation with lysine. Biochem Biophys Res Commun *9*, 38-43.

Greenberg, W. A., Varvak, A., Hanson, S. R., Wong, K., Huang, H., Chen, P., and Burk, M. J. (2004). Development of an efficient, scalable, aldolase-catalyzed process

for enantioselective synthesis of statin intermediates. Proc Natl Acad Sci U S A *101*, 5788-5793.

Grogan, D. W. (1989). Phenotypic characterization of the archaebacterial genus Sulfolobus: comparison of five wild-type strains. J Bacteriol *171*, 6710-6719.

Grogan, D. W. (1991). Evidence that beta-Galactosidase of Sulfolobus solfataricus Is Only One of Several Activities of a Thermostable beta-d-Glycosidase. Appl Environ Microbiol *57*, 1644-1649.

Haapalainen, A. M., Merilainen, G., and Wierenga, R. K. (2006). The thiolase superfamily: condensing enzymes with diverse reaction specificities. Trends Biochem Sci *31*, 64-71.

Hagler, A. T., Huler, E., and Lifson, S. (1974). Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. J Am Chem Soc *96*, 5319-5327.

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*, 383-394.

Haseltine, C., Rolfsmeier, M., and Blum, P. (1996). The glucose effect and regulation of alpha-amylase synthesis in the hyperthermophilic archaeon Sulfolobus solfataricus. J Bacteriol *178*, 945-950.

Heath, R. J., and Rock, C. O. (2002). The Claisen condensation in biology. Nat Prod Rep *19*, 581-596.

Heathcock, C. H., Buse, C. T., Kleschick, W. a., Pirrung, M. C., Sohn, J. E., and Lampe, J. (1980). Acyclic Stereoselection.7. Stereoselective Synthesis of 2-Alkyl-3-Hydroxy Carbonyl-Compounds by Aldol Condensation. Journal of Organic Chemistry *45*, 1066-1081.

Heimer, R., and Meyer, K. (1956). Studies on sialic acid of maxillary mucoid. Proc Natl Acad Sci USA *42*, 728-734.

Heine, A., DeSantis, G., Luz, J. G., Mitchell, M., Wong, C. H., and Wilson, I. A. (2001). Observation of Covalent Intermediates in an Enzyme Mechanism at Atomic Resolution. Science *294*, 369-374.

Heine, A., Luz, J. G., Wong, C. H., and Wilson, I. A. (2004). Analysis of the class I aldolase binding site architecture based on the crystal structure of 2-deoxyribose-5-phosphate aldolase at 0.99 angstrom resolution. Journal of Molecular Biology *343*, 1019-1034.

Hendry, E. J., Buchanan, C. L., Russell, R. J. M., Hough, D. W., Reeve, C. D., Danson, M. J., and Taylor, G. L. (2000). Preliminary crystallographic studies of an extremely thermostable KDG aldolase from Sulfolobus sofataricus. Acta Crystallographica Section D Biological Crystallography *D56*, 1437-1439.

Hester, G., Brenner-Holzach, O., Rossi, F. A., Struck-Donatz, M., Winterhalter, K. H., Smit, J. D., and Piontek, K. (1991). The crystal structure of fructose-1,6-bisphosphate aldolase from Drosophila melanogaster at 2.5 A resolution. FEBS Lett *292*, 237-242.

Heyer, N. I. (1999) Glucose dehydrogenase from the hyperthermophilic archaeaon *Sulfolobus solfataricus*, University of Bath, PhD. thesis.

Hirth, G., and Walther, W. (1985). Synthesis of the (R)-Glycerol and (S)-Glycerol Acetonides - Determination of the Optical Purity. Helvetica Chimica Acta *68*, 1863-1871.

Hochstein, L. I. (1974). The metabolism of carbohydrates by extremely halophilic bacteria: glucose metabolism via a modified Entner-Doudoroff pathway. Can J Microbiol *20*, 1085-1091.

Hoffmann, T., Zhong, G., List, B., Shabat, D., Anderson, J., Gramatikova, S., Lerner, R. A., and Barbas, C. F., 3rd (1998). Aldolase antibodies of remarkable scope. Journal of American Chemical Society *120*, 2768-2779.

Horecker, B. L., Tsolas, O., and Lai, C. Y. (1972). Aldolases, In The Enzymes, P. D. Boyer, ed. (New York; London: Academic Press), pp. 213-253.

Hough, D. W., and Danson, M. J. (1999). Extremozymes. Curr Opin Chem Biol *3*, 39-46.

Huang, X. Q., and Miller, W. (1991). A Time-Efficient, Linear-Space Local Similarity Algorithm. Advances in Applied Mathematics *12*, 337-357.

Hutton, C. A., Southwood, T. J., and Turner, J. J. (2003). Inhibitors of lysine biosynthesis as antibacterial agents. Mini Rev Med Chem *3*, 115-127.

Hsu, C. C., Hong, Z., Wada, M., Franke, D., and Wong, C. H. (2005). Directed evolution of D-sialic acid aldolase to L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase. Proc Natl Acad Sci U S A *102*, 9122-9126.

Iwabuchi, T., and Harayama, S. (1998). Biochemical and genetic characterization of trans-2'-carboxybenzalpyruvate hydratase-aldolase from a phenanthrene-degrading Nocardioides strain. J Bacteriol *180*, 945-949.

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*, 3849-3856.

Izard, T., Lawrence, M. C., Malby, R. L., Lilley, G. G., and Colman, P. M. (1994). The three-dimensional structure of N-acetylneuraminate lyase from Escherichia coli. Structure *2*, 361-369.

Jakubowski, H. (2002). Biochemistry Online: An Approach Based on Chemical Logic (http://employees.csbsju.edu/hjakubowski/classes/ch331/bcintro/default.html).

Jaschke, A. (2001). RNA-catalyzed carbon-carbon bond formation. Biol Chem *382*, 1321-1325.

Jeffcoat, R., Hassall, H., and Dagley, S. (1969). Purification and properties of D-4-deoxy-5-oxoglucarate hydro-lyase (decarboxylating). Biochem J *115*, 977-983.

Jia, J., Schorken, U., Lindqvist, Y., Sprenger, G. A., and Schneider, G. (1997). Crystal structure of the reduced Schiff-base intermediate complex of transaldolase B from Escherichia coli: mechanistic implications for class I aldolases. Protein Sci *6*, 119-124.

Joerger, A. C., Mayer, S., and Fersht, A. R. (2003). Mimicking natural evolution in vitro: An N-acetylneuraminate lyase mutant with an increased dihydrodipicolinate synthase activity. Proc Natl Acad Sci U S A *100*, 5694-5699.

Johnsen, U., Selig, M., Xavier, K. B., Santos, H., and Schonheit, P. (2001). Different glycolytic pathways for glucose and fructose in the halophilic archaeon Halococcus saccharolyticus. Arch Microbiol *175*, 52-61.

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991). Improved Methods for Building Protein Models in Electron-Density Maps and the Location of Errors in These Models. Acta Crystallographica Section A *47*, 110-119.

Jorgensen, T. J. D., Roepstorff, P., and Heck, A. J. R. (1998). Direct determination of solution binding constants for noncovalent complexes between bacterial cell wall peptide analogues and vancomycin group antibiotics by electrospray ionization mass spectrometry. Analytical Chemistry *70*, 4427-4432.

Kabsch, W. (1976). Acta Crystallogr *A32*, 922-923.

Kabsch, W., and Sander, C. (1983). Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. Biopolymers *22*, 2577-2637.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., *et al.* (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, Sulfolobus tokodaii strain7. DNA Res *8*, 123-140.

Kerscher, L., Nowitzki, S., and Oesterhelt, D. (1982). Thermoacidophilic archaebacteria contain bacterial-type ferredoxins acting as electron acceptors of 2-oxoacid:ferredoxin oxidoreductases. Eur J Biochem *128*, 223-230.

Kim, H., Certa, U., Dobeli, H., Jakob, P., and Hol, W. G. (1998). Crystal structure of fructose-1,6-bisphosphate aldolase from the human malaria parasite Plasmodium falciparum. Biochemistry *37*, 4388-4396.

Kim, M. J., Hennen, W. J., Sweers, H. M., and Wong, C. H. (1988). Enzymes in Carbohydrate Synthesis - N-Acetylneuraminic Acid Aldolase Catalyzed-Reactions and Preparation of N-Acetyl-2-Deoxy-D-Neuraminic Acid-Derivatives. Journal of the American Chemical Society *110*, 6481-6486.

Kleywegt, G. J., and Brunger, A. T. (1996). Checking your imagination: applications of the free R value. Structure *4*, 897-904.

Koerner, T. A., Jr., Cary, L. W., Bhacca, N. S., and Younathan, E. S. (1973). Tautomeric composition of D-fructose phosphates in solution by Fourier transform carbon-13 nuclear magnetic resonance. Biochem Biophys Res Commun *51*, 543-550.

Kragl, U., Godde, A., Wandrey, C., Lubin, N., and Auge, C. (1994). New Synthetic Applications of Sialic-Acid Aldolase, a Useful Catalyst for Kdo Synthesis - Relation between Substrate Conformation and Enzyme Stereoselectivity (Vol 1, Pg 119, 1994). Journal of the Chemical Society-Perkin Transactions 1, 909-909.

Kuo, D. J., and Rose, I. A. (1985). Chemical trapping of complexes of dihydroxyacetone phosphate with muscle fructose-1,6-bisphosphate aldolase. Biochemistry *24*, 3947-3952.

Kydd, C. L. (1999) A novel aldolase from the hyperthermophilic archaeon *Sulfolobus solfataricus*, University of Bath.

Lamble, H. J. (2004) 2-keto-3-deoxygluconate Aldolase from *Sulfolobus solfataricus* - Fundamental Studies and Biotechnological Application, University of Bath, Bath.

Lamble, H. J., Heyer, N. I., Bull, S. D., Hough, D. W., and Danson, M. J. (2003). Metabolic pathway promiscuity in the archaeon Sulfolobus solfataricus revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase. J Biol Chem *278*, 34066-34072.

Lamble, H. J., Milburn, C. C., Taylor, G. L., Hough, D. W., and Danson, M. J. (2004). Gluconate dehydratase from the promiscuous Entner-Doudoroff pathway in Sulfolobus solfataricus. FEBS Lett *576*, 133-136.

Lamble, H. J., Danson, M. J., Hough, D. W., and Bull, S. D. (2005a). Engineering stereocontrol into an aldolase-catalysed reaction. Chem Commun (Camb), 124-126.

Lamble, H. J., Theodossis, A., Milburn, C. C., Taylor, G. L., Bull, S. D., Hough, D. W., and Danson, M. J. (2005). Promiscuity in the part-phosphorylative Entner-Doudoroff pathway of the archaeon Sulfolobus solfataricus. FEBS Lett *579*, 6865-6869.

Lamzin, V. S., and Wilson, K. S. (1993). Automated Refinement of Protein Models. Acta Crystallographica Section D-Biological Crystallography *49*, 129-147.

Langer, D., Hain, J., Thuriaux, P., and Zillig, W. (1995). Transcription in archaea: similarity to that in eucarya. Proc Natl Acad Sci U S A *92*, 5768-5772.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). J App Cryst *26*, 283.

Lawrence, M. C., Barbosa, J. A., Smith, B. J., Hall, N. E., Pilling, P. A., Ooi, H. C., and Marcuccio, S. M. (1997). Structure and mechanism of a sub-family of enzymes related to N-acetylneuraminate lyase. J Mol Biol *266*, 381-399.

Lee, B., and Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. J Mol Biol *55*, 379-400.

Legon, a. C., and Millen, D. J. (1987). Directional Character, Strength, and Nature of the Hydrogen-Bond in Gas-Phase Dimers. Accounts of Chemical Research *20*, 39-46.

Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data, In Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography.

Liang, Z. X., Lee, T., Resing, K. A., Ahn, N. G., and Klinman, J. P. (2004). Thermal-activated protein mobility and its correlation with catalysis in thermophilic alcohol dehydrogenase. Proc Natl Acad Sci U S A *101*, 9556-9561.

Lilley, G. G., Barbosa, J. A., and Pearce, L. A. (1998). Expression in Escherichia coli of the putative N-acetylneuraminate lyase gene (nanA) from Haemophilus influenzae: overproduction, purification, and crystallization. Protein Expr Purif *12*, 295-304.

Lin, C. H., Sugai, T., Halcomb, R. L., Ichikawa, Y., and Wong, C. H. (1992). Unusual stereoselectivity in sialic acid aldolase-catalysed aldol condensations: synthesis of both enantiomers of high-carbon monosaccharides. J Am Chem Soc *114*, 10138-10145.

Lokanath, N. K., Shiromizu, I., Ohshima, N., Nodake, Y., Sugahara, M., Yokoyama, S., Kuramitsu, S., Miyano, M., and Kunishima, N. (2004). Structure of aldolase from Thermus thermophilus HB8 showing the contribution of oligomeric state to thermostability. Acta Crystallogr D Biol Crystallogr *60*, 1816-1823.

Loo, J. A. (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. Mass Spectrom Rev *16*, 1-23.

Lorentzen, E., Pohl, E., Zwart, P., Stark, A., Russell, R. B., Knura, T., Hensel, R., and Siebers, B. (2003). Crystal structure of an archaeal class I aldolase and the evolution of (betaalpha)8 barrel proteins. J Biol Chem *278*, 47253-47260.

Lorentzen, E., Siebers, B., Hensel, R., and Pohl, E. (2004). Structure, function and evolution of the Archaeal class I fructose-1,6-bisphosphate aldolase. Biochem Soc Trans *32*, 259-263.

Lorentzen, E., Siebers, B., Hensel, R., and Pohl, E. (2005). Mechanism of the Schiff base forming fructose-1,6-bisphosphate aldolase: structural analysis of reaction intermediates. Biochemistry *44*, 4222-4229.

Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, 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*, 437-450.

Machajewski, T. D., and Wong, C. H. (2000). The Catalytic Asymmetric Aldol Reaction. Angew Chem Int Ed Engl *39*, 1352-1375.

Marsh, J. J., and Lebherz, H. G. (1992). Fructose-bisphosphate aldolases: an evolutionary history. Trends Biochem Sci *17*, 110-113.

Matthews, B. W. (1968). Solvent content of protein crystals. J Mol Biol *33*, 491-497.

Mattos, C. (2002). Protein-water interactions in a dynamic world. Trends Biochem Sci *27*, 203-208.

Maurady, A., Zdanov, A., de Moissac, D., Beaudry, D., and Sygusch, J. (2002). A conserved glutamate residue exhibits multifunctional catalytic roles in D-fructose-1,6-bisphosphate aldolases. J Biol Chem *277*, 9474-9483.

Mavridis, I. M., Hatada, M. H., Tulinsky, A., and Lebioda, L. (1982). Structure of 2-keto-3-deoxy-6-phosphogluconate aldolase at 2. 8 A resolution. J Mol Biol *162*, 419-444.

McDonald, I. K., and Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. J Mol Biol *238*, 777-793.

McElheny, D., Schnell, J. R., Lansing, J. C., Dyson, H. J., and Wright, P. E. (2005). Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. Proc Natl Acad Sci U S A *102*, 5032-5037.

McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996). The pKa of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a 13C-NMR study of bacillus circulans xylanase. Biochemistry *35*, 9958-9966.

Midelfort, C. F., Gupta, R. K., and Meloche, H. P. (1977). Specificity of 2-Keto-3-Deoxygluconate-6-P Aldolase for Open-Chain Form of 2-Keto-3-Deoxygluconate-6-P - Kinetic and C-13 Nuclear Magnetic-Resonance Studies. Journal of Biological Chemistry *252*, 3486-3492.

Midelfort, C. F., Gupta, R. K., and Rose, I. A. (1976). Fructose 1,6-bisphosphate: isomeric composition, kinetics, and substrate specificity for the aldolases. Biochemistry *15*, 2178-2185.

Midelfort, C. F., Gupta, R. K., and Meloche, H. P. (1977). Specificity of 2-Keto-3-Deoxygluconate-6-P Aldolase for Open-Chain Form of 2-Keto-3-Deoxygluconate-6-P - Kinetic and C-13 Nuclear Magnetic-Resonance Studies. Journal of Biological Chemistry *252*, 3486-3492.

Mirwaldt, C., Korndorfer, I., and Huber, R. (1995). The crystal structure of dihydrodipicolinate synthase from Escherichia coli at 2.5 A resolution. J Mol Biol *246*, 227-239.

Moll, R., and Schäfer, G. (1988). Chemiosmotic H+ cycling across the plasma membrane of the thermoacidophilic archaebacterium Sulfolobus acidocaldarius. FEBS Letters *232*, 359-363.

Morris, A. J., Davenport, R. C., and Tolan, D. R. (1996). A lysine to arginine substitution at position 146 of rabbit aldolase A changes the rate-determining step to Schiff base formation. Protein Eng *9*, 61-67.

Morris, A. J., and Tolan, D. R. (1993). Site-directed mutagenesis identifies aspartate 33 as a previously unidentified critical residue in the catalytic mechanism of rabbit aldolase A. J Biol Chem *268*, 1095-1100.

Morris, A. J., and Tolan, D. R. (1994). Lysine-146 of rabbit muscle aldolase is essential for cleavage and condensation of the C3-C4 bond of fructose 1,6-bis(phosphate). Biochemistry *33*, 12291-12297.

Morrison, R. T., and Boyd, R. N. (1992). Organic chemistry, 6th edition (London: Prentice-Hall International).

Mulder, F. A., Mittermaier, A., Hon, B., Dahlquist, F. W., and Kay, L. E. (2001). Studying excited states of proteins by NMR spectroscopy. Nat Struct Biol *8*, 932-935.

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica Section D-Biological Crystallography *53*, 240-255.

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*, 536-540.

Nagano, N., Orengo, C. A., and Thornton, J. M. (2002). One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J Mol Biol *321*, 741-765.

Navaza, J. (1994). Amore - an Automated Package for Molecular Replacement. Acta Crystallographica Section A *50*, 157-163.

Nielsen, J. E., and McCammon, J. A. (2003). Calculating pKa values in enzyme active sites. Protein Sci *12*, 1894-1901.

Nishide, K., Shibata, K., Fujita, T., Kajimoto, T., Wong, C. H., and Node, M. (2000). An asymmetric total synthesis of a potent immunosuppressant, mycestericins D and F, through an aldol reaction using L-threonine aldolase. Heterocycles *52*, 1191-+.

O'Brien, P. J., and Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. Chem Biol *6*, R91-R105.

Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., and Thornton, J. M. (1997). CATH--a hierarchic classification of protein domain structures. Structure *5*, 1093-1108.

Plantierroyon, R., Anker, D., and Robertbaudouy, J. (1991a). New Synthesis of 3-Deoxy-D-Erythro-2-Hexulosonic Acid (2-Keto-3-Deoxygluconic Acid, Kdg) from D-Glucose. Journal of Carbohydrate Chemistry *10*, 239-249.

Plantierroyon, R., Cardona, F., Anker, D., Condemine, G., Nasser, W., and Robertbaudouy, J. (1991b). New Synthesis of 3-Deoxy-D-Erythro-2-Hexulosonic Acid (Kdg) from D-Glucono-1,5-Lactone - Synthesis and Nmr-Study on O-Methylated Derivatives of Kdg. Journal of Carbohydrate Chemistry *10*, 787-811.

Potters, M. B., Solow, B. T., Bischoff, K. M., Graham, D. E., Lower, B. H., Helm, R., and Kennelly, P. J. (2003). Phosphoprotein with phosphoglycerate mutase activity from the archaeon Sulfolobus solfataricus. J Bacteriol *185*, 2112-2121.

Pricer, W. E., Jr., and Horecker, B. L. (1960). Deoxyribose aldolase from Lactobacillus plantarum. J Biol Chem *235*, 1292-1298.

Radzicka, A., and Wolfenden, R. (1995). A proficient enzyme. Science *267*, 90-93.

Reetz, M. T. (2004). Controlling the enantioselectivity of enzymes by directed evolution: practical and theoretical ramifications. Proc Natl Acad Sci U S A *101*, 5716-5722.

Reymond, J., and Chen, Y. (1995). Catalytic, enantioselective aldol reaction using antibodies against a quaternary ammonium ion with a primary amine cofactor. Tetrahedron Letters *36*, 2575-2578.

Richards, F. M. (1977). Areas, volumes, packing and protein structure. Annu Rev Biophys Bioeng *6*, 151-176.

Ringe, D., and Petsko, G. A. (2003). The 'glass transition' in protein dynamics: what it is, why it occurs, and how to exploit it. Biophys Chem *105*, 667-680.

Robertson, C. E., Harris, J. K., Spear, J. R., and Pace, N. R. (2005). Phylogenetic diversity and ecology of environmental Archaea. Curr Opin Microbiol *8*, 638-642.

Rose, I. A., O'Connell, E. L., and Mehler, A. H. (1965). Mechanism of the Aldolase Reaction. J Biol Chem *240*, 1758-1765.

Rose, I. A., and Rieder, S. V. (1958). Studies on the mechanism on the aldolase reaction; isotope exchange reactions of muscle and yeast aldolase. J Biol Chem *231*, 315-329.

Rose, I. A., and Warms, J. V. (1985). Complexes of muscle aldolase in equilibrium with fructose 1,6-bisphosphate. Biochemistry *24*, 3952-3957.

Rose, I. A., Warms, J. V., and Kuo, D. J. (1987). Concentration and partitioning of intermediates in the fructose bisphosphate aldolase reaction. Comparison of the muscle and liver enzymes. J Biol Chem *262*, 692-701.

Rossmann, M. G., and Arnold, E., eds. (2001). International Tables of Crystallography, Volume F: Crystallography of biological macromolecules, 1 edn (1st Edition: Springer).

Russell, R. J., and Taylor, G. L. (1995). Engineering thermostability: lessons from thermophilic proteins. Curr Opin Biotechnol *6*, 370-374.

Rutter, W. J. (1964). Evolution of Aldolase. Fed Proc *23*, 1248-1257.

Sakuraba, H., Tsuge, H., Shimoya, I., Kawakami, R., Goda, S., Kawarabayasi, Y., Katunuma, N., Ago, H., Miyano, M., and Ohshima, T. (2003). The first crystal structure of archaeal aldolase. Unique tetrameric structure of 2-deoxy-d-ribose-5-phosphate aldolase from the hyperthermophilic archaea Aeropyrum pernix. J Biol Chem *278*, 10799-10806.

Schafer, G. (1996). Bioenergetics of the archaebacterium Sulfolobus. Biochim Biophys Acta *1277*, 163-200.

Schultz, P. G., and Lerner, R. A. (1995). From molecular diversity to catalysis: lessons from the immune system. Science *269*, 1835-1842.

Sensen, C. W., Charlebois, R. L., Chow, C., Clausen, I. G., Curtis, B., Doolittle, W. F., Duguet, M., Erauso, G., Gaasterland, T., Garrett, R. A., *et al.* (1998). Completing the sequence of the Sulfolobus solfataricus P2 genome. Extremophiles *2*, 305-312.

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., *et al.* (2001). The complete genome of the crenarchaeon Sulfolobus solfataricus P2. Proc Natl Acad Sci U S A *98*, 7835-7840.

Shedlarski, J. G., and Gilvarg, C. (1970). The pyruvate-aspartic semialdehyde condensing enzyme of Escherichia coli. J Biol Chem *245*, 1362-1373.

Sheldrick, G. M. (1990). Phase annealing in SHELX-90: direct methods for larger structures. Acta Crystallogr *A46*, 467-473.

Sheldrick, G. M., and Schneider, T. R. (1997). Shelxl: High-Resolution Refinement. Macromolecular Crystallography, Pt B *277*, 319-343.

Shelton, M. C., Cotterill, I. C., Novak, S. T. A., Poonawala, R. M., Sudarshan, S., and Toone, E. J. (1996). 2-keto-3-deoxy-6-phosphogluconate Aldolases as Catalysts for Stereocontrolled Carbon-Carbon Bond Formation. Journal of American Chemical Society *118*, 2117-2125.

Shneier, a., Kleanthous, C., Deka, R., Coggins, J. R., and Abell, C. (1991). Observation of an Imine Intermediate on Dehydroquinase by Electrospray Mass-Spectrometry. Journal of the American Chemical Society *113*, 9416-9418.

Siebers, B., Brinkmann, H., Dorr, C., Tjaden, B., Lilie, H., van der Oost, J., and Verhees, C. H. (2001). Archaeal fructose-1,6-bisphosphate aldolases constitute a new family of archaeal type class I aldolase. J Biol Chem *276*, 28710-28718.

Siebers, B., Tjaden, B., Michalke, K., Dorr, C., Ahmed, H., Zaparty, M., Gordon, P., Sensen, C. W., Zibat, A., Klenk, H. P., *et al.* (2004). Reconstruction of the central carbohydrate metabolism of Thermoproteus tenax by use of genomic and biochemical data. J Bacteriol *186*, 2179-2194.

Sievers, A., Beringer, M., Rodnina, M. V., and Wolfenden, R. (2004). The ribosome as an entropy trap. Proc Natl Acad Sci U S A *101*, 7897-7901.

Silvestri, M. G., DeSantis, G., Mitchell, M., and Wong, C.-H. (2003). Asymmetric aldol reactions using aldolases. Top Stereochem *23*, 267-342.

Smith, B. J., Lawrence, M. C., and Barbosa, J. A. (1999). Substrate-Assisted Catalysis in Sialic Acid Aldolase. J Org Chem *64*, 945-949.

Smith, L. D., Budgen, N., Bungard, S. J., Danson, M. J., and Hough, D. W. (1989). Purification and characterization of glucose dehydrogenase from the thermoacidophilic archaebacterium Thermoplasma acidophilum. Biochem J *261*, 973-977.

Stampfer, W., Kosjek, B., Faber, K., and Kroutil, W. (2003). Biocatalytic oxidative kinetic resolution of sec-alcohols: stereocontrol through substrate-modification. Tetrahedron-Asymmetry *14*, 275-280.

Stevenson, D. E., Feng, R., and Storer, A. C. (1990). Detection of covalent enzyme-substrate complexes of nitrilase by ion-spray mass spectroscopy. FEBS Lett *277*, 112-114.

St-Jean, M., Lafrance-Vanasse, J., Liotard, B., and Sygusch, J. (2005). High resolution reaction intermediates of rabbit muscle fructose-1,6-bisphosphate aldolase: substrate cleavage and induced fit. J Biol Chem *280*, 27262-27270.

Stryer, L. (1981). Biochemistry, 2nd edn (San Francisco: W.H. Freeman).

Stura, E. A., Ghosh, S., Garcia-Junceda, E., Chen, L., Wong, C. H., and Wilson, I. A. (1995). Crystallization and preliminary crystallographic data for class I deoxyribose-5-phosphate aldolase from Escherichia coli: an application of reverse screening. Proteins *22*, 67-72.

Sygusch, J., Beaudry, D., and Allaire, M. (1987). Molecular architecture of rabbit skeletal muscle aldolase at 2.7-A resolution. Proc Natl Acad Sci U S A *84*, 7846-7850.

Takayama, S., McGarvey, G. J., and Wong, C. H. (1997). Microbial aldolases and transketolases: new biocatalytic approaches to simple and complex sugars. Annu Rev Microbiol *51*, 285-310.

Tanaka, F., Fuller, R., and Barbas, C. F., 3rd (2005). Development of small designer aldolase enzymes: catalytic activity, folding, and substrate specificity. Biochemistry *44*, 7583-7592.

Theodossis, A., Walden, H., Westwick, E. J., Connaris, H., Lamble, H. J., Hough, D. W., Danson, M. J., and Taylor, G. L. (2004). The structural basis for substrate promiscuity in 2-keto-3-deoxygluconate aldolase from the Entner-Doudoroff pathway in Sulfolobus solfataricus. J Biol Chem *279*, 43886-43892.

Thomson, G. J., Howlett, G. J., Ashcroft, A. E., and Berry, A. (1998). The dhnA gene of Escherichia coli encodes a class I fructose bisphosphate aldolase. Biochem J *331 (Pt 2)*, 437-445.

Tolstrup, N., Sensen, C. W., Garrett, R. A., and Clausen, I. G. (2000). Two different and highly organized mechanisms of translation initiation in the archaeon Sulfolobus solfataricus. Extremophiles *4*, 175-179.

Tudor, D. W., Lewis, T., and Robins, D. J. (1993). Synthesis, 1061-1062.

Turner, N. J. (2003). Controlling chirality. Curr Opin Biotechnol *14*, 401-406.

Uchida, Y., Tsukada, Y., and Sugimori, T. (1984). Purification and properties of N-acetylneuraminate lyase from Escherichia coli. J Biochem (Tokyo) *96*, 507-522.

Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr D Biol Crystallogr *55 (Pt 1)*, 191-205.

Valentin-Hansen, P., Boetius, F., Hammer-Jespersen, K., and Svendsen, I. (1982). The primary structure of Escherichia coli K12 2-deoxyribose 5-phosphate aldolase. Nucleotide sequence of the deoC gene and the amino acid sequence of the enzyme. Eur J Biochem *125*, 561-566.

Verhees, C. H., Kengen, S. W., Tuininga, J. E., Schut, G. J., Adams, M. W., De Vos, W. M., and Van Der Oost, J. (2003). The unique features of glycolytic pathways in Archaea. Biochem J *375*, 231-246.

Vieille, C., and Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev *65*, 1-43.

Vimr, E. R., and Troy, F. A. (1985). Identification of an inducible catabolic system for sialic acids (nan) in Escherichia coli. J Bacteriol *164*, 845-853.

Voet, D., and Voet, J. G. (1995). Chapter 16: Glycolysis, In Biochemistry 2nd Ed. (John Wiley & Sons Inc.), pp. 450, 451.

von Itzstein, M., Wu, W. Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Phan, T. V., Smythe, M. L., White, H. F., Oliver, S. W., *et al.* (1993). Rational Design of Potent Sialidase-Based Inhibitors of Influenza-Virus Replication. Nature *363*, 418-423.

Wagner, J., Lerner, R. A., and Barbas, C. F., 3rd (1995). Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. Science *270*, 1797-1800.

Wahl, M. C., and Sundaralingam, M. (1997). C-H.O hydrogen bonding in biology. Trends Biochem Sci *22*, 97-102.

Walden, H. (2001) Probing the Structural Basis of Protein Thermostability, PhD, University of St Andrews, St Andrews.

Warshel, A. (1998). Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. J Biol Chem *273*, 27035-27038.

Westheimer, F. H. (1995). Coincidences, Decarboxylation, and Electrostatic Effects. Tetrahedron *51*, 3-20.

Whitford, D. (2005). Proteins: structure and function (Chichester: Wiley).

Williams, G. J., Woodhall, T., Nelson, A., and Berry, A. (2005). Structure-guided saturation mutagenesis of N-acetylneuraminic acid lyase for the synthesis of sialic acid mimetics. Protein Eng Des Sel *18*, 239-246.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A *74*, 5088-5090.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A *87*, 4576-4579.

Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E. Z., and Kern, D. (2004). Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nat Struct Mol Biol *11*, 945-949.

Wolfenden, R., and Snider, M. J. (2001). The depth of chemical time and the power of enzymes as catalysts. Acc Chem Res *34*, 938-945.

Wong, C.-H., and Whitesides, G. M. (1994). Enzymes in synthetic organic chemistry, 1st edn (Oxford, U.K: Pergamon; Elsevier Science).

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*, 1-9.

Yamagata, Y., Ogasahara, K., Hioki, Y., Lee, S. J., Nakagawa, A., Nakamura, H., Ishida, M., Kuramitsu, S., and Yutani, K. (2001). Entropic stabilization of the tryptophan synthase alpha-subunit from a hyperthermophile, Pyrococcus furiosus. X-ray analysis and calorimetry. J Biol Chem *276*, 11062-11071.

Yugari, Y., and Gilvarg, C. (1965). The condensation step in diaminopimelate synthesis. J Biol Chem *240*, 4710-4716.

Zhang, X., and Houk, K. N. (2005). Acid/base catalysis by pure water: the aldol reaction. J Org Chem *70*, 9712-9716.

Zillig, W., Stetter, K., Wunderl, S., Schulz, W., Priess, H., and Scholz, L. (1980). Arch Microbiol *125*, 259-269.

Zuckerkandl, E., and Pauling, L. (1965). Molecules as documents of evolutionary history. J Theor Biol *8*, 357-366.