Integrated molecular analysis of **Sugar Metabolism** of *Sulfolobus solfataricus*

Stan J.J. Brouns

Integrated molecular analysis of **sugar metabolism** of *Sulfolobus solfataricus*

Stan J.J. Brouns

Promotoren

prof. dr. Willem M. de Vos
hoogleraar in de Microbiologie
Wageningen Universiteit

prof. dr. John van der Oost
persoonlijk hoogleraar Microbiologie en Biochemie
Wageningen Universiteit

Leden van de promotie commissie

prof. dr. Ton J.W.G. Visser
bijzonder hoogleraar Microspectroscopie in de Biochemie
Wageningen Universiteit

prof. dr. Arnold J.M. Driessen
hoogleraar Moleculaire Microbiologie
Rijksuniversiteit Groningen

dr. Loren L. Looger
Howard Hughes Medical Institute
Ashburn (VA), Verenigde Staten

dr. Thijs Kaper
Genencor International
Palo Alto (CA), Verenigde Staten

Dit onderzoek is uitgevoerd binnen de onderzoekschool VLAG

Integrated molecular analysis of **sugar metabolism** of *Sulfolobus solfataricus*

Stan J.J. Brouns

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus van
Wageningen Universiteit, prof. dr. M.J. Kropff,
in het openbaar te verdedigen
op dinsdag 2 oktober 2007
des namiddags te half twee in de Aula

Cover Boiling water: the habitat of a hyperthermophile (photo: S.J.J. Brouns)
Printing Gildeprint (Enschede)
Sponsoring Hellma Benelux, bioMérieux

Brouns, S.J.J. - Integrated molecular analysis of sugar metabolism of *Sulfolobus solfataricus*

in Dutch
Geïntegreerde moleculaire analyse van het suikermetabolisme van *Sulfolobus solfataricus*

PhD Thesis Wageningen University, Wageningen, Netherlands (2007)
176 p. - with summary in Dutch

ISBN **978-90-8504-713-1**

Voor mijn ouders
en Marloes

DANKWOORD

Met trots ligt ligt hier nu een boekje. Natuurlijk kon het niet totstandkomen zonder de hulp van anderen. In dit stukje wil ik die mensen bedanken.

Allereerst mijn promotor John. Ik ben blij dat ik 6 jaar in je groep heb mogen werken; eerst als student, toen als AIO en nu als postdoc. Je nieuwe ideeën, onbegrensd positivisme, excellente netwerkkwaliteiten en vertrouwen in iedere AIO bieden een de basis voor je succesvolle onderzoeksgroep en geven genoeg ruimte voor de persoonlijke ontwikkeling. Ik denk dat we menig idee succesvol hebben uitgewerkt. Ik kan me niet voorstellen dat ik met mijn volgende baas weer zo'n goede band kan opbouwen als met jou. Bedankt!

Ook wil ik mijn promotor Willem bedanken voor inzichten, steun en supervisie op de achtergrond. Ondanks het feit dat je tientallen AIO's onder je hoede hebt, weet je toch altijd weer tijd te vinden om iedereen van advies te voorzien.

Paranimfen Harmen en Thijs E., ik wil jullie bedanken voor jullie optreden op mijn grote dag. Mede door jullie ging ik altijd met plezier naar mijn werk. Het geintjes uithalen bij elkaar en het eindeloos discussiëren bij de koffie ga ik zeker missen. Alle congressen waren wetenschappelijke vakanties en soms vol met avontuur, zoals de legendarische overnachting op de Strompoli....

Jasper Akerboom, waar zijn we samen niet geweest is eigenlijk de vraag. Onze Berlijn trips waren ongeëvenaard. Het boekje had er heel anders uitgezien als de eiwitkristallen niet tegen 100 dB aan geluid en trillingen hadden gekund, die ontstond nadat we een stuk uitlaat op de snelweg waren verloren.

Door mijn langere verblijf bij BacGen heb ik twee generaties AIO's en Postdocs kunnen meemaken, waarvan sommigen elkaar niet eens meer kennen. Ik wil de oude garde (Corné V., Johan, Leon, Ana, Don, en Arjen) bedanken voor de gezelligheid en goede sfeer binnen de groep. Thijs K., leuk dat je naast de begeleider van mijn afstudeervak, nu ook opponent bent in mijn promotiecommissie. Waarschijnlijk doe ik veel dingen in het lab onbewust nog steeds "op jouw manier".

Telkens wanneer iemand vertrok dacht iedereen dat de sfeer wel een stuk minder zou worden. Dat bleek echter niet zo, slechts de gemiddelde leeftijd werd minder. Jasper W., ik wil je bedanken voor je bijdrage aan dit boekje. Jouw stage naar Zweden heeft een kentering teweeg gebracht in mijn promotieonderzoek. Hao, thanks for all your hard work. This book would not have been the same without you! Serve en Ans, jullie vaste waarde in de groep is van groot belang voor de groep en bedankt voor de gezelligheid op de werkkamer. Wim, Nees, Jannie en Francis, zonder jullie zou het lab niet draaien. Bedankt voor jullie hulp en soepele manier van omgang, iets wat Microbiologie zo kenmerkt.

Ook wil ik de nieuwe garde bedanken voor hun positieve bijdrage aan de goede sfeer in het lab. Ronnie, Krisztina, Susanne, Odette, Marcel, Mark, Colin, Marco, Pawel, Marke, Corné v.d. K., Lei, veel succes de komende jaren. Bart en Fabian, ik hoop dat jullie het mysterie rond MBF1 kunnen ophelderen. Matthijs, het was zeer prettig samenwerken aan ons' hot-topic en ik vind het jammer dat daar een eind aan gaat komen. Edze en Stieneke het allerbeste in Bologna!

Ik ben mijn studenten dankbaar voor de inzet die zij tijdens hun afstudeervak getoond hebben en ben trots dat de namen van een aantal van jullie op publicaties terecht zijn gekomen. Ester, Marco, Ivor, Nicole, Hanneke, Petra en Marjon, het ga jullie goed.

De geïntegreerde manier van onderzoek was alleen mogelijk door samen te werken met andere onderzoeksgroepen in Nederland, Groot-Brittannië, Zweden, Duitsland, Frankrijk en de Verenigde Staten.

Beste Bram, jouw naam staat op drie van de negen hoofdstukken wat aangeeft dat je een grote rol van betekenis hebt gespeeld in het onderzoek. Onze complementaire expertise heeft duidelijk een win-win situatie gecreëerd waardoor we het suikermetabolisme op een meer systems biology manier in beeld konden brengen. Bedankt voor alles. Phil, thanks for making it possible.

Beste Thomas, toen ik naar huis fietste nadat je me de donut structuur had opgestuurd, besefte ik hoe bijzonder het is om als één van de eersten ooit een compleet nieuwe eiwitstructuur onder ogen te krijgen. Ook al is hoofdstuk vijf het enige niet-gepubliceerde verhaal van dit boekje, ik denk dat het een grote toegevoegde waarde heeft voor het geheel. We kunnen er trots op zijn en ik wil je bedanken voor je grote bijdrage. Daar moeten we op 2 oktober bij onze eerste ontmoeting, maar eens het glas op heffen!

Beste Robert, na de Protein Engineering course in 2003 is er toch een mooi samenwerkingsverband ontstaan tussen Wageningen en Utrecht wat heeft geresulteerd in een aantal fraaie publicaties. Bedankt voor je bijdrage op het gebied van de natieve massaspectrometrie. Hortense, thanks for your contribution.

Laurent, although we never met, you have been a great help from the moment I contacted you. We could never have achieved the results described in chapter 3 and 5 without your knowledge of organic chemistry and biochemistry. Merci beaucoup!

Rolf, Magnus, Peter and Anders, the arrays have played a vital role in the discovery of the degradation pathway for pentoses, and have created a number of interesting spin-off projects described in this thesis. I am thankful that you have been willing to collaborate with us, especially at such an early stage of DNA microarray development.

Ken, I appreciate it that you taught me all the tricks with *Sulfolobus*, which have helped me tremendously during my PhD. Besides work I think we had great times as well.

Mark, thanks for having me in your lab. It was an unforgettable time. I'm glad that it has been the start of a good collaboration between Wageningen and Bozeman. Blake, good to hear that your staying in the CRISPR field. When are we going to play some more fussball and drink

Margarita's?

Azide, thanks mate. Your name is on three chapters. Apart from raising the level of scientific impact of the work enormously, the X-ray trips to Berlin and Oxford were great fun as well.

Tot slot wil ik mijn ouders, Imke en naaste familie bedanken voor hun motiverende woorden, morele steun en interesse die zij altijd getoond hebben, ook al was het helaas niet altijd mogelijk om jullie precies uit te leggen wat ik in het lab deed.

Lieve Marloes, van dichtbij heb je kunnen meemaken hoe het onderzoek van een AIO gepaard gaat met successen en teleurstellingen. Door jouw onvoorwaardelijke steun werden slechte dagen gerelativeerd en goede dagen gevierd. Hopelijk kunnen we nog lang veel voor elkaar blijven betekenen.

Stan
oktober 2007

TABLE OF CONTENTS

	Aim and outline of this thesis	**1**
Chapter 1	Introduction	**5**
Chapter 2	Reconstruction of central carbon metabolism in *Sulfolobus solfataricus* using a two-dimensional gel electrophoresis map, stable isotope labeling and DNA microarray analysis	**15**
Chapter 3	Identification of the missing links in prokaryotic pentose oxidation pathways: evidence for enzyme recruitment	**31**
Chapter 4	Crystal structure and biochemical properties of the D-arabinose dehydrogenase from *Sulfolobus solfataricus*	**49**
Chapter 5	Structural insight into substrate binding and catalysis of a novel 2-keto-3-*deoxy*-D-arabinonate dehydratase illustrates common mechanistic features of the FAH superfamily	**63**
Chapter 6	Identification of a novel α-galactosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*	**81**
Chapter 7	DNA family shuffling of hyperthermostable β-glycosidases	**95**
Chapter 8	Engineering a selectable marker for hyperthermophiles	**111**
Chapter 9	Summary and concluding remarks	**129**
	References	**133**
	Appendix I: Color figures	**151**
	Appendix II: Co-author affiliations	**161**
	Nederlandse samenvatting	**162**
	Educational statement of graduate school VLAG	**164**
	Curriculum vitae	**165**
	List of publications	**166**

Aim and outline of this thesis

This thesis presents the results of an integrated molecular analysis of the sugar metabolism of the hyperthermophilic archaeon *Sulfolobus solfataricus*. The primary aim of this work has been to obtain functional insight into several aspects of sugar metabolism of *S. solfataricus* using a wide range of techniques. These aspects include the discovery of new sugar degrading enzymes, the elucidation of unknown metabolic pathways, and the functional and structural characterization of the enzymes involved in these new metabolic routes. Additionally, the last two chapters present the work of laboratory evolution studies in an effort to alter protein properties such as enzyme activity and protein stability. The research performed in this work is both fundamental and applied, and was conducted in the course of the European Union project SCREEN.

Chapter 1 - Introduction

The first chapter gives an historical overview of the discovery of thermophiles, as well as their lifestyle, classification and genomic features. The micro-organism *S. solfataricus* is presented as a model archaeon for studying sugar metabolism. Special attention is given to the currently available genetic tools for *S. solfataricus*, its glycoside hydrolases and the method of directed evolution.

Chapter 2 – Reconstruction of central carbon metabolism in *Sulfolobus solfataricus* using a two-dimensional gel electrophoresis map, stable isotope labeling and DNA microarray analysis

The second chapter focuses on the expression of genes involved in three main pathways of central carbon metabolism: glycolysis, gluconeogenesis and the tricarboxylic acid cycle. The aim was to see whether these pathways would fluctuate during growth on either amino acids or sugars by analyzing the relevant enzymes at the mRNA and protein level.

Chapter 3 – Identification of the missing links in prokaryotic pentose oxidation pathways: evidence for enzyme recruitment

The third chapter presents the work of a functional genomics study using a multidisciplinary approach directed to elucidate the metabolism of pentoses in *S. solfataricus*. After transcriptome and proteome analyses, the activities of enzymes suspected to be involved in the degradation of D-arabinose were shown using biochemistry, and the catabolic pathway in which they were involved was uncovered. Bioinformatics was employed to find and predict the occurrence of this uncommon pentose oxidation pathway in other prokaryotes, and to deduce evolutionary implications that shed light on metabolic pathway genesis.

Chapter 4 – Crystal structure and biochemical properties of the D-arabinose dehydrogenase from *Sulfolobus solfataricus*

The fourth chapter provides detailed insight into structure-function relationships of the first enzyme of the pentose oxidation pathway: the D-arabinose dehydrogenase. The structure of the enzyme provides molecular clues as to why the enzyme prefers NADP$^+$ as a cofactor and D-arabinose and L-fucose-like sugars as substrates.

Chapter 5 – Structural insight into substrate binding and catalysis of a novel 2-keto-3-*deoxy*-D-arabinonate dehydratase illustrates common mechanistic features of the FAH superfamily.

The crystal structure of a completely new dehydratase involved in the pentose oxidation pathway is described in the fifth chapter. Insight into substrate binding was obtained from co-crystal structures with a divalent metal ion and a substrate analog bound, on the basis of which two catalytic mechanisms were proposed. The structure was compared to several functionally unknown structural homologs.

Chapter 6 – Identification of a novel a-galactosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*

The goal of the sixth chapter was to find new enzymes involved in the degradation of α–linked galactose oligosaccharides. Classical biochemistry and reverse genetics methods were used to identify an unusual α-galactosidase with similarities to plant alkaline α-galactosidases. At present, this enzyme represents the only enzyme of this type in the domain of the Archaea.

Chapter 7 – DNA family shuffling of hyperthermostable β-glycosidases

The research described in the seventh chapter was aimed at the optimization of existing enzymes for their application potential in lactose valorization. Following a DNA family shuffling approach using the β-glucosidase CelB from *Pyrococcus furiosus* and the β-galactosidase LacS from *S. solfataricus* as parents, a hybrid β-glycosidase was obtained containing the desired features of both parental enzymes.

Chapter 8 – Engineering a selectable marker for hyperthermophiles

Genetic modification of hyperthermophiles has been hampered by the lack of antibiotic selection markers. The eighth chapter aimed at developing an antibiotic resistance marker for hyperthermophiles. A directed evolution approach was chosen by making use of error-prone PCR, and the selection of thermostabilized resistance markers in the thermophilic bacterium Thermus thermophilus.

Chapter 9 – Summary & concluding remarks

The final chapter is a brief summary of the previous chapters and reflects on the results obtained.

1

Introduction

Partly adapted from

Brouns, S.J.J., T.J.G. Ettema, K.M. Stedman, J. Walther, H. Smidt, A.P.L. Snijders, M.J. Young, R. Bernander, P.C. Wright, B. Siebers, and J. van der Oost (2005) The hyperthermophilic archaeon *Sulfolobus*: from exploration to exploitation, p. 261-276. In W.P. Inskeep and T.R. McDermott (ed.), *Geothermal biology and geochemistry in Yellowstone national park*. Thermal Biology Institute, Montana State University, Bozeman.

and

Van der Oost, J., J. Walther, S.J.J. Brouns, H.J.G. van de Werken, A.P.L. Snijders, P.C. Wright, A. Andersson, R. Bernander, and W.M. de Vos (2006) Functional genomics of the thermo-acidophilic archaeon *Sulfolobus solfataricus*, p. 201-232. In F. Rainey and A. Oren (ed.), *Extremophiles, Methods in Microbiology*, vol. 35. Elsevier.

Until the late 1960's life was generally considered impossible at temperatures above 60 °C. This assumption was adjusted in 1969 when Thomas Brock and colleagues discovered microorganisms capable of growing at temperatures of 75 °C and higher in thermal springs in Yellowstone National Park (USA). The first organism they found was a gram-negative bacterium termed *Thermus aquaticus* [37]. Nearly twenty years later this thermophilic bacterium proved to be a good source of DNA polymerase, which led to the revolutionary invention of DNA amplification using the polymerase chain reaction by Kary Mullis and co-workers [328]. After finding thermophilic bacteria, Brock isolated a remarkable thermo-acidophilic microorganism called *Sulfolobus acidocaldarius* in acid hot springs rich in sulfur (pH < 3, 65-90 °C) [36]. Based on morphological analysis, *Sulfolobus* was initially classified as a thermo-acidophilic bacterium with some remarkable properties (Fig. 1.1) [354]. The molecular classification system introduced by Carl Woese and colleagues in the late 1970's, however, indicated that *Sulfolobus* does not belong to the Bacteria, but rather to a distinct prokaryotic domain: the Archaea [282,432].

The pioneering work set the stage for further exploration of thermal ecosystems world-wide. This has resulted in an ever-growing collection of thermophiles (both Archaea and Bacteria; optimal growth temperature 60-80 °C) and hyperthermophiles (mainly Archaea and some Bacteria; optimal growth temperature > 80 °C) isolated from terrestrial hot springs and marine ecosystems ranging from shallow vents to deep-sea black smokers [160,324,381,382]. Because of the high water pressure in deep-sea ecosystems, the water temperature can exceed 100 °C before the boiling point is reached. Moreover, at these temperatures the oxygen solubility is dramatically decreased, explaining why mostly anaerobes are found under these conditions. The most extreme hyperthermophiles described to date are crenarchaea of the family of *Pyrodictiaceae*, such as *Pyrolobus fumarii* (optimum growth temperature 106 °C, growth up to 113 °C; [31]) and a related iron-reducing isolate "strain 121" (optimum 106 °C, growth between 113 and 121 °C; [187]). Remarkably, vegetative cultures of *Pyrolobus* and *Pyrodictium* have been demonstrated to survive even autoclaving.

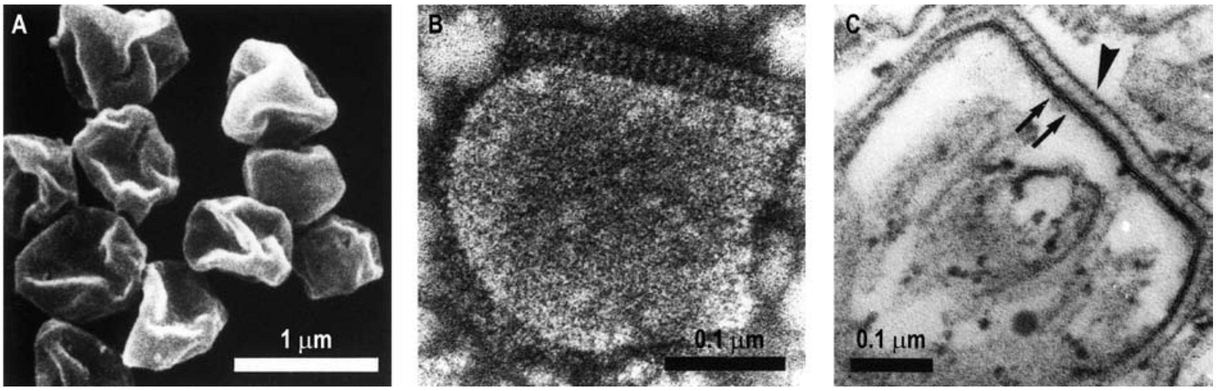


Figure 1.1
Electron micrographs of *Sulfolobus* cells. A) scanning electron micrograph B) negatively stained (uranyl acetate) transmission electron micrograph, C) transmission electron micrograph of an ultra-thin section. Arrows indicate cell-membrane, arrowhead marks the surface layer. Images provided by Dr. Mathias Lübben, Ruhr-Universität Bochum, Germany.

Thermophile genomics

Since the discovery of thermophilic microorganisms attempts have been made to reveal the secrets of their thermal lifestyle. Initially this was done physiological, biochemical and genetic analyses of these microbes [9], followed by X-ray crystallographic studies of their extremely thermostable proteins [212,419,420]. In addition, the sequencing of the genomes of many thermophiles during the last decade has allowed for a series of genome-based research lines. Several thermophilic Archaea and Bacteria have been selected for complete genomic sequencing (Table 1.1). Not only has this resulted in the discovery of various robust biocatalysts, but it also has led to insight into (i) thermophile physiology (unique metabolic enzymes and pathways, especially in Archaea [111,244]), (ii) the molecular basis of thermostability of bio-molecules (*e.g.* enhanced numbers of charged residues at surface and subunit/domain interfaces [49]), and (iii) the evolution

Table 1.1 - Sequenced thermophiles and hyperthermophiles

Organism	Lifestyle	T_{opt} (°C)	Genome Mb,GC%	Reference
Archaea				
Aeropyrum pernix (C)	AE, H	90	1.7, 56	190
Archaeoglobus fulgidus (E)	AN, FA	83	2.2, 49	202
Hyperthermus butylicus (C)	AN, H	101	1.7, 54	40
Methanocaldococcus jannaschii (E)	AN, A	85	1.7, 31	47
Methanopyrus kandleri (E)	AN, A	98	1.7, 61	363
Methanosaeta thermophile (E)	AN, A	61	1.9, 54	366
Methanothermobacter thermoautotrophicus (E)	AN, A	65-70	1.8, 50	364
Nanoarchaeum equitans	AN, P	90	0.5, 32	428
Picrophilus torridus (E)	AE, H	60	1.6, 36	131
Pyrobaculum aerophilum (C)	FAN, FA	100	2.2, 51	119
Pyrobaculum calidifontis (C)	FAN, H	90-95	2.0, 57	a
Pyrobaculum islandicum (C)	AN, FA	100	1.8, 50	a
Pyrococcus abyssi (E)	AN, H	96	1.8, 45	63
Pyrococcus furiosus (E)	AN, H	100	1.9, 41	318
Pyrococcus horikoshii (E)	AN, H	98	1.7, 42	191
Staphylothermus marinus (C)	AN, H	92	1.6, 36	a
Sulfolobus acidocaldarius (C)	AE, H	80	2.2, 37	57
Sulfolobus solfataricus (C)	AE, H	80	3.0, 36	353
Sulfolobus tokodaii (C)	AE, H	80	2.7, 33	189
Thermococcus kodakaraensis (E)	AN, H	95	2.1, 52	130
Thermofilum pendens (C)	AN, H	88	1.8, 58	a
Thermoplasma acidophilum (E)	FAN, H	60	1.6, 46	325
Thermoplasma volcanium (E)	FAN, H	60	1.6, 40	192
Thermoproteus tenax (C)	AN, FA	88	1.8, 55	359
Bacteria				
Aquifex aeolicus	FAN, A	95	1.6, 43	86
Carboxydothermus hydrogenoformans	AN, A	78	2.4, 42	436
Geobacillus kaustophilus	AE, H	60	3.6, 52	391
Geobacillus thermodenitrificans	FAN, H	65	3.6, 49	115
Rubrobacter xylanophilus	AE, H	60	3.2, 71	a
Symbiobacterium thermophilum	AE, H	60	3.6, 69	405
Thermoanaerobacter tengcongensis	AN, H	75	2.7, 38	21
Thermotoga maritima	AN, H	80	1.9, 46	270
Thermus thermophilus	AE, H	75	2.1, 69	152

C Crenarchaeota, *E* Euryarchaeota, AE aerobic, AN anaerobic, FAN facultative (an)aerobic, FA facultative autotrophic, P parasite. [a] genome paper not available of the eukaryotic cell (fusion of archaeal and bacterial cell [317]).

The first complete genome analysis of a hyperthermophilic archaeon, *Methanocaldococcus jannaschii* [47] (Table 1.1), confirmed the monophyletic position of the Archaea, with respect to the Bacteria and the Eukaryotes. In addition, Archaea appeared to possess a bacterial-like compact chromosomal organization with clustering of genes as polycistronic units (operons), and with only few interrupted genes (introns). Moreover, the archaeal systems that drive the flow of genetic information (transcription, translation, replication, DNA repair) correspond to the core of the eukaryal counterparts. These initial observations of bacterial-like "information storage" and eukaryal-like "information processing" have been confirmed by the analyses of subsequently sequenced hyperthermophilic model Archaea: the euryarchaea *Pyrococcus* spp. (*P. furiosus, P. abyssi, P. horikoshii*) as well as the crenarchaea *Sulfolobus* spp. (*S. solfataricus, S. tokodaii, S. acidocaldarius*) (reviewed by [244]). The comparative analysis of the genome of the hyperthermophilic bacterium *Thermotoga maritima* with that of *Pyrococcus furiosus* (both isolated from shallow thermal vents at the same beach (Vulcano, Italy), led to the conclusion that horizontal (or lateral) gene transfer substantially contributes to the apparent high degree of genome flexibility [206,270]. In addition, the comparison of closely related *Pyrococcus* species revealed a high degree of genome plasticity; moreover, it was proposed that the lateral gain as well as the loss of genes is a modular event [110]. Horizontal gene transfer has also been proposed to explain the relatively high degree of homology between genome fragments of the euryarchaeon *Thermoplasma acidophilum* and the crenarchaeon *Sulfolobus solfataricus*, phylogenetically distant Archaea but inhabiting the same environment (65-85 °C, pH 2). The *Sulfolobus*-like genes in the *T. acidophilum* genome are clustered into at least five discrete regions, again indicating recombination of larger DNA fragments [122].

After the genome sequence of thermophiles was established, comparative genomics analyses have been performed to assign potential functions for the identified open reading frames. As in all studied genomes, many unique and conserved hypothetical genes (typically half of the total number of genes) were found for which a function could not reliably be predicted. The main challenge of the post-genome era is therefore to improve the functional annotation of genes by integrating classical approaches (physiology, biochemistry and molecular genetics) with genomics-based high-throughput approaches (comparative, functional and structural genomics). Obvious targets of comparative and functional analysis of thermophile genomes are the numerous missing links in metabolic pathways as well as the largely unknown regulatory systems of RNA and protein turnover [111,244].

Sulfolobus

The Domain Archaea has two main Kingdoms, the Crenarchaeota (*e.g. Sulfolobus, Pyrobaculum, Pyrodictium*) and the Euryarchaea (*e.g. Pyrococcus, Methanococcus, Halobacterium*) (Fig. 1.2). Although initially the majority of the Archaea were isolated from extreme environments (high temperature, high salt concen-

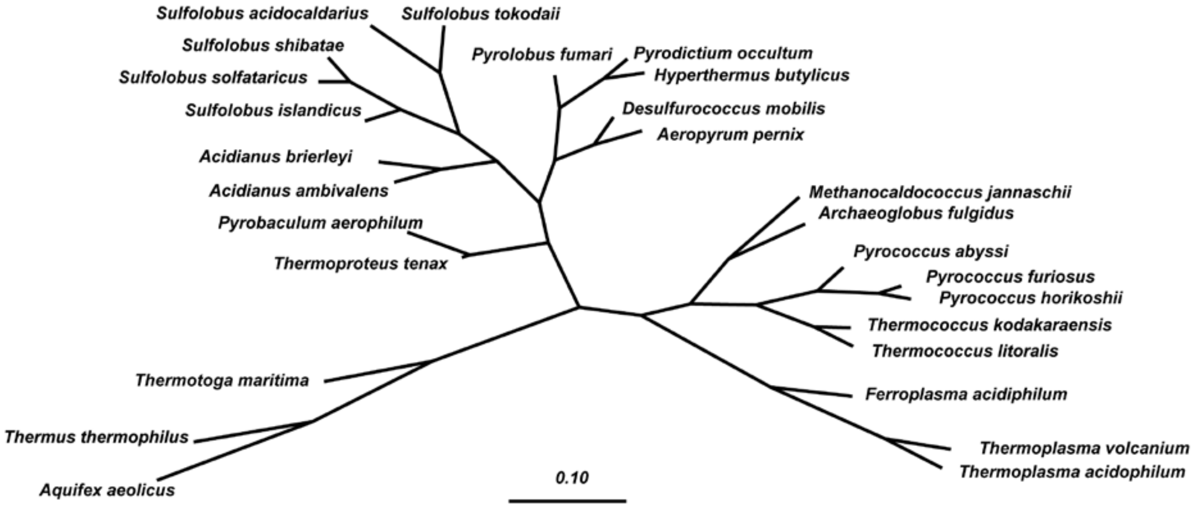

Sulfolobus acidocaldarius
Sulfolobus tokodaii
Sulfolobus shibatae
Sulfolobus solfataricus
Sulfolobus islandicus
Acidianus brierleyi
Acidianus ambivalens
Pyrobaculum aerophilum
Thermoproteus tenax
Pyrolobus fumari
Pyrodictium occultum
Hyperthermus butylicus
Desulfurococcus mobilis
Aeropyrum pernix
Methanocaldococcus jannaschii
Archaeoglobus fulgidus
Pyrococcus abyssi
Pyrococcus furiosus
Pyrococcus horikoshii
Thermococcus kodakaraensis
Thermococcus litoralis
Ferroplasma acidiphilum
Thermoplasma volcanium
Thermoplasma acidophilum
Thermotoga maritima
Thermus thermophilus
Aquifex aeolicus
0.10

Figure 1.2
Phylogenetic tree based on thermophilic archaeal and bacterial 16S rRNA sequences. Alignment and phylogenetic analysis were performed with the ARB software, and the three was constructed using the neighbor-joining method. The reference bar indicates the branch length that represents 10% diversity.

tration, extreme pH), it has become clear that Archaea also thrive in non-extreme environments: Archaea are abundant in many ecosystems ranging from soil to oceans [225,282]. Since its original discovery in Yellowstone, species of the genus *Sulfolobus* have been isolated from various solfataric fields, such as Pisciarelli (Italy; *S. solfataricus*) [446], Sogasel (Iceland; *S. islandicus*) [445], and Beppu (Japan; *S. tokodaii*) [388], Yang-Ming (Taiwan, *S. yangmingensis*) [168] and Tengchong (China, *S. tengchongensis*) [437]. *Sulfolobus* species are aerobic heterotrophs that oxidize a wide range of carbohydrates, yeast extract and peptide mixtures to CO_2 [144,344]. In addition, both autotrophic (oxidation of $S_2O_3^{2-}$, $S_4O_6^{2-}$, S^0 and S^{2-} to sulfuric acid, and of H_2 to water) and heterotrophic growth have been described for *S. acidocaldarius* [344,354], and it has been suggested that anaerobic respiration (*e.g.* reduction of NO) by certain *Sulfolobales* might be possible [353]. Other closely related genera that have been relatively well-characterized include *Acidianus* (order *Sulfolobales*; obligatory chemolithoautotroph; aerobic S^0 oxidation to sulphuric acid, anaerobic S^0 reduction coupled to H_2 oxidation), *Hyperthermus* (order *Desulfurococcales*; anaerobic, amino acid fermentation), and *Aeropyrum* (order *Desulfurococcales*; aerobe; heterotroph on starch and peptides) (Fig. 1.2). *Aeropyrum pernix* was the first Crenarchaeote for which the complete genome was sequenced [190] (Table 1.1). One of the most striking observations of the overall comparative analyses of multiple genomes has been their unexpected plasticity. In the case of many *Sulfolobales*, mobile genetic elements are anticipated to play an important role in the ever-changing genome composition [41]. The relatively large genome of *S. solfataricus* (3.0 Mb) (Table 1.1) contains an unprecedented high number of insertion sequences (IS elements, 344), some of which have been demonstrated to actively move through the genome [248]. Interestingly, the *S. tokodaii* genome (2.7 Mb) contains significantly fewer IS elements (95) [189], while the *S. acidocaldarius* genome (2.2 Mb) does not contain a single one [57]. The *Sulfolobus* genomes are also well known for their unusually high number of Clustered Regularly Interspaced Palindromic Repeats (CRISPR, >400), which constitute up to 1% of the genome, and are thought to encode an immune system against foreign genetic elements [22,169,230,243,262]. Possibly the high number of CRISPR in the genome reflects the abundance of viruses and plasmids in the thermoacidic ecological niche of *Sulfolobus* species.

The best characterized archaeal virus is SSV1 isolated from *S. shibatae* [316,337,438]. Upon infection of *S. shibatae* as well as *S. solfataricus*, the viral genome was generally found to be integrated into a specific tRNA$^{\mathrm{Arg}}$ gene of the host chromosome, indicating that site-specific integration is involved in establishing the lysogenic state [337]. The pRN family of small plasmids from the *Sulfolobales* includes pSSVx from *S. islandicus* REY 15/4, a hybrid between a plasmid and a fusellovirus. In the presence of the helper virus, SSV2, it can spread as a virus satellite [16]. Most conjugative plasmids have been isolated from a collection of *S. islandicus* strains. As well as spreading through *S. islandicus* cultures, these plasmids have also been reported to be capable of replicating in *S. solfataricus* (pNOB, pING, pSOG) [306,376]. In addition, exchange of marker genes has been well documented for *S. acidocaldarius* [315,340], suggesting a Hfr-like system as present in some conjugative bacteria. In line with this, a 30 kb fragment of the *S. tokodaii* chromosome strongly resembles conjugative plasmids like pSOG [105].

Genetic systems

One of the main reasons why *Sulfolobus solfataricus* has become a model archaeon for studying metabolism, transcription, translation and DNA replication has been its ability to grow aerobically on a wide variety of substrates in liquid media as well as on plates [23,353]. Moreover, Sulfolobus isolates from all over the world appeared to possess a wide array of extrachromosomal genetic elements [444]. These findings have encouraged researchers to develop tools for genetic manipulation.

The currently available genetic tools for *Sulfolobus* are summarized in Table 1.2. Many attempts have been made to generate shuttle vectors that replicate stably in both *Escherichia coli* and *Sulfolobus*, while maintaining a relatively high copy number in both organisms. Hybrid vectors were constructed containing the replicase gene from the euryarchaeal rolling circle plasmid pGT5 [104], which could be transformed to *S. acidocaldarius* by electroporation and maintained by applying butanol or benzylalcohol tolerance selection [1,13]. In the absence of selectable markers, the highly efficient self spreading capabilities of the conjugative plasmid pNOB8 [336] and *S. shibatae* fusellovirus SSV1 [316] were exploited to generate a population in which the majority of the cells contained the recombinant plasmid [100,375]. A thermostable hygromycin phosphotransferase mutant gene enabled selection of *S. solfataricus* Gθ transformants containing plasmid pEXSs, a small shuttle vector replicating from a 1.7 kb SSV1 origin [50,51]. Most success however, has been obtained by using β-galactosidase (*lacS*) deficient strains derived from spontaneous insertion-element disruption [338,435] or large chromosomal deletions at this locus [25,335]. At the phenotype level, β-galactosidase complementation allows blue/

Table 1.2 - Genetic tools for *Sulfolobus*

Vector (kb)	Selectable property / phenotypic marker	Archaeal replicon	*E. coli* replicon	Hosts	Features	Ref.
pCSV1 (6.1)	none	pGT5	pUC19	*P. furiosus* *S. acidocaldarius*	low copy number	1
pAG21 (6.5)	Sso adh, butanol and benzylalcohol tolerance	pGT5	pBR322	*P. furiosus* *S. acidocaldarius*	low copy number	13
pBS-KK (48)	conjugation blue/white screening	pNOB8	pBS	*S. solfataricus* P1 derivative PH1 *lacS*$^-$	Growth retardation, recombination into host genome	100
pEXSs (6.35)	hph$_{mut}$ Hygromycin^R	SSV1ori	pGEM5Zf(-)	*S. solfataricus* Gθ deriv. GθW, Δ*lacS*	low copy number, sometimes high background	25,50,51,68
pSSVrt (8.4)	spreading by SSV2 or SSV1	pSSVx	pUC19	*S. solfataricus* P2, *S. solfataricus* MT4, *S. solfataricus* GθW	high copy number	18
pKMSD48 (18.5)	self spreading plaque assay screening	SSV1	pBSII	*S. solfataricus* P1,P2 *S. solfataricus* Gθ	Single integration site and present as plasmid (5-40/cell)	375
pMJ03 (21.8)	self spreading Uracil auxotrophy compl. (*pyrEF*)	SSV1	pBSII	*S. solfataricus* P1 derivatives PH1 (*pyrEF*$^-$, *lacS*$^-$)	Single integration site and present as plasmid	175,248
pSVA5/6	see pMJ03	SSV1	pBSII	see pMJ03	Inducible Ara promoter, C-terminal StrepII and His tag	8
Knockout vector	Lactose auxotrophy compl. (*lacS*)	none	pNEB193, pUC19	*S.solfataricus* 98/2 *lacS*$^-$ derivatives, Δ*lacS* PBL2002/25	Homologous recombination suicide vector	7,335,435
pG, pJlacS	Uracil/Lactose auxotrophy compl. (*pyrEF/lacS*)	pRN1	pBSII	*S. solfataricus* *S. acidocaldarius*	Stable in both hosts, no rearrangements or integration	29

white activity screening using chromogenic compound 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-β-Gal), but has in some cases allowed selection for lactose utilization as a sole carbon and energy source [25,435]. The latter property has led to a successful homologous recombination system for *S. solfataricus* strain 98/2, in which a specific gene can be disrupted by a suicide vector [335,435]. For the first time in hyperthermophilic crenarchaeota, this system has enabled researchers to construct knock-out mutants and to study gene function *in vivo*. A system that allows the generation of gene knockouts has recently also been developed in the hyperthermophilic euryarchaeote *Thermococcus kodakaraensis* [252,332]. Significant progress has also been made with the construction of viral shuttle vector pMJ03. This SSV1-derived vector, which includes *lacS* for blue/white screening, also contains genes coding for orotidine-5'-monophosphate pyrophosphorylase (*pyrE*) and orotidine-5'-monophosphate decarboxylase (*pyrF*), which serve as a selectable marker complementing uracil auxotrophic *Sulfolobus* strains [175,248]. This selection marker appears to be crucial for the stability of the shuttle vector; without selective pressure, rearrangement of the shuttle vector was frequently observed during prolonged cultivation [175]. The shuttle vector was recently upgraded by the implementation of an inducible arabinose promoter and optional affinity tags, which allows for the efficient purification of recombinant proteins from *Sulfolobus* [8]. Small non-viral shuttle vectors based on plasmid *S. islandicus* plasmid pRN1 are the latest and most promising addition to the growing list of genetic tools. The vectors replicate stably in both *E. coli* and *Sulfolobus* spp, without genome integration or rearrangements [29]. The availability of these vectors will be instrumental in the genetic analysis of this extremophilic model organism.

Glycoside hydrolases

Within the domain of Archaea, *Sulfolobus* species are well known for their saccharolytic capacity. This feature is particularly well reflected by the number of glycoside hydrolases (GHs) and transglycosidases that have been found by either purification of their activity or by homology search methods. Several enzymes that are involved in the breakdown of starch have been described, such as extracellular surface-layer associated α-amylase [435], and an intracellular maltase that releases glucose from maltodextrins [106,319]. These dextrins are also the substrate for the enzyme TreY (Sso2095) and TreZ (Sso2093) that successively convert it into trehalose [249], which appears to be the main compatible solute in *Sulfolobus* spp. [331].

A gene cluster involved in glycogen synthesis and breakdown can be found in the genome. Apart from a glycogen synthase (Sso0987), this cluster contains a glycogen associated α-amylase (Sso0988) [52], a glucoamylase (Sso0990) [198] and a putative glycogen debranching enzyme (Sso0991).

Besides the potential to degrade α-1,4-linked glucosides, the genome of *Sulfolobus* has revealed several enzymes that are involved in the degradation of other plant polysaccharides. An interrupted α-fucosidase gene (Sso11867/Sso3060) was found, of which the mRNA was shown to be correctly translated by the ribosome via programmed -1 frameshifting [61]. A β-galactosidase (Sso3019) involved in the degradation of lactose and cellobiose has been described [300], which not only has served as a thermostable model protein in engineering studies [179], but also has allowed the development of genetic manipulation systems for *S. solfataricus* [175,435]. An α-xylosidase gene (Sso3022) was uncovered in the genomic vicinity of the β-galactosidase gene, which may suggest a cooperative function of the two enzymes in the degradation of xyloglucan oligosaccharides [264]. The same gene cluster also harbors a bifunctional β-D-xylosidase/α-L-arabinosidase (Sso3032) that is specifically induced in the presence of xylan polysaccharides [266]. Finally, two cellulases were found (Sso1949, Sso2534) that could operate in the extracellular environment of *Sulfolobus* which is characterized by extreme acidity and high temperature [159,231]. These and other *Sulfolobus* enzymes may have potentially interesting industrial applications in the microbial conversion of the world's most abundant plant polymer cellulose to biofuels such as ethanol and hydrogen.

Directed evolution

Directed evolution has emerged in just a decade as one of the most effective approaches to adapt proteins. It mimics the processes of Darwinian evolution *in vitro* [79], by combining random mutagenesis and/or recombination with screening or selection for protein variants that have desired properties [211]. This revolution in customizing protein properties was largely started by the introduction of DNA shuffling by Pim Stemmer in 1994 [377]. Soon after, enzyme improvements in activity [378], stability [140] but also adaptations in substrate specificity [442] were reported. Even whole operons containing three genes were successfully shuffled [70]. The DNA shuffling method consists of the reassembly of fragmented, randomly point-mutated genes in a self-priming PCR. Recombination of beneficial mutations occurs through template switching, creating a pool of mutated genes as a source of sequence diversity for a library. Although random mutagenesis at low error-rates can be very effective, DNA shuffling allows the generation new combinations of mutations, providing a more complete spectrum of genes with multiple substitutions. The major advantage of these approaches over site-directed mutagenesis is that, apart from the fact that the effect of a site-directed mutation is often unpredictable, no structural information of the protein is required. An adaptation called DNA family shuffling allows two or more homologous sequences to be used as starting material [71]. In this method homologous parent genes are mixed, randomly fragmented and recombined in an assembly PCR using conditions that permit annealing and extension of non-identical complementary strands. Following library screening, improved clones can be reshuffled to recombine useful adaptations in additive or synergistic ways, in effect mimicking the process of natural sexual recombination. The screening of libraries is the major bottleneck of directed evolution compared to site-directed mutagenesis. In directed evolution approaches the most desired mutants are only found when the first law of directed evolution is obeyed: “You get what you screen for” [440].

2

Reconstruction of central carbon metabolism in *Sulfolobus solfataricus* using a two-dimensional gel electrophoresis map, stable isotope labeling and DNA microarray analysis

Ambrosius P.L. Snijders*
Jasper Walther*
Stefan Peter
Iris Kinnman
Marjon G.J. de Vos
Harmen J.G. van de Werken
Stan J.J. Brouns
John van der Oost
Phillip C. Wright

*contributed equally

Proteomics (2006) **6**: 1518-1529

Abstract

In the last decade, an increasing number of sequenced archaeal genomes have become available, opening up the possibility for functional genomic analyses. Here, we reconstructed the central carbon metabolism in the hyperthermophilic crenarchaeon *Sulfolobus solfataricus* (glycolysis, gluconeogenesis and tricarboxylic acid cycle) on the basis of genomic, proteomic, transcriptomic and biochemical data. A 2-DE reference map of *S. solfataricus* grown on glucose, consisting of 325 unique ORFs in 255 protein spots, was created to facilitate this study. The map was then used for a differential expression study based on ^{15}N metabolic labeling (yeast extract + tryptone grown cells vs. glucose grown cells). In addition, the expression ratio of the genes involved in carbon metabolism was studied using DNA microarrays. Interestingly, only 3% and 14% of the genes and proteins respectively involved in central carbon metabolism showed a greater than two fold change in expression level. All results are discussed in the light of the current understanding of central carbon metabolism in *S. solfataricus* and will help to obtain a systems wide understanding of this organism.

INTRODUCTION

Sulfolobus solfataricus is a thermoacidophilic crenarchaeon that grows between 70 and 90 °C and in a pH range of 2-4 [446]. Its preference for environments hostile to many other organisms makes it an interesting source for novel, thermostable enzymes. *S. solfataricus* has been an attractive crenarchaeal model organism since its isolation in the early 1980's, and the completion of the genomic sequence in 2001 [353] has further increased its popularity. Currently, 1941 genes (53.11%) in TIGR's comprehensive microbial resource (CMR) database have no known function [297]. Of the 2977 Open Reading Frames (ORFs) originally identified in the genome of *S. solfataricus*, 40% of the genes are archaea specific, 12% are bacteria specific and 2.3% are shared exclusively with eukaryotes. Currently, genetic tools are under development that will contribute to our understanding of fundamental processes in *Sulfolobus* [51,68,175,375,435]. In order to fully exploit its potential for metabolic engineering, a deeper understanding of the central energy and precursor generating pathways is necessary.

The central metabolic pathways in archaea contain many unique features compared to the classical pathways in bacteria and eukaryotes [2,417]. In *S. solfataricus*, glucose degradation proceeds via a non-phosphorylated version of the Entner-Doudoroff (ED) pathway [344,334,84]. In this pathway, glucose is converted into pyruvate through the action of glucose dehydrogenase, gluconate dehydratase, 2-keto-3-*deoxy*-D-gluconate (KDG) aldolase, glyceraldehyde dehydrogenase, glycerate kinase, enolase and pyruvate kinase. Recently, experimental evidence has been provided for the operation of the semi-phosphorylated ED pathway in *S. solfataricus* in which KDG is phosphorylated [4]. Gluconeogenesis via a reversed ED pathway is unlikely, since the key enzymes in this pathway do not seem to be able to distinguish between glucose and galactose derivatives. In this case, gluconeogenesis via a reversed ED pathway would result in a mixture of glucose and galactose [213]. Instead, *in silico* analysis of the *Sulfolobus* genomes as well as experimental evidence has revealed the presence of a near complete set of proteins involved in the Embden-Meyerhof-Parnas (EMP) pathway [417], suggested to be active in the gluconeogenic direction rather than in the glycolytic direction [213].

In this study, we reconstructed central carbon metabolism and the TCA cycle on the basis of biochemical, computational, proteomic and DNA microarray data, obtained from cell extracts of *S. solfataricus* grown on sugars and peptides. First of all, a two-Dimensional gel Electrophoresis (2-DE) map was created to provide a global overview of protein expression under glucose degrading conditions. This map was then used to investigate the relative abundance of proteins involved in sugar metabolism under minimal or rich media through a ^{15}N metabolic labeling approach. Moreover, DNA microarray analysis was performed to compare mRNA expression under the same conditions. In the last few years, similar transcriptome studies have been conducted with several archaea that utilize different types of glycolysis. These organisms include: *Pyrococcus furiosus* [346] an obligate anaerobic hyperthermophile with an EMP-like pathway and *Haloferax vol-*

canii [441] a facultative anaerobic halophile using an ED-like glycolysis. However, there are relatively few studies that combine transcriptomics and proteomics, and none have so far been published for archaea.

Here, we present a study in which both quantitative proteomics and transcriptomics were used to analyze the expression of the genes involved in the central carbon metabolism of *Sulfolobus solfataricus*.

RESULTS AND DISCUSSION

Generation and application of a two-dimensional gel electrophoresis map

Figure 2.1 shows an image of the 2-DE reference map for *S. solfataricus*. With Coomassie Brilliant blue G250, approximately 500 spots were visualized. The highest spot count was obtained in the region pI 5-9, and proteins ranged in size from 15-123 kDa (predicted values). In total, 255 spots were selected for Mass Spectrometry (MS) analysis on the basis of their relative high abundance. In addition, faint spots were selected to test the sensitivity of the MS method. In total, 325 unique proteins in 255 spots were identified, with even the faintest spots yielding significant Molecular Weight Search (MOWSE) scores (> 51). All 255 spots were found on the triplicate gels. The complete dataset is presented in the supplementary material online. A subset, representing key elements of central energy metabolism and other relevant proteins is discussed more extensively in this paper. The highest MOWSE score, 1362, was achieved for elongation factor 2 (Sso0728, spot 26). Generally, one peptide (intact mass and tandem mass spectrometry (MS/MS) ion spectrum) was sufficient for confident identification of a *S. solfataricus* protein against the full Mass Spectrometry protein sequence Database (MSDB). In most cases, however, multiple peptides of the same protein were recovered from a spot. On average, the sequence coverage was 30%. The highest sequence coverage (75%) was found for the α-subunit of the proteasome (Sso0738) in spot 213. There was no correlation between the sequence coverage and the protein size. However, larger proteins usually resulted in higher MOWSE scores. This is due to the fact that larger proteins generate a larger number of unique peptides after tryptic digestion. For example, MOWSE scores greater than 800 were only obtained for proteins larger than 48 kDa.

The number of proteins that matched to ORFs that are either hypothetical or conserved hypothetical proteins was 157 (48%). This is similar compared to the expected 53%, on the basis of the genome composition. This was also found in a similar study on the *Methanocaldococcus jannaschii* proteome [138]. Interestingly, there were only two hypothetical proteins amongst the 20 most intense spots, (Sso0029, Sso0099 relating to spots 130 and 224 respectively). The relatively high abundance of those proteins suggests an important function.

Another important observation is that a number of proteins were found in more than one spot. Interestingly, this was true for a large number of proteins involved in the TCA cycle (*e.g.* 2-oxoacid:ferredoxin oxidoreductase (Sso2815) was found in eight different spots). There are a number of explanations for this: (1) Isoforms or post-translationally modified ver-

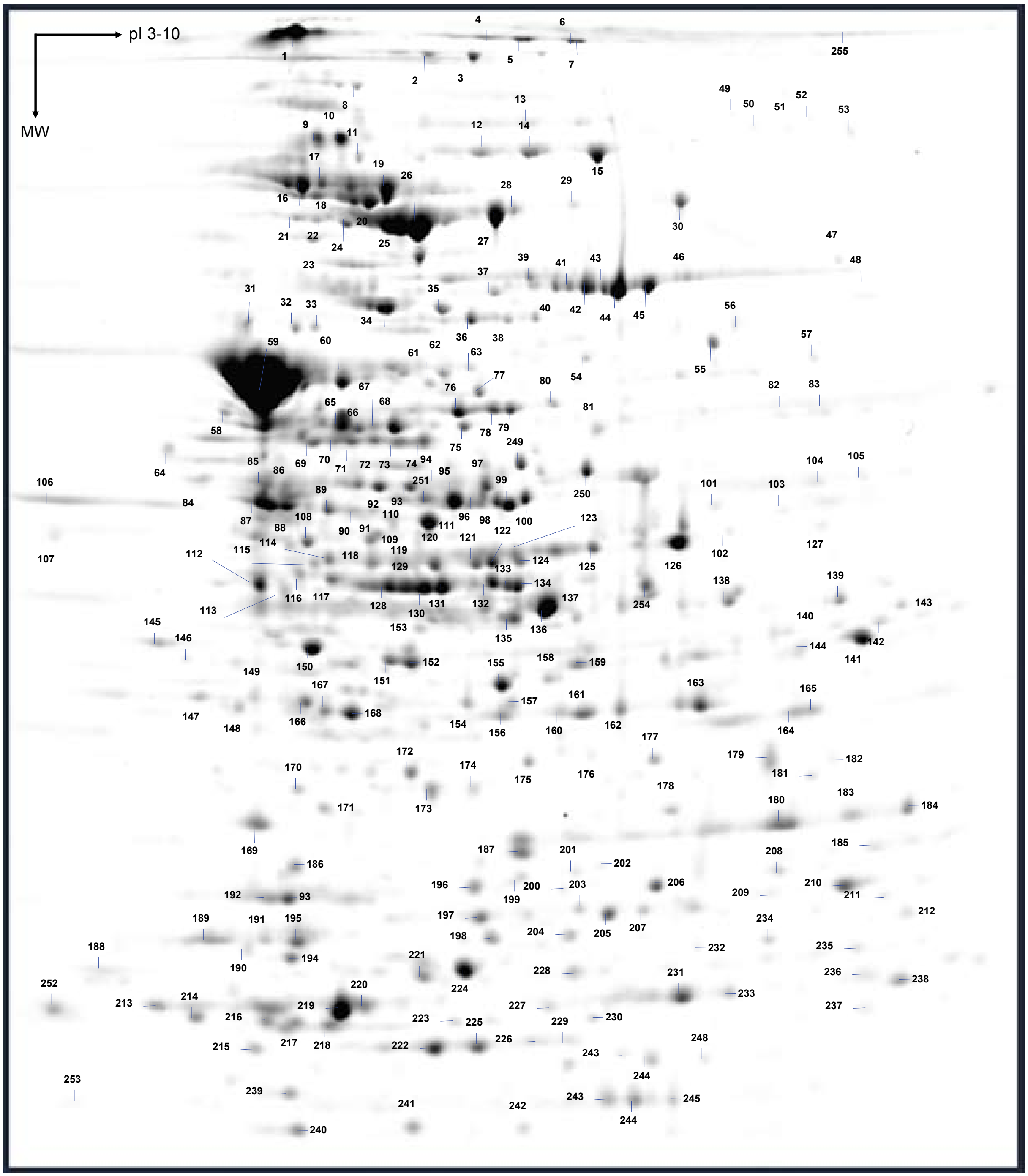

pI 3-10
MW
4
6
1
255
5
7
2
3
13
49
52
50
51
53
8
10
9
12
14
11
17
19
15
26
29
28
16
18
20
30
21
22
27
25
24
47
23
39
41
43
46
48
37
31
35
32
33
56
34
40
42
44
45
57
59
60
36
38
62
61
63
55
77
80
54
82
83
67
76
68
65
66
81
58
78
79
249
75
104
105
85
94
69
70
71
72
73
74
97
64
86
95
106
251
99
101
103
89
250
84
92
93
110
96
98
123
87
108
111
100
88
90
91
109
120
121
122
127
114
119
112
115
118
124
102
126
107
129
133
125
139
138
134
116
117
137
143
128
131
132
254
113
130
136
140
145
153
146
135
144
142
158
159
141
150
152
155
149
151
167
163
165
161
157
147
168
166
148
154
156
160
162
164
177
172
179
182
174
170
176
181
175
178
183
180
171
173
184
187
201
185
169
208
202
186
196
206
210
200
203
209
192
93
199
211
212
197
234
189
191
195
207
204
205
198
188
232
235
221
194
228
231
236
190
238
252
220
224
214
233
213
219
227
237
216
223
225
229
230
248
226
215
217
218
222
243
244
253
239
243
245
241
242
244
240

Figure 2.1
2-DE reference map for S. solfataricus grown on glucose. All numbered spots were subjected to LC-MS-MS analysis. Results are available online as supplementary table 1.

sions of the protein might be present in the cell (2) the protein was modified during protein extraction or during 2-DE (*e.g.* proteolysis, methionine oxidation,), (3) the protein does not resolve well on the gel and therefore “smears” out over a large pH or mass range, or (4) the denaturating conditions are not strong enough to completely break protein associations. The presence of a protein in multiple spots was also observed in similar proteomic studies [138]. To find post-translational modifications, all mass spectra were searched again but this time with phosphorylation of serine or threonine, and with methylation set as variable modifications.

Unfortunately, no consistent results were obtained, and therefore more specific studies targeted to identify post-translational modifications are necessary.

In a number of cases multiple proteins per spot were found. Often these proteins have similar molecular weights (MW) and iso-electric points (pI) indicating that the resolution on the gel was insufficient to resolve these proteins into single protein spots. In other cases however, proteins in the same spot differ significantly in MW and pI. These represent biologically interesting cases since these could indicate stable protein associations. An example was found in spot 1, where subunits α, β en γ of aldehyde oxidoreductase (Sso2636, Sso2637, Sso2639) were found.

Protein quantitation was performed on the basis of ^{15}N metabolic labeling as recently described. With this method a number of problems associated with 2-DE (*e.g.* multiple proteins per spot) can be avoided. Moreover, the reproducibility of gel staining becomes of lesser importance since protein quantitation takes place on the MS [368].

Figure 2.2 shows an example of a TOF-MS spectrum containing both the light and the heavy version of the peptide IFGSLSSNYVLTK. This peptide is derived from the 2-keto-3-*deoxy*-D-gluconate aldolase (Sso3197). The light peptide at m/z 714.99 corresponds to the yeast extract + tryptone (YT) grown cells and the heavy peptide at m/z 722.47 corresponds to the glucose (G) grown cells. The relative abundance of the heavy and light peptide can now be calculated by determining the ratio of the peak areas. Note that the difference in mass between the heavy and light version of the peptide corresponded exactly to the number of nitrogen atoms in the peptide, in this case 15 atoms ($\Delta M/z = 7.5$). Table 2.1 summarizes the differential proteomic data obtained in this way, as well as the corresponding transcriptomic data.

Exploration of the transcriptome

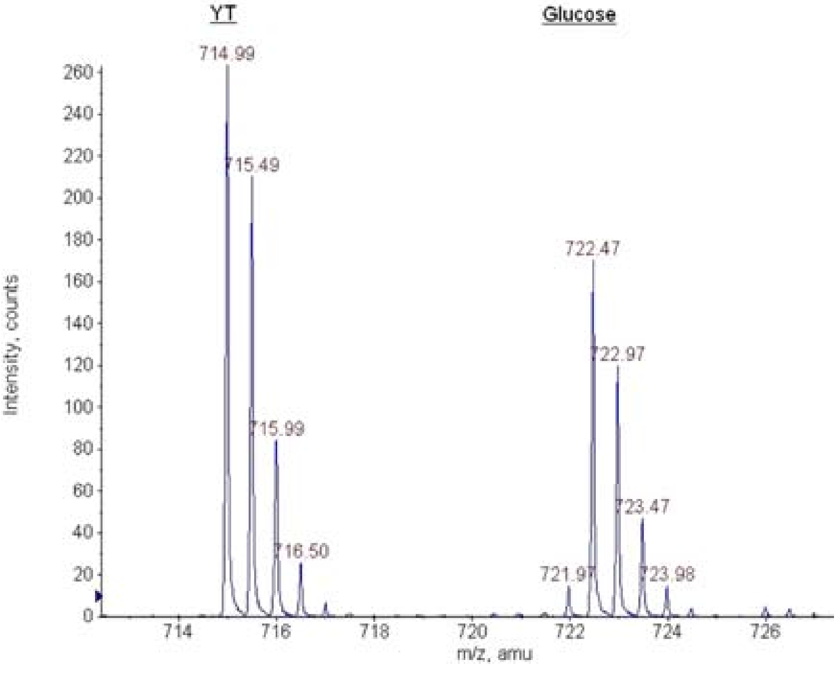


Figure 2.2
Peptide quantitation. TOF MS spectrum of a ^{15}N labeled and an unlabeled peptide. The peak on the left at *m/z* 714.99 represents the unlabeled version of the peptide (protein from cells grown on yeast extract+tryptone (YT)). The peak at the right at *m/z* 722.47 represents the ^{15}N labeled version of the peptide (protein from cells grown on glucose). This peptide was identified as IFGSLSSNYVLTK, corresponding to 2-keto-3-*deoxy*-D-gluconate aldolase (Sso3197). The ratio between the areas of the heavy and light versions of this peptide was 1.56.

In total, 1581 of the 2315 genes printed on the microarray were used in the analysis (selected, according to criteria described above). There were 184 significantly differentially expressed genes ($p<0.05$; p is the statistical certainty that the observed change in ratio is not caused by a biological effect). In total, 135 and 49 genes are up-regulated under glucose and YT conditions, respectively. Of these up-regulated genes 23% and 20% were annotated as either hypothetical or conserved hypothetical. Interestingly, these percentages are lower than the expected 53%. Genes involved in amino acid biosynthesis were regulated under both glucose and YT conditions. This was 16% and 10%, of the total amount of upregulated genes, in the case of glucose and YT, respectively. Regulation in this functional group was expected since amino acids are synthesized under glucose conditions and predominantly degraded under YT conditions. This data, therefore, provides an excellent starting point for amino acid metabolism reconstruction. Future biochemical and proteomic studies are necessary to confirm the exact composition and direction of the responsible pathways.

In addition, three genes involved in nitrogen metabolism were regulated: (1) glutamate synthase (Sso0684, 0.15) (2) glutamine synthase (Sso0366; 0.27) and (3) glutamate dehydrogenase (Sso2044; 6.29), absolute ratios are given as YT/G). These results show that cells which grow on glucose assimilate nitrogen by the sequential action of glutamine synthase and glutamate synthase. Under YT conditions glutamate dehydrogenase produces free ammonium by converting glutamate into 2-oxoglutarate. This is necessary because there is an excess of nitrogen bound to carbon when grown in the presence of YT.

Transport and binding proteins are also a major group of up-regulated genes (12% and 8% for glucose and YT, respectively). Previously, it was shown that both glucose and YT grown cells have the capacity to transport glucose [99]. This is reflected by the fact that the genes involved in glucose transport were not differentially expressed (Sso2847, Sso2848, Sso2849, Sso2850). In addition, genes involved in dipeptide transport were up-regulated under YT conditions (Sso1282; 2.01 / Sso2615; 1.74 / Sso2616; 1.57). Interestingly, genes involved in maltose transport were slightly up-regulated under glucose conditions (Sso3053; 0.36 / Sso3058; 0.50 / Sso3059; 0.53).

Metabolic pathway reconstruction

During the last two decades, the main metabolic pathways in *Sulfolobus* spp. have been the subject of extensive experimental research. This has led to a profound understanding of the enzymes and protein complexes that are involved in the glycolysis, the tricarboxylic acid cycle (TCA) and related metabolic conversions [77,417]. The availability of the genome sequences of *S. solfataricus* [353], *S. tokodaii* [189] and *S. acidocaldarius* [57] has recently allowed for the identification of the genes encoding these proteins by matching full-length or N-terminal protein sequences to the predicted proteomes. A reconstruction of the central carbon metabolic pathways in *Sulfolobus solfataricus* was performed (Fig. 2.3). The results should be taken with a degree of caution since significant differences exist in the physiology between the three *Sulfolobus* species [334]. Al-

Table 2.1 - Relative abundances of mRNA and protein levels of the genes involved in central metabolic pathways of *Sulfolobus solfataricus* grown on yeast extract and tryptone (YT) compared to glucose (G).

Locus	Enzyme description	EC	COG	Transcriptomics[a]	Proteomics[a]	Ref.
Glycolysis						
Sso3003	Glucose-1-dehydrogenase	1.1.1.47	1063	NS	NF	213
Sso2705	Gluconolactonase	3.1.1.17	3386	1.15 ± 0.07	NF	417
Sso3041	Gluconolactonase	3.1.1.17	3386	NF	NF	
Sso3198	Gluconate dehydratase	4.2.1.39	4948	1.00 ± 0.07	1.42 ± 0.14	200,214
Sso3197	2-keto-3-*deoxy*-D-gluconate aldolase	4.1.2.-	0329	0.96 ± 0.19	1.55 ± 0.05	46
Sso3195	2-keto-3-*deoxy*-D-gluconate kinase	2.7.1.45	0524	1.19 ± 0.15	NF	417
Sso3194	Glyceraldehyde-3-phosphate dehydrogenase (non-phosphorylating)	1.2.1.3	1012	0.87 ± 0.10	0.66 ± 0.07	4,45
Sso2639[c]	Aldehyde oxidoreductase, α-subunit	1.2.7.-	1529	0.65 ± 0.01[b]	4.51 ± 0.78	185
Sso2636[c]	Aldehyde oxidoreductase, β-subunit		1319	0.55 ± 0.13[b]	4.89 ± 0.40	
Sso2637[c]	Aldehyde oxidoreductase, γ-subunit		2080	0.62 ± 0.14[b]	4.22 ± 1.03	
Sso0666	Glycerate kinase	2.7.1.-	2379	0.70 ± 0.24	NF	84,417
Sso0981	Pyruvate kinase	2.7.1.40	0469	0.98 ± 0.10	NF	345
Glycolysis / Gluconeogenesis						
Sso0417	Phosphoglycerate mutase	5.4.2.1	3635	1.03 ± 0.13	1.55 ± 0.14	414
Sso2236	Phosphoglycerate mutase	5.4.2.1	0406	NS	NF	
Sso0913	Enolase	4.2.1.11	0148	1.36 ± 0.35	1.59 ± 0.23	290
Gluconeogenesis						
Sso0883	Phospho*enol*pyruvate synthase	2.7.9.2	0574	1.62 ± 0.08[b]	1.77 ± 0.22	162
Sso0527	Phosphoglycerate kinase	2.7.2.3	0126	1.26 ± 0.26	2.30 ± 0.28	154
Sso0528	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)	1.2.1.12	0057	1.07 ± 0.20	1.16 ± 0.02	326
Sso2592	Triose-phosphate isomerase	5.3.1.1	0149	NF	1.17 ± 0.12	204
Sso3226	Fructose-bisphosphate aldolase	4.1.2.13	1830	NS	1.84 ± 0.10	357
Sso0286	Fructose-bisphosphatase	3.1.3.11	1980	1.24 ± 0.18	1.32 ± 0.05	272
Sso2281	Glucose-6-phosphate isomerase	5.3.1.9	0166	1.01 ± 0.13	1.51 ± 0.10	149
Sso0207	Phosphoglucomutase	5.4.2.2	1109	1.03 ± 0.32	1.55 ± 0.01	372
Tricarboxylic acid cycle						
Sso2589	Citrate synthase	2.3.3.1	0372	0.84 ± 0.09	1.02 ± 0.03	232,367
Sso1095	Aconitase	4.2.1.3	1048	1.05 ± 0.14	1.11 ± 0.03	406
Sso2182	Isocitrate dehydrogenase	1.1.1.42	0538	1.34 ± 0.65	1.18 ± 0.03	48
Sso2815[d]	2-oxoacid:ferredoxin oxidoreductase α/γ-subunit	1.2.7.1 1.2.7.3	0674 1014	0.89 ± 0.07	0.56 ± 0.05	128,196, 443
Sso2816[d]	2-oxoacid:ferredoxin oxidoreductase β-subunit		1013	0.85 ± 0.31	0.60 ± 0.02	
Sso2482	Succinate-CoA ligase, α-subunit	6.2.1.5	0074	0.93 ± 0.25	0.54 ± 0.04	78
Sso2483	Succinate-CoA ligase, β-subunit		0045	0.94 ± 0.30	0.51± 0.05	
Sso2356	Succinate dehydrogenase, subunit A	1.3.99.1	1053	NS	0.58 ± 0.4	170
Sso2357	Succinate dehydrogenase, subunit B		0479	0.75 ± 0.28	NF	
Sso2358	Succinate dehydrogenase, subunit C		2048	0.94 ± 0.27	NF	
Sso2359	Succinate dehydrogenase, subunit D			0.89 ± 0.16	NF	
Sso1077	Fumarate hydratase	4.2.1.2	0114	1.08 ± 0.10	1.53 ± 0.09	64,307
Sso2585	Malate dehydrogenase	1.1.1.37	0039	0.82 ± 0.27	0.69 ± 0.01	151
Glyoxylate shunt						
Sso1333	Isocitrate lyase	4.1.3.1	2224	0.30 ± 0.07[b]	NF	407
Sso1334	Malate synthase	2.3.3.9	2225	1.11 ± 0.47	1.18 ± 0.04	
C3/C4 interconversions						
Sso2869	Malic enzyme	1.1.1.38	0281	1.05 ± 0.24	1.92 ± 0.15	24
Sso2537	Phosphoenolpyruvate carboxykinase	4.1.1.32	1274	1.42 ± 0.42	NF	129
Sso2256	Phosphoenolpyruvate carboxylase	4.1.1.31	1892	0.83 ± 0.18	0.88 ± 0.17	112,329

NF not found, NS no significant signal. [a] Relative abundance ratio with standard deviation Yeast extract + Tryptone grown cells / Glucose grown cells (YT/G), [b] Probability value (*p*) smaller than 0.05, [c] Enzyme complex has broad substrate specificity for aldehydes, [d] exhibits pyruvate, 2-oxoglutarate and 2-oxobutyrate oxidoreductase activity.

most all proteins involved in this scheme have been experimentally verified in either *Sulfolobus* spp. or other hyperthermophilic Archaea, such as *Thermoproteus tenax*, *Archaeoglobus fulgidus*, *Thermoplasma acidophilum*, *Pyrococcus furiosus*, *Thermococcus kodakaraensis*, *Methanothermus fervidus* and *Methanocaldococcus jannaschii*. Moreover, the vast majority of the anticipated proteins in *Sulfolobus solfataricus* were found on the 2-DE reference map (Fig. 2.1). On average, the TCA cycle proteins made up approximately 12% of the total staining intensity.

Glycolysis and gluconeogenesis

The genus *Sulfolobus* is known to degrade glucose according to a modified version of the Entner-Doudoroff (ED) pathway. While in most cases phosphorylation in the bacterial ED pathway occurs at the level of glucose, gluconate or 2-keto-3-*deoxy*-D-gluconate (KDG), *S. solfataricus* has been reported to utilize a non-phosphorylated version of the ED pathway, which phosphorylates only at the level of glycerate [84,348]. Recent experimental findings, however, indicated the presence of a semi-phosphorylated ED pathway, in which KDG is phosphorylated and subsequently cleaved forming pyruvate and glyceraldehyde-3-phosphate (GAP) by the action of the KDG kinase (Sso3195) and the KDG aldolase (Sso3197) respectively. GAP is then oxidized by a non-phosphorylating GAP dehydrogenase (GAPN, Sso3194) forming 3-phosphoglycerate (3PG) [4]. The only net difference between the non- and semi-phosphorylated pathways is the fact that either reduced ferredoxin (Fd_R) or NADPH is produced, since neither pathway directly yields ATP by substrate level phosphorylation.

The intrinsic irreversibility of several ED enzymes, such as the gluconate dehydratase and GAPN, prevents the ED to operate in the gluconeogenic direction, which is, for instance, required to store energy in the form of glycogen [362]. Another important role for the gluconeogenic EMP pathway is the production of fructose-6-phosphate (F6P), which has been proposed to be the main precursor for the Pentose Phosphate Pathway (PPP) [417]. Except for three kinases (GK glucokinase, PFK phosphofructokinase and PK pyruvate kinase), the catabolic Embden-Meyerhof-Parnas (EMP) pathway consists of reversible enzymes. Although the genes encoding a GK and PFK were absent, the genes encoding the reversible EMP enzymes were all found in the genome of *Sulfolobus*. Moreover, a gene encoding a fructose-1,6-bisphosphatase (FBPase) was also detected. Because it is known that the catabolic EMP pathway is not operational in *Sulfolobus* [348], it is likely that these EMP enzymes serve a gluconeogenic role. The simultaneous operation of both the ED and a gluconeogenic EMP pathway, however, requires a strict control of the metabolic flux through the pathway in order to prevent an energetically futile cycle. Allosteric regulation, post-translational protein modification and regulation at the transcriptional level are common strategies to modulate the activity and abundance of key enzymes, such as the fructose-1,6-bisphosphatase.

Although glycolysis in *Sulfolobus* is well studied, there are still unconfirmed genes and activities in the pathway. For instance, the transcriptome analysis revealed the expression of one of two putative gluconolactonases

(Sso2705) that have generally been omitted in the analysis of the ED pathway, since the reaction from gluconolactone to gluconate also occurs spontaneously [333]. The expression of the enzyme, however, would suggest a functional role in the metabolism of *Sulfolobus*. Additionally, only one of two phosphoglycerate mutases (Sso0417) that were found in its genome was expressed in both the proteome and transcriptome, while the other type (Sso2236) remained undetected. Expression of the predicted glycerate kinase (Sso0666) was only detected at the mRNA level.

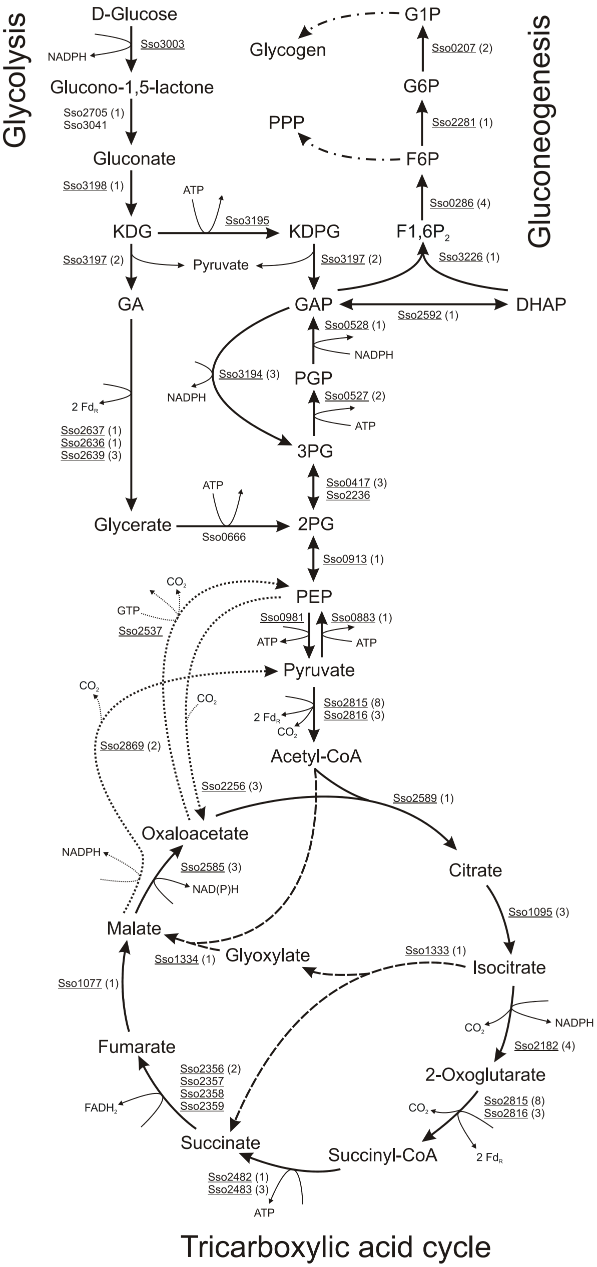

Glycolysis
Gluconeogenesis
D-Glucose
Sso3003
NADPH
Glucono-1,5-lactone
Sso2705 (1)
Sso3041
Gluconate
Sso3198 (1)
ATP
Sso3195
KDG
KDPG
Sso3197 (2)
Pyruvate
Sso3197 (2)
GA
GAP
Glycogen
G1P
Sso0207 (2)
G6P
Sso2281 (1)
PPP
F6P
Sso0286 (4)
F1,6P₂
Sso3226 (1)
Sso2592 (1)
DHAP
Sso0528 (1)
NADPH
Sso3194 (3)
NADPH
PGP
Sso0527 (2)
ATP
2 Fd_R
Sso2637 (1)
Sso2636 (1)
Sso2639 (3)
3PG
Sso0417 (3)
Sso2236
ATP
Glycerate
Sso0666
2PG
Sso0913 (1)
CO₂
GTP
Sso2537
PEP
Sso0981
Sso0883 (1)
ATP
ATP
Pyruvate
CO₂
CO₂
Sso2815 (8)
Sso2816 (3)
2 Fd_R
CO₂
Sso2869 (2)
Acetyl-CoA
Sso2256 (3)
Sso2589 (1)
Oxaloacetate
NADPH
Sso2585 (3)
NAD(P)H
Citrate
Sso1095 (3)
Malate
Sso1334 (1)
Glyoxylate
Sso1333 (1)
Isocitrate
Sso1077 (1)
CO₂
NADPH
Sso2182 (4)
Fumarate
Sso2356 (2)
Sso2357
Sso2358
Sso2359
2-Oxoglutarate
FADH₂
CO₂
Sso2815 (8)
Sso2816 (3)
Succinate
Succinyl-CoA
2 Fd_R
Sso2482 (1)
Sso2483 (3)
ATP
Tricarboxylic acid cycle

Figure 2.3
Reconstruction of the central metabolic pathways in *Sulfolobus solfataricus*. Genes involved in the glycolysis, gluconeogenesis and tricarboxylic acid cycle were surveyed and are indicated by their locus name. Underlined genes were experimentally verified in *Sulfolobus* or related hyperthermophilic Archaea (Table 2.1). The number of spots that were found on the 2-DE reference map is indicated between brackets. The glyoxylate shunt is shown by dashed arrows, while the three to four carbon interconversions are depicted by dotted arrows. Mixed dashed and dotted arrows indicate that the exact pathway to glycogen and pentoses is unknown. The following abbreviations were used: KD(P)G 2-keto-3-*deoxy*-D-gluconate-(6-phosphate), GA(P) glyceraldehyde-(3-phosphate), PGP 1,3-bisphosphoglycerate, 3PG 3-phosphoglycerate, 2PG 2-phosphoglycerate, PEP phosphoenolpyruvate, DHAP dihydroxyacetonephosphate, F1,6P$_2$ fructose-1,6-bisphosphate, F6P fructose-6-phosphate, G6P glucose-6-phosphate, G1P glucose-1-phosphate, Fd$_R$ reduced ferredoxin, PPP pentose phosphate pathway. NAD(P)H indicates that both NAD$^+$ and NADP$^+$ can be used as a cofactor. Arrows represent the presumed physiologically relevant direction of catalysis and are not indicative of enzymatic reversibility.

Tricarboxylic acid cycle

Sulfolobus spp. is an obligate aerobe that primarily obtains energy by the oxidation of organic molecules and elemental sulphur [36]. This oxidation results in the formation of reduced electron carriers, such as NAD(P)H, Fd$_R$ and FADH$_2$. The majority of these reducing equivalents are generated in the tricarboxylic acid (TCA) cycle. Per round of the cycle, the succinate-CoA ligase of *Sulfolobus* generates one molecule of ATP, instead of the commonly produced GTP [78]. Apart from being the main metabolic converter of chemical energy, the TCA cycle intermediates serve an important role as biosynthetic precursors for many cellular components, such as amino acids. Consequently, when too many intermediates are withdrawn from the cycle, they need to be replenished by anaplerotic enzyme reactions. The phospho*enol*pyruvate carboxylase (PEPC), which forms oxaloacetate from phospho*enol*pyruvate, is the only anaplerotic enzyme from *Sulfolobus* that has been described to date [112,329]. A gene product with high similarity to known pyruvate carboxylases could not be detected in the predicted proteome of *Sulfolobus*. In the glyoxylate shunt, which is normally only active during growth on acetate, isocitrate and acetyl-CoA are converted into succinate and malate by the action of the isocitrate lyase and the malate synthase. Interestingly, the isocitrate lyase of glucose-grown *S. acidocaldarius* cells co-purified with the aconitase [406,407]. Not only would this suggest a cytosolic association of the enzymes, but it also suggests that the glyoxylate shunt operates under saccharolytic conditions. This pathway may therefore constitute another way of replenishing four-carbon TCA cycle intermediates.

When there is an excess of TCA intermediates, for instance during growth on proteinaceous substrates, both malate and oxaloacetate can be decarboxylated to pyruvate by the malic enzyme [24]. Oxaloacetate can also be converted to phospho*enol*pyruvate by the GTP-dependent carboxykinase [129]. These four-to-three carbon conversions then provide the precursors that are required in, for instance, the gluconeogenesis pathway. In contrast to aerobic bacteria and eukaryotes, *Sulfolobus* uses ferredoxin instead of NAD^+ as a cofactor in the formation of acetyl-CoA from pyruvate and succinyl-CoA from 2-oxoglutarate [196]. The protein complex responsible for both conversions was shown to consist of two subunits; a fused α/γ subunit (Sso2815) and a β subunit (Sso2816) [128,443]. The genome sequences of the three *Sulfolobus* species, however, revealed several paralogs of ferredoxin-dependent 2-oxoacid oxidoreductases, which might also be involved in these conversions.

What is also evident from this reconstruction is that almost all dehydrogenases in the central carbon metabolism of *Sulfolobus* show a clear cofactor preference for $NADP^+$ over NAD^+ [24,48,78,213,326,353]. The only exception to this rule seems to be the malate dehydrogenase, which, at least *in vitro*, uses both electron acceptors equally well [151]. In bacteria and eukaryotes, most NADPH is usually formed in the PPP and used for reductive biosynthesis purposes. In *Sulfolobus*, the apparent enzyme preference for $NADP^+$ would suggest a more general role of its reduced form, in energy conservation by oxidative phosphorylation. Interestingly, as noted by She *et al.* [353], all genes encoding the NAD(P)H dehydrogenase complex are present in the genome, except the three that encode the subunits which are required for NAD(P)H binding and oxidation. It has been proposed that the reducing equivalents are first transferred to ferredoxin by a NADPH:ferredoxin oxidoreductase, before entering the respiratory chain [353].

Regulation of the main metabolic pathways

Insight was obtained into the regulation of the genes anticipated in glycolysis, gluconeogenesis and TCA cycle by measuring the relative abundance of their mRNA and protein levels by using a whole-genome DNA microarray and a quantitative proteomics approach, respectively (Table 2.1). In the measurements, 35 out of 41 transcripts ratios were determined, while 29 out of 41 protein ratios were analysed on 2-DE gels. On average the proteomic and transcriptomic data correlate reasonably well. For 26 genes both proteomic data and transcriptomic data are presented. In general, changes at proteomic and transcriptomic level show a similar trend, however, proteomic changes tend to be more pronounced. In only 3 cases the proteomic data contradict the transcriptome data. This concerns the three subunits for aldehyde dehydrogenase (Sso2639, Sso2636 and Sso2637). However, the fact that these clustered genes show a similar ratio at proteomic or transcriptomic level indicates the reliability of the data. Interestingly, all three subunits were found in the same protein spot on the gel, suggesting that a strong (non-covalent) interaction exists between them. The stability of the protein complex might be affected by stabilizing factors such as cofactors that may lead to different degrees of aggregation under different growth conditions. In terms of regulatory effects, the glyceraldehyde-3-phosphate dehydrogenase (non-phosphorylating; GAPN) was up-regulated under glucose conditions, or alternatively, down-regulated during growth in YT media. This is not surprising, since GAP is the crucial intermediate between the ED and gluconeogenic EMP, and too much of the strictly catabolic GAPN would be likely to interfere with gluconeogenesis. The enzymes involved in gluconeogenesis were all slightly up-regulated during growth on YT media, in agreement with expectations. Especially the phospho*enol*pyruvate synthase and the phosphoglycerate kinase, key enzymes of the pathway appeared to be most differentially expressed.

The expression levels of the TCA-cycle genes were only marginally different under the two conditions. Under glucose conditions, several enzymes of the TCA cycle were slightly induced at proteomic level, including the 2-oxoacid:ferredoxin oxidoreductase, the succinate-CoA ligase, the succinate dehydrogenase and the malate dehydrogenase. This was also true for the enzymes that replenish the four-carbon TCA cycle intermediates, such as the isocitrate lyase and the phospho*enol*pyruvate carboxylase. This ensures that sufficient oxaloacetate is present to serve as biosynthetic precursor and as an acceptor molecule for acetyl-CoA. The differences may be due to the fact that glucose catabolism mainly results in acetyl-CoA and oxaloacetate formation, whereas peptide degradation probably yields various central intermediates of carbon metabolism, such as pyruvate (Ala, Cys, Trp, Thr, Ser, Gly), acetyl-CoA (Phe, Tyr, Ile, Leu, Lys, Trp, Thr), 2-oxoglutarate (Arg, Gln, His, Pro, Glu), succinyl-CoA (Ile, Met, Val, Thr), fumarate (Phe, Tyr, Asp) and oxaloacetate (Asn, Asp).

Concluding remarks

In this study, we have created a proteome reference map for *Sulfolobus solfataricus* consisting of 325 proteins in 255 spots, and have reconstructed its central carbon metabolic pathways. The expression of the genes in these pathways was analysed by measuring the relative abundance of mRNA and protein under peptide- or sugar-degrading conditions. Although most observed differences were small, the expression of some key enzymes in glycolysis, gluconeogenesis and TCA cycle was significantly altered. Apart from looking at abundance levels, proteomics studies are now ongoing that focus on the modulation of enzyme activity by protein post-translational modification. These studies will provide additional clues that will reveal the details of regulation of the central carbon metabolism in *Sulfolobus solfataricus*.

EXPERIMENTAL PROCEDURES

Cell growth and harvest

Sulfolobus solfataricus P2 (DSM1617) was grown aerobically in a rotary shaker at 80 °C in a medium of pH 3.5-4.0 which contained: 2.5 g L^{-1} $(NH_4)_2SO_4$, 3.1 g L^{-1} KH_2PO_4, 203.3 mg L^{-1} $MgCl_2 \bullet 6 H_2O$, 70.8 mg L^{-1} $Ca(NO_3)_2 \bullet 4 H_2O$, 2 mg L^{-1} $FeSO_4 \bullet 7 H_2O$, 1.8 mg L^{-1} $MnCl_2 \bullet 4 H_2O$, 4.5 mg L^{-1} $Na_2B_4O_7 \bullet 2 H_2O$, 0.22 mg L^{-1} $ZnSO_4 \bullet 7 H_2O$, 0.06 mg L^{-1} $CuCl_2 \bullet 2 H_2O$, 0.03 mg L^{-1} $Na_2MoO_4 \bullet 2 H_2O$, 0.03 mg L^{-1} $VOSO_4 \bullet 2 H_2O$, 0.01 mg L^{-1} $CoCl_2 \bullet 6 H_2O$. The medium was supplemented with Wollin vitamins, and either 0.3% to 0.4% D-glucose (G) or 0.1% Yeast extract and 0.2% Tryptone (YT). The Wollin vitamin stock (100x) contained 2 mg L^{-1} D-Biotin, 2 mg L^{-1} Folic acid, 10 mg L^{-1} Pyridoxine-HCl, 10 mg L^{-1} Riboflavin, 5 mg L^{-1} Thiamine-HCl, 5 mg L^{-1} Nicotinic acid, 5 mg L^{-1} DL-Ca-Pantothenate, 0.1 mg L^{-1} Vitamin B12, 5 mg L^{-1} p-Aminobenzoic acid, 5 mg L^{-1} Lipoic acid. Cell growth was monitored by measuring the turbidity at 530 or 600 nm. Cells for the proteome reference map were harvested by centrifugation in the late exponential growth phase at an A_{530} of 1.0. Cells were washed twice with a 10 mM Tris-HCl Buffer (pH 7). Subsequently, cells were stored at -20 °C until required. During this whole process, considerable care was taken to ensure that culture to culture variation was minimized, and cultures were prepared in at least triplicate. In the case of the ^{15}N labeling experiment, $(^{15}NH_4)_2SO_4$ was used as the nitrogen source. Cells were incubated with ^{15}N ammonium sulphate for at least 8 doubling times to allow for full incorporation of the label. After this, the ^{14}N and ^{15}N growth experiments were set up simultaneously. When the optical density reached a value of 0.5, the cultures were mixed. To ensure that equal amounts of biomass were mixed, slight corrections in volume were made in case the A_{530} was not exactly 0.5. Previously, we have demonstrated that this approach leads to accurate mixing [370]. Next, cells were pelleted by centrifugation, washed twice with a 10 mM Tris-HCl Buffer (pH 7) and stored at -20 °C. Preparation of cell extracts, 2-DE and protein identification was performed in exactly the same manner for the labeled/unlabeled cells as for the unlabeled cells.

Preparation of cell extracts

The -20 °C frozen cells were thawed and immediately resuspended in 1.5 ml of 10 mM Tris-HCl buffer (pH 7), and 25 µl of a protease-inhibitor cocktail (Sigma) was added. Cells were disrupted by sonication for 10 min on ice (Soniprep 150, Sanyo). Insoluble cell material was removed by centrifugation at 13,000 rpm for 10 min. The protein concentration of the supernatant was determined using the Bradford Protein Assay (Sigma). The supernatant was subsequently stored at -80 °C.

Two-dimensional gel electrophoresis

Gels for the reference map were prepared in triplicate. The extract was mixed with a rehydration buffer containing 50 mM DTT

(Sigma), 8 M Urea (Sigma), 2% CHAPS (Sigma), 0.2% (w/v) Pharmalyte ampholytes pH 3-10 (Fluka) and Bromophenol Blue (trace) (Sigma). This mixture was designated as the sample mix. Three IPG strips (pH 3-10) (Bio-Rad) were rehydrated with 300 μl (400 μg) of this sample mix. Strips were allowed to rehydrate overnight. IEF was performed using a 3-step protocol at a temperature of 20 °C using a Protean IEF cell (Bio-Rad). In the first step, the voltage was linearly ramped to 250 V over 30 min to desalt the strips. Next, the voltage was linearly ramped to 1000 V over 2.5 h. Finally, the voltage was rapidly ramped to 10,000 V for 40,000 V/h to complete the focusing. At this stage, the strips were stored overnight at -20 °C. Focused strips were first incubated for 15 min in a solution containing 6 M Urea, 2% SDS, 0.375 M Tris-HCl (pH 8.8), 20% glycerol, and 2% (w/v) DTT. After this, the solution was discarded and the strips were incubated in a solution containing 6 M Urea, 2% SDS, 0.375 M Tris-HCl (pH 8.8), 20% Glycerol, and 4% Iodoacetamide. After equilibration, proteins were separated in the second dimension using SDS-PAGE performed using a Protean II Multicell (Bio-Rad) apparatus on 10% T, 2.6% C gels (17 cm x 17 cm x 1 mm). Electrophoresis was carried out with a constant current of 16 mA/gel for 30 min; subsequently the current was increased to 24 mA/gel for another 7 h.

Protein visualization and image analysis

Gels were stained using Coomassie Brilliant Blue G250 (Sigma). Gels were scanned using a GS-800 densitometer (Bio-Rad) at 100 microns resolution. All spot detection and quantification was performed with PDQUEST 7.1.0 (Bio-Rad). Staining intensity was normalized against the total staining intensity on the gel. 255 spots were selected for mass spectrometric analysis. For protein quantitation, metabolic labeling was used, and for this gel image was matched to the reference map and protein spots of interest were selected for MS analysis and quantitation.

Protein isolation and identification by MS

Spots of interest were excised from the stained 2-DE gels by hand, destained with 200 mM ammonium bicarbonate with 40% acetonitrile. The gel pieces were incubated overnight in a 0.4 μg trypsin solution (Sigma) and 50 μl of 40 mM ammonium bicarbonate in 9% acetonitrile. The next day, peptides were extracted in three subsequent extraction steps using 5 μl of 25 mM NH_4HCO_3 (10 min, room temperature), 30 μl acetonitrile (15 min, 37 °C), 50 μl of 5% formic acid (15 min, 37 °C) and finally with 30 μl acetonitrile (15 min, 37 °C). All extracts were pooled and dried in a vacuum centrifuge, then stored at -20 °C.

The lyophilized peptide mixture was resuspended in 0.1% formic acid in 3% acetonitrile. This mixture was separated on a PepMap C-18 RP capillary column (LC Packings, Amsterdam, Netherlands) and eluted in a 30 min gradient via a LC Packings Ultimate nanoLC directly onto the mass spectrometer. Peptides were analyzed using an Applied Biosystems QStarXL electrospray ionization quadrupole time of flight tandem mass spectrometer (ESI qQ-TOF). The data acquisition on the MS was performed in the positive ion mode using Information Dependent Acquisition (IDA). Peptides with charge states 2 and

3 were selected for tandem mass spectrometry. IDA data were submitted to Mascot for database searching in a sequence query type of search (www.matrixscience.com). The peptide tolerance was set to 2.0 Da and the MS/MS tolerance was set to 0.8 Da. A carbamidomethyl modification of cysteine was set as a fixed modification and methionine oxidation was set as a variable modification. Up to 1 missed cleavage site by trypsin was allowed. The search was performed against the Mass Spectrometry protein sequence DataBase (MSDB; ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.nam). Molecular Weight Search (MOWSE) [287] scores greater than 50 were regarded as significant.

Peptide quantitation

In the metabolic labeling experiments, peptide identification of the light (^{14}N) version of the peptide was performed as described above. After this the heavy ^{15}N version of the peptide could be identified by changing the isotope abundance of ^{15}N nitrogen to 100% in the Analyst software data dictionary. Next, the peak area of both version of the same peptide was integrated over time using LC-MS reconstruction tool in the Analyst software. In addition, an extracted ion chromatogram (XIC) was constructed for each peptide. The XIC is an ion chromatogram that shows the intensity values of a single mass (peptide) over a range of scans. This tool was used to check for chromatographic shifts between heavy and light versions of the same peptide.

RNA extraction and probe synthesis

Early-log phase cultures (A_{600} 0.1-0.2) of *S. solfataricus* grown on 0.1% yeast extract and 0.2% tryptone (YT) or 0.3% D-glucose (G) were quickly cooled in ice-water and harvested by centrifugation at 4 °C. The RNA extraction was done as described previously [35]. Preparation of cDNA was done as follows: to 15 µg of RNA, 5 µg of random hexamers (Qiagen) was added in a total volume of 11.6 µl. This was incubated for ten min at 72 °C after which the mixture was cooled on ice. Next, dATP, dGTP and dCTP (5 µM final concentration) were added, together with 4 µM aminoallyl dUTP (Sigma), 1 µM dTTP, 10 mM dithiotreitol (DTT), 400 U superscript II (Invitrogen) and the corresponding 5x RT buffer in a final volume of 20 µl. The reverse transcriptase reaction was carried out at 42 °C for one h. To stop the reaction and to degrade the RNA, 2 µl 200 mM EDTA and 3 µl 1 M NaOH were added to the reaction mixture, after which it was incubated at 70 °C for 15 min. After neutralization by the addition of 3 µl 1 M HCl, the cDNA was purified using a Qiagen MinElute kit according to the manufacturer's instructions, except that the wash buffer was replaced with 80% (v/v) ethanol. The cDNA was then labeled using post labeling reactive CyDye packs (Amersham Biosciences), according to the protocol provided by the company. Differentially labeled cDNA derived from *S. solfataricus* cells grown on either YT or G media was pooled (15 µg labeled cDNA of each sample) and excess label was removed by cDNA purification using the MinElute kit.

DNA microarray hybridization, scanning and data analysis

The design and construction of the microarray, as well as the hybridization was performed as described previously [12,236]. After hybridization, the microarrays were scanned at a resolution of 5 microns with a Genepix 4000B scanner (Axon Instruments) using the appropriate laser and filter settings. Spots were analyzed with the Genepix pro 5.0 software package (Axon Instruments). Low-quality spots were excluded using criteria that were previously described [236]. ^{2}Log transformed ratios (^{2}log(YT/G)) from the replicate slides were averaged after first averaging the duplicate spots on the array. Statistical significance for the observed ratios was calculated by doing a Significance Analysis of Microarrays (SAM) analysis [404]. Each ^{2}log value represents 2 hybridization experiments, performed in duplicate by using cDNA derived from four different cultures of *S. solfataricus*: two grown on YT media and two grown on glucose media. The result of each ORF therefore consisted of 8 pairwise comparisons. The ORFs were categorized according to the 20 functional categories of the comprehensive microbial resource (CMR) [297].

Metabolic pathway reconstruction based on biochemical and genomic data

The reconstruction of the main metabolic pathways was performed with BLASTP and PSI-BLAST programs [10] on the non-redundant (NR) database of protein sequences (National Center for Biotechnology Information) by using full-length or N-terminal protein sequences. All the sequences were derived from verified enzymatic activities of thermophilic or hyperthermophilic archaea unless stated otherwise. The sequences from *Sulfolobus acidocaldarius* were analysed by BLASTP program using the complete genome sequence [57]. All the assigned enzymatic functions for the proteins of *Sulfolobus solfataricus* P2 were checked with the annotations in public protein databases, such as the BRaunschweig ENzyme DAtabase (BRENDA) [343], Clusters of Orthologous Groups of proteins (COG) [395], InterPro [267] and the fee-based ERGO bioinformatics suite [281]. The reconstructed pathways were compared with previous reports [163,321,417] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) [178].

ACKNOWLEDGEMENTS

The authors wish to thank: Anders Andersson and Peter Nilsson (Department of Biotechnology, Royal Institute of Technology Stockholm, Sweden) Magnus Lundgren and Rolf Bernander (Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, Uppsala, Sweden) for supplying the micro-array, technical assistance and advice. We thank the University of Sheffield and UK's SRIF Infrastructure funds for support. APLS thanks the University of Sheffield and the EPSRC for a scholarship. MGJV thanks the European Union's ERASMUS Program. PCW thanks the UK's Engineering and Physical Sciences Research Council (EPSRC) for provision of an Advanced Research Fellowship (GR/A11311/01). This work was partly supported by a grant from the European Union in the framework of the SCREEN project.

3

Identification of the missing links in prokaryotic pentose oxidation pathways: evidence for enzyme recruitment

Stan J.J. Brouns
Jasper Walther
Ambrosius P.L. Snijders
Harmen J.G. van de Werken
Hanneke L.D.M. Willemen
Petra Worm
Marjon G.J. de Vos
Anders Andersson
Magnus Lundgren
Hortense F.M. Mazon
Robert H.H. van den Heuvel
Peter Nilsson
Laurent Salmon
Willem M. de Vos
Phillip C. Wright
Rolf Bernander
John van der Oost

J Biol Chem (2006) **281**: 27378-27388

Abstract

The pentose metabolism of archaea is largely unknown. Here, we have employed an integrated genomics approach including DNA microarray and proteomics analyses to elucidate the catabolic pathway for D-arabinose in *Sulfolobus solfataricus*. During growth on this sugar, a small set of genes appeared to be differentially expressed compared to growth on D-glucose. These genes were heterologously overexpressed in *Escherichia coli*, and the recombinant proteins were purified and biochemically studied. This showed that D-arabinose is oxidized to 2-oxoglutarate by the consecutive action of a number of previously uncharacterized enzymes, including a D-arabinose dehydrogenase, a D-arabinonate dehydratase, a novel 2-keto-3-*deoxy*-D-arabinonate dehydratase and a 2,5-dioxopentanoate dehydrogenase. Promoter analysis of these genes revealed a palindromic sequence upstream of the TATA-box, which is likely to be involved in the concerted transcriptional control of these genes. Integration of the obtained biochemical data with genomic context analysis strongly suggests the occurrence of pentose oxidation pathways in both archaea and bacteria, and predicts the involvement of additional enzyme components. Moreover, it revealed striking genetic similarities between the catabolic pathways for pentoses, hexaric acids and hydroxyproline degradation, which support the theory of metabolic pathway genesis by enzyme recruitment.

INTRODUCTION

Pentose sugars are a ubiquitous class of carbohydrates with diverse biological functions. Ribose and *deoxy*-ribose are major constituents of nucleic acids, while arabinose and xylose are building blocks of several plant cell wall polysaccharides. Many prokaryotes, as well as yeasts and fungi, are able to degrade these polysaccharides, and use the released five-carbon sugars as a sole carbon and energy source. At present, three main catabolic pathways have been described for pentoses. The first is present in bacteria and uses isomerases, kinases and epimerases to convert D- and L-arabinose (Ara) and D-xylose (Xyl) into D-xylulose-5-phosphate (Xu-5P) (Fig. 3.1A), which is further metabolized by the enzymes of the phosphoketolase or pentose phosphate pathway. The genes encoding the pentose-converting enzymes are often located in gene clusters in bacterial genomes, for example the *araBAD* operon for L-Ara [222], the *xylAB* operon for D-Xyl [323] and the *darK-fucPIK* gene cluster for D-Ara [101]. The second catabolic pathway for pentoses converts D-Xyl into Xu-5P as well, but the conversions are catalyzed by reductases and dehydrogenases instead of isomerases and epimerases (Fig. 3.1B). This pathway is commonly found in yeasts, fungi, mammals and plants, but also in some bacteria [58,120,433]. In a third pathway, pentoses such as L-Ara, D-Xyl, D-ribose (D-Rib) and D-Ara are metabolized non-phosphorylatively to either 2-oxoglutarate (2OG) or to pyruvate and glycolaldehyde (Fig. 3.1C). The conversion to 2OG, which is a tricarboxylic acid (TCA) cycle intermediate, proceeds via the subsequent action of a pentose dehydrogenase, a pentonolactonase, a pentonic acid dehydratase,

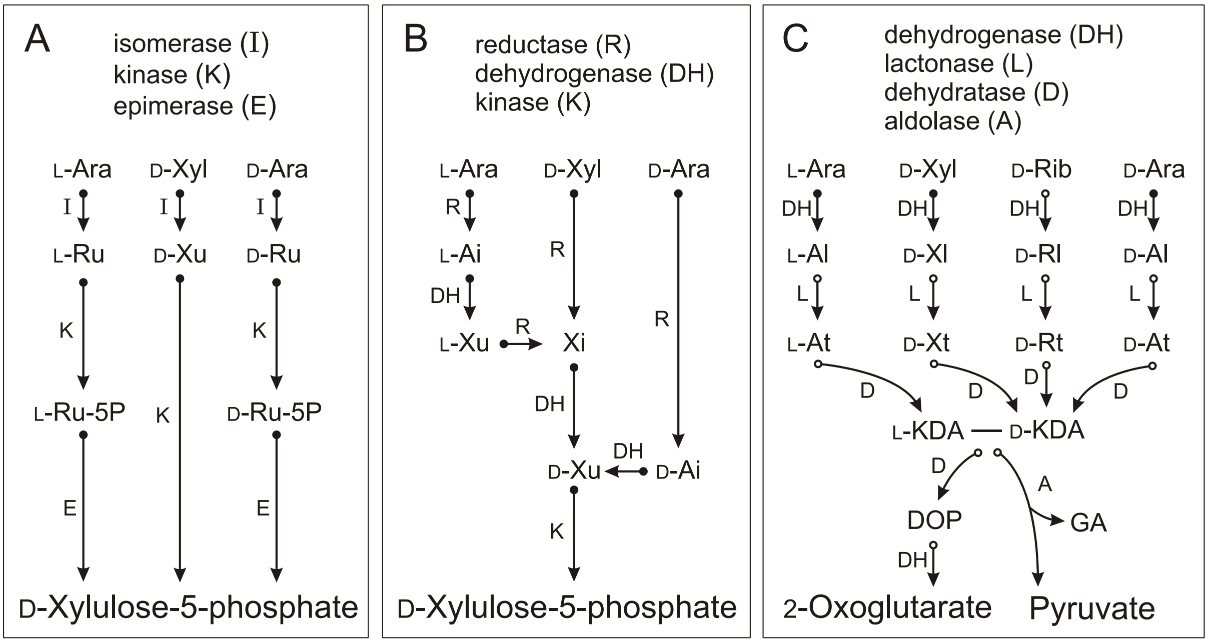

A
isomerase (I)
kinase (K)
epimerase (E)
L-Ara
D-Xyl
D-Ara
I
L-Ru
D-Xu
D-Ru
K
L-Ru-5P
D-Ru-5P
E
D-Xylulose-5-phosphate
B
reductase (R)
dehydrogenase (DH)
kinase (K)
L-Ara
D-Xyl
D-Ara
R
L-Ai
DH
L-Xu
Xi
D-Xu
D-Ai
K
D-Xylulose-5-phosphate
C
dehydrogenase (DH)
lactonase (L)
dehydratase (D)
aldolase (A)
L-Ara
D-Xyl
D-Rib
D-Ara
DH
L-Al
D-Xl
D-Rl
D-Al
L
L-At
D-Xt
D-Rt
D-At
D
L-KDA
D-KDA
A
DOP
GA
2-Oxoglutarate
Pyruvate

Figure 3.1
Schematic representation of three types of pentose degrading pathways (A, B, C). Arrows with an open or closed arrowtail represent enzymatic steps that are performed by unknown proteins or known proteins, respectively. Abbreviations: Ara arabinose, Xyl xylose, Rib ribose, Ru ribulose, Xu xylulose, Ai arabinitol, Xi xylitol, Al arabinonolactone, Xl xylonolactone, Rl ribonolactone, At arabinonate, Xt xylonate, Rt ribonate, KDA 2-keto-3-*deoxy*-arabinonate (also called 2-oxo-4,5-dihydroxypentanoate), DOP 2,5-dioxopentanoate (also called 2-oxoglutarate semialdehyde), GA glycolaldehyde.

a 2-keto-3-*deoxy*-pentonic acid dehydratase and a 2,5-dioxopentanoate dehydrogenase. This metabolic pathway has been reported in several aerobic bacteria, such as strains of *Pseudomonas* [75,429,430], *Rhizobium* [93,94], *Azospirillum* [276] and *Herbaspirillum* [251]. Alternatively, some *Pseudomonas* and *Bradyrhizobium* species have been reported to cleave the 2-keto-3-*deoxy*-pentonic acid with an aldolase to yield pyruvate and glycolaldehyde [76,285,291]. Despite the fact that these oxidative pathway variants have been known for more than five decades, surprisingly, the majority of the responsible enzymes and genes remain unidentified.

Sulfolobus spp. are obligately aerobic crenarchaea that are commonly found in acidic geothermal springs. Among the archaea, this genus is well known for its broad saccharolytic capacity, which is reflected in their ability to utilize several pentoses and hexoses, as well as oligosaccharides and polysaccharides as a sole carbon and energy source [144]. Although the catabolism of hexoses is well studied (reviewed in [358]), the pathways for pentose degradation have neither been established in *S. solfataricus*, nor in any other member of the archaea [174].

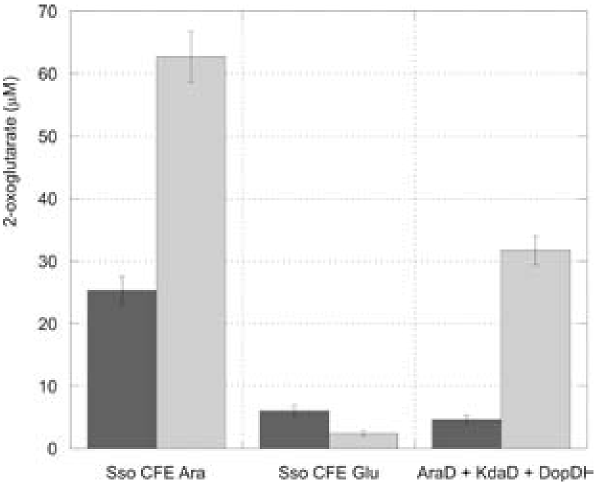


Figure 3.2
The formation of 2-oxoglutarate from D-arabinonate by extracts of S. solfataricus and by the recombinant enzymes AraD, KdaD and DopDH, in the presence of 0.4 mM NAD$^+$ (dark gray bars) or NADP$^+$ (light gray bars).

RESULTS AND DISCUSSION

Sulfolobus solfataricus is a model archaeon for studying metabolism and information processing systems, such as transcription, translation and DNA replication [237,353]. Several halophilic and thermophilic archaea have been reported to assimilate pentose sugars, but neither the catabolic pathways for these 5-carbon sugars nor the majority of its enzymes are known [144,174]. To close this knowledge gap, we have studied the growth of *S. solfataricus* on the pentose D-Ara using a multi-disciplinary genomics approach, and compared the results to growth on the hexose D-Glu. Both culture media supported growth to cell densities of approximately $2 \cdot 10^9$ cells ml^{-1} (A_{600} 2.5) with similar doubling times of around 6 h.

Several enzyme activity assays were performed with CFEs from both cultures to establish a mode of D-Ara degradation (Fig. 3.1). A 12.3-fold higher NADP$^+$-dependent D-Ara dehydrogenase activity (45.5 mU mg^{-1}) was detected in D-Ara CFE (Table 3.1), which indicated the presence of an inducible D-Ara dehydrogenase. D-Ara reductase, D-arabinitol dehydrogenase and D-Ara isomerase activity were not detected. Activity assays using D-arabinonate indicated that D-Ara CFE contained a 13.9-fold higher D-arabinonate dehydratase activity (7.4 mU mg^{-1}) than D-Glu CFE (Table 3.1). Moreover, the multi step conversion of D-arabinonate to 2OG could readily be demonstrated with D-Ara CFE in the presence of NADP$^+$ (Fig. 3.2). The formation of pyruvate

Table 3.1 - Properties of the D-Ara degrading enzymes of *S. solfataricus*

	AraDH	**AraD**	**KdaD**	**DopDH**
Description	D-arabinose 1-dehydrogenase	D-arabinonate dehydratase	2-keto-3-*deoxy*-D-arabinonate dehydratase	2,5-dioxopentanoate dehydrogenase
EC number	1.1.1.117	4.2.1.5	4.2.1.-	1.2.1.26
Locus ID	Sso1300	Sso3124	Sso3118	Sso3117
Uniprot ID	Q97YM2	Q97U96	Q97UA0	Q97UA1
Genbank ID	15898142	15899830	15899826	15899825
COG	1064	4948	3970	1012
PFAM	00107	01188	01557	00171
Sp. act. *S.s* extracts mU mg^{-1} (fold A/G)	45.5 (12.3)	7.4 (13.9)	ND	255 (3.6)
ARA-box present	yes	yes	yes	yes
Subunit size (kDa)	37.3	42.4	33.1	52.3
Oligomerization	Tetramer	Octamer	Tetramer	Tetramer
Substrate specificity, turnover rate (s^{-1})	D-arabinose (23.8) L-fucose (26.8) D-ribose (17.7) L-galactose (17.7)	D-arabinonate (1.8)	ND	dioxopentanoate (8.6) glycolaldehyde (5.3) glyceraldehyde (4.8)
Cofactor	NADP$^+$, Zn^{2+}	Mg^{2+}	ND	NADP$^+$
App. pH optimum (>50% activity)	8.2 (7.3-9.3)	6.7 (5.2-10.2)	ND	7.8 (6.7-8.2)
Apparent T-opt (°C) (>50% activity)	91 (74->100)	85 (75-92)	ND	ND
Thermal inactivation half-life time	42 min at 85 °C 26 min at 90 °C	18 min at 85 °C <10 min at 90 °C	ND	ND

ND not determined

as one of the products from D-arabinonate was not observed, while control reactions with both CFEs and D-gluconate as a substrate did yield pyruvate (data not shown), indicating that the enzymes of the Entner-Doudoroff pathway were operative. In the final step of the pathway, D-Ara CFE contained a 3.6-fold higher activity (255 mU mg^{-1}) towards the aldehyde 2,5-dioxopentanoate (DOP) using NADP$^+$ as a cofactor. The data suggest that *S. solfataricus* employs an inducible enzyme set that converts D-Ara into the TCA cycle intermediate 2OG via the pentose oxidation pathway (Fig. 3.1C).

Transcriptomics

The global transcriptional response of *S. solfataricus* growing exponentially on D-Ara or D-Glu was determined by DNA microarray analysis. The transcriptome comparison between both growth conditions showed that a small set of genes was differentially expressed 3-fold or more (Table 3.2). The highly expressed genes under D-Ara conditions included all four subunits of the Ara ABC transporter (Sso3066-3069) [99], a putative sugar permease for D-Ara (Sso2718), five out of six subunits of the SoxM quinol oxidase complex (Sso2968-2973) [205] and five metabolic genes with general function predictions only (Sso1300, Sso3124, Sso3117, Sso3118 and Sso1303). The differential expression of the gene for the remaining SoxM subunit, *i. e.* the sulfocyanin subunit SoxE (Sso2972), was just below the threshold level (online Suppl. Table 2). While the expression of the ABC-type transport system genes had been shown to be induced in Ara media previously [99,235], the differential expression of the SoxM gene cluster was not anticipated.

The genes that were upregulated under D-Glu conditions encode seven uncharacterized proteins (Sso3073, Sso3089, Sso3104, Sso1312, Sso2884, Sso3085 and Sso3100), the SoxB subunit of the SoxABCD quinol oxidase complex (Sso2657) [141] and a glutamate dehydrogenase (Sso2044) [66] (Table 3.2). The Glu ABC transporter was not differentially expressed, confirming previous observations [235]. The difference in gene expression of subunits SoxA (Sso2658), SoxC (Sso2656) and SoxD (Sso10828) was just below the threshold level (online Suppl. Table 2). Next to the SoxABCD genes, a small gene cluster containing the Rieske iron-sulfur cluster protein SoxL (Sso2660) and Sso2661 to Sso2663 appeared to be expressed with a 2 to 3-fold difference (online Suppl. Table 2). It thus appears that under D-Glu conditions, the SoxABCD quinol oxidase complex is preferentially used, whereas under D-Ara conditions the SoxM-mediated terminal quinol oxidation is favored. Differential use of both oxidase complexes was recently also found in *Metallosphaera sedula*.

Table 3.2 - Differentially expressed genes (> 3-fold change)

Locus name	Microarray ^{2}log fold(A/G) ± SD (q-value[a])	Proteomics fold(A/G) ± SD	Description	Ref.
High expression on D-Ara				
Sso3066	4.02 ± 0.58 (1.1)	>20	Arabinose ABC transporter, arabinose binding protein (AraS)	99
Sso3068	3.71 ± 0.97 (1.1)	ND	Arabinose ABC transporter, permease	99
Sso1300	3.64 ± 0.95 (1.1)	>20	Alcohol dehydrogenase IV[b] D-arabinose 1-dehydrogenase (AraDH)	This study
Sso3067	3.37 ± 0.49 (1.1)	ND	Arabinose ABC transporter, permease	99
Sso3069	2.97 ± 0.18 (2.5)	ND	Arabinose ABC transporter, ATP binding protein	99
Sso2968	2.56 ± 1.30 (1.1)	ND	Quinol oxidase subunit (SoxM complex), SoxI	205
Sso3124	2.44 ± 1.13 (1.1)	>20	Mandelate racemace / muconate lactonizing enzyme[b] D-arabinonate dehydratase (AraD)	This study
Sso3117	2.39 ± 0.62 (1.1)	>20	Aldehyde dehydrogenase[b] 2,5-dioxopentanoate dehydrogenase (DopDH)	This study
Sso2971	2.24 ± 1.17 (1.1)	ND	Quinol oxidase (SoxM complex), SoxF (Rieske Fe-S protein)	205
Sso2973	2.10 ± 1.48 (2.6)	ND	Quinol oxidase (SoxM complex), SoxM (I + III)	205
Sso2970	2.09 ± 1.53 (2.6)	ND	Quinol oxidase (SoxM complex), SoxG (cytochr. a_{587})	205
Sso2718	1.99 ± 1.10 (1.1)	ND	put. sugar permease	-
Sso3046	1.89 ± 0.87 (1.1)	ND	put. ABC sugar transporter, ATP binding protein	-
Sso2969	1.86 ± 1.09 (1.1)	ND	Quinol oxidase (SoxM complex), SoxH subunit (II)	205
Sso3118	1.78 ± 0.45 (1.1)	>20	Conserved hypothetical protein[b] 2-keto-3-*deoxy*-D-arabinonate dehydratase (KdaD)	This study
Sso1303	1.77 ± 0.57 (1.1)	ND	put. pentonic acid dehydratase	-
Sso1333	NDE	3.48 ± 0.62	Isocitrate lyase	407
Sso0527	NDE	3.45 ± 0.72	Phosphoglycerate kinase	154
Sso2869	NDE	3.06 ± 1.06	Malic enzyme	24
High expression on D-Glu				
Sso3073	-2.59 ± 0.71 (1.1)	ND	put. Sugar permease	-
Sso3089	-2.12 ± 1.12 (1.1)	ND	Hypothetical membrane protein component	-
Sso3104	-2.04 ± 0.31 (1.1)	ND	Hypothetical protein	-
Sso1312	-2.02 ± 0.52 (1.1)	ND	put. Ring oxidation compl./phenylacetic acid degradation rel. pr.	-
Sso2884	-1.87 ± 0.37 (1.1)	ND	put. 4-carboxymuconolactone decarboxylase	-
Sso2657	-1.77 ± 0.87 (1.1)	ND	Quinol oxidase (SoxABCD complex), SoxB (cytochr. *aa*3)	141
Sso3085	-1.63 ± 0.86 (1.1)	ND	Conserved hypothetical membrane protein	-
Sso3100	-1.60 ± 0.88 (1.1)	ND	Hypothetical membrane protein component	-
Sso2044	-1.60 ± 0.48 (1.1)	ND	L-Glutamate dehydrogenase	-

NDE not differentially expressed, ND not determined

[a] q-value < 5 indicates statistically significant differential expression, [b] original annotation

Here the SoxABCD genes were expressed at high levels during growth on sulfur, while heterotrophic growth on yeast extract induced the production of the SoxM complex [183]. Since the aeration and cell-density of the D-Ara and D-Glu cultures was similar, the trigger for the differential expression of the two oxidase complexes in *S. solfataricus* is currently unknown.

Proteomics

Protein expression in the soluble proteomes of D-Ara and D-Glu grown *S. solfataricus* cells was compared using a combination of two-dimensional gel electrophoresis, stable isotope labeling and tandem mass spectrometry. By employing this strategy, five proteins were found with more than a 20-fold difference in expression level, including the Ara binding protein from the Ara ABC transporter (AraS, Sso3066) [99], Sso1300, Sso3124, Sso3118 and Sso3117 (Table 3.2). Interestingly, the difference in expression level of these genes at the protein level appeared to be more pronounced than at the mRNA level, which ranged from 3.4 to 16-fold. Three other proteins were also produced in higher amounts during growth on D-Ara, albeit only up to a 3-fold difference (Table 3.2). These were the isocitrate lyase (Sso1333) [407], the phosphoglycerate kinase (Sso0527) [154] and the malic enzyme (Sso2869) [24].

Promoter motif analysis

The promoters of the differentially expressed genes were analyzed for the occurrence of DNA sequence motifs that could play a role as *cis*-acting elements in the coordinated transcriptional control of these genes. The analysis indeed revealed the presence of a palindromic motif (consensus: AACATGTT) in the promoters of Sso3066 (*araS*), Sso1300, Sso3124, Sso3118 and Sso3117 genes (Fig. 3.3). This motif was designated the ARA-box and it was always located upstream of the predicted TATA-box with a separation of 10 bases. A conserved transcription factor B (TFB) recognition element (BRE) appeared to be absent from the interspaced sequence between both boxes. Additional copies of the ARA-box were identified further upstream of both Sso3066 and Sso1300. Although primer extension analysis was only performed for the *araS* gene [235], the promoter architecture suggests that the

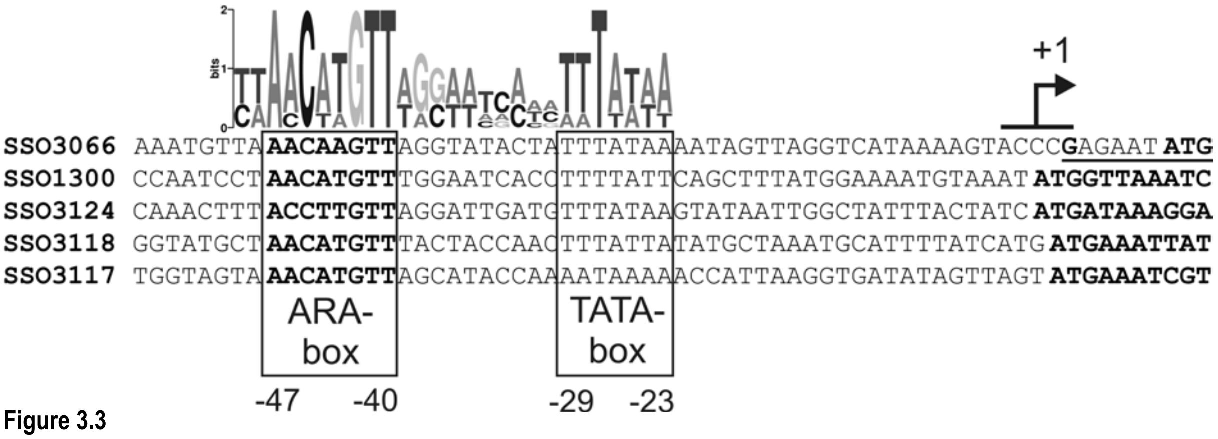

bits
2
1
0
+1
SSO3066 AAATGTTA AACAAGTT AGGTATACTA TTTATAA AATAGTTAGGTCATAAAAGTACCC GAGAAT ATG
SSO1300 CCAATCCT AACATGTT TGGAATCACC TTTTATT CAGCTTTATGGAAAATGTAAAT ATGGTTAAATC
SSO3124 CAAACTTT ACCTTGTT AGGATTGATG TTTATAA GTATAATTGGCTATTTACTATC ATGATAAAGGA
SSO3118 GGTATGCT AACATGTT TACTACCAAC TTTATTA TATGCTAAATGCATTTTATCATG ATGAAATTAT
SSO3117 TGGTAGTA AACATGTT AGCATACCAA AATAAAA ACCATTAAGGTGATATAGTTAGT ATGAAATCGT
ARA-box
TATA-box
-47
-40
-29
-23

Figure 3.3
Putative *cis*-regulatory element ARA-box and TATA-box in upstream sequences of the D-Ara induced genes. The predicted transcription start site is indicated by +1. Transcripts are underlined [235]. Coding sequences are in bold. Additional ARA-boxes were found for *Sso3066* at -90 to -83 and *Sso1300* at -235 to -228 relative to the transcription start sites. Note: a single ARA-box is present in the intergenic region between the divergently oriented genes *Sso3118* and *Sso3117*.

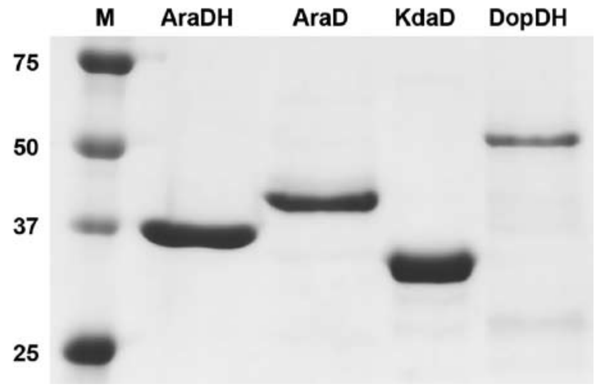


Figure 3.4
A digital photograph is shown of a Coomassie Blue stained 8% SDS-PAGE gel that was loaded with purified recombinantly produced enzymes from the D-Ara oxidation pathway of *S. solfataricus*. Marker sizes are indicated in kDa.

transcript leader of Sso1300, Sso3124, Sso3118 and Sso3117 will either be very short, or absent. This is in good agreement with the fact that a large proportion of the *S. solfataricus* genes is predicted to be transcribed without a leader [233]. The inducibility of the *araS* promoter has recently been exploited in viral expression vectors that enable recombinant protein production in *S. solfataricus* [8].

Characterization of induced proteins

The genes that were differentially expressed and contained an Ara-box in their promoter were selected and cloned in an *E. coli* expression vector. The resulting proteins were overproduced, purified and characterized to investigate their role in the metabolism of D-Ara.

Arabinose dehydrogenase (AraDH)

The putative zinc-containing, medium-chain alcohol dehydrogenase encoded by Sso1300 was efficiently produced and purified using a single step of affinity chromatography (Fig. 3.4). The enzyme was most active on L-fucose (6-*deoxy*-L-galactose) (k_{cat} 26.8 s^{-1}), followed by D-Ara (k_{cat} 23.8 s^{-1}), using preferentially $NADP^+$ (K_m 0.04 ± 0.01 mM) over NAD^+ (K_m 1.25 ± 0.45 mM) as a cofactor. This enzyme was thus likely to account for the elevated D-Ara dehydrogenase activities in *S. solfataricus* CFE. AraDH could also oxidize L-galactose and the D-Ara C2-epimer D-ribose with similar rates (k_{cat} 17.7 s^{-1}) (Table 3.1). Enzyme activity towards other sugars remained below 7% of the highest activity. Similar substrate specificities and affinities have been found previously for mammalian and bacterial L-fucose or D-Ara dehydrogenases, although these enzymes prefer NAD^+ as a cofactor [60,261]. AraDH was more than 50% active in a relatively narrow pH range from 7.3 to 9.3, with optimal catalysis proceeding at pH 8.2. The thermophilic nature of the enzyme is apparent from its optimal catalytic temperature of 91 °C. The enzyme maintained a half-life of 42 and 26 min at 85 and 90 °C, respectively, indicating that the enzyme is thermostable at physiological growth temperatures of *S. solfataricus*. Native mass spectrometry experiments showed that the intact recombinant AraDH has a molecular mass of 149,700 ± 24 Da. Comparing these data with the expected mass on the basis of the primary sequence (37,291 Da) clearly showed that the protein has a tetrameric structure and contains two zinc atoms per monomer. This is in good agreement with the tetrameric structure that has been reported for another alcohol dehydrogenase from *S. solfataricus* (Sso2536), which has a 33% identical protein sequence [107]. This dehydrogenase, however, prefers aromatic or aliphatic alcohols as a substrate, and NAD^+ over $NADP^+$ as a cofactor. A structural study of AraDH is currently ongoing to explain the observed differences in substrate and cofactor selectivity.

Arabinonate dehydratase (AraD)

The protein encoded by gene Sso3124 was originally annotated as a member of the mandelate racemase and muconate lactonizing enzyme family. This superfamily, which additionally comprises aldonic acid dehydratases, is mechanistically related by their common ability to abstract α-protons from carboxylic acids [19]. Production of the enzyme in *E. coli* yielded approximately 10% soluble recombinant protein, which was purified using anion-exchange and size exclusion chromatography (Fig. 3.4). The enzyme was shown to catalyze the strictly Mg^{2+}-dependent dehydration reaction of D-arabinonate to 2-keto-3-*deoxy*-D-arabinonate (KDA) (Fig. 3.5). It is therefore conceivable that this enzyme is largely responsible for the increased levels of D-arabinonate dehydratase activity in *S. solfataricus* extracts. AraD displayed a maximum turnover rate of 1.8 s^{-1} at a substrate concentration of 8 mM, while higher substrate concentrations imposed severe inhibitory effects on the enzyme. No activity was measured with D-gluconate up to 20 mM. More than 50% enzyme activity was observed in a broad pH range of 5.2 to 10.2 with an optimum at pH 6.7 (Table 3.1). The enzyme was optimally active at 85 °C during which it maintained a half-life time of 18 min. Native mass spectrometry revealed that the protein had a molecular mass of 340,654 ± 63 Da, which corresponds well to an octameric protein assembly (expected monomeric mass is 42,437 Da). The native D-gluconate dehydratase from *S. solfataricus* (GnaD, Sso3198), which has a 23% identical protein sequence, was found to be an octamer as well [200]. Interestingly, AraD was only produced as an octamer when the media was supplemented with 20 mM Mg^{2+} during protein overexpression. Without this divalent cation, the recombinant protein was inactive and appeared to be monomeric. Sequence alignment analysis as well as 3D modeling based on a *Agrobacterium tumefaciens* protein with unknown function (Atu3453, PDB-ID 1RVK) showed that Asp199, Glu225 and Glu251 are likely to be involved in binding the divalent metal ion, which is required to stabilize the enolate reaction intermediate [19].

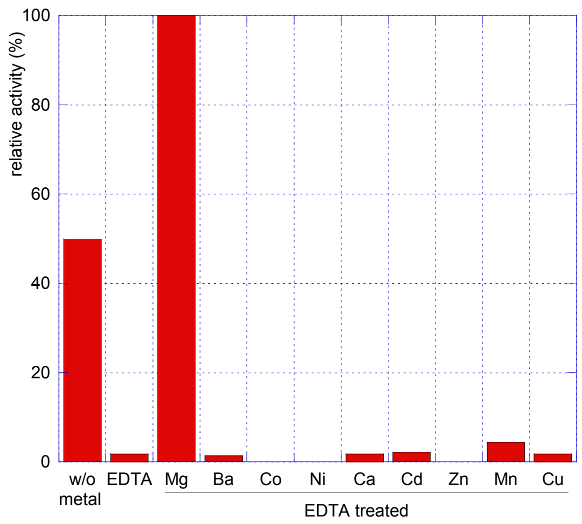


Figure 3.5
Magnesium ion dependence of AraD

D-KDA dehydratase (KdaD)

To investigate the possible role of Sso3118, the protein was overproduced in *E. coli*, and subsequently purified (Fig. 3.4). Surprisingly, although the predicted pI of the enzyme is 5.9, the vast majority of protein did not bind to the anion exchange column at a pH of 8. Moreover, the protein had a tendency to precipitate, which could be reversed and effectively prevented by the addition of 0.5 mM of DTT to all buffers. Native mass spectrometry under reducing conditions revealed that the protein had a molecular mass of 132,850 ± 47 Da, which corresponds with a tetrameric quaternary structure (expected monomeric mass of 33,143 Da). The catalytic activity of the protein was investigated by performing indirect enzyme assays using AraD with D-arabinonate as a substrate. A 50% decrease in the yield of KDA was observed when both enzymes were co-incubated in the presence of Mg^{2+}, but this did not result in the formation of either 2OG or pyruvate. Given the fact that D-arabinonate is converted to 2OG in D-Ara CFE, this enzyme was anticipated to be responsible for the dehydration of D-KDA to the aldehyde DOP. However, due to the unavailability of D-KDA, it was not possible to show this in a direct enzyme assay. We therefore employed an indirect assay using AraD, the putative D-KDA dehydratase (KdaD) and the predicted aldehyde dehydrogenase. The results of this assay are described below in the DopDH section.

According to the Clusters of Orthologous Groups of proteins classification system, the putative KDA dehydratase belongs to COG3970. The catalytic domain of these proteins resembles that of the eukaryal fumarylacetoacetate hydrolase (FAH); an enzyme that catalyzes the Mg^{2+}- or Ca^{2+}-dependent hydrolytic cleavage of fumarylacetoacetate to yield fumarate and acetoacetate as the final step of phenylalanine and tyrosine degradation [27]. In humans, several mutations in the FAH gene will lead to hereditary tyrosinemia type I, which is mainly characterized by liver defects [383]. Members of COG3970 are also homologous to the C-terminal decarboxylase domain of the bifunctional enzyme HpcE from *E. coli*, which in addition consists of an N-terminal isomerase domain [394]. This enzyme is active in the homoprotocatechuate pathway of aromatic compounds and is responsible for the Mg^{2+}-dependent decarboxylation of 2-oxo-5-carboxy-hept-3-ene-1,7-dioic acid to 2-hydroxy-hepta-2,4-diene-1,7-dioic acid and its subsequent isomerization to 2-oxo-hept-3 -ene-1,7-dioic acid [394]. Although the function of these enzyme classes is rather diverse, their structures have revealed similarities in terms of a fully conserved metal ion binding site and a relatively conserved active site architecture. Multiple sequence alignment analysis of KdaD indicated the presence of a metal binding site consisting of Glu143, Glu145 and Asp164, which may implicate a metal dependent activity as well. Further structural studies of KdaD are currently ongoing.

DOP dehydrogenase (DopDH)

The putative aldehyde dehydrogenase encoded by Sso3117 was overproduced in *E. coli*, which resulted in the formation of approximately 5% soluble protein. This protein fraction was purified using affinity and size exclusion chromatography (Fig. 3.4). From native mass spectrometry experiments we could determine a molecular mass of 210,110 Da, which is in reasonable agreement with the expected mass of the tetramer on the basis of the primary sequence (52,290 Da). The measured mass may be somewhat higher due to the binding of small molecules to the protein oligomer. The determined oligomerization state corresponds to that of the closely related aldehyde dehydrogenase ALDH-T from *Geobacillus stearothermophilus* [164]. The aldehyde dehydrogenase was tested for the activity towards different aldehydes and cofactors (Table 3.1). This indicated that the enzyme preferred $NADP^+$ over NAD^+, and that it oxidized several hydrophilic aldehydes with the highest activity towards 2,5-dioxopentanoate (DOP) followed by glycolaldehyde and DL-glyceraldehyde. More than 50% enzyme activity was observed in a pH range of 6.7 to 8.2, with an optimum at pH 7.8. The enzyme was also tested in conjunction with AraD and KdaD for the production of 2OG or pyruvate. Similar to the activities in D-Ara CFE, these three enzymes were able to form 2OG and not pyruvate, from D-arabinonate using preferably $NADP^+$ as a cofactor (Fig. 3.2). Omission of either the cofactor, AraD, KdaD or DopDH prevented the formation of 2OG, indicating that all components were essential for the enzymatic conversions, and that KdaD was most likely responsible for the dehydration of D-KDA to DOP.

Extensive kinetic characterization of DopDH proved to be rather complicated, since the enzyme lost nearly all its activity within one day after its purification, even in the presence of high concentrations of reducing agents, such as DTT or β-mercaptoethanol. This could be due to the fact that this class of enzymes contains a catalytic cysteine residue (in DopDH Cys293), which can become irreversibly oxidized, leading to a total loss of enzymatic activity. A rapid inactivation was also observed with ALDH-T from *G. stearothermophilus* [164].

Central carbohydrate metabolism

Some central metabolic routes, such as the glycolysis, gluconeogenesis and the tricarboxylic acid (TCA) cycle have been studied extensively in *S. solfataricus*, *S. acidocaldarius* and other archaea. The availability of their genome sequences [57,353] as well as the genome sequence of *S. tokodaii* [189], has recently allowed a reconstruction of the genes involved in these pathways [371]. The effect of the introduction of excess 2OG resulting from the D-Ara oxidative pathway led to the differential expression of only a few additional genes in these central carbon metabolic routes (Table 3.2, Fig 3.6). The isocitrate lyase, the phosphoglycerate kinase and the malic enzyme were upregulated at the protein level under D-Ara conditions. The induction of the malic enzyme might indicate that the main proportion of 2OG is converted to malate, which is then decarboxylated to pyruvate and acetyl-CoA, respectively, and is then fully oxidized to two molecules of CO_2 in one round of the TCA cycle. Although this may seem energetically unfavorable, the net difference in yield between the full degradation of one molecule of D-Glu or D-Ara to CO_2 is only one NADPH reduction equivalent in favor of D-Glu, since both degradation schemes lead to 6 reduced ferredoxins, 2 $FADH_2$, 2 ATP and 6 or 5 NADPH molecules, respectively. It is therefore not surprising that the growth rates under both conditions are similar. The phosphoglycerate kinase may be indicative of increased gluconeogenic activities that are required under D-Ara conditions. The isocitrate lyase is normally operative in the glyoxylate shunt, but high production levels of the enzyme have also been observed during growth on L-glutamate compared to D-Glu [368]. Oxidative deamination of L-glutamate leads to the formation of 2OG as well, which may inhibit the isocitrate dehydrogenase activity leading to an accumulation of isocitrate. This could trigger the production of the isocitrate lyase, which can bypass this step without the loss of CO_2.

Pentose oxidation gene clusters

The comprehensive analysis of conserved gene clustering in multiple genome sequences is becoming an increasingly important tool to predict functionally or physically associated proteins in prokaryotic cells (reviewed in [278]). Genomic context analysis of the genes involved in the D-Ara oxidative pathway of *S. solfataricus* showed that *kdaD* and *dopDH* gene orthologs are often located adjacently in prokaryotic genomes. This finding supports the proposed enzymatic functions of an aldehyde producing and an aldehyde oxidative activity. In addition, the analysis uncovered the presence of putative pentose oxidative gene clusters in the genomes of several aerobic proteobacteria, such as members of the genera *Burkholderia*, *Rhizobium*, *Bradyrhizobium*, *Agrobacterium* and *Pseudomonas*. In some cases, the presence of such a gene cluster correlates well with the organism's ability to

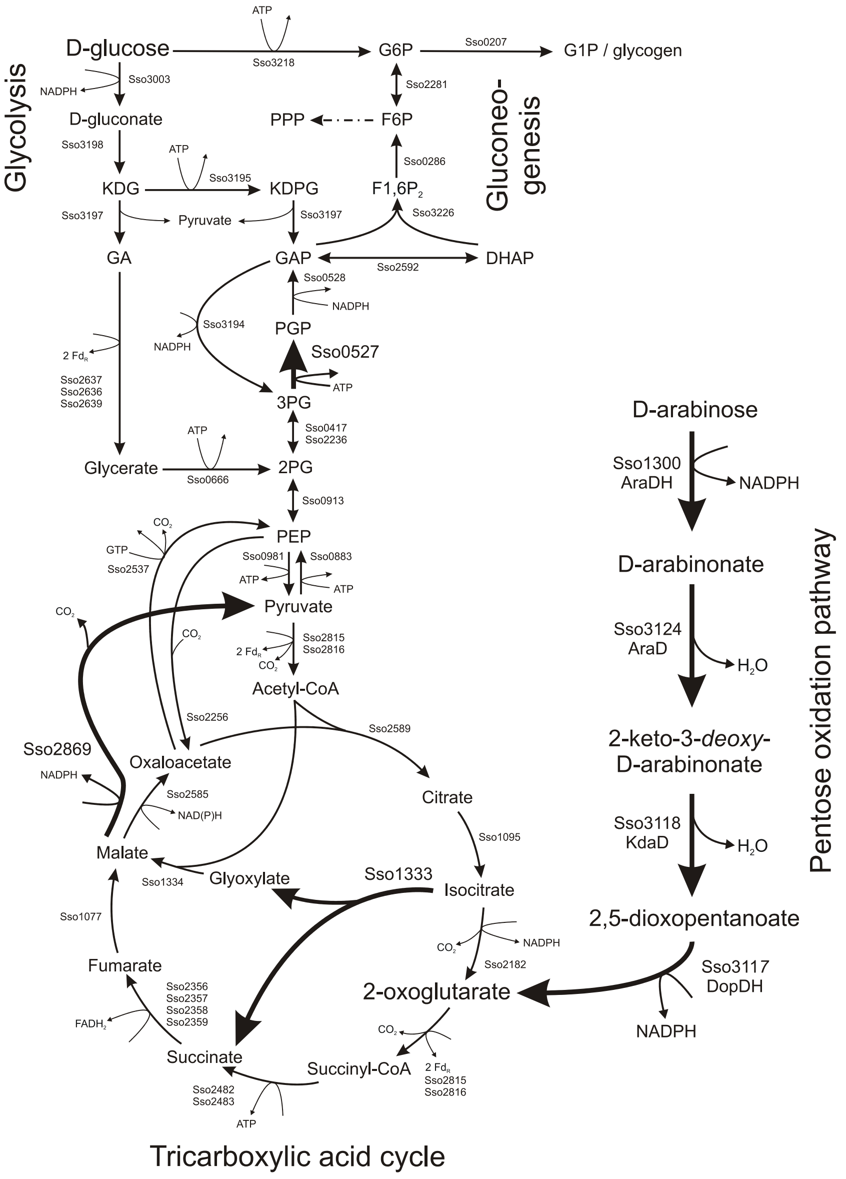

Glycolysis
Gluconeo-genesis
ATP
D-glucose
Sso3218
G6P
Sso0207
G1P / glycogen
Sso3003
NADPH
Sso2281
D-gluconate
PPP
F6P
Sso3198
ATP
Sso0286
Sso3195
KDG
KDPG
F1,6P2
Sso3197
Pyruvate
Sso3197
Sso3226
GA
GAP
Sso2592
DHAP
Sso0528
NADPH
Sso3194
PGP
NADPH
Sso0527
2 FdR
ATP
Sso2637
Sso2636
Sso2639
3PG
Sso0417
Sso2236
ATP
Glycerate
Sso0666
2PG
Sso0913
CO2
PEP
GTP
Sso2537
Sso0981
Sso0883
ATP
ATP
CO2
Pyruvate
CO2
Sso2815
Sso2816
2 FdR
CO2
Acetyl-CoA
Sso2256
Sso2589
Sso2869
Oxaloacetate
NADPH
Sso2585
NAD(P)H
Citrate
Sso1095
Malate
Sso1334
Glyoxylate
Sso1333
Isocitrate
Sso1077
CO2
NADPH
Sso2182
Fumarate
Sso2356
Sso2357
Sso2358
Sso2359
2-oxoglutarate
FADH2
CO2
Succinate
Succinyl-CoA
2 FdR
Sso2815
Sso2816
Sso2482
Sso2483
ATP
Tricarboxylic acid cycle
D-arabinose
Sso1300
AraDH
NADPH
D-arabinonate
Sso3124
AraD
H2O
2-keto-3-deoxy-D-arabinonate
Sso3118
KdaD
H2O
2,5-dioxopentanoate
Sso3117
DopDH
NADPH
Pentose oxidation pathway

Figure 3.6
Enzymes involved in the degradation of D-Glu and D-Ara in *S. solfataricus* [273,314,371]. The following abbreviations were used: KD(P)G 2-keto-3-*deoxy*-D-gluconate (6-phosphate), GA(P) D-glyceraldehyde (3-phosphate), PGP 1,3-bisphosphoglycerate, 3PG 3-phosphoglycerate, 2PG 2-phosphoglycerate, PEP phosphoenolpyruvate, DHAP dihydroxyacetonephosphate, F1,6P2 D-fructose 1,6-bisphosphate, F6P D-fructose 6-phosphate, G6P D-glucose 6-phosphate, G1P D-glucose 1-phosphate, FdR reduced ferredoxin, PPP pentose phosphate pathway. Thick arrows are indicative of increased expression of the enzyme under D-Ara conditions.

assimilate pentoses and with enzymatic activities present in cell extracts [75,93,94,429,430], while in other cases biochemical data is not available. Nonetheless, a few of these characteristic gene clusters have been demonstrated genetically to be linked to pentose degradation. Combined with the findings in *S. solfataricus*, this allows the identification of additional enzymatic components in the pentose oxidation pathway and prediction of their enzymatic functions (Fig. 3.7A).

A putative operon of five genes was found in the genome of the oligotrophic α-proteobacterium *Caulobacter crescentus*, which was 2.8 to 11.6 fold upregulated during growth on D-Xyl as compared to D-Glu [158]. Reporter fusion constructs of the CC0823 promoter to the β-galactosidase gene (*lacZ*) from *E. coli* confirmed that this promoter is highly induced during growth on D-Xyl, and repressed on D-Glu or proteinaceous media [158,256]. Moreover, the disruption of the CC0823 gene prevented the *C. crescentus* from growth on D-Xyl as a single carbon source [256].

A second pentose degradation gene cluster involved in L-Ara uptake and utilization was found on chromosome II of the pathogenic β-proteobacterium *Burkholderia thailandensis*. This cluster consisting of nine genes was proposed to be responsible for the L-Ara degradation to 2OG (Fig. 3.7A) [263]. Disruption of the *araA*, *araC*, *araE* and *araI* gene led to an L-Ara negative phenotype. Reporter gene insertions showed that the expression of the *araC* and the *araE* gene was repressed during growth in D-Glu media, and induced in L-Ara media. The transfer of the gene cluster to the related bacterium *B. pseudomallei* enabled this organism to utilize L-Ara as a sole carbon source also [263]. Interestingly, an L-Ara dehydrogenase with 80% sequence identity to AraE has recently been characterized from *Azospirillum brasiliense* [423]; an organism which is known to degrade L-Ara to 2OG [276]. The flanking sequences of this gene revealed close homologs of the *B. thailandensis araD* and *araE*, which would indicate a similar gene cluster in *A. brasiliense* [423].

Apart from several bacteria, putative pentose oxidation clusters are also present in some archaea. In the halophile *Haloarcula marismortui*, a gene cluster was found on chromosome I that seems to contain all the necessary components for D-Xyl oxidation, including a gene that has been identified as a D-Xyl dehydrogenase [174] (Fig. 3.7A).

Components of pentose oxidation pathways

Careful inspection of the different pentose oxidation gene clusters shows that the gene encoding the final enzymatic step, from DOP to 2OG, is fully conserved between the different pentose oxidation gene clusters. The remaining analogous enzymatic steps that convert D-Ara, D-Xyl or L-Ara into DOP are performed by enzymes from distinct COGs (Clusters of Orthologous Groups of proteins) [396] (Fig. 3.7, pentose panels). While some of this variation in enzyme use may simply be explained by substrate differences, other variations may be due to the individual adaptation of existing enzymes with similar reaction chemistry, such as the pentose dehydrogenases.

A striking difference between the set of enzymes responsible for D-Ara degradation in *S. solfataricus* on the one hand, and the predicted sets for D-Xyl degradation in *C. crescentus* and *H. marismortui* and L-Ara degradation in *B. thailandensis* on the other hand, is the apparent absence of an upregulated lactonase in the hyperthermophile. This enzyme is responsible for the hydrolysis of the lactone, yielding the corresponding linear pentonic acid. Such ring opening reactions are reported

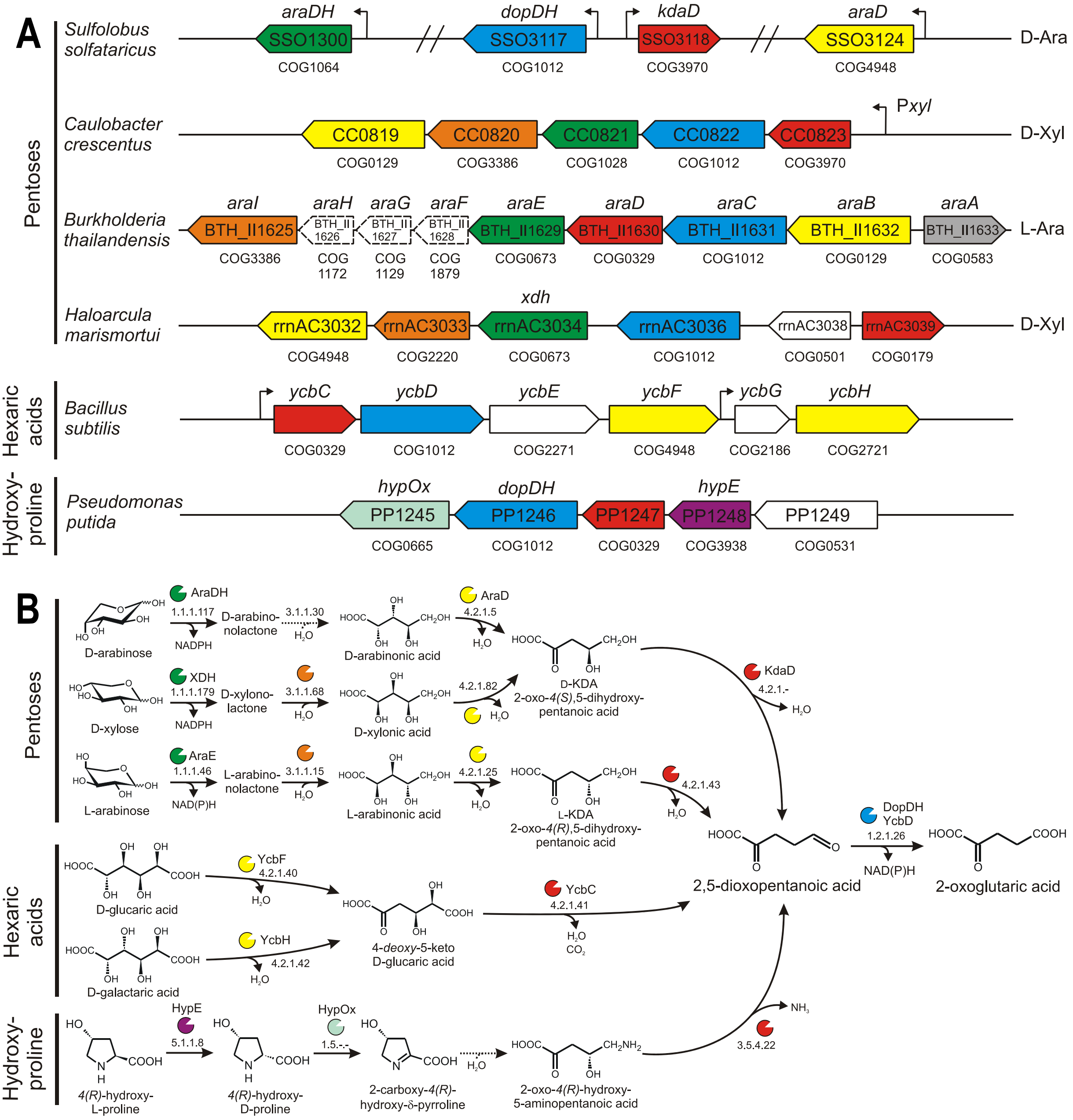

A
Pentoses
Sulfolobus solfataricus
araDH
SSO1300
COG1064
dopDH
SSO3117
COG1012
kdaD
SSO3118
COG3970
araD
SSO3124
COG4948
D-Ara
Caulobacter crescentus
Pxyl
CC0819
COG0129
CC0820
COG3386
CC0821
COG1028
CC0822
COG1012
CC0823
COG3970
D-Xyl
Burkholderia thailandensis
araI
BTH_II1625
COG3386
araH
BTH_II 1626
COG 1172
araG
BTH_II 1627
COG 1129
araF
BTH_II 1628
COG 1879
araE
BTH_II1629
COG0673
araD
BTH_II1630
COG0329
araC
BTH_II1631
COG1012
araB
BTH_II1632
COG0129
araA
BTH_II1633
COG0583
L-Ara
Haloarcula marismortui
rrnAC3032
COG4948
rrnAC3033
COG2220
xdh
rrnAC3034
COG0673
rrnAC3036
COG1012
rrnAC3038
COG0501
rrnAC3039
COG0179
D-Xyl
Hexaric acids
Bacillus subtilis
ycbC
COG0329
ycbD
COG1012
ycbE
COG2271
ycbF
COG4948
ycbG
COG2186
ycbH
COG2721
Hydroxy-proline
Pseudomonas putida
hypOx
PP1245
COG0665
dopDH
PP1246
COG1012
PP1247
COG0329
hypE
PP1248
COG3938
PP1249
COG0531
B
Pentoses
D-arabinose
AraDH
1.1.1.117
NADPH
D-arabino-nolactone
3.1.1.30
H₂O
D-arabinonic acid
AraD
4.2.1.5
H₂O
D-KDA
2-oxo-4(S),5-dihydroxy-pentanoic acid
KdaD
4.2.1.-
H₂O
D-xylose
XDH
1.1.1.179
NADPH
D-xylono-lactone
3.1.1.68
H₂O
D-xylonic acid
4.2.1.82
H₂O
L-arabinose
AraE
1.1.1.46
NAD(P)H
L-arabino-nolactone
3.1.1.15
H₂O
L-arabinonic acid
4.2.1.25
H₂O
L-KDA
2-oxo-4(R),5-dihydroxy-pentanoic acid
4.2.1.43
H₂O
2,5-dioxopentanoic acid
DopDH
YcbD
1.2.1.26
NAD(P)H
2-oxoglutaric acid
Hexaric acids
D-glucaric acid
YcbF
4.2.1.40
H₂O
D-galactaric acid
YcbH
4.2.1.42
H₂O
4-deoxy-5-keto D-glucaric acid
YcbC
4.2.1.41
H₂O
CO₂
Hydroxy-proline
4(R)-hydroxy-L-proline
HypE
5.1.1.8
4(R)-hydroxy-D-proline
HypOx
1.5.-.-
2-carboxy-4(R)-hydroxy-δ-pyrroline
H₂O
2-oxo-4(R)-hydroxy-5-aminopentanoic acid
3.5.4.22
NH₃

Figure 3.7 (in color on p.152)
A) Cartoon of the organization of conserved gene clusters involved in the pentose, hexaric acid and hydroxyproline degradation. Proposed analogous gene functions are indicated in the same color (green: pentose dehydrogenase, orange: pentonolactonase, yellow: aldonic acid dehydratase, red: 2-keto-3-*deoxy*-aldonic acid dehydratase, blue: 2,5-dioxopentanoate dehydrogenase. Dashed genes are displayed smaller than their relative size. Protein family numbers are displayed below each gene according to Clusters of Orthologous Groups of proteins classification system (COG) [396]. The genes indicated in white or gray encode the following putative functions: araA: transcriptional regulator, *araF-araH*: L-Ara ABC transporter (periplasmic L-Ara binding protein, ATP binding protein, permease), *rrnAC3038*: heat shock protein X, *ycbE*: glucarate/galactarate permease, *ycbG*: transcriptional regulator, *PP1249*: hydroxyproline permease. **B)** Schematic representation of the convergence of catabolic pathways for pentoses, hexaric acids [75,157,172,352] and hydroxyproline [311,361,439] at the level of 2,5-dioxopentanoate. Enzymatic activities are indicated by their EC number. Dashed lines indicate proposed spontaneous reactions.

to proceed spontaneously at ambient temperatures, albeit at slow rates [284]. Overexpressing a lactonase may therefore be advantageous at mesophilic growth temperatures, whereas at 80 °C the spontaneous reaction may well proceed rapidly enough not to be rate limiting. The pentose oxidation gene clusters seem to be predominated by lactonases of COG3386, which are often annotated as “senescence marker protein-30 family proteins”. The genome of *S. solfataricus* contains two of these genes (Sso2705 and Sso3041), but they were not differentially expressed, indicating that they are either not involved or that their basal expression level is sufficient for arabinonolactone hydrolysis. The putative xylonolactonase from *H. marismortui*, however, is homologous to metal-dependent β-lactamases belonging to COG2220, which catalyze similar lactame-ring opening reactions [88].

Other non-orthologous enzyme components of the pentose oxidation pathway include the pentonic acid dehydratases. While the D-arabinonate dehydratase from *S. solfataricus* belongs to COG4948, the same function seems to be performed by members of COG0129 which are commonly annotated as dihydroxyacid dehydratases (IlvD) or 6-phosphogluconate dehydratases (Edd) [95]. A member of this family has recently been characterized from *S. solfataricus* (DHAD, Sso3107), which revealed a broad substrate specificity for aldonic acids [199]. However, this gene was not differentially expressed according to the transcriptome or proteome analysis.

The 2-keto-3-*deoxy*-D-arabinonate dehydratase (COG3970) appears to be present in D-Ara and D-Xyl degradation gene clusters. Interestingly, in several *Burkholderia* species, and in *A. brasiliense*, this gene is replaced by a member of the dihydrodipicinolate synthase family (COG0329, *B.th araD*). Members of this family catalyze either aldolase or dehydratase reactions via a Schiff-base dependent reaction mechanism by a strictly conserved lysine residue. Interestingly, a detailed study of an L-KDA dehydratase involved in the L-Ara metabolism of *P. saccharophila* was reported a few decades ago, but unfortunately, neither the amino terminal sequence of the protein nor the gene sequence was determined [304,384]. The authors found that this enzyme was enantioselective for L-KDA (2-oxo-*4(R)*,5-dihydroxypentanoate), and that the reaction proceeds via a Schiff-base intermediate. The enzyme activity was not affected by the presence of 1 mM EDTA, which suggests a divalent metal ion independent reaction. It seems likely that this enzyme is encoded by homologs of the *B. thailandensis araD* gene, and that the apparent enantioselectivity of this enzyme does not allow a function in the degradation of D-Ara or D-Xyl, which results in a 2-keto-3-*deoxy*-pentonic acid with the *S*-configuration (Fig. 3.7B).

The aldehyde dehydrogenase from COG1012 is fully conserved in the pentose oxidation gene clusters (Fig. 3.7A). Strikingly, close homologs of this gene can also be found in hexaric acid degradation gene clusters of *Bacillus* species (*ycbC–ycbI*) [157,352] (Fig. 3.7A). The same holds for a gene cluster in *P. putida* (PP1245–PP1249) that is likely to be involved in the breakdown of L-hydroxyproline, which is a major constituent of collagen and plant cell wall proteins [311,439] (Fig. 3.7B). Apparently, since the degradation of both hexaric acids and L-hydroxyproline is also known to proceed through DOP [75], the genetic information for the conversion of DOP to 2OG has been shared between multiple metabolic pathways during evolution (Fig. 3.7). Apart from the *dopDH* gene, orthologs of the D-glucarate dehydratase gene (*ycbF*, COG4948) are observed in the pentose degradation gene clusters of both *S. solfataricus* and *H. marismortui*, while remarkably, the keto-*deoxy*-acid dehydratase of COG0329 is found in all three pathways. In the hydroxyproline degradation pathway, this enzyme might function as a deaminase instead [361].

The apparent mosaic of orthologous and non-orthologous proteins involved in the pentose oxidation pathway suggests that some of these enzymatic steps may have evolved by recruitment events between enzymes from the hexaric acid or hydroxyproline degradation pathways, which too make use of DOP as an intermediate and produce 2OG as the final product [173,341]. The low number of enzymes required, their common cofactor usage and the large gain of obtaining the hub metabolite 2OG as the end product of pentose oxidation, may have been the driving force in the creation of this pathway in aerobically respiring bacteria and archaea.

EXPERIMENTAL PROCEDURES

All chemicals were of analytical grade and purchased from Sigma, unless stated otherwise. Oligonucleotide primers were obtained from MWG Biotech AG (Ebersberg, Germany).

Growth of Sulfolobus species

Sulfolobus solfataricus P2 (DSM1617) was grown in media containing either 3 g L^{-1} D-Ara or D-Glu as previously described [38].

Transcriptomics

Whole-genome DNA microarrays containing gene-specific tags representing >90% of the *S. solfataricus* P2 genes [12] were used for global transcript profiling of cultures grown on D-Ara as compared to D-Glu. Total RNA extraction, cDNA synthesis and labeling, hybridization and scanning were performed as previously described, as were data filtration, normalization and statistical evaluation [236,371].

Quantitative proteomics

The proteome of *S. solfataricus* P2 was studied with a combination of two-dimensional gel electrophoresis (2-DE), ^{15}N metabolic labeling and tandem mass spectrometry as previously described [368,370]. Two separate growth experiments were set up; (1) *S. solfataricus* with D-Ara as the carbon source and $(^{14}NH_4)_2SO_4$ as the nitrogen source, and (2) *S. solfataricus* with D-Glu as the carbon source and $(^{15}NH_4)_2SO_4$ as the nitrogen source. Next, the ^{14}N and ^{15}N cultures were mixed in equal amounts on the basis of optical density (A_{530}) measurements, proteins were extracted and separated by 2-DE. For the localization of proteins, a previously described 2-DE reference map was used [371]. Spots were excised from the gel, and peptides were quantified on the basis of their relative intensity in the Time-Of-Flight mass spectrum, according to established methods [371].

Synthesis of organic compounds

D-Arabinonate was synthesized from D-Ara as previously described [373]. The aldehyde 2,5-dioxopentanoate was synthesized from 1,4-dibromobutane according to reported procedures [72,209,241].

Gene cloning and protein overexpression

The genes *araDH* (Sso1300), *araD* (Sso3124), *kdaD* (Sso3118) and *dopDH* (Sso3117) were amplified by PCR from genomic DNA using *Pfu* TURBO polymerase (Stratagene) and cloned in expression vector pET24d (Novagen) (online Suppl. Table 1). The resulting plasmids were harvested from *Escherichia coli* HB101 transformants by Miniprep (Qiagen), sequenced by Westburg Genomics (Leusden, Netherlands), and transformed to *E. coli* expression strain BL21(DE3) containing the tRNA accessory plasmid pRIL (Stratagene).

All proteins were produced according to standard procedures in four 1 L shake-flasks containing LB medium, but with some exceptions. When the culture A_{600} reached 0.5, the cultures were cold-shocked by placing them on ice for 30 min to induce host chaperones [38]. After that, the expression was started by adding 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and the cultures were incubated for 12 to 16 h at 37 °C after which they were spun down (10 min, 5,000 xg, 4 °C). At the time of induction, the AraDH and the AraD overexpression culture were supplemented with 0.25 mM $ZnSO_4$ [107] and 20 mM of $MgCl_2$, respectively.

Protein purification

Pelleted *E. coli* and *S. solfataricus* cells were resuspended in buffer and disrupted by sonication at 0 °C. Afterwards, insoluble cell material was spun down (30 min, 26,500 xg, 4 °C) and the *E. coli* supernatants were subjected to a heat-treatment of 30 min at 75 °C. Denatured proteins were removed by centrifugation (30 min, 26,500 xg, 4 °C) yielding the heat stable cell-free extract (HSCFE).

AraDH - HSCFE in 20 mM Tris-HCl (pH 7.5) supplemented with 50 mM NaCl was applied to a 20 ml Matrex RedA affinity column (Amicon). After washing the bound protein with two column volumes of buffer, the recombinant protein was eluted by a linear gradient of 2 M of NaCl.

AraD - HSCFE in 50 mM HEPES-KOH (pH 8.0) supplemented with 50 mM NaCl was applied to a 70 ml Q-Sepharose Fast Flow (Amersham) anion exchange column, and eluted by a 2 M NaCl gradient. Fractions containing the recombinant protein were pooled, concentrated with a 30 kDa cut-off filter (Vivaspin), and purified by size exclusion chromatography using a Superdex 200 HR 10/30 column (Amersham) and 50 mM HEPES-KOH buffer (pH 8.0) supplemented with 100 mM NaCl as an eluent.

KdaD - HSCFE in 25 mM NaP_i buffer (pH 6.8) was applied to a 70 ml Q-Sepharose Fast Flow (Amersham) anion exchange column. Flow-through fractions containing KdaD were collected, loaded onto a 46 ml Bio-Gel Hydroxyapatite column (Bio-Rad), and eluted by a linear gradient of 0.5 M NaP_i buffer (pH 6.8). Fractions containing the recombinant proteins were pooled, and dialyzed overnight in 50 mM HEPES-KOH (pH 8.0) supplemented with 0.5

mM dithiothreitol (DTT).

DopDH - HSCFE in 20 mM HEPES-KOH (pH 8.0) supplemented with 200 mM NaCl and 7.5 mM DTT was purified by affinity chromatography, as described for AraDH. Fractions containing the protein were pooled, concentrated using a 30 kDa cut-off membrane (Vivaspin), and purified by size exclusion chromatography as described for AraD.

Enzyme assays

Unless stated otherwise, all enzymatic assays were performed in degassed 100 mM HEPES-KOH buffer (pH 7.5) at 70 °C. The optimal pH of catalysis was determined using a 30 mM citrate-phosphate-glycine buffer system which was adjusted in the range of pH 3 to 11 at 70 °C. Thermal inactivation assays were performed by incubating 50 µg ml^{-1} of enzyme at 70, 80, 85 and 90 °C and drawing aliquots at regular intervals during 2 h followed by a standard activity assay.

Dehydrogenase assays - Sugar dehydrogenase activity was determined on a Hitachi U-1500 spectrophotometer in a continuous assay using 10 mM of D- and L-arabinose, D- and L-xylose, D-ribose, D-lyxose, D- and L-fucose, D- and L-galactose, D-mannose and D-glucose as a substrate, and 0.4 mM of NAD^+ or $NADP^+$ as a cofactor. Aldehyde dehydrogenase reactions were performed using 5 mM of 2,5-dioxopentanoate, glycolaldehyde, DL-glyceraldehyde, acetaldehyde and propionaldehyde in the presence of 10 mM DTT. Initial enzymatic activity rates were obtained from the increase in absorption at 340 nm (A_{340}), and calculated using a molar extinction coefficient of 6.22 $mM^{-1}\cdot cm^{-1}$.

Dehydratase assay - Standard reactions were performed using 10 mM potassium D-arabinonate in the presence of 1 mM $MgCl_2$. The formation of 2-keto-3-*deoxy*-acid reaction products was determined with the thiobarbiturate (TBA) assay at 549 nm using a molar extinction coefficient of 67.8 $mM^{-1}\cdot cm^{-1}$ [200,422]. The effect of different divalent cations on enzymatic activity was investigated by a pretreatment of the enzyme with 1 mM EDTA for 20 min at 70 °C, followed by a standard assay in the presence of 2 mM of divalent metal ions.

Formation of 2-oxoglutarate and pyruvate - Enzyme reactions were performed with cell-free extract (CFE) from *S. solfataricus* cultures grown on either D-Ara or D-Glu which were harvested at mid-exponential phase. The reaction was started by adding 25 µl of 3.5 mg ml^{-1} CFE to a mixture containing 10 mM of potassium D-arabinonate, 1 mM $MgCl_2$, and either 0.4 mM NAD^+ or $NADP^+$. After an incubation of 2 h at 75 °C, the reactions were stopped by placing the tubes on ice. Identical reactions were setup in which the CFE was replaced by the purified enzymes AraD (4.2 µg), KdaD (13.4 µg) and DopDH (3.8 µg). The amount of 2-oxoglutarate in these mixtures was then determined by the reductive amination of 2-oxoglutarate to L-glutamate using purified recombinant *Pyrococcus furiosus* glutamate dehydrogenase (GDH) at 60 °C [217]. The detection reaction was started by the addition of 5 U of GDH to a sample that was supplemented with 10 mM NH_4Cl and 0.12 mM NADPH.

The formation of pyruvate was determined at 30 °C using 4 U of chicken heart lactate dehydrogenase and 0.1 mM NADH. The conversion of 2-oxoglutarate or pyruvate was continuously monitored on a Hitachi U-1500 spectrophotometer by following the decrease in A_{340} until substrate depletion occurred. Changes in concentrations of NAD(P)H were calculated as described above.

Determination of protein oligomeric state

The oligomerization state of AraDH, AraD, KdaD and DopDH was determined by nanoflow electrospray ionization mass spectrometry. For this, the protein was concentrated in the range of 5 to 15 μM and the buffer was exchanged to 50 or 200 mM ammonium acetate (pH 6.7 or 7.5) by using an Ultrafree 0.5 ml centrifugal filter device with a 5 kDa cut-off (Millipore). Protein samples were introduced into the nanoflow electrospray ionization source of a Micromass LCT mass spectrometer (Waters), which was modified for high mass operation and set in positive ion mode. Mass spectrometry experiments were performed under conditions of increased pressure in the interface region between sample and extraction cone of 8 mbar by reducing the pumping capacity of the rotary pump [339,390]. Capillary and sample cone voltages were optimized for the different proteins and were in the range of 1.4-1.6 kV and 75-150 V, respectively.

Bioinformatic analyses

Upstream sequences of the differentially expressed genes were extracted between -200 and +50 nucleotides relative to the open reading frame translation start site. These sequences were analyzed using the Gibbs Recursive Sampler algorithm [399]. Possible sequence motifs were checked against all upstream sequences and the complete genome of *S. solfataricus*. A diagram of the sequence motif was created using the WebLogo server.

Protein sequences were retrieved from the National Center for Biotechnology Information (NCBI) and analyzed using PSI-BLAST on the non-redundant database, and RPS-BLAST on the conserved domain database. Multiple sequence alignments were built using either ClustalX or TCoffee software. Gene neighborhood analyses were performed using various webserver tools: STRING at the EMBL, Gene Ortholog Neighborhoods at the Integrated Microbial Genomes server of the Joint Genome Institute, and pinned regions at the ERGO bioinformatics suite.

ACKNOWLEDGEMENTS

This work was supported by a grant from the European Union in the framework of the SCREEN project, and by a dedicated functional genomics grant from the Swedish Research Council to RB. APLS thanks the University of Sheffield and the UK's Engineering and Physical Sciences Research Council (EPSRC) for a scholarship. PCW thanks the EPSRC for provision of an Advanced Research Fellowship (GR/A11311/01) and funding (GR/S84347/01).

4

Crystal structure and biochemical properties of the D-arabinose dehydrogenase from *Sulfolobus solfataricus*

Stan J.J. Brouns
Andrew P. Turnbull
Hanneke L.D.M. Willemen
Jasper Akerboom
John van der Oost

J Mol Biol (2007) **371**: 1249-1260

Abstract

Sulfolobus solfataricus **metabolizes the 5-carbon sugar D-arabinose to 2-oxoglutarate by an inducible pathway consisting of dehydrogenases and dehydratases. Here we report the crystal structure and biochemical properties of the first enzyme of this pathway: the D-arabinose dehydrogenase. The AraDH structure was solved to a resolution of 1.80 Å by single-wavelength anomalous diffraction and phased using the two endogenous zinc ions per subunit. The structure revealed a catalytic and cofactor binding domain, typically present in mesophilic and thermophilic alcohol dehydrogenases. Cofactor modeling showed the presence of a phosphate binding pocket sequence motif (SRS-X_2-H), which is likely to be responsible for the enzyme's preference for $NADP^+$. The homo-tetrameric enzyme is specific for D-arabinose, L-fucose, L-galactose and D-ribose. These sugars are optimally converted at pH 8.2 and 91 °C. The protein displays a half-life of 42 and 26 minutes at 85 and 90 °C, respectively, indicating that the enzyme is thermostable at physiological operating temperatures of 80 °C. The structure represents the first crystal structure of an $NADP^+$-dependent member of the medium-chain dehydrogenase/reductase (MDR) superfamily from Archaea.**

INTRODUCTION

Alcohol dehydrogenases (ADHs, EC 1.1.1.1) are a widespread and functionally diverse class of enzymes that catalyze the oxidation of alcohols and the reduction of aldehydes or ketones in the presence of a cofactor [59]. The intrinsic reversibility of this reaction allows these enzymes to play distinct functional roles in biosynthetic pathways, fermentations and detoxification reactions.

D-Arabinose dehydrogenase (EC 1.1.1.116/1.1.1.117) catalyzes the nicotinamide cofactor-dependent oxidation of the pentose D-arabinose. In the hyperthermophilic archaeon *Sulfolobus solfataricus* an enzyme with such substrate specificity was overproduced to a more than 20-fold higher level during growth in D-arabinose containing media. This enzyme, termed AraDH, was shown to be responsible for the first of four enzymatic steps metabolizing D-arabinose into 2-oxoglutarate [39]. Following the action of AraDH, the D-arabinonate dehydratase (AraD) converts D-arabinonate into 2-keto-3-*deoxy*-D-arabinonate (KDA). A second water elimination reaction is then performed by the KDA dehydratase (KdaD), yielding the aldehyde 2,5-dioxopentanoate (DOP). The DOP dehydrogenase oxidizes the substrate under the formation of the citric acid cycle intermediate 2-oxoglutarate and a second molecule of NADPH. This uncommon oxidative pathway, which involves no phosphorylated intermediates, results in a central metabolic intermediate that has retained all five carbon atoms of D-arabinose. Analogous pentose oxidation pathways for L-arabinose or D-xylose have recently also been demonstrated in bacteria such as *Azospirillum brasiliense* and *Caulobacter crescentus*, respectively, although differences in enzyme machineries exist [39,380,423,425,426]. The promiscuous Entner-Doudoroff pathway for D-glucose and D-galactose in *S. solfataricus* involves a 2-keto-3-*deoxy*-gluconate aldolase and ultimately produces two pyruvate molecules per hexose, which are fed into the citric acid cycle as acetyl-CoA [46]. Neither the pentose oxidation pathway nor the Entner-Doudoroff pathway in *S. solfataricus* yields net ATP by substrate level phosphorylation, strongly suggesting that all energy must come from aerobic respiration.

AraDH belongs to the medium-chain dehydrogenase/reductase (MDR) superfamily. This group of enzymes is characterized by subunits of approximately 340 to 375 amino acids and typically contains a structural as well as catalytic zinc ion. The horse liver ADH is by far the best studied member of this class, both structurally and mechanistically, leading to a current total of over thirty structures of complexes and mutants of this enzyme in the protein structure databank [59,148].

RESULTS AND DISCUSSION

Overall structure

The AraDH monomer comprises 344 amino acids and is composed of two domains: a catalytic domain (residues 1-154 and 292-344) and a nucleotide binding domain (residues 155-291) (Fig. 4.1). The active site resides in a deep cleft between the domains, which can accommodate the nicotinamide adenine dinucleotide cofactor and a substrate. The catalytic zinc ion is bound in this cleft by residues in α-helix 1

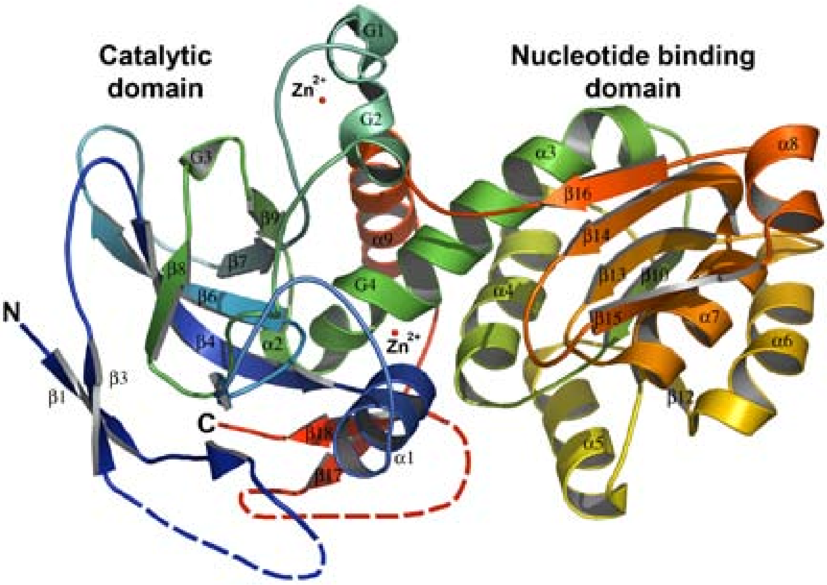


Figure 4.1 (in color on p.153)
Schematic representation of the *Sulfolobus solfataricus* AraDH structural model of the monomer showing the catalytic and nucleotide binding domain. The N-terminus of the polypeptide chain (N, in blue) follows a color gradient towards the C-terminus (C, in red). Interruptions in the structure due to missing residues corresponding to regions of weak electron density are indicated by dashed lines. The structural (top) and the catalytic (bottom) coordinated zinc ions are displayed as gray spheres.

and 3_{10}-helix 4 (G4), and the loop between β5 and 6. A second zinc ion is coordinated in a protrusion from the catalytic domain that forms the small helical elements G1 and G2. The core of the catalytic domain is formed by a seven stranded mixed β-sheet of five anti-parallel β-strands (β4, 6-9) and two parallel β-strands (β17, 18). This core is flanked by two anti-parallel β-sheets (β1, 2, and β3, 5) and several helices (α1, G1-3). The catalytic and the nucleotide binding domains are interconnected by three helical elements (5-7) that together span more than half of the protein molecule. A typical kink is induced by the middle 3_{10}-helix preceding the position of the catalytic zinc-binding aspartate. The nucleotide binding domain adopts a classic Rossmann fold [312] comprising a six stranded parallel β-sheet (β10-16) surrounded by five α-helices (α4-8). The amino terminal end of α-helix 4 harbors the characteristic glycine-rich sequence motif that is typical for the Rossmann cofactor binding domain. A flexible region called the hinge (G269-R274) is located between β-strand 14 and 15, which in the horse liver ADH mediates a 10° interdomain rotation upon cofactor and substrate binding [310]. This rotation results in a "closed" interdomain cleft, which reverts to the "open" conformation after catalysis. The extent of relative domain movement appears to differ for each enzyme and can be as little as 2.5° [107,207]. Two neighboring polypeptide backbone sections of the catalytic domain (Glu14-Ile18 and Asp322-Asp337) were omitted in the final model due to weak electron density. These parts of the protein are expected to form a turn and an α-helix respectively, similar to all known ADH structures.

Quaternary structure

The homo-tetramer of AraDH, built by applying the crystallographic symmetry operators, is composed of a dimer of dimers (Fig. 4.2), in which the AB or CD subunit oligomerization is found in dimeric eukaryotic ADHs (Fig. 4.2A) [32,97,289]. The AB interface buries a surface area of approximately 3800 Å^2 per dimer due to extensive hydrogen bonding and hydrophobic interactions between two β-strands (β15, 16) and α-helix 8 of the nucleotide binding domain. While the β16 strand is an antiparallel intersubunit continuation of the main β-sheet of the Rossman-fold, the β15 strand consisting of Val275, Ser276 and Leu277 is merely involved in a two-stranded antiparallel β-sheet between both subunits, creating a second layer of interacting β-stands. The structural loop containing the zinc ion forms another point of contact with the nucleotide binding domain of the adjacent subunit in the AB dimer. A total of 36 hydrogen bonds and 4 salt bridges are formed between the two subunits. Most notable among these interactions are the two intersubunit surface ion pairs (Glu250-Arg274 and Lys104-Glu262). Both the relatively high occurrence of these oppositely charged amino acid pairs in protein structures from hyperthermophiles [186] and experimental evidence [418] have suggested that these interactions contribute to the structural integrity of a protein at elevated temperature.

The AD dimer (Fig. 4.2B) buries a solvent accessible surface area of ~1500 Å^2 per dimer through interactions of the C-terminal parts of α-helices 3 and 4 of the nucleotide binding domain with their symmetry related counterparts. Additionally, α-helix 9 interacts with the loop between α-helix 3 and β-strand 10 of the adjacent subunit. The AD interface is sustained by 12 intersubunit hydrogen bonds and two ion pairs (Glu171-Arg305).

The AC dimer (Fig. 4.2C) is mainly held together by interactions between the structural loops containing the zinc ion, and those of the loop with the N-terminal part of α-helix 9. A few minor areas of contact of the AC dimer are formed by residues on two loops that are flanked by β-strands 3 and 4, and 6 and 7. The AC dimer interface shields ~1200 Å^2 of solvent accessible surface area of the protein and thus forms the smallest intersubunit contact surface of the AraDH tetramer. Eight hydrogen bonds and two ion pairs interconnect the A and C subunits. Through this subunit organization, a large internal cavity is formed with approximate dimensions of 22, 17 and 12 Å at the core of the tetramer. The cavity connects to the bulk solvent by four narrow water channels which are located at each junction between three subunits. The pore is delineated by the struc-

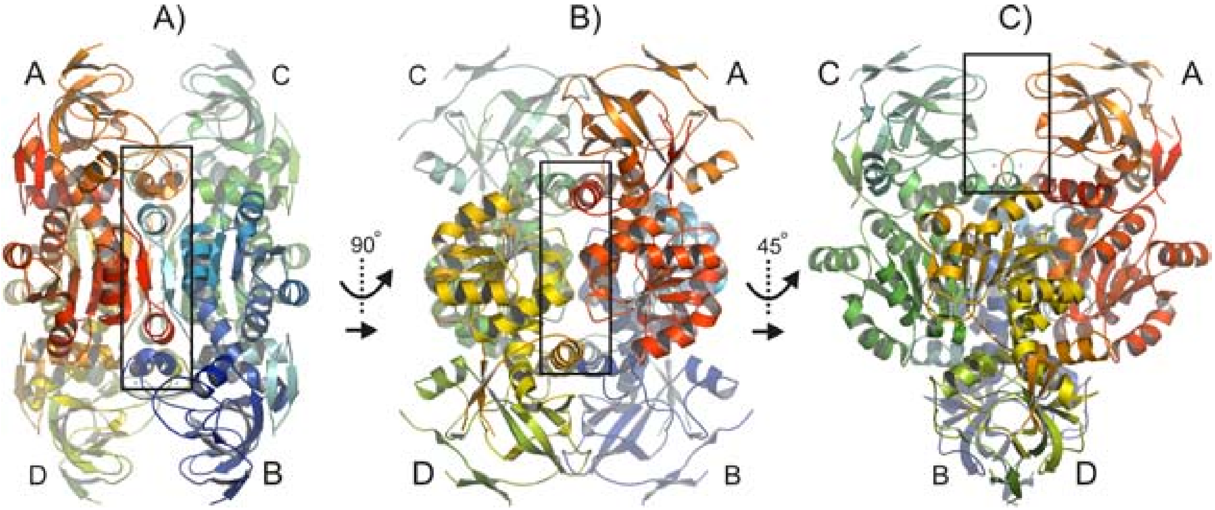


Figure 4.2 (in color on p.152)
Schematic diagrams of the *S. solfataricus* AraDH homo-tetramer generated by crystallographic symmetry operations. View perpendicular to **A)** the AB, **B)** the AD, and **C)** the AC dimer interface showing the intersubunit contacts (boxed).

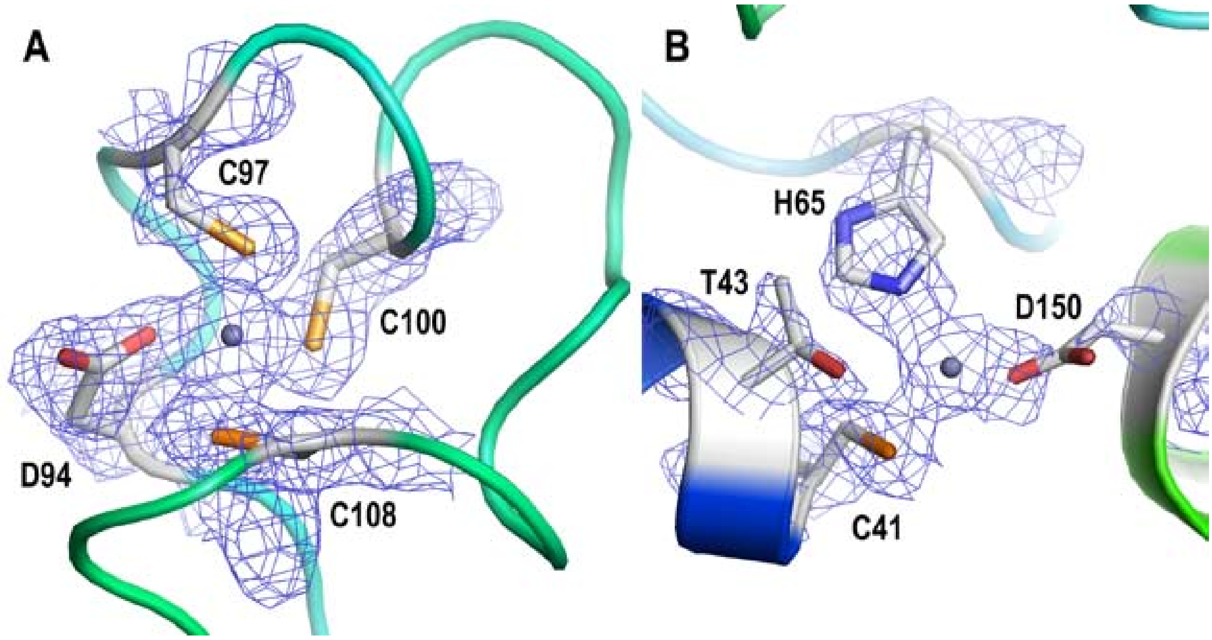


Figure 4.3 (in color on p.154)
View of the composite omit map calculated using the program SFCHECK 409 contoured at 1.0σ around **A)** the structural zinc ion and **B)** the active site zinc ion. **A)** The structural zinc ion is tetrahedrally coordinated by Asp94 Oδ1 at a distance of 1.84 Å, Cys97 Sγ at 2.30 Å, Cys100 Sγ at 2.30 Å and Cys108 Sγ at 2.34 Å. **B)** The active site zinc ion is coordinated by Cys41 Sγ at a distance of 2.36 Å, His65 Nε2 at 2.35 Å and Asp150 Oδ2 at 1.84 Å. A water molecule is located 4.55 Å from the active site zinc ion (not shown).The threonine residue (Thr43) involved in the proton relay mechanism is also shown.

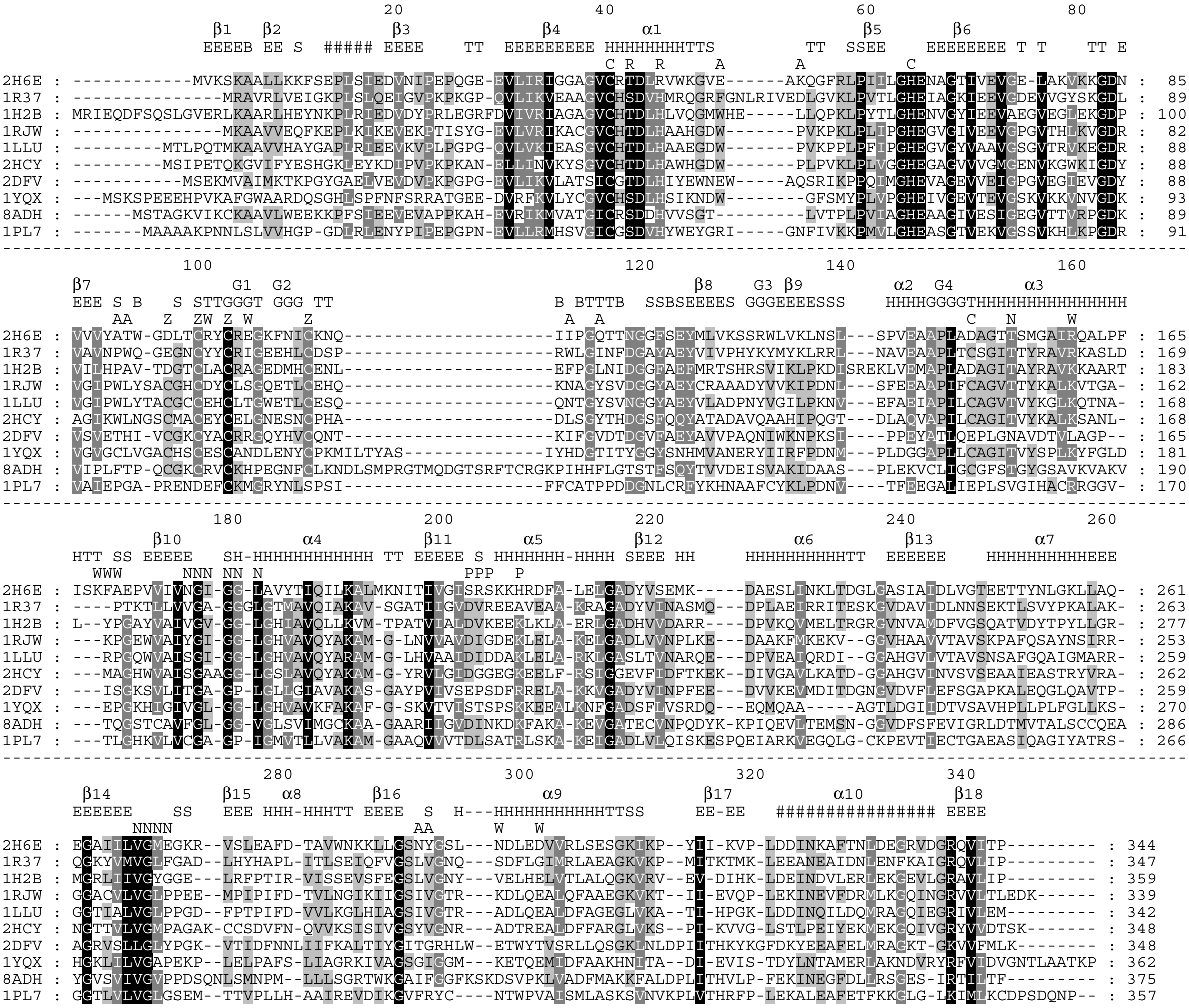
tural zinc loop of subunit A (Arg98, Glu102), the loop between α-helix 3 and β-strand 10 of subunit B (Lys168, Phe169, Ala170), and α-helix 3 and 9 of subunit C (Arg160 and Asn298, Asp302, respectively).

Structural zinc loop

A zinc ion is bound in the loop that protrudes from the catalytic domain. This zinc is tetrahedrally-coordinated by residues Asp94, Cys97, Cys100 and Cys108 (Fig. 4.3A) similar to the zinc ion present in the *Aeropyrum pernix* ADH structure [147]. The aspartate is substituted for a glutamate in the structure of the *S. solfataricus* ADH [108], while in most non-archaeal ADH structures this residue is replaced by a fourth cysteine (Fig. 4.4). Since the removal of the ion is accompanied by a loss of thermostability, the presence of the zinc ion has been associated with structural stability [242]. In addition, the introduction of a zinc ion binding site by mutagenesis can lead to an increase of thermostability [424].

Active site

The active site harbors a zinc ion that is believed to play a crucial role in the catalytic mechanism of this ADH type (Fig. 4.3B). Similar to the horse liver ADH, catalysis by AraDH is expected to proceed via a proton relay system in which the substrate's C1 hydroxyl group is bound to the zinc ion, allowing its deprotonation by Thr43 and formation of a zinc-bound alkoxide ion. A hydride from C1 is then transferred to the cofactor's nicotinamide ring by a hydrogen tunneling event [203], leaving the characteristic lactone product [310]. The product and the reduced cofactor then leave the active site, completing the ordered bi-bi mechanism. The zinc ion is bound in a tetrahedral coordination



```
                                      20                   40                        60                   80
                     β1   β2           β3             β4       α1                    β5       β6
                    EEEEB EE S  ##### EEEE    TT  EEEEEEEEE HHHHHHHHTTS         TT  SSEE    EEEEEEEE T T    TT E
                                                            C R  R     A       A          C
2H6E : ------------MVKSKAALLKKFSEPLSIEDVNIPEPQGE-EVLIRIGGAGVCRTDLRVWKGVE------AKQGFRLPIILGHENAGTIVEVGE-LAKVKKGDN :  85
1R37 : ---------------MRAVRLVEIGKPLSLQEIGVPKPKGP-QVLIKVEAAGVCHSDVHMRQGRFGNLRIVEDLGVKLPVTLGHEIAGKIEEVGDEVVGYSKGDL :  89
1H2B : MRIEQDFSQSLGVERLKAARLHEYNKPLRIEDVDYPRLEGRFDVIVRIAGAGVCHTDLHLVQGMWHE-----LLQPKLPYTLGHENVGYIEEVAEGVEGLEKGDP : 100
1RJW : ---------------MKAAVVEQFKEPLKIKEVEKPTISYG-EVLVRIKACGVCHTDLHAAHGDW-------PVKPKLPLIPGHEGVGIVEEVGPGVTHLKVGDR :  82
1LLU : ----------MTLPQTMKAAVVHAYGAPLRIEEVKVPLPGPG-QVLVKIEASGVCHTDLHAAEGDW-------PVKPPLPFIPGHEGVGYVAAVGSGVTRVKEGDR :  88
2HCY : ----------MSIPETQKGVIFYESHGKLEYKDIPVPKPKAN-ELLINVKYSGVCHTDLHAWHGDW-------PLPVKLPLVGGHEGAGVVVGMGENVKGWKIGDY :  88
2DFV : ------------MSEKMVAIMKTKPGYGAELVEVDVPKPGPG-EVLIKVLATSICGTDLHIYEWNEW-----AQSRIKPPQIMGHEVAGEVVEIGPGVEGIEVGDY :  88
1YQX : ---MSKSPEEEHPVKAFGWAARDQSGHLSPFNFSRRATGEE-DVRFKVLYCGVCHSDLHSIKNDW--------GFSMYPLVPGHEIVGEVTEVGSKVKKVNVGDK :  93
8ADH : ------MSTAGKVIKCKAAVLWEEKKPFSIEEVEVAPPKAH-EVRIKMVATGICRSDDHVVSGT---------LVTPLPVIAGHEAAGIVESIGEGVTTVRPGDK :  89
1PL7 : -------MAAAAKPNNLSLVVHGP-GDLRLENYPIPEPGPN-EVLLRMHSVGICGSDVHYWEYGRI-----GNFIVKKPMVLGHEASGTVEKVGSSVKHLKPGDR :  91
----------------------------------------------------------------------------------------------------------------------
                  100                                         120                 140                    160
       β7              G1  G2                                        β8    G3 β9          α2  G4       α3
       EEE S B   S STTGGGT GGG TT                      B BTTTB  SSBSEEEES GGGEEEESSS    HHHHGGGTHHHHHHHHHHHHHHHH
           AA   Z  ZW Z W     Z                         A  A                                    C   N     W
2H6E : VVVYATW-GDLTCRYCREGKFNICKNQ---------------------IIPGQTTNGGFSEYMLVKSSRWLVKLNSL---SPVEAAPLADAGTTSMGAIRQALPF : 165
1R37 : VAVNPWQ-GEGNCYYCRIGEEHLCDSP---------------------RWLGINFDGAYAEYVIVPHYKYMYKLRRL---NAVEAAPLTCSGITTYRAVRKASLD : 169
1H2B : VILHPAV-TDGTCLACRAGEDMHCENL---------------------EFPGLNIDGGFAEFMRTSHRSVIKLPKDISREKLVEMAPLADAGITAYRAVKKAART : 183
1RJW : VGIPWLYSACGHCDYCLSGQETLCEHQ---------------------KNAGYSVDGGYAEYCRAAADYVVKIPDNL---SFEEAAPIFCAGVTTYKALKVTGA- : 162
1LLU : VGIPWLYTACGCCEHCLTGWETLCESQ---------------------QNTGYSVNGGYAEYVLADPNYVGILPKNV---EFAEIAPILCAGVTVYKGLKQTNA- : 168
2HCY : AGIKWLNGSCMACEYCELGNESNCPHA---------------------DLSGYTHDGSFQQYATADAVQAAHIPQGT---DLAQVAPILCAGITVYKALKSANL- : 168
2DFV : VSVETHI-VCGKCYACRRGQYHVCQNT---------------------KIFGVDTDGVFAEYAVVPAQNIWKNPKSI----PPEYATLQEPLGNAVDTVLAGP-- : 165
1YQX : VGVGCLVGACHSCESCANDLENYCPKMILTYAS--------------IYHDGTITYGGYSNHMVANERYIIRFPDNM---PLDGGAPLLCAGITVYSPLKYFGLD : 181
8ADH : VIPLFTP-QCGKCRVCKHPEGNFCLKNDLSMPRGTMQDGTSRFTCRGKPIHHFLGTSTFSQYTVVDEISVAKIDAAS---PLEKVCLIGCGFSTGYGSAVKVAKV : 190
1PL7 : VAIEPGA-PRENDEFCKMGRYNLSPSI--------------------FFCATPPDDGNLCRFYKHNAAFCYKLPDNV----TFEEGALIEPLSVGIHACRRGGV- : 170
----------------------------------------------------------------------------------------------------------------------
                     180                  200                  220                      240                  260
               β10            α4          β11       α5         β12              α6         β13           α7
       HTT SS EEEEE   SH-HHHHHHHHHHHH TT EEEEE S HHHHHHH-HHHH SEEE HH     HHHHHHHHHHTT  EEEEEE    HHHHHHHHHHEEE
         WWW      NNN NN N                    PPP  P
2H6E : ISKFAEPVVIVNGI-GG-LAVYTIQILKALMKNITIVGISRSKKHRDFA-LELGADYVSEMK-----DAESLINKLTDGLGASIAIDLVGTEETTYNLGKLLAQ- : 261
1R37 : ----PTKTLLVVGA-GGGLGTMAVQIAKAV-SGATIIGVDVREEAVEAA-KRAGADYVINASMQ---DPLAEIRRITESKGVDAVIDLNNSEKTLSVYPKALAK- : 263
1H2B : L--YPGAYVAIVGV-GG-LGHIAVQLLKVM-TPATVIALDVKEEKLKLA-ERLGADHVVDARR----DPVKQVMELTRGRGVNVAMDFVGSQATVDYTPYLLGR- : 277
1RJW : ---KPGEWVAIYGI-GG-LGHVAVQYAKAM-G-LNVVAVDIGDEKLELA-KELGADLVVNPLKE---DAAKFMKEKV--GGVHAAVVTAVSKPAFQSAYNSIRR- : 253
1LLU : ---RPGQWVAISGI-GG-LGHVAVQYARAM-G-LHVAAIDIDDAKLELA-RKLGASLTVNARQE---DPVEAIQRDI--GGAHGVLVTAVSNSAFGQAIGMARR- : 259
2HCY : ---MAGHWVAISGAAGG-LGSLAVQYAKAM-G-YRVLGIDGGEGKEELF-RSIGGEVFIDFTKEK--DIVGAVLKATD-GGAHGVINVSVSEAAIEASTRYVRA- : 262
2DFV : ---ISGKSVLITGA-GP-LGLLGIAVAKAS-GAYPVIVSEPSDFRRELA-KKVGADYVINPFEE---DVVKEVMDITDGNGVDVFLEFSGAPKALEQGLQAVTP- : 259
1YQX : ---EPGKHIGIVGL-GG-LGHVAVKFAKAF-G-SKVTVISTSPSKKEEALKNFGADSFLVSRDQ---EQMQAA-----AGTLDGIIDTVSAVHPLLPLFGLLKS- : 270
8ADH : ---TQGSTCAVFGL-GG-VGLSVIMGCKAA-GAARIIGVDINKDKFAKA-KEVGATECVNPQDYK-KPIQEVLTEMSN-GGVDFSFEVIGRLDTMVTALSCCQEA : 286
1PL7 : ---TLGHKVLVCGA-GP-IGMVTLLVAKAM-GAAQVVVTDLSATRLSKA-KEIGADLVLQISKESPQEIARKVEGQLG-CKPEVTIECTGAEASIQAGIYATRS- : 266
----------------------------------------------------------------------------------------------------------------------
                          280                     300                    320                  340
        β14           β15   α8       β16               α9               β17          α10         β18
       EEEEEE    SS   EEE HHH-HHHTT EEEE  S  H---HHHHHHHHHHHTTSS     EE-EE   ################ EEEE
             NNNN                        AA      W   W
2H6E : EGAIILVGMEGKR--VSLEAFD-TAVWNKKLLGSNYGSL---NDLEDVVRLSESGKIKP--YII-KVP-LDDINKAFTNLDEGRVDGRQVITP--------- : 344
1R37 : QGKYVMVGLFGAD--LHYHAPL-ITLSEIQFVGSLVGNQ---SDFLGIMRLAEAGKVKP--MITKTMK-LEEANEAIDNLENFKAIGRQVLIP--------- : 347
1H2B : MGRLIIVGYGGE---LRFPTIR-VISSEVSFEGSLVGNY---VELHELVTLALQGKVRV--EV-DIHK-LDEINDVLERLEKGEVLGRAVLIP--------- : 359
1RJW : GGACVLVGLPPEE--MPIPIFD-TVLNGIKIIGSIVGTR---KDLQEALQFAAEGKVKT--II-EVQP-LEKINEVFDRMLKGQINGRVVLTLEDK------ : 339
1LLU : GGTIALVGLPPGD--FPTPIFD-VVLKGLHIAGSIVGTR---ADLQEALDFAGEGLVKA--TI-HPGK-LDDINQILDQMRAGQIEGRIVLEM--------- : 342
2HCY : NGTTVLVGMPAGAK-CCSDVFN-QVVKSISIVGSYVGNR---ADTREALDFFARGLVKS--PI-KVVG-LSTLPEIYEKMEKGQIVGRYVVDTSK------- : 348
2DFV : AGRVSLLGLYPGK--VTIDFNNLIIFKALTIYGITGRHLW--ETWYTVSRLLQSGKLNLDPIITHKYKGFDKYEEAFELMRAGKT-GKVVFMLK-------- : 348
1YQX : HGKLILVGAPEKP--LELPAFS-LIAGRKIVAGSGIGGM---KETQEMIDFAAKHNITA--DI-EVIS-TDYLNTAMERLAKNDVRYRFVIDVGNTLAATKP : 362
8ADH : YGVSVIVGVPPDSQNLSMNPM--LLLSGRTWKGAIFGGFKSKDSVPKLVADFMAKKFALDPLITHVLP-FEKINEGFDLLRSGES-IRTILTF--------- : 375
1PL7 : GGTLVLVGLGSEM--TTVPLLH-AAIREVDIKGVFRYC----NTWPVAISMLASKSVNVKPLVTHRFP-LEKALEAFETFKKGLG-LKIMLKCDPSDQNP-- : 357
```

Figure 4.4

Structural sequence alignment of medium chain ADHs. 2H6E *S. solfataricus* AraDH, 1R37 *S. solfataricus* ADH [107], 1H2B *Aeropyrum pernix* ADH [147], 1RJW *Geobacillus stearothermophilus* ADH [55], 1LLU *Pseudomonas aeruginosa* ADH [226], 2HCY *Saccharomyces cerevisiae* ADH1 [309], 2D8A *Pyrococcus horikoshii* threonine dehydrogenase [166], 1YQX *Populus tremuloides* sinapyl ADH [32], 8ADH Horse liver ADH [65], 1PL7 Human Sorbitol dehydrogenase [289]. Amino acid residue numbering of AraDH and secondary structure elements according to the Definition of Secondary Structure of Proteins (DSSP 177; H α-helix, E extended β-strand, B isolated β-bridge, G 3_{10}-helix, T H-bonded turn, S bend, # unstructured) are displayed. Individual amino acid functions are indicated by the following single letter codes: C catalytic zinc binding, Z structural zinc binding, R proton relay, A active site pocket, N cofactor pocket, P adenyl phosphate pocket, F flexible region (hinge), W water channel (pore).

by three ligands: Cys41, His65 and Asp150. The cysteine and histidine are nearly invariant in closely related ADH structures, but the aspartate is often replaced by a second cysteine (Fig. 4.4).

An interesting feature of the AraDH structure is that a highly conserved histidine, which has been suggested to be involved in the proton relay system, is substituted by an arginine (Arg46) (Fig. 4.4). Although there is some debate, it is generally believed that during substrate oxidation, the abstracted proton from the substrate is passed on to the solvent by this histidine, after it has been shuttled through a hydrogen bond network consisting of Thr43 and the 2' and 3' hydroxyl groups of the nicotinamide ribose moiety of the cofactor. The positive charge of the guanidinium group arginine (pK_a 12.5), however, would not allow it to function as a hydrogen bond acceptor at the slightly basic catalytic pH range of the enzyme. Similar exceptions can be observed in the cod liver ADH, where a tyrosine is found instead of a histidine [308], and the mouse ADH2 which has a glutamine at the corresponding position [389]. Mutagenesis studies with the human liver ADH showed that the histidine indeed acts as the base catalyst, but that it can be functionally replaced by a glutamine when a general base catalyst is present in the solvent [96]. Further studies are required to elucidate the exact role of Arg46 in AraDH, and the details of the proton relay system of ADH enzymes in general.

Computational substrate and cofactor modeling studies revealed that a relatively hydrophilic substrate binding pocket is formed by residues in several loops of both the catalytic and the nucleotide binding domain (Fig. 4.1 and Fig. 4.4). The pocket consists of Glu52, Lys54, Ala90, Thr91, Ile113, Gln116, Met270, Asn293, Tyr294 and is occupied by 6 water molecules in the apo AraDH structure. The low level of conservation of these residues between different ADHs as well as the fact that they are located in loops probably reflects the diverse substrate range of ADHs, and perhaps also allows new specificities to evolve by mutagenesis without affecting the structural integrity of the enzyme. The cofactor binding site, on the other hand, proved to be very well conserved. The main cofactor binding residues are located in α-helix 3 (Thr154), between β-strand 10 and α-helix 4 (Asn177, Gly178, Ile179, Gly180, Gly181, Leu182), and in the hinge region between β-strand 14 and 15 (Val268, Gly269, Met270, Glu271) (Fig. 4.4). Interestingly, a small cluster of residues is present in AraDH with sequence and strong spatial resemblance to the phosphate binding pocket of the NADP$^+$-dependent sinapyl ADH from aspen (*Populus tremuloides*) [32]. The plant enzyme uses the amino acid sequence STS-X_2-K to stabilize the adenyl phosphate group of the cofactor that distinguishes NADP$^+$ from NAD$^+$. Likewise, the *Sulfolobus* enzyme seems to employ the sequence SRS-X_2-H (Ser203-His208) (Fig. 4.4). This sequence is located in the peptide region that joins β-strand 11 and α-helix 5 and is most commonly replaced with a sequence starting with an aspartate followed by a hydrophobic residue in NAD$^+$ utilizing enzymes [20]. The aspartate is the primary determinant for NAD$^+$-cofactor specificity, since its sidechain hydrogen bonds with the 2' and 3' hydroxyl groups of the adenine ribose moiety. The fact that the NADP$^+$-specificity determining residues from aspen are relatively conserved among enzymes from hyperthermophilic Archaea may not only be useful for genome annotation purposes, but also for enzyme engineering studies. This phosphate binding motif is distinct from the GSR-X_{17}-Y motif, which is present in NADP$^+$-dependent dehydrogenases from mesophilic and thermophilic bacteria [207].

Comparison to the MDR superfamily

The AraDH structure superimposes well on structures of members of the MDR superfamily, despite the fact that the sequence identity to these proteins does not exceed 33%. The structures include the tetrameric ADHs from other hyperthermophilic archaea (*S. solfataricus*, *Aeropyrum pernix* and *Pyrococcus horikoshii*), thermophilic and mesophilic bacteria (*Geobacillus stearothermophilus* and *Pseudomonas aeruginosa*), yeast (*Saccharomyces cerevisiae*), and a number of dimeric ADHs from plants and mammals (Table 4.1). Remarkably, although the AraDH structure was solved without the substrate and cofactor, the lowest backbone RMSD was observed between AraDH and the *S. solfataricus* ADH structure with 2-ethoxyethanol and NADH bound, suggesting that the AraDH structure resembles the closed conformation. An unusual feature of the *S. solfataricus* ADH concerns a loop that covers the active site (Gly50-Val55), which is absent in AraDH and most structural homologs (Fig. 4.4). Also notable is the largely hydrophobic active site of most ADH structures, which probably reflects their preference for aliphatic and aromatic substrates. In contrast, the AraDH active site is very hydrophilic and is therefore much better suited for the binding and conversion of sugars. A unique feature of the AraDH structure seems to be an extension of α-helix 3 and elongation of the connecting loop towards β-strand 10. This enhances the subunit interface contacts between subunits A and D, possibly reflecting an adaptation to increased thermostability of the tetramer.

Enzyme kinetics

The enzyme was tested for its activity towards various pentoses and hexoses in combination with either NAD$^+$ or NADP$^+$ as a cofactor (Fig. 4.5). This showed that AraDH is specific for sugars with D-arabinose-type hydroxyl group stereo-configurations at C3 and C4. This configuration is shared among the hexoses L-fucose (6-*deoxy*-L-galactose) and L-galactose, and the pentose D-ribose. Although the highest activity was observed with L-fucose in combination with NAD$^+$, the remaining sugars were oxidized at higher rates using NADP$^+$. The Michaelis-Menten constants were determined for its preferred substrates and this showed that the enzyme has a 6.7-fold higher catalytic

Table 4.1 - Comparison to ADH structures of the MDR family

Source / description	PDB ID	Form / mutation	Res. (Å)	Seq. ID (%)	RMSD (Å)[a]	Ref.
S. solfataricus D-arabinose DH	2H6E	Apo	1.8	-	-	This study
S. solfataricus ADH	1JVB	Apo	1.85	33	1.67	108
	1R37	NADH, 2-ethoxyethanol	2.3		1.41	107
Aeropyrum pernix ADH	1H2B	NADH, octanoic acid	1.62	33	1.78	147
Geobacillus stearothermophilus ADH	1RJW	Apo	2.35	30	1.99	55
Pseudomonas aeruginosa ADH	1LLU	NADH, ethylene glycol	2.3	26	1.82	226
Saccharomyces cerevisiae ADH1	2HCY	8-Iodo-NAD$^+$, trifluorethanol	2.44	24	1.89	309
Pyrococcus horikoshii threonine DH	2DFV	NAD$^+$	2.05	25	2.32	166
Populus tremoloides sinapyl ADH	1YQX	Asp140Asn	2.5	23	2.49	32
	1YQD	NADP$^+$, Asp140Asn	1.65		2.47	32
Horse liver ADH	8ADH	Apo	2.4	21	2.97	65
	6ADH	NAD$^+$, dimethylsulfoxide	2.9		3.37	97
Human sorbitol dehydrogenase	1PL7	Apo	2.2	18	2.39	289

[a] Averaged backbone superimposition root-mean-square deviation values to all chains

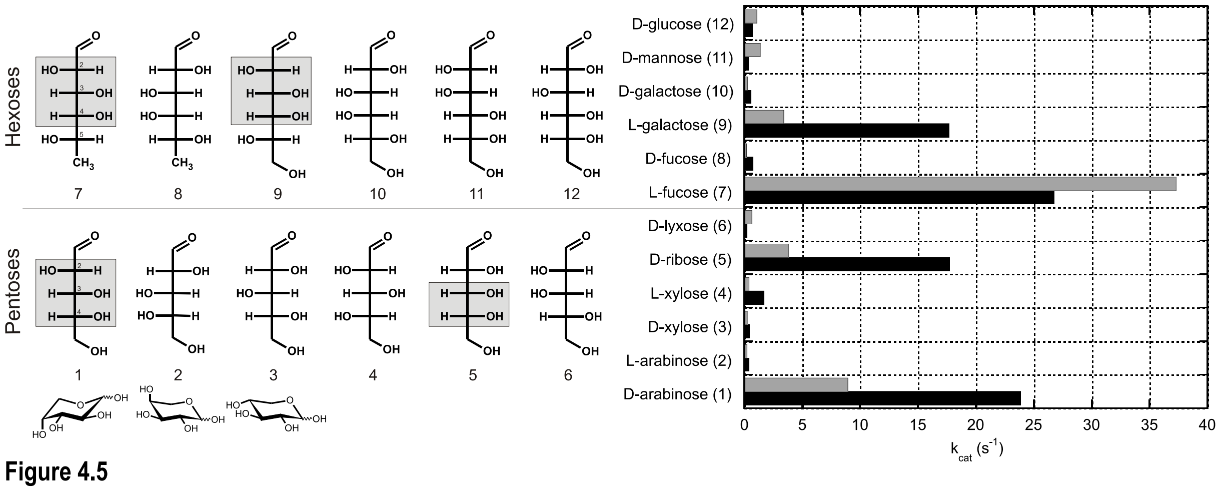


Figure 4.5
Relationship between substrate hydroxyl group configuration (Fisher projection) and turnover rate (k_{cat}). Hexoses and pentoses were tested at a concentration of 10 mM in combination with 0.4 mM of cofactor (NAD^+ gray bars, $NADP^+$ black bars). Gray boxes indicate the preferred stereo configurations at the C2, C3 and C4 atoms of the sugar. D-arabinose, L-arabinose and D-xylose are also shown in their chair conformation.

efficiency for L-fucose compared to its physiological substrate D-arabinose (Table 4.2). This dual specificity is common among several different protein types, including kinases [288], isomerases [347] and lectins [327], and is caused by the high degree of similarity between the conformers and stereochemistry of the hydroxyl groups of both sugars [30].

The catalytic efficiency of $NADP^+$ in combination with D-arabinose was 36-fold higher than that of NAD^+, which clearly indicates the preference for the phosphorylated cofactor. Similar substrate specificities and affinities have been found previously for bacterial and yeast D-arabinose dehydrogenases, although these enzymes favor NAD^+ as a cofactor [11,60]. In yeast, the concerted action of this enzyme and a D-arabinono-1,4-lactone oxidase produces D-erythroascorbic acid, which is believed to be an antioxidant molecule [161].

The *S. solfataricus* AraDH is more than 50% active in a relatively narrow pH range from 7.3 to 9.3, with optimal catalytic rates at pH 8.2 (Fig. 4.6A). The thermophilicity of the enzyme is apparent from its optimal catalytic temperature of 91 °C (Fig. 4.6B). The enzyme is relatively stable at 70 and 80 °C and maintains a half-life of 42 and 26 min at 85 and 90 °C, respectively, indicating that the enzyme is thermostable at physiological growth temperatures of *S. solfataricus* of approximately 80 °C [144] (Fig. 4.6C).

Substrate docking

To investigate the structural basis of substrate and cofactor binding the monosaccharides L-fucose and D-arabinose were docked in the AraDH active site in their preferred chair conformer, *i.e.* β-L-fucopyranose and α-D-arabinopyranose, respectively. During *in silico* docking analyses, the equatorial C1O was assumed to form the fourth ligand of the zinc ion, in agreement with the generally accepted mechanism of oxidation by ADHs [310]. The

Table 4.2 - Kinetic parameters

Substrate / Cofactor	K_m (mM)	V_{max} ($U\ mg^{-1}$)	$k_{cat} K_m^{-1}$ ($s^{-1}\ mM^{-1}$)
D-arabinose[a]	1.40 ± 0.09	23.8 ± 0.4	10.6
L-fucose[a]	0.22 ± 0.03	23.7 ± 0.2	67.0
$NADP^{+b}$	0.038 ± 0.010	32.1 ± 0.8	525
NAD^{+b}	1.25 ± 0.45	29.1 ± 8.0	14.5

Determined with [a] 0.4 mM $NADP^+$, and [b] 10 mM D-arabinose results from these docking simulations consistently indicated that the energetically most favorable orientation of both substrates is such that the C3 and C4 hydroxyl groups of the substrate are hydrogen bonded to Lys54 (Fig. 4.7). In addition, the C3 hydroxyl also appears to be hydrogen bonded to Asn293, whereas the substrate's sole axial hydroxyl substituent at C4 forms a hydrogen bond with the side chain of Gln116. The important role of the stereoconfiguration at C3 and C4 is supported by the substrate specificity profile of AraDH, which indicates that the enzyme only accepts substrates with this specific configuration at the C3 and C4 positions, such as D-arabinose, L-fucose, L-galactose and D-ribose (Fig. 4.5). The methyl group of L-fucose, which distinguishes it from D-arabinose, is stacked against the phenyl ring of Tyr294 and forms more hydrophobic contacts with the aliphatic side chains of Ala90, Thr91 and Ile113 (latter not shown). Perhaps these hydrophobic interactions account for the higher affinity and catalytic efficiency of the enzyme for L-fucose as compared to D-arabinose (Table 4.2).

Sugar metabolism

Members of the genus *Sulfolobus* are unique among hyperthermophilic Archaea, because they can metabolize hexoses (glucose, galactose, fructose and mannose) as well as pentoses (arabinose, xylose and ribose) [144,168,437]. Recently, D-arabinose was shown to be oxidized to 2-oxoglutarate by an induced pathway, in which AraDH catalyzes the initial step [39]. Both glucose and galactose are degraded by promiscuous enzymes according to a slightly modified scheme that was originally proposed by Entner and Doudoroff (ED) for *Pseudomonas saccharophila* [84,103,213]. Instead of employing a glucokinase or hexokinase [273], *Sulfolobus* seems to postpone substrate phosphorylation to the level of 2-keto-3-*deoxy*-D-gluconate [4].

The structure of the first enzyme of the ED pathway, the glucose dehydrogenase (GDH, Sso3003) which also belongs to the MDR superfamily, was recently elucidated [258]. AraDH shares 19% sequence identity with this enzyme, and their sugar substrate specificity range appears to be very distinct. Whilst GDH preferentially catalyzes the oxidation of D-glucose, D-galactose, D-fucose, L-arabinose and D-xylose [213], the activity of AraDH towards these substrates is negligible (Fig. 4.5). In contrast, AraDH favors L-fucose, L-galactose, D-arabinose and D-ribose, the latter two sug-

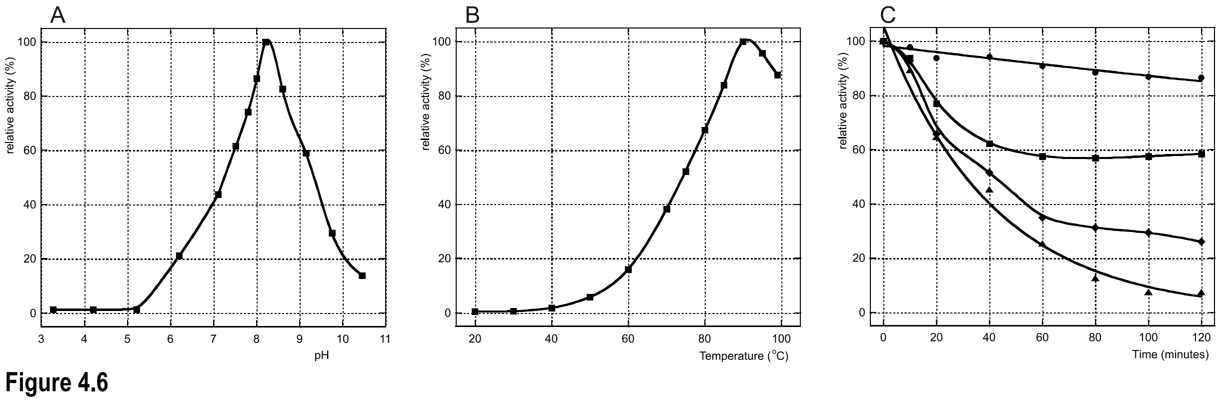


Figure 4.6
Catalytic properties of AraDH. Relative reaction rates as a function of **A)** pH and **B)** temperature. **C)** Thermal inactivation curves. Residual activity is given after a pre-incubation at 70 °C (dot), 80 °C (square), 85 °C (diamond) and 90 °C (triangle).

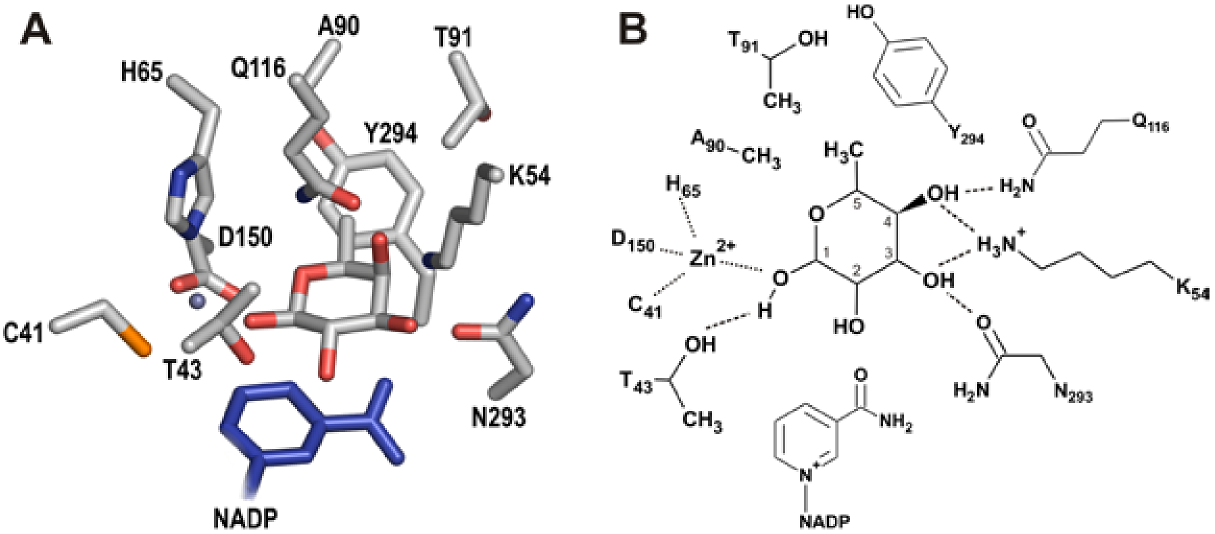


Figure 4.7 (in color on p.154)
Docking solution of the Michaelis complex for α-L-fucopyranoside (L-fucose) and the nicotinamide cofactor in **A)** stick and **B)** schematic representation. The C1O of the substrate forms the fourth ligand of the tetrahedrally coordinated zinc ion (gray sphere). The C6 methyl group of L-fucose that distinguishes it from D-arabinose is stacked against the phenyl ring of Tyr294 and forms additional hydrophobic contacts with the aliphatic side chains of Ala90, Thr91 and Ile113 (latter not shown). Hydroxyl groups at C3 and C4 form extensive hydrogen bonds with Lys54, Gln116 and Asn293. The C1H is positioned correctly to allow hydride transfer to the nicotinamide cofactor.

ars which in turn form no substrates for GDH (L-fucose and L-galactose not tested [213]). At the molecular level, the substrate specificity of GDH has been elegantly explained by the co-crystal structures of GDH mutant Thr41Ala with both D-glucose and D-xylose in the presence of the NADP$^+$ cofactor [258]. It shows that the enzyme binds the β-anomeric pyranose form of the sugar in the typical D-glucose and D-xylose-preferred chair conformer. AraDH, however, seems to opt for monosaccharides that adopt the opposite chair conformation (Fig. 4.5) [30]. A comparison of the active site of GDH and AraDH by superimposition shows that the sugars in both enzymes occupy approximately the same position and plane in the active site pocket (not shown). However, the pyranose ring appears to be flipped 180° relative to the other, as can be deduced from the position of the ring-oxygen of the sugar. Besides the residues involved in zinc ion binding (Cys and His) and proton-relay (Thr), none of the amino acids that form hydrogen bonds with the hydroxyl groups of the substrate are conserved between the two enzymes, further supporting their distinct modes of substrate binding that results in different substrate specificity. GDH exhibits a high catalytic efficiency for D-xylose, which suggests that the enzyme could play a role in an uncharacterized D-xylose degrading pathway in *S. solfataricus* [258]. Given the presence of D-xylonate dehydratase activity [199], this indeed appears to be plausible since both the oxidation of D-xylose and D-arabinose yield the intermediate 2-oxo-*4(S)*,5-dihydroxypentanoate (also called 2-keto-3-*deoxy*-D-arabinonate/xylonate). In this pathway intermediate the chiral differences between the two pentoses at the C2 and C3 atoms have been eliminated by a pentonate dehydratase. The final two conversions to 2-oxoglutarate could then be conducted by the enzymes 2-keto-3-*deoxy*-D-arabinonate dehydratase (KdaD) and the 2,5-dioxopentanoate dehydrogenase (DopDH) [39].

Apart from the complementary substrate range of AraDH and GDH, it seems likely that because of the unusually high number of paralogs (GDH: Sso3042, Sso3204; and AraDH: Sso2441, Sso1646, Sso1220, Sso0472, Sso2494, Sso0764, Sso2501, Sso2878, Sso2334, Sso2800, Sso2717 and Sso3237) even more sugar dehydrogenases with yet different substrate specificities are encoded in the *S. solfataricus* P2 genome [353]. This large set of enzymes may be indicative of its aerobic sugar and alcohol degrading lifestyle in acidic hot springs.

EXPERIMENTAL PROCEDURES

Gene cloning and protein overexpression

The genomic fragment corresponding to Sso1300 (GenBank ID: 15898142) was PCR-amplified from *S. solfataricus* P2 genomic DNA using *Pfu* TURBO polymerase (Stratagene) and primers 5'-GGATACCATGGTTAAATCAAAAGCAGCCCTTC-3' and 5'-CTAGAGGATCCTTATGGTGTTATTACTTGTCTTCCG-3' (NcoI and BamHI underlined). The amplified genes were cloned into vector pET24d (Novagen) using *Escherichia coli* HB101 as a host. The resulting plasmid (pWUR284) was harvested by Miniprep (Qiagen), sequenced by Westburg genomics (Leusden, Netherlands), and transformed into *E. coli* expression strain BL21(DE3) (Novagen) containing the tRNA accessory plasmid pRIL (Stratagene). The transformants were grown in four 1 L shake-flasks containing LB medium until the optical cell density (A_{600}) had reached 0.5. The cultures were then cold-shocked to induce host chaperones by placing them on ice [38]. After 30 min, the protein overproduction was started by adding 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). A final concentration of 0.25 mM $ZnSO_4$ was added to ensure a full occupancy of the zinc binding sites of the enzyme [139]. The culture was incubated for 12 h at 37 °C after which it was spun down (10 min, 5,000 xg, 4 °C).

Protein purification

Pelleted *E. coli* cells were resuspended in 20 mM Tris-HCl buffer (pH 7.5) supplemented with 50 mM NaCl, and disrupted by sonication at 0 °C. Insoluble cell material was spun down (30 min, 26,500 xg, 4 °C) and the *E. coli* supernatant was subjected to heat-treatment for 30 min at 75 °C. Denatured proteins were removed by centrifugation (30 min, 26,500 xg, 4 °C) yielding a heat stable cell-free extract. This fraction was filtered (0.2 μm) and applied to a 20 ml Matrex RedA affinity column (Amicon) connected to an FPLC. After washing with 2 column volumes of buffer, the recombinant protein was eluted by a linear gradient of 10 column volumes of buffer supplemented with 2 M of NaCl.

Enzyme assays

All enzymatic assays were performed in degassed 0.1 M HEPES-KOH buffer (pH 7.5) at 70 °C. Sugar dehydrogenase activity was determined on a Hitachi U-1500 spectrophotometer in a continuous assay using 10 mM of D- and L-arabinose, D- and L-xylose, D-ribose, D-lyxose, D- and L-fucose, D- and L-galactose, D-mannose and D-glucose as a substrate, and 0.4 mM of NAD^+ or $NADP^+$ as a cofactor. Initial enzymatic activity rates were obtained from the increase in absorption at 340 nm (A_{340}) and calculated using a molar extinction coefficient of 6.22 mM^{-1}·cm^{-1}. The optimal pH of catalysis was determined using a 30 mM citrate-phosphate-glycine buffer system which was adjusted in the range of pH 3 to 11 at 70 °C. Thermal inactivation assays were performed by incubating 50 μg ml^{-1} of enzyme at 70, 80, 85 and 90 °C and drawing aliquots at regular intervals during 2 h followed by a standard activity assay.

Protein crystallization and data collection

The protein was crystallized by the hanging drop method of vapor diffusion in 0.1 M HEPES-NaOH (pH 8.55) and 10% v/v Jeffamine M-600 which was mixed 1:1 with 1 μl of protein solution at a concentration of 4 mg ml^{-1}. Tetragonal bipyrimidal crystals of 50-150 μm appeared after 3 weeks at 20 °C. Data were collected at the K-edge anomalous peak wavelength of zinc (1.2827 Å) from a single flash-frozen (100 K) native crystal to 1.8 Å resolution using a MAR345 imaging plate at the Protein Structure Factory beamline BL-14.2 of the Free University of Berlin at BESSY (Berlin, Germany). All data were reduced using HKL2000 [280]. The crystal used for data collection had unit cell parameters *a,b*= 84.019 Å, *c*= 194.979 Å, α,β,γ= 90° and belonged to space group $I4_122$ with a monomer in the asymmetric unit (Table 4.3). The solvent content of the crystal has been estimated as approximately 47% (v/v) with a Matthews coefficient (V_M) of 2.3 Å^3·Da^{-1} for a monomer in the asymmetric unit [253].

Structure solution and refinement

The structure of AraDH was determined by single-wavelength anomalous dispersion (SAD) and phased by making use of the endogenous zinc ions in each subunit. The two zinc sites were located using the program SOLVE, corresponding to the single AraDH molecule in the asymmetric unit [397]. Subsequently, density modification and automated model building was performed using RESOLVE [397]. The initial model was then improved using the free-atom refinement method together with automatic

Table 4.3 - Data collection and refinement

X-ray data collection statistics	
X-ray source	Synchrotron BESSY, Beamline 14.2
Wavelength (Å)	1.2827
Space group	$I4_122$
Unit cell dimensions and angles	a=b=84.02, c=194.98 α=β=γ=90°
Subunits per asymmetric unit	1
Matthews coefficient (Å^3 Da^{-1})	2.3
Resolution limits (Å)	50 - 1.80 (1.86 – 1.80)
Total measurements	403473
Unique reflections	32525
Completeness (%)	99.1 (93.0)
I / $\sigma(I)$	31.7 (3.0)
R_{sym} [a]	0.076 (0.452)
Redundancy	12.4 (7.0)
Refinement statistics	
Resolution (Å)	42 – 1.80
Reflections in refinement	30781
R_{work} [b]	0.195
R_{free} [c]	0.232
RMSD bond distances (Å)	0.011
RMSD bond angles (°)	1.285
Total number of protein atoms	2483
Av. protein B value (Å^2)	27
Number of zinc ion atoms	2
Av. zinc ion B value (Å^2)	44
Number of solvent molecules	215
Av. solvent B value (Å^2)	41

[a] $R_{sym} = \Sigma_{hkl}\,\Sigma_i\,|I - \langle I_i \rangle| \,/\, \Sigma_{hkl}\,\Sigma_i\,I_i$, where I_i is the intensity of a given measurement and the sums are over all measurements and reflections. Values in parentheses refer to the highest resolution shell.

[b] $R_{work} = \Sigma\,||F(\text{obs})|-|F(\text{calc})|| \,/\, \Sigma\,|F(\text{obs})|$ for the 95% of the reflection data used in refinement.

[c] $R_{free} = \Sigma\,||F(\text{obs})|-|F(\text{calc})|| \,/\, \Sigma\,|F(\text{obs})|$ for the remaining 5%.

model tracing in ARP/wARP [294]. TLS (translation/libration/screw) parameters were determined and TLS/positional-restrained refinement was performed using REFMAC5 [268]. Several rounds of iterative model building and refinement followed and water molecules were added automatically using ARP/wARP [294]. Coot was used throughout for manual model building and for viewing electron density maps [102].

The final model (comprising 324 amino acids, 215 water molecules and 2 zinc ions), refined using data between 50 and 1.8 Å resolution, has an R- and free R-factor of 19.5 and 23.2%, respectively, with good geometry (Table 4.3). Alternative side chain conformations were observed for residues Arg98, Arg132, Leu182, Asn196, Asp235, Ser240, Glu249, Met270, Glu301, Arg305 and Ser309. Residues Glu14 to Ile18 and Asp322 to Asp337 are not visible in the electron density map and have therefore been excluded from the model. Additionally, the side chains of Lys3, Lys5, Lys11, Ser13, Glu19, Arg42, Arg58, Lys79, Arg204, Lys232, Lys273 and Lys313 have been truncated in the final model. The stereochemical quality of the model and the model fit to the diffraction data were analyzed with the programs PROCHECK [216] and SFCHECK [409]. The coordinates and experimental structure factors have been deposited in the Protein Data Bank with accession code 2H6E.

Computational methods

Structural homologs were identified using blastp searches against the pdb sequence database at the National Center for Biotechnology Information (NCBI), and by VAST and DALI searches against the structural database. Oligomeric structure analysis was performed at the PISA (Protein Interfaces, Surfaces and Assemblies) webserver at EMBL-EBI. A structural multiple sequence alignment of these homologous protein sequences was generated using the Expresso program. Substrate and cofactor docking studies were performed with GOLD v3.1.1 (Genetic Optimization for Ligand Docking, CCDC software, Cambridge Crystallographic Data Centre) [416]. Figures of the protein structures were prepared with PyMol.

ACKNOWLEDGEMENTS

This work was supported by a grant from the European Union in the framework of the SCREEN project.

Structural insight into substrate binding and catalysis of a novel 2-keto-3-*deoxy*-D-arabinonate dehydratase illustrates common mechanistic features of the FAH superfamily

Stan J.J. Brouns*
Thomas R.M. Barends*
Petra Worm
Jasper Akerboom
Andrew P. Turnbull
Laurent Salmon
John van der Oost

*contributed equally

Submitted for publication

Abstract

The archaeon *Sulfolobus solfataricus* converts D-arabinose to 2-oxoglutarate by an enzyme set consisting of two dehydrogenases and two dehydratases. The third step of the pathway is catalyzed by a novel 2-keto-3-*deoxy*-D-arabinonate dehydratase (KdaD). In this study, the crystal structure of the enzyme has been solved to 2.1 Å resolution. The enzyme forms a tetrameric oval-shaped ring of identical subunits, each consisting of an N-terminal domain with a four stranded β-sheet flanked by two α-helices, and a C-terminal catalytic domain with a fumarylacetoacetate hydrolase (FAH) fold. Crystal structures of complexes of the enzyme with magnesium or calcium ions and either a substrate analog 2-oxobutyrate, or the aldehyde enzyme product 2,5-dioxopentanoate were obtained by soaking native protein crystals. The divalent metal ion in the active site is octahedrally coordinated by three conserved carboxylate residues, a water molecule, and both the carboxylate and the oxo groups of the substrate molecule. An enzymatic mechanism for base-catalyzed dehydration is proposed based on the binding mode of the substrate to the metal-ion, which suggests that the enzyme enhances the acidity of the protons α to the carbonyl group, facilitating their abstraction by glutamate-114. A comprehensive structural comparison of members of the FAH superfamily is presented and their evolution is discussed, providing a basis for functional investigations of this largely unexplored protein superfamily.

INTRODUCTION

Pentoses are ubiquitous five-carbon sugars that occur in various polysaccharides and in nucleic acids. Several aerobic archaea and proteobacteria metabolize these sugars to the citric acid cycle intermediate 2-oxoglutarate by employing an enzyme set that consists of a pentose dehydrogenase, a pentonolactonase, a pentonate dehydratase, a 2-keto-3-*deoxy*-pentonate dehydratase and a 2,5-dioxopentanoate dehydrogenase (also called α-ketoglutarate semialdehyde dehydrogenase) [39,380,423,425-427]. Starting from stereochemically diverse pentoses such as D- and L-arabinose, D-xylose and D-ribose, these enzymes cancel out the chiral differences between the hydroxyl groups of sugars, funneling them to the final and shared aldehyde oxidation step. This pathway is an alternative for the well-known xylulose-5-phosphate generating pathways, in which the conversions are carried out by isomerases, epimerases and kinases in bacteria, and reductases, dehydrogenases and kinases in fungi [58,222,323].

The hyperthermophilic archaeon *Sulfolobus solfataricus* catabolizes D-arabinose to 2-oxoglutarate. The penultimate step of this pathway, the elimination of a water molecule from 2-keto-3-*deoxy*-D-arabinonate (D-KDA), is catalyzed by the D-KDA dehydratase (KdaD) (Fig. 5.1) [39]. The C-terminal domain of this enzyme resembles the catalytic domain of members of the fumarylacetoacetate hydrolase (FAH) protein family, an enzyme class involved in the catabolism of aromatic compounds in mammals and bacteria. The mammalian FAH enzyme catalyzes the last step in tyrosine degradation; the hydrolytic cleavage of fumarylacetoacetate yielding fumarate and acetoacetate [400]. The catalytic domain of KdaD is also homologous to the hydratase MhpD and the bifunctional decarboxy-lase/isomerase HpcE, both involved in two distinct *meta*-fission pathways of hydroxyphenyl-related compounds in *Escherichia coli* [90].

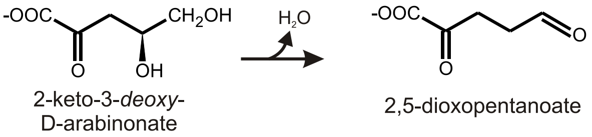


Figure 5.1 - Chemical reaction catalyzed by the 2-keto-3-*deoxy*-D-arabinonate dehydratase

In this study we have elucidated the structure of KdaD, established its mode of substrate binding, and obtained insight into its catalytic mechanism. A comprehensive structural comparison of KdaD, MhpD, HpcE and FAH with five functionally unassigned homologs is presented, providing a basis for uncovering the functions of these homologs in archaea, bacteria and eukaryotes.

RESULTS

Overall structure

The KdaD monomer comprises 293 amino acids and is composed of two domains: an N-terminal domain (N-domain, residues 1-69), and a C-terminal catalytic domain (residues 70-293) (Fig. 5.2). The N-domain consists of a four-stranded anti-parallel β-sheet (β_A 1-2-3-4) flanked on either side by an α-helix (α1 and 2). The core of the catalytic domain adopts a mixed β-sandwich roll fold typical of the fumarylacetoacetate hydrolase (FAH) enzyme [400]. This fold is made up of two mostly antiparallel β-sheets ($\beta_{B,C}$) consisting of six and five β-strands, respectively (β_B 6-5-13-8-9-10

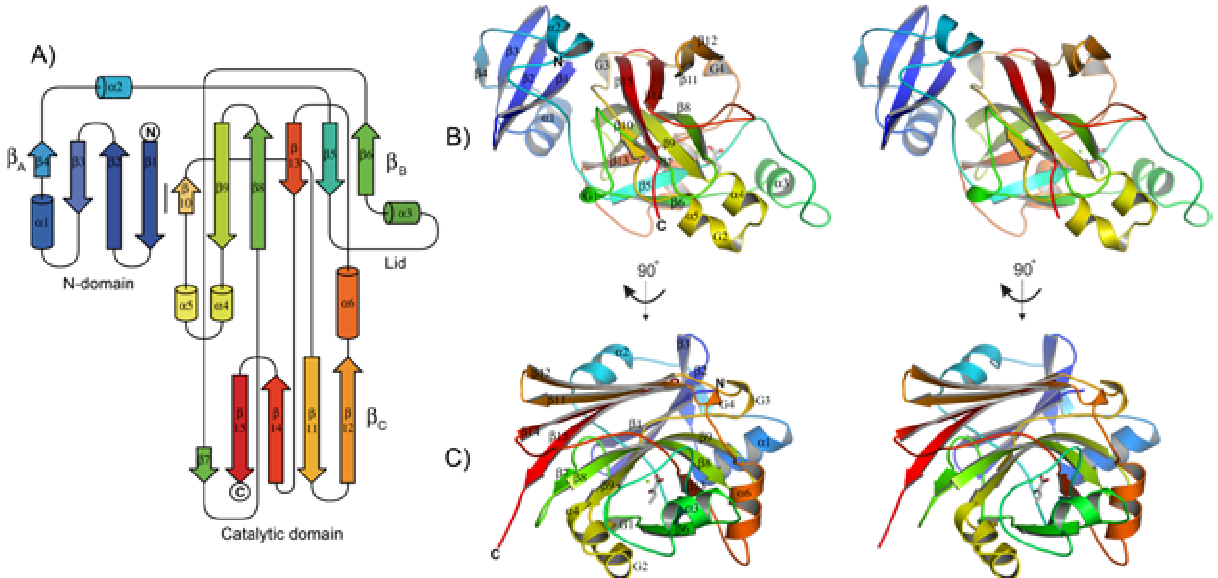


Figure 5.2 (in color on p.155)
A) Topology diagram showing the connectivity of secondary structure elements and domain organization of the *S. solfataricus* KdaD monomer. The N-terminus of the polypeptide chain (N, in blue) follows a color gradient towards the C-terminus (C, in red). **B)** and **C)** Stereo ribbon diagrams at two viewing angles of the KdaD monomer with Mg^{2+} and 2-oxobutyrate bound.

and β_C 12-11-14-15-7), which are intertwined by five crossovers of the polypeptide from sheet to sheet (Fig. 5.2). The two β-sheets are flanked by four α-helices (α3 to 6) and four small 3_{10}-helices (G1 to 4). The β_B-sheet displays a pronounced twist delineating the active site pocket by forming a barrel-like structure (Fig. 5.2C). The boundaries of the catalytic site are further defined by the helical elements α4, G2 and α5, which seem to serve as a platform from which residues point towards the active site cavity. On the opposite site, α-helix 3 marks the boundary of the active center. Interestingly, this α-helix is preceded by a loop, termed the lid (Met85-Asn94) that covers the active site entrance. The loop appears to be flexible based on its high B-factors in the KdaD-Mg^{2+}/Ca^{2+}-2OB structures, and on its disordered structure in the apo and DOP-bound KdaD structures.

Quaternary structure

A homo-tetramer of KdaD can be generated by applying crystallographic symmetry operators and is composed of a dimer of dimers (Fig. 5.3). The observation of a tetramer is consistent with the oligomeric state of KdaD determined by electrospray ionization mass-spectrometry [39,413]. Both the N-domain and catalytic domains are involved in intermolecular interfaces resulting in the formation of an elongated donut-shaped tetramer with a central hole of approximately 26 x 43 Å (Fig. 5.3). The largest protein-protein interface in the tetramer is formed by the catalytic domains of two monomers, burying a total of approximately 3700 Å^2 of solvent accessible protein surface per dimer. This subunit contact is mediated by thirteen hydrogen bonds and six salt bridges. The residues involved in these intersubunit interactions are located in the secondary structure elements β5, α3, β6, G1, α4, G2, α5, and in loops between β7-8, α5-β10, and α6-β13. Two catalytic-domain dimers form a tetramer through interactions between the N-terminal domains (Fig. 5.3). This subunit interaction shields approximately

1300 Å^2 of accessible protein surface area from the solvent, and involves six hydrogen bonds and six salt bridges per dimer. The subunit interface includes residues that are located mainly in β-strand 2, 3 and α-helix 1.

Active site

On the surface of the catalytic domain, a deep pocket is observed containing a number of acidic residues. Inside the pocket, a phosphate ion was observed in the SeMet structure. The presence of a phosphate ion in a presumably negatively charged pocket seems unlikely, but given the pH of the experiment (4.8) and the

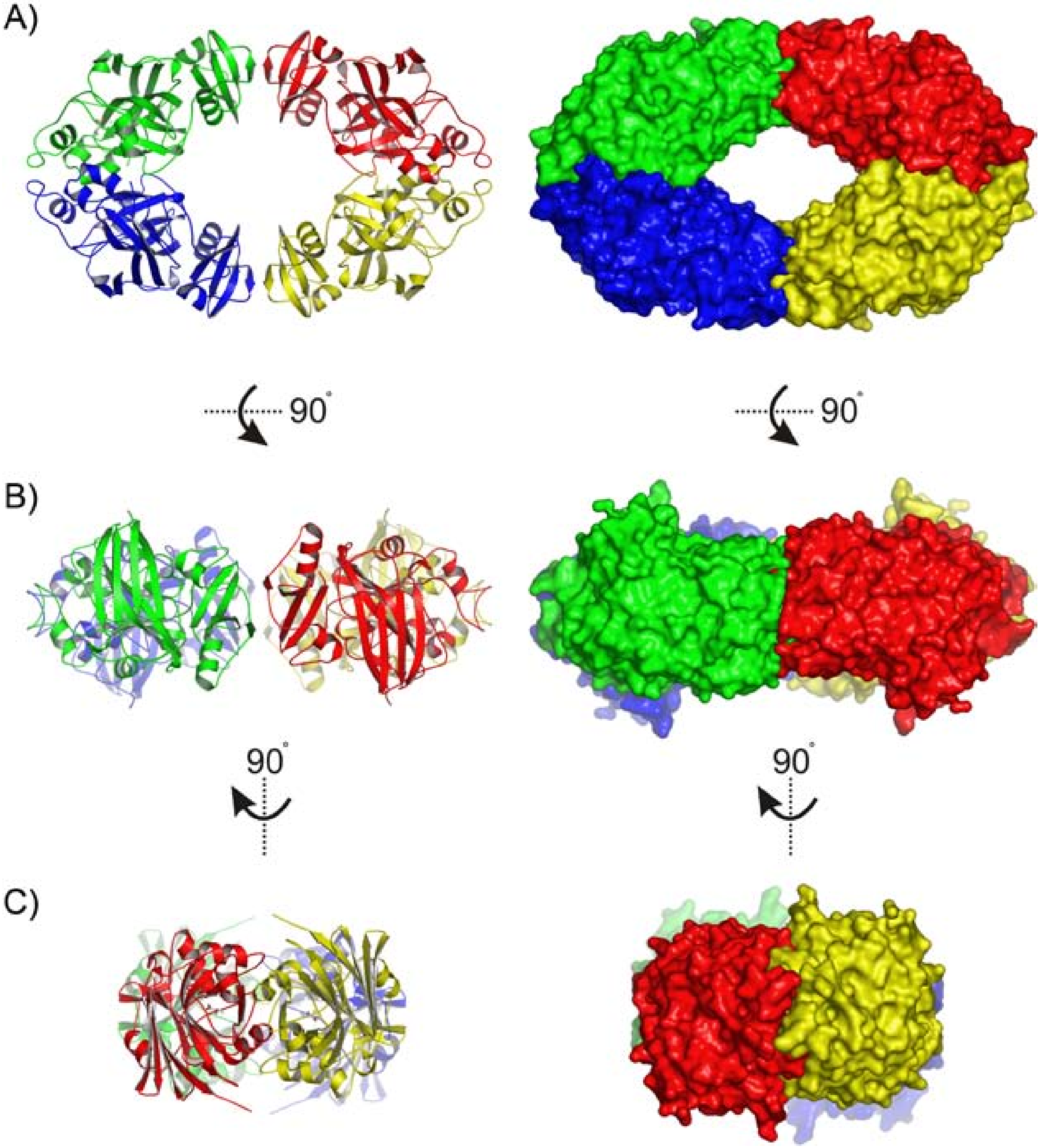


Figure 5.3 (in color on p.155)
Schematic ribbon and surface diagrams are shown of the *S. solfataricus* KdaD homo-tetramer generated by crystallographic symmetry operations. View **A)** perpendicular to the tetrameric ring, **B)** along the N-domain dimer-dimer interface, and **C)** along the catalytic domain dimer interface.

presence of a lysine residue (Lys182) in the pocket interacting with this ion, it is likely that this is a dihydrogen phosphate ion. Next to the lysine, three acidic residues (Glu143, Glu145, Asp164) emerging from β-strand 8 and 9, respectively, form the bottom of the pocket in a constellation reminiscent of a metal ion binding site, reinforcing the notion that this pocket is the active site. This prompted us to prepare crystals of the holoenzyme, *i.e.* with magnesium or calcium bound. To this end, native protein crystals obtained in high phosphate concentrations were cross-linked with glutaraldehyde, and subsequently washed with buffer solutions containing these divalent metal ions. Since glutaraldehyde binds covalently to lysines, a dataset was collected from a cross-linked crystal washed in a solution without phosphate or divalent metal ions. The resulting apo-enzyme electron density maps did not show any additional density associated with Lys182, indicating that the cross-linking procedure did not introduce glutaraldehyde adducts in the pocket (not shown).

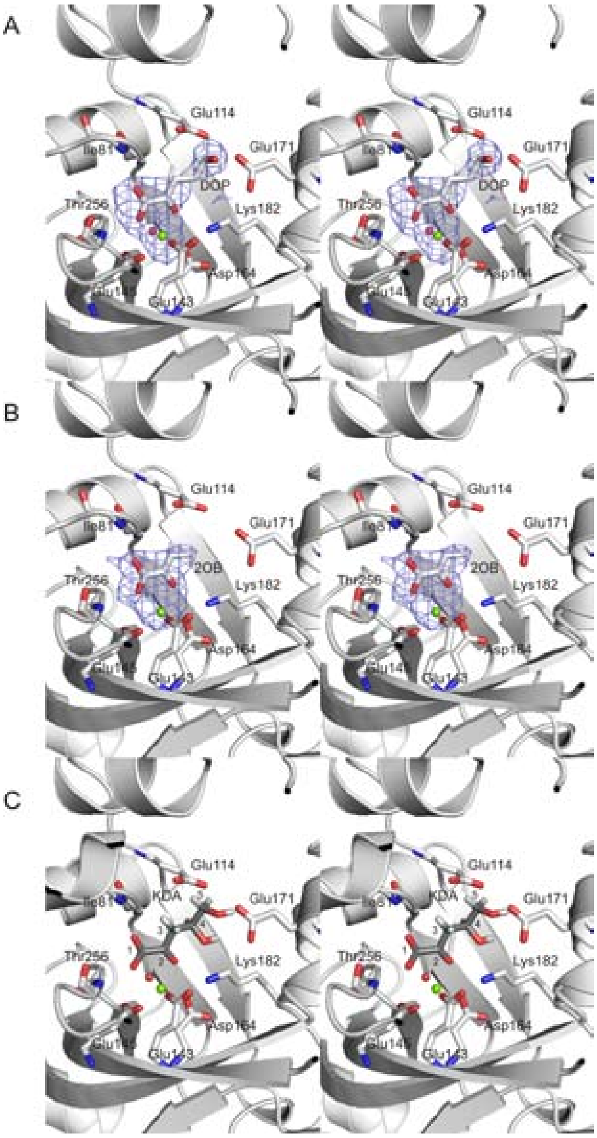


Figure 5.4 (in color on p.156)
Stereo diagrams are shown of the KdaD active site with Mg^{2+} (green spheres), water molecules (red spheres) and **A)** 2,5-dioxopentanoate or **B)** 2-oxobutyrate bound. Simulated annealing F_o-DF_c omit electron density maps for the ligand and the metal ion are displayed at a 3σ contour level. **C)** Stereo image of a model of the Michaelis complex of KdaD with its substrate 2-keto-3-*deoxy*-D-arabinonate (D-KDA) bound. Hydrogen atoms for D-KDA are shown as well as the carbon atom numbers.

Soaking cross-linked, washed crystals in solutions with magnesium and either the product 2,5-dioxopentanoate (DOP) or substrate analogue 2-oxobutyrate (2OB) resulted in complexes of the enzyme with these compounds as evidenced by σ_A-weighted [313] simulated annealing difference density maps (Fig. 5.4AB). Either compound was found inside the pocket, ligating the metal ion in the same bidentate way, with the *anti*-orbital of one of its carboxylate oxygens and with its 2-oxo atom. The 2-oxo atom is also hydrogen bonded to Lys182. In the KdaD-DOP and KdaD-2OB structures, additional interactions are made between the carboxylate moiety of the ligand and the backbone amide groups of Ile81 and Thr256. In the KdaD-DOP complex, the 5-oxo atom of the product is hydrogen bonded to the Glu171 side chain, whereas the C5 methylene group seems involved in CH-O hydrogen bonds to both Oε atoms of the Glu114 (distances 3.2 and 3.3 Å) (Fig. 5.4A and Fig. 5.5A).

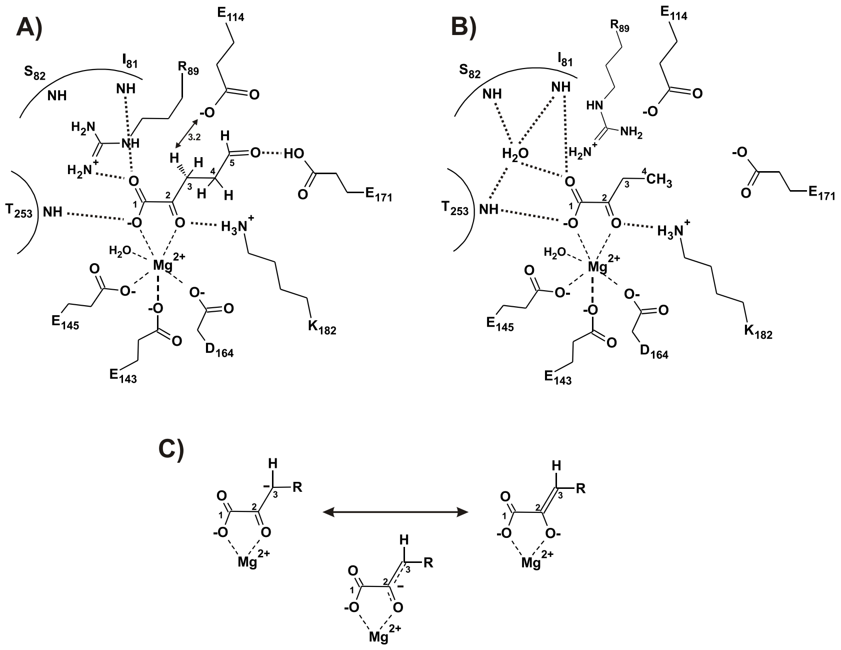


Figure 5.5
Schematic diagrams are shown illustrating the binding of Mg^{2+} and **A)** 2,5-dioxopentanoate and **B)** 2-oxobutyrate. **C)** Enolate resonance structures.

The metal ion is hexacoordinated with octahedral geometry by the acidic residues in the pocket (Glu143, Glu145 and Asp164), two oxygen atoms of the organic ligand, and a water molecule. An essentially identical complex was obtained with 2-oxobutyrate when calcium was used instead of magnesium (Fig. 5.4B and Fig. 5.5B).

The electron density for the lid (Met85–Asn94) became sufficiently visible in the 2-oxobutyrate complex to include the loop in the model. Although the B-factors in this loop refined to high values, the peptide chain could be traced with good confidence. Apart from closing off the active site pocket, the lid places the side chain of Arg89 deeply inside the pocket where it binds in close proximity to the inhibitor and interacts with Glu143 and Glu171.

Interestingly, nine ordered water molecules are present in a cavity at the interface between two catalytic domains (Fig. 5.6). Five of these water molecules interconnect the carboxylic acid side chains of the active site residue Glu114 of both subunits by a water-filled tunnel of approximately 12.5 Å. The wall of the tunnel is lined by hydrophobic residues Phe116, Pro175, Leu176, Leu178 and Pro179, from β-strand 6, and helices G2 and α5.

DISCUSSION

Catalytic mechanism

The structures of KdaD with Mg^{2+} and 2OB or DOP yield significant insight into substrate

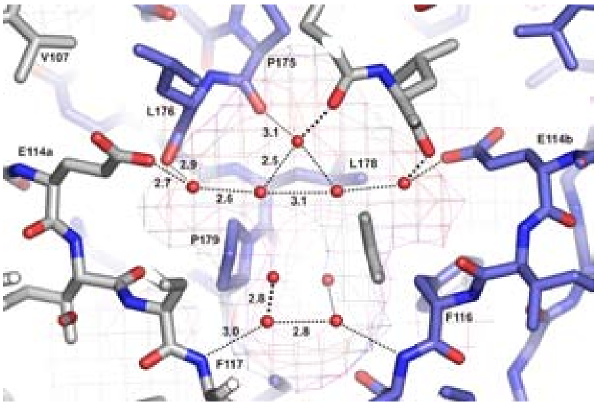


Figure 5.6 (in color on p.157)
A schematic representation of the catalytic dimer interface is shown, indicating the hydrogen bond connectivity between Glu114a and Glu114b residues of both active sites through a mostly hydrophobic water tunnel. Interatomic distances are displayed in Å.

binding, and allow for speculation on the catalytic mechanism of water elimination by the 2-keto-3-*deoxy*-D-arabinonate dehydratase. The most important observation of the complex structures is the bidentate chelation of the ligands to the metal ion. As indicated in Figure 5.5C, this increases the acidity of the protons α to the carbonyl group by mesomery, allowing for easy proton abstraction from the C3 carbon atom by a base such as Glu114. Furthermore, the KdaD-DOP structure shows that the C5 aldehyde oxygen atom is hydrogen bonded to residue Glu171. A third observation can be made from modeling the substrate 2-keto-3-*deoxy*-D-arabinonate (D-KDA) into the active site (Fig. 5.4C). This places the C4 hydroxyl group of the substrate KDA in close proximity of Lys182.

Based on these observations, two tentative base-catalyzed dehydration mechanisms are proposed (Fig. 5.7). In the $E_2\alpha\beta$ mechanism (Fig. 5.7A), one of the C3 protons is abstracted by Glu114, while Lys182 assists in the expulsion of the C4-hydroxyl group by acting as a proton donor, resulting in intermediate 3. Whether a metal cofactor-stabilized enolate intermediate exists (intermediate 2), or whether elimination would occur concertedly, is presently unknown. The next stage of the proposed mechanism is a rearrangement of the intermediate 3 into the enol compound 4, involving proton abstraction from the C5 carbon atom by Glu114, which must be deprotonated for this to take place. Interestingly, Glu114 is hydrogen bonded to the buried water cluster at the interface between the

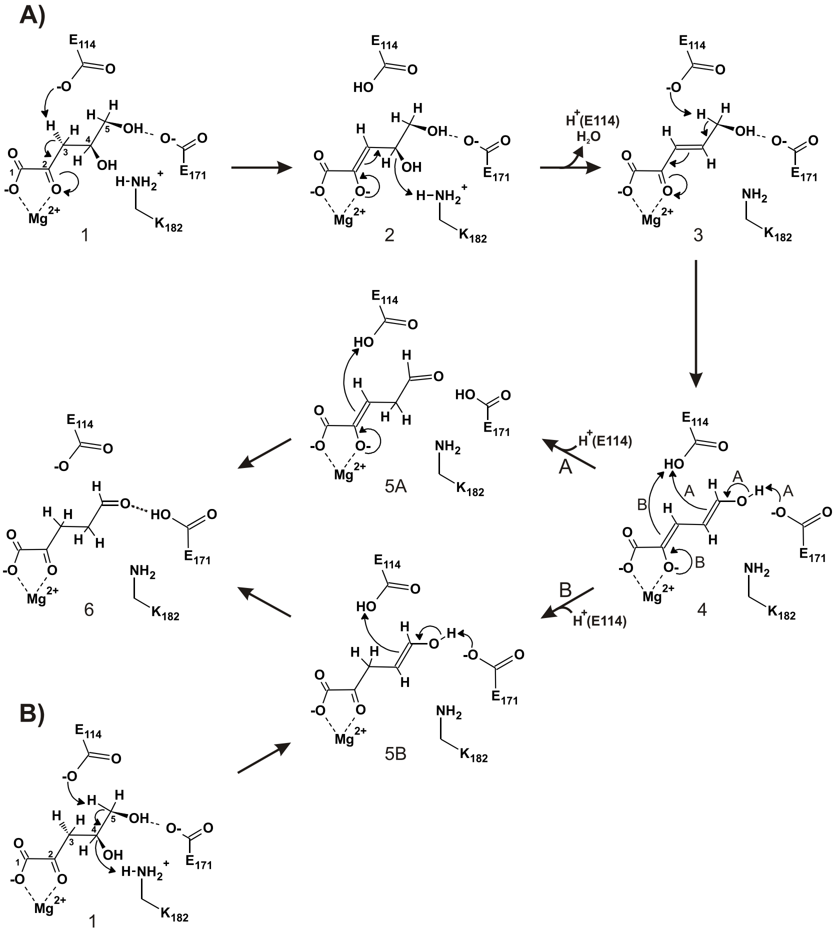


Figure 5.7
Proposed mechanism of base-catalyzed dehydration according to a **A)** $E_2\alpha\beta$ and **B)** $E_2\beta\gamma$ mechanism.

catalytic domains. In fact, the two active sites in a catalytic-domain dimer are connected *via* hydrogen bonds through this interaction. We propose that this water cluster serves to deprotonate Glu114, enabling it to abstract a proton from C5. From here, the reaction could continue *via* two paths; either direct tautomerization of the enol 4 occurs, leading to the enolate intermediate 5A followed by protonation at C3 by Glu114, or the sequence of these two events is reversed, *i.e.* C3 is protonated first leading to intermediate 5B, and tautomerization occurs subsequently.

A second and much simpler reaction mechanism would be an $E_2\beta\gamma$ reaction mechanism (Fig. 5.7B). In this mechanism, Glu114 acts as the base by abstracting a proton from the C5 methylene directly, while Lys182 protonates the C4 hydroxyl group causing it to leave as a water molecule, leading to the enol form of the product (intermediate 5B).

The main difference of the two mechanisms concerns the abstraction of the C4 hydroxyl moiety, which is accompanied by proton abstraction from either the C3 methylene ($E_2\alpha\beta$, Fig. 5.7A), or from the C5 methylene ($E_2\beta\gamma$, Fig. 5.7B). Following the considerations about the increased acidity of the C3 methylene protons given above, it seems clear that the former option is the most probable one. An argument in favor of the latter mechanism is the very close interaction between the C5 methylene and Glu114 observed in the KdaD-DOP complex. In both cases, the mechanism requires Glu114 to function as a relatively strong base, which may be favored by the exclusion of the bulk solvent by the closing of the lid. Conceivably, the interaction between Arg89 and Glu171 serves as an anchor for the lid. The side chain of residue Arg168 forms a second layer of positive charge at the active site entrance. This residue is highly conserved throughout the FAH superfamily (Fig. 5.8), and may serve to exclude water from the active center, as has been suggested for FAH [26].

Catalytic domain

The catalytic domain of KdaD (residues 76-293) is homologous to several structures in the protein databank belonging to the fumarylacetoacetate hydrolase (FAH) protein superfamily (Table 5.1, Fig. 5.8 and Fig 5.9). The domain superimposes relatively well on these structures with RMSD values of 2.3 to 8.1 Å, despite a rather low sequence identity of 12 to 18% in this polypeptide region (Table 5.1). A structure-based sequence alignment of these proteins reveals a conserved core of amino acid residues, which are interrupted by peptide insertions in the various protein structures, most notably in the FAH enzyme (Fig. 5.8). Interestingly, a region corresponding to the lid in KdaD is disordered in many structural homologs, including 1WZO, 2DFU, 1I7O, 1NKQ, 1NR9 and 1SAW (Fig. 5.8 and Fig. 5.9). This reinforced the notion that the flexibility of this loop allows closing and opening of the active site entrance, and could be a requirement for catalytic activity in other FAH-like enzymes.

N-domain

The N-terminal domain of KdaD consisting of a four-stranded antiparallel β-sheet appears to be present only in the structures 1WZO and 2DFU, and in the FAH crystal structures (Fig.

5.8 and Fig 5.9). While both 1WZO and 2DFU structures lack α-helix 1, FAH is decorated with five extra α-helices surrounding the β-sheet core of the *Sulfolobus* enzyme. Interestingly, although all these proteins contain this additional domain, only KdaD and 1WZO seem to form ring-like tetramers through association of their N-terminal domains. The ring-like structure of 1WZO (not shown), however, is much more compact and spherical than the planar KdaD structure, which is caused by the deletion of α-helix 1 that allows both dimers to associate at an angle. The fact that the only two FAH homologs that appear to form tetramers through N-domain association are present in thermophiles, suggests that having this domain might be a thermostabilizing feature. Possibly this feature has become redundant in mesophiles, allowing the N-domain to be completely replaced, discarded or to evolve into a new functional domain, which may be of regulatory nature [400].

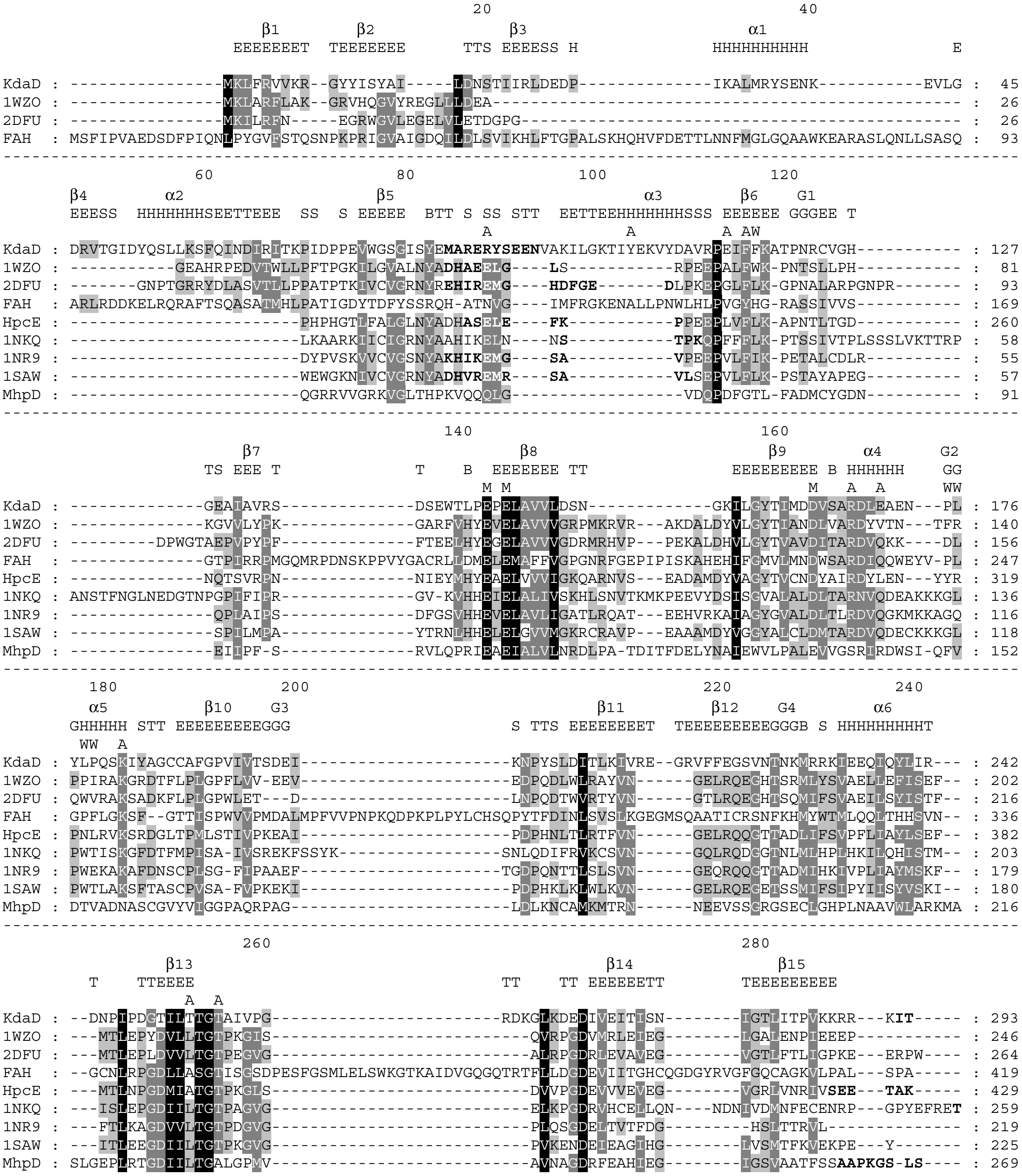
FAH superfamily

Several members of the FAH superfamily for which a crystal structure is available, have unknown functions (Table 5.1). These structures include two putative enzymes from *Thermus thermophilus* HB8 (1WZO and 2DFU) [259,260], the protein YcgM (1NR9) from *E. coli* [201], and

Table 5.1 - Members of the FAH superfamily

Source / description	PDB ID	Form / mutation	Res. (Å)	Seq. ID (%)	RMSD (Å)[a]	Ref.
S. solfataricus, KdaD, Sso3118	2Q18	SeMet, phosphate, tetramer	2.1	-	-	This study
	2Q1A	Mg^{2+}, 2-oxobutyrate, tetramer	2.5			
	2Q1D	Mg^{2+}, 2,5-dioxopentanoate, tetramer	2.7			
T. thermophilus, TTHA0958[c]	1WZO	tetramer	1.9	18	2.46 (40-246)	259[b]
T. thermophilus, TTHA0809[c]	2DFU	dimer	2.2	19	2.86 (44-264)	260[b]
M. musculus, FAH, fumarylacetoacetate hydrolase	1QQJ	Ca^{2+}, acetate, dimer	1.55	13	8.05 (118-419)	400
	1QCO	Ca^{2+}, fumarate, acetoacetate, dimer	1.9			
	1QCN	Ca^{2+}, acetate, dimer	1.9			
	1HYO	Mg^{2+}, Ca^{2+}, HBU[d] , acetate, dimer	1.3			27
	2HZY	Mn^{2+}, Na^{+}, DHJ, dimer	1.35			26
E. coli, HpcE OPET[d] decarboxylase / isomerase	1I7O	Ca^{2+}, monomer	1.7	17	3.06 (219-429)	394
	1GTT	Ca^{2+}, monomer	1.7			
S. cerevisiae YNQ8_YEAST[c]	1NKQ	Ca^{2+}, acetate, dimer	2.2	13	5.14 (6-259)	109[b]
E. coli YcgM EC1262 (APC5008)[c]	1NR9	Mg^{2+}, dimer	2.7	17	2.25 (13-219)	201[b]
H. sapiens, FLJ36880[c]	1SAW	Mg^{2+}, dimer	2.2	18	3.13 (15-225)	246
E. coli MhpD HPDA[d] hydratase	1SV6	apo, pentamer	2.9	12	ND (54-271)	114[b]

[a] Averaged backbone superimposition RMSD values to the catalytic domain of all chains. Catalytic domain residue numbers in parenthesis; [b] reference to PDB entry; [c] function unassigned; [d] Abbeviations: OPET 2-oxo-5-carboxy-hept-3-ene-1,7-dioic acid; HBU 4-(hydroxylmethylphosphinoyl)-3-oxo-butanoic acid; DHJ 4-(2-carboxyethyl)(hydroxyl)phosphoryl))-3-oxo-butanoic acid; HPDA 2-hydroxypenta-2,4-dienoic acid

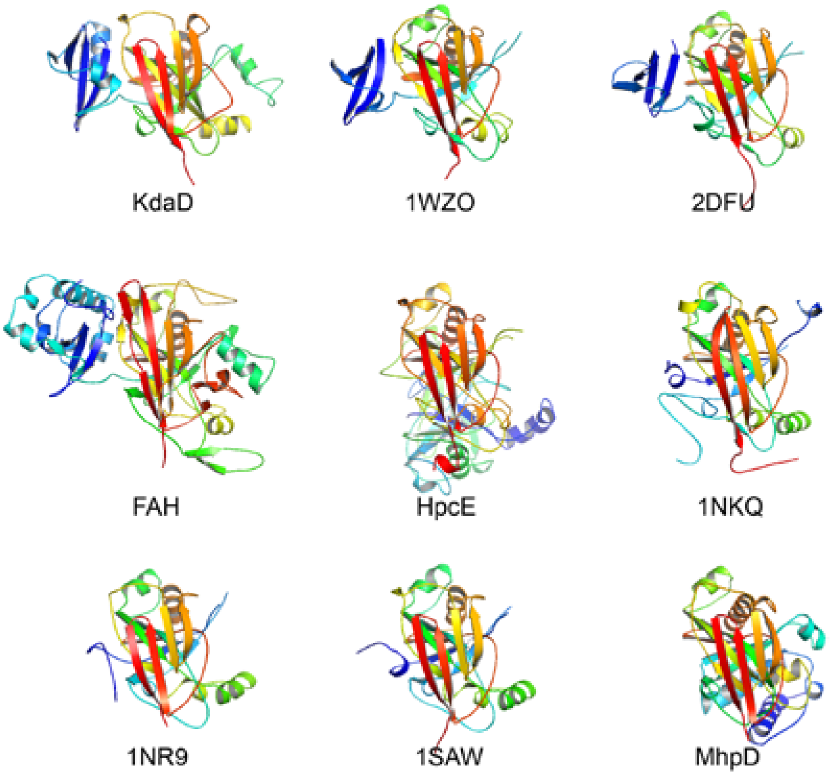


```
                                                 20                                40
                           β1        β2              β3                       α1
                        EEEEEEET  TEEEEEEE      TTS EEEESS H              HHHHHHHHHH               E

KdaD : ----------------MKLFRVVKR--GYYISYAI-----LDNSTIIRLDEDP--------------IKALMRYSENK-----------EVLG :  45
1WZO : ----------------MKLARFLAK--GRVHQGVYREGLLLDEA------------------------------------------------- :  26
2DFU : ----------------MKILRFN-----EGRWGVLEGELVLETDGPG---------------------------------------------- :  26
FAH  : MSFIPVAEDSDFPIQNLPYGVFSTQSNPKPRIGVAIGDQILDLSVIKHLFTGPALSKHQHVFDETTLNNFMGLGQAAWKEARASLQNLLSASQ :  93
----------------------------------------------------------------------------------------------------------
                    60                   80                 100                 120
       β4        α2                    β5                          α3        β6    G1
       EEESS  HHHHHHHSEETTEEE  SS  S EEEEE  BTT S SS STT  EETTEEHHHHHHHSSS EEEEEE GGGEE T
                                                  A              A         A AW
KdaD : DRVTGIDYQSLLKSFQINDIRITKPIDPPEVWGSGISYEMARERYSEENVAKILGKTIYEKVYDAVRPEIFFKATPNRCVGH----------- : 127
1WZO : -----------GEAHRPEDVTWLLPFTPGKILGVALNYADHAEELG----LS-----------RPEEPALFWK-PNTSLLPH----------- :  81
2DFU : -------GNPTGRRYDLASVTLLPPATPTKIVCVGRNYREHIREMG----HDFGE-------DLPKEPGLFLK-GPNALARPGNPR------- :  93
FAH  : ARLRDDKELRQRAFTSQASATMHLPATIGDYTDFYSSRQH-ATNVG----IMFRGKENALLPNWLHLPVGYHG-RASSIVVS----------- : 169
HpcE : ------------------------PHPHGTLFALGLNYADHASELE----FK-----------PPEEPLVFLK-APNTLTGD----------- : 260
1NKQ : ------------------------LKAARKIICIGRNYAAHIKELN----NS-----------TPKQPFFFLK-PTSSIVTPLSSSLVKTTRP :  58
1NR9 : ------------------------DYPVSKVVCVGSNYAKHIKEMG----SA-----------VPEEPVLFIK-PETALCDLR---------- :  55
1SAW : ------------------------WEWGKNIVCVGRNYADHVREMR----SA-----------VLSEPVLFLK-PSTAYAPEG---------- :  57
MhpD : ------------------------QGRRVVGRKVGLTHPKVQQQLG------------------VDQPDFGTL-FADMCYGDN---------- :  91
----------------------------------------------------------------------------------------------------------
                                              140                              160
                         β7                           β8                        β9        α4      G2
                     TS EEE T              T    B  EEEEEEE TT               EEEEEEEEE B HHHHHH    GG
                                                  M M                               M   A  A      WW
KdaD : --------------GEAIAVRS--------------DSEWTLPEPELAVVLDSN-------------GKILGYTIMDDVSARDLEAEN---PL : 176
1WZO : --------------KGVVLYPK--------------GARFVHYEVELAVVVGRPMKRVR---AKDALDYVLGYTIANDLVARDYVTN---TFR : 140
2DFU : ---------DPWGTAEPVPYPF--------------FTEELHYEGELAVVVGDRMRHVP---PEKALDHVLGYTVAVDITARDVQKK----DL : 156
FAH  : --------------GTPIRRPMGQMRPDNSKPPVYGACRLLDMELEMAFFVGPGNRFGEPIPISKAHEHIFGMVLMNDWSARDIQQWEYV-PL : 247
HpcE : --------------NQTSVRPN--------------NIEYMHYEAELVVVIGKQARNVS---EADAMDYVAGYTVCNDYAIRDYLEN---YYR : 319
1NKQ : ANSTFNGLNEDGTNPGPIFIPR--------------GV-KVHHEIELALIVSKHLSNVTKMKPEEVYDSISGVALALDLTARNVQDEAKKKGL : 136
1NR9 : ---------------QPLAIPS--------------DFGSVHHEVELAVLIGATLRQAT---EEHVRKAIAGYGVALDLTLRDVQGKMKKAGQ : 116
1SAW : ---------------SPILMPA--------------YTRNLHHELELGVVMGKRCRAVP---EAAAMDYVGGYALCLDMTARDVQDECKKKGL : 118
MhpD : ---------------EIIPF-S--------------RVLQPRIEAEIALVLNRDLPA-TDITFDELYNAIEWVLPALEVVGSRIRDWSI-QFV : 152
----------------------------------------------------------------------------------------------------------
         180                 200                                         220                 240
         α5          β10     G3                                β11         β12     G4        α6
       GHHHHH STT EEEEEEEEEGGG                       S TTS EEEEEEEET  TEEEEEEEEEGGGB S HHHHHHHHHT
        WW  A
KdaD : YLPQSKIYAGCCAFGPVIVTSDEI----------------------KNPYSLDITLKIVRE--GRVFFEGSVNTNKMRRKIEEQIQYLIR--- : 242
1WZO : PPIRAKGRDTFLPLGPFLVV-EEV----------------------EDPQDLWLRAYVN------GELRQEGHTSRMLYSVAELLEFISEF-- : 202
2DFU : QWVRAKSADKFLPLGPWLET---D----------------------LNPQDTWVRTYVN------GTLRQEGHTSQMIFSVAEILSYISTF-- : 216
FAH  : GPFLGKSF--GTTISPWVVPMDALMPFVVPNPKQDPKPLPYLCHSQPYTFDINLSVSLKGEGMSQAATICRSNFKHMYWTMLQQLTHHSVN-- : 336
HpcE : PNLRVKSRDGLTPMLSTIVPKEAI----------------------PDPHNLTLRTFVN------GELRQQGTTADLIFSVPFLIAYLSEF-- : 382
1NKQ : PWTISKGFDTFMPISA-IVSREKFSSYK-----------------SNLQDIFRVKCSVN------GQLRQDGGTNLMLHPLHKILQHISTM-- : 203
1NR9 : PWEKAKAFDNSCPLSG-FIPAAEF---------------------TGDPQNTTLSLSVN------GEQRQQGTTADMIHKIVPLIAYMSKF-- : 179
1SAW : PWTLAKSFTASCPVSA-FVPKEKI----------------------PDPHKLKLWLKVN------GELRQEGETSSMIFSIPYIISYVSKI-- : 180
MhpD : DTVADNASCGVYVIGGPAQRPAG-----------------------LDLKNCAMKMTRN------NEEVSSGRGSECLGHPLNAAVWLARKMA : 216
----------------------------------------------------------------------------------------------------------
                         260                                                 280
                 β13                                           β14               β15
         T    TTEEEE                                TT    TT EEEEEETT        TEEEEEEEEE
                   A  A
KdaD : --DNPIPDGTILTTGTAIVPG------------------------RDKGLKDEDIVEITISN--------IGTLITPVKKRR---KIT----- : 293
1WZO : ---MTLEPYDVLLTGTPKGIS---------------------------QVRPGDVMRLEIEG--------LGALENPIEEEP----------- : 246
2DFU : ---MTLEPLDVVLTGTPEGVG---------------------------ALRPGDRLEVAVEG--------VGTLFTLIGPKE---ERPW---- : 264
FAH  : --GCNLRPGDLLASGTISGSDPESFGSMLELSWKGTKAIDVGQGQTRTFLLDGDEVIITGHCQGDGYRVGFGQCAGKVLPAL---SPA----- : 419
HpcE : ---MTLNPGDMIATGTPKGLS---------------------------DVVPGDEVVVEVEG--------VGRLVNRIVSEE---TAK----- : 429
1NKQ : ---ISLEPGDIILTGTPAGVG---------------------------ELKPGDRVHCELLQN----NDNIVDMNFECENRP---GPYEFRET : 259
1NR9 : ---FTLKAGDVVLTGTPDGVG---------------------------PLQSGDELTVTFDG---------HSLTTRVL-------------- : 219
1SAW : ---ITLEEGDIILTGTPKGVG---------------------------PVKENDEIEAGIHG--------LVSMTFKVEKPE---Y------- : 225
MhpD : SLGEPLRTGDIILTGALGPMV---------------------------AVNAGDRFEAHIEG--------IGSVAATFSSAAPKGS-LS---- : 269
```

Figure 5.8
Structure-based sequence alignment of members of the FAH superfamily: KdaD *S. solfataricus* 2-keto-3-*deoxy*-D-arabinonate dehydratase; 1WZO *Thermus thermophilus* HB8 TTHA0958 [259]; 2DFU *T. thermophilus* HB8 TTHA0809 [260]; FAH *Mus musculus* fumarylacetoacetate hydrolase [400]; HpcE *Escherichia coli* 2-oxo-5-carboxy-hept-3-ene-1,7-dioic acid decarboxylase-domain (residues 219-429) [394]; 1NKQ *Saccharomyces cerevisiae* YNQ8_YEAST (res. 6-259) [109]; 1NR9 E. coli YcgM (res. 13-219) [201]; 1SAW *Homo sapiens* FLJ36880 (res. 15-225) [246]; MhpD *E. coli* 2-hydroxypenta-2,4-dienoic acid hydratase (res. 54-271) [114,303]. Amino acid numbering follows the KdaD protein sequence. Secondary structure elements of KdaD are displayed according to the Definition of Secondary Structure of Proteins (DSSP [177]; H α-helix, E extended β-strand, B isolated β-bridge, G 3_{10}-helix, T H-bonded turn, S bend). Residues in bold indicate disordered peptide regions in the structure. Individual amino acid functions are marked by the following single letter codes: M metal ion ligand, A active site, W water tunnel.

Figure 5.9 (in color on p.157)
A ribbon representation is shown of members of the FAH superfamily as monomers. See Table 5.1 for details.



both a human FLJ36880 (1SAW) [246] and a yeast protein YNQ8_YEAST (1NKQ) [109].

A biochemically characterized member includes the bifunctional decarboxylase/isomerase HpcE (1I7O, 1GTT, EC 4.1.1.68/5.3.3.10) from *E. coli* C, which is involved in two consecutive steps of the homoprotocatechuate *meta*-fission pathway of aromatic compounds such as 3- and 4-hydroxyphenylacetate that ultimately leads to pyruvate and succinate [90,394]. This monomeric enzyme performs the Mg^{2+}-dependent decarboxylation of 2-oxo-5-carboxy-hept-3-ene-1,7-dioic acid to 2-hydroxy-hepta-2,4-diene-1,7-dioic acid, which is then possibly channeled to the N-terminal isomerase domain and converted into 2-oxo-hept-3-ene-1,7-dioic acid. Interestingly, HpcE seems to be the result of a gene duplication of the decarboxylase domain, followed by a gene fusion event and subsequent diversification leading to the isomerase domain. The fold of the two-domain protein resembles the dimeric assembly other FAH family members.

The enzyme MhpD from *E. coli* is the second enzyme that is biochemically studied and from which the crystal structure is known [114,303] (1SV6, EC 4.2.1.80, COG3971 [396]). This protein is involved in the *meta*-fission pathway of the aromatic compounds 3-hydroxyphenylpropionic acid and 3-hydroxycinnamic acid, which are degraded to pyruvate and acetyl-CoA [90]. The enzyme efficiently hydrates 2-hydroxypenta-2,4-dienoic acid to 4-hydroxy-2-ketopentanoic acid using a divalent metal ion, preferably manganese. Oxalate is a potent inhibitor of catalysis, suggesting that *cis*-2-keto-pent-3-enoic acid is a transition state reaction intermediate [303]. Remarkably, the crystal structure suggests that this protein forms ring-shaped pentamers [114], instead of the common dimeric unit that is observed in other FAH family members.

The prototype of the protein family is the fumarylacetoacetate hydrolase (FAH, EC 3.7.1.2) [26,27,400]. The dimeric FAH enzyme catalyzes oxo-hydrolytic cleavage of a carbon-carbon bond of the substrate and produces fumarate and acetoacetate as the last step in the mammalian tyrosine catabolism. A genetic loss of enzyme function in humans leads to an autosomal recessive disorder called hereditary tyrosinemia type I (HTI). The catalytic mechanism of FAH involves both a divalent metal ion and a catalytic triad consisting of a water molecule, a histidine and a glutamate. After substrate binding, the diketo group of the acetoacetate moiety forms a bidentate chelate of the metal ion. As with KdaD, this binding mode allows for the stabilization of negative charges developing on the substrate as the reaction proceeds [400].

Common mechanistic features

There are remarkable similarities in terms of conserved active site residues among members of the FAH protein superfamily (Fig. 5.8). The hallmark of these enzymes is the nearly invariant divalent metal ion binding site normally consisting of two glutamates and one aspartate residue (in KdaD Glu143, Glu145 and Asp164). These carboxylate groups form three of the six ligands of the octahedrally-coordinated divalent metal ion. The only structures with alterations in the EED triad are the *E. coli* MhpD enzyme in which the aspartate is replaced by a third glutamate, and the FAH enzyme in which the water ligand is replaced for an additional aspartate residue (DEED). A valine instead of an aspartate (DEEV) will lead to tyrosinemia in humans, which implies a critical role for this amino acid residue in enzyme function [28].

Another highly conserved residue in the FAH protein family seems to be Lys182 of KdaD. In both KdaD and FAH this residue is thought to act as a proton donor for the C4 hydroxyl leaving group of D-KDA, and the acetoacetate carbanion, respectively. Moreover, this lysine residue stabilizes the negative charges developing during the reaction in both enzymes (Fig. 5.7A). Only in MhpD is this amino acid not conserved, suggesting that the hydration reaction catalyzed by this enzyme occurs without a proton donation step by this lysine.

The two glutamate residues predicted to function as catalytic base and acid (Glu114 and Glu171) are not conserved in the distant structural homologs mentioned above (Fig. 5.8), indicating that these proteins are involved in a different type of enzymatic reaction. However, multiple sequence alignment of thirty-nine KdaD orthologs belonging to COG3970 [396] from a variety of microbial sources, including pathogens from the genera *Yersinia*, *Burkholderia* and *Bordetella*, reveals that Glu114 and Glu171 are completely conserved at these positions (not shown). This is consistent with their important functional roles in all enzymes that belong to this functional category.

Analysis of the active site architecture further reveals that the unknown enzymatic activity of 2DFU from *T. thermophilus* (TTHA0809), 1NR9 from *E. coli* (YcgM), 1NKQ from yeast (YNL168C), and 1SAW from *H. sapiens* (FLJ36880) is expected to be very similar. Interestingly, a genomic context survey of prokaryotic genomes starting with the *Thermus* TTHA0809 gene (2DFU) shows that this gene often co-occurs next to a glutamyl-tRNA synthetase gene, which may suggest a role in glutamate metabolism. The other *Thermus* gene sequence (TTHA0958, 1WZO) is commonly found in gene clusters of the homoprotocatechuate pathway for the degradation of aromatic compounds, strongly suggesting a decarboxylase activity similar to *E. coli* HpcE.

Despite the variation in active site composition, the structural evidence clearly suggests that these proteins have evolved divergently from a common ancestor. We speculate that the ancestral enzyme had a metal binding site, capable of the binding of either a 2- or 3-oxoacid substrate moiety, which allows for the stabilization of negative charges developing during the reaction.

We here described a metal ion dependent D-KDA dehydratase as a key enzyme in the degradation of D-arabinose. Interestingly, the degradation pathway for L-arabinose in *Azospirillum brasiliense* was recently shown to contain a 2-keto-3-*deoxy*-L-arabinonate (L-KDA) dehydratase (AraD) belonging to this Schiff-base forming enzyme class [426]. A detailed kinetic and mechanistic study of a similar L-KDA dehydratase involved in the L-arabinose metabolism of *Pseudomonas saccharophila* has been described a few decades ago, but unfortunately, neither the amino terminal sequence of the protein nor the gene sequence was reported [304,384]. The authors found that this enzyme was enantioselective for L-KDA and that the reaction also proceeds via a Schiff-base intermediate. Because of this enantioselectivity, this enzyme cannot operate in the degradation pathways for D-arabinose or D-xylose, which result in the intermediate D-KDA [39]. In those cases a KdaD-ortholog seems to be required, which is consistent with the gene being present in catabolic gene clusters for the degradation of D-arabinose in *S. solfataricus* and D-xylose in *Caulobacter crescentus* [158,256,380]. The complementary substrate range of the KdaD and the *A. brasiliense* AraD enzyme therefore seem to form independent evolutionary solutions for the dehydration of the chiral molecule KDA. These two enzymes enable microbes to degrade a wider variety of pentose sugars, and provide them with a selective advantage over competing microbes in their ecological niche.

EXPERIMENTAL PROCEDURES

Gene cloning and protein overexpression

The genomic fragment corresponding to Sso3118 was PCR-amplified from *S. solfataricus* P2 genomic DNA using *Pfu* TURBO polymerase (Stratagene) and primers 5'-GTTCCGTCATGAAATTATTTAGAGTTGTAAAAAG-3' and 5'-GACGGAAGCTTCTAAGTTATTTTTCTTCTTTTTTTAAC-3' (BspHI and HindIII underlined). The amplified genes were cloned into vector pET24d (Novagen) using *E. coli* HB101 as a host [33]. The resulting plasmid (pWUR286) was harvested by Miniprep (Qiagen), sequenced by Westburg genomics (Leusden, Netherlands), and used to transform *E. coli* expression strain BL21(DE3) (Novagen) containing the tRNA accessory plasmid pRIL (Stratagene). The transformants were grown in four 1 L shake-flasks containing LB medium until the optical cell density (A_{600}) reached

0.5. The cultures were then cold-shocked by placing them on ice to induce host chaperones [38]. After 30 min, the protein overproduction was started by adding 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The culture was incubated for 12 h at 37 °C after which the cells were spun down (10 min, 5,000 xg, 4 °C) and stored at -20 °C.

The selenium-labeled protein was obtained from a 1 L culture of BL21(DE3)-pRIL transformants grown overnight at 37 °C in Luria-Bertani media. The cells were spun down (8 min, 5,000 xg , 22 °C), resuspended in M9 media supplemented with Seleno-L-methionine (SeMet) (Fluka) [6], and used as an inoculum for a 5 L batch fermenter. Protein overexpression was induced by adding 1 mM IPTG, after which the cells were cultivated for 8 h at 35.5 °C with a stirrer speed of 182 rpm, before being spun down and frozen at -20 °C.

Protein purification

***Native* KdaD** - Pelleted *E. coli* cells were resuspended in 25 mM NaP_i buffer (pH 6.8) and disrupted by sonication at 0 °C. Insoluble cell material was spun down (30 min, 26,500 xg, 4 °C) and the *E. coli* supernatant was subjected to a heat-treatment of 30 min at 75 °C. Denatured proteins were removed by centrifugation (30 min, 26,500 xg, 4 °C) yielding a heat stable cell-free extract (HSCFE). This fraction was filtered (0.2 μm) and applied to a 70 ml Q-Sepharose Fast Flow (GE Healthcare) anion exchange column connected to an FPLC. Flowthrough fractions containing KdaD were collected, loaded onto a 46 ml Bio-Gel Hydroxyapatite column (Bio-Rad), and eluted by a linear gradient of 0.5 M NaP_i buffer (pH 6.8). Fractions containing the recombinant proteins were combined, and dialyzed overnight to 50 mM HEPES-KOH (pH 8.0) supplemented with 0.5 mM dithiothreitol (DTT).

***SeMet* KdaD** - The selenium-labeled protein was purified as above with the following modifications. The HSCFE (25 min, 60 °C) was loaded directly onto the Hydroxyapatite column. The recombinant protein containing fractions were dialyzed overnight to 20 mM HEPES-KOH (pH 8.0), concentrated with a 30 kDa cut-off spin filter (Vivaspin), supplemented with 1 mM DTT and 100 mM NaCl, and loaded onto a Superdex 200 HR 10/30 column (GE Healthcare). Peak fractions corresponding to the KdaD tetramer were then loaded onto a MonoQ 5/50 GL column (GE Healthcare). The vast majority of KdaD was collected in the flowthrough, effectively separated from contaminating proteins.

Protein crystallization and data collection

The native protein was crystallized by the hanging drop vapor diffusion method by mixing 1 μl of 15 mg ml^{-1} protein solution with 1 μl of crystallization solution at 20 °C. Large tetragonal-bipyramidal protein crystals up to 0.5 mm were observed after 3 weeks in conditions containing: (1) 0.1 M Na-HEPES, 0.8 M NaH_2PO_4, 0.8 M KH_2PO_4 (pH 4.8); (2) 0.1 M Tris-HCl, 2 M $(NH_4)H_2PO_4$ (pH 4.8); and after 6 to 12 weeks in condition (3) 0.1 M MES, 1.6 M $MgSO_4$ (pH 6.5). The SeMet protein crystallized in condition 2 at similar protein concentrations.

Data were collected at the peak wavelength of selenium (0.9795 Å) from a single flash-frozen (100 K) SeMet crystal to 2.1 Å resolution using a MARdtb imaging plate at the BESSY synchrotron source (beamline 14.2) of the Protein Structure Factory of the Free University of Berlin (Germany). Crystals were cryo-protected by soaking them in condition 1 supplemented with 35% glycerol (v/v). Data collection statistics are reported in Table 5.2.

Phasing, model building and refinement

A SAD dataset to 2.1 Å resolution was recorded from a SeMet-labeled crystal. Using XPREP, $\Delta|F_{ANO}|$ coefficients were extracted from the Se-SAD data, which were used to find four selenium sites with SHELXD [342]. Given the sequence, which shows five methionines per monomer, the detection of only four sites suggested the presence of only one molecule in the asymmetric unit, indicating a very high solvent content of approximately 82%. Using the four sites for phasing followed by density modification in SHELXE resulted in an excellent, readily interpretable experimental elec-

Table 5.2 - Data collection and refinement

Structure	**SeMet, phosphate**	**Apoenzyme**	**Mg^{2+}, 2-oxobutyrate**	**Ca^{2+}, 2-oxobutyrate**	**Mg^{2+}, 2,5-dioxopent.**
PDB ID	2Q18	2Q19	2Q1A	2Q1C	2Q1D
X-ray source	BESSY, beamline 14.2	Rigaku MicroMax	Rigaku MicroMax	Rigaku MicroMax	Rigaku MicroMax
Wavelength (Å)	0.9795	1.5418	1.5418	1.5418	1.5418
Space group, cell dimensions (Å)	$I4_122$, $a=b=127.8$ $c=222.4$ $\alpha=\beta=\gamma=90°$	$I4_122$, $a=b=128.9$ $c=225.1$ $\alpha=\beta=\gamma=90°$	$I4_122$, $a=b=128.8$ $c=224.4$ $\alpha=\beta=\gamma=90°$	$I4_122$, $a=b=128.8$ $c=224.7$ $\alpha=\beta=\gamma=90°$	$I4_122$, $a=b=128.3$ $c=223.5$ $\alpha=\beta=\gamma=90°$
Resol.limits (Å)	111-2.1	15-3.0	20-2.5	15-2.8	20-2.7
Unique reflections	53699	18610	31603	22992	25301
Completeness (%)	99.4 (98.7)	95.9 (99.5)	97.0 (95.7)	97.1 (99.3)	97.3 (93.4)
Redundancy	10.4	4.4	2.4	3.4	3.5
R_{sym}^{a}	0.094 (0.582)	0.116 (0.381)	0.11 (0.450)	0.123 (0.446)	0.071 (0.281)
$I / \sigma (I)$	23.1 (3.4)	9.5 (3.8)	5.7 (2.2)	8.1 (3.0)	11.4 (4.1)
Resolution range refinement (Å)	111-2.1	15-3.0	15-2.5	15-2.8	20-2.7
Reflections in ref.	50918	17680	30010	21834	24400
Protein atoms	2236	2235	2317	2236	2236
Water molecules	285	48	83	79	49
Ligand atoms	5 (phosphate ion)	Not applicable	7 (analogue) 1 (Mg^{2+}-ion)	7 (analogue) 1 (Ca^{2+}-ion)	9 (product) 1 (Mg^{2+}-ion)
R_{work}^{b}	0.195	0.208	0.220	0.224	0.210
R_{free}^{c}	0.213	0.236	0.240	0.252	0.243
Rms deviations					
Bond lengths (Å)	0.007	0.006	0.008	0.006	0.010
Bond angles (°)	1.093	0.956	1.139	0.938	1.30
Ramachandran regions (%)					
Most favored	87.7	87.2	85.0	86.2	85.1
Additional and generously allowed	11.5	11.9	14.6	13.0	14.4
Disallowed	0.8	0.8	0.4	0.8	0.4

[a] $R_{sym} = \Sigma_{hkl} \Sigma_i |I_i - \langle I_i \rangle| / \Sigma_{hkl} \Sigma_i I_i$, where I_i is the intensity of a given measurement and the sums are over all measurements and reflections. Values in parentheses refer to the highest resolution shell.

[b] $R_{work} = \Sigma ||F(obs)|-|F(calc)|| / \Sigma |F(obs)|$ for the 95% of the reflection data used in refinement.

[c] $R_{free} = \Sigma ||F(obs)|-|F(calc)|| / \Sigma |F(obs)|$ for the remaining 5%.

tron density map. ARP/wARP [294,295] was used to trace most of the main-chain, after which the sequence docking feature of XtalView/Xfit [255] was used to build the majority of the side chains semi-automatically, using the density and the experimentally determined selenium positions as a guide. Simulated annealing [43] was performed with CNS [44]. The final model was obtained by iterative cycles of rebuilding with XtalView/Xfit and refinement with Refmac 5.0 [268]. During refinement, R-factors, B-factors and difference electron density maps showed that the structure of the selenium-labeled crystals is best modeled with sulfur atoms in the methionines rather than with selenium atoms, indicating incomplete labeling of the protein. The structure was therefore refined as an unlabeled molecule. No electron density was observed for residues 85-94 in the SeMet structure. These residues include Met85, explaining why only four out of five methionines are observed. A Met-85 selenium atom could also not be observed in an anomalous difference Fourier synthesis. Towards the end of the refinement of the SeMet structure, a large, tetrahedral blob of electron density was found in a pocket on the surface of the molecule. Considering the crystallization conditions, this density was modeled as a phosphate ion.

Preparation of complex crystals

For the preparation of complexes with Mg^{2+} or Ca^{2+} and 2-oxobutyrate (2OB) or 2,5-dioxopentanoate (DOP), the phosphate from the crystallization buffer had to be removed in order to avoid precipitation with the alkaline earth metals and to allow for competition of the ligands for the active site. To this end, crystals were cross-linked using glutaraldehyde vapor as described [238]. Briefly, hanging-drop setups were opened and a sitting-drop rod (Hampton Research, Laguna Niguel, CA., USA) was placed in the reservoir. On the sitting-drop rod, 6 μl of 20-23% glutaraldehyde (Sigma) was placed and the setup was closed again and left to incubate for 1 h. A crystal was fished out with a cryoloop after crosslinking and washed three times for 30 s in freshly made drops of soaking solutions consisting of 30% polyethylene glycol 400 in 0.1 M HEPES-NaOH buffer (pH 7.5) supplemented with 0.1 M $MgCl_2$ or $CaCl_2$, and 0.1 M 2-oxobutyrate (Sigma) or 0.25 M 2,5-dioxopentanoate. Similarly, a crystal was treated with only 0.1 M of 2-oxobutyrate, without magnesium or calcium ions. After an incubation of 1 h in their respective soaking solutions, the crystals were cryo-cooled in liquid nitrogen and diffraction data were collected on a Rigaku 007HF rotating anode equipped with Osmic mirrors and a MAR345 image plate. Data were processed with XDS [176].

Computational methods

Structural homologs were identified using blastp searches against the pdb sequence database at the National Collaborative Bioinformatics Institute (NCBI), and by DALI searches against the structural database [155]. Oligomeric structure analysis was performed at the PISA (Protein Interfaces, Surfaces and Assemblies) webserver at EMBL-EBI [210]. A structure guided multiple sequence alignment of these homologous protein sequences was generated using the Expresso program [14]. PyMol was used to analyze structures and to prepare figures [87].

Gene neighborhood analyses were performed using various webserver tools: STRING at the EMBL, Gene Ortholog Neighborhoods at the Integrated Microbial Genomes server of the Joint Genome Institute, and pinned regions at the ERGO bioinformatics suite.

ACKNOWLEDGEMENTS

This work was supported by a grant from the European Union in the framework of the SCREEN project. TRMB is grateful to Prof. Ilme Schlichting (MPI-Heidelberg) for very generous support.

6

Identification of a novel α-galactosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*

Stan J.J. Brouns
Nicole Smits
Hao Wu
Ambrosius P.L. Snijders
Phillip C. Wright
Willem M. de Vos
John van der Oost

J Bacteriol (2006) **188**: 2392-2399

Abstract

Sulfolobus solfataricus is an aerobic crenarchaeon that thrives in acidic volcanic pools. In this study, we have purified and characterized a thermostable α-galactosidase from cell-extracts of Sulfolobus solfataricus P2 grown on the trisaccharide raffinose. The enzyme, designated GalS, is highly specific for α-linked galactosides, which are optimally hydrolyzed at pH 5 and 90 °C. The protein has 74.7 kDa subunits and has been identified as the gene-product of open reading frame _Sso3127_. Its primary sequence is most related to plant-enzymes of Glycoside Hydrolase family 36, which are involved in the synthesis and degradation of raffinose and stachyose. Both the _galS_ gene from _S. solfataricus_ P2 and an orthologous gene from _S. tokodaii_ have been cloned and functionally expressed in _Escherichia coli_ and their activity was confirmed. At present, these _Sulfolobus_ enzymes not only constitute a distinct type of thermostable α-galactosidases within Glycoside Hydrolase Clan D but also represent the first members from Archaea.

INTRODUCTION

Alpha-galactosidases (αGals, EC 3.2.1.22) are a widespread class of enzymes that liberate galactose from the non-reducing end of sugars. In bacteria, yeasts and fungi these enzymes are usually involved in the degradation of various plant saccharides, which can then serve as a carbon and energy source for growth. Plants synthesize α-galactosides such as raffinose and stachyose as the major energy storage molecules in leaves, roots and tubers. Moreover, these oligosaccharides have been associated with cold and desiccation tolerance of seeds [193]. Both raffinose and stachyose are produced by two specialized synthases (raffinose synthase EC 2.4.1.82 and stachyose synthase EC 2.4.1.67) that use galactinol as a galactosyl donor [296]. The oligosaccharides are degraded during seed germination by the action of two distinct types of αGal that differ in their optimal pH of catalysis. While the acid αGal type is most likely active in the acidic environment of the vacuole and the apoplasm, the alkaline αGal type probably catalyses galactose release in the more neutral or alkaline cytoplasm [53,132,223]. Mammals express an αGal and an alpha-*N*-acetylgalactosaminidase (α-NAGal) in lysosomal bodies to degrade glycolipids, glycoproteins and oligosaccharides. In humans, mutations in the X-chromosomal αGal gene can lead to an accumulation of these α-linked galactosides in tissues, which results in a recessive disorder called Fabry disease [113]. Mutations in the related α-NAGal gene, which is located on chromosome 22, lead to either Schindler or Kanzaki disease.

Glycoside Hydrolases (GHs) have been classified into families based on their primary sequence similarities [153]. Except for some rare cases [229], the majority of αGals can be found in GH Clan D (GH-D), which comprises family 27 and 36. The two families share a fairly conserved catalytic domain and hydrolyze the glycosidic bond with retention of configuration of the liberated D-galactose [42]. Substantial insight into the molecular mechanisms of substrate recognition and catalysis of these enzymes has been acquired when the crystal structures of the human, rice and a fungal αGal [127,133,142], as well as the chicken α-NAGal [134] were obtained.

In this study, we have purified an unusual intracellular αGal from the hyperthermophilic crenarchaeon *Sulfolobus solfataricus* P2, an aerobic micro-organism that lives in terrestrial volcanic pools of high acidity (75-85 °C, pH 2-4) [36].

RESULTS

Purification and identification of the αGal from *S. solfataricus*

Sulfolobus solfataricus P2 is capable of using α-linked galactosides such as melibiose and raffinose as a sole carbon and energy source. The capability to do so not only requires the presence of a suitable sugar uptake system, but also of an intracellular αGal and an efficient metabolic pathway for the step-wise oxidation of the released monosaccharides. Activity assays using cell-free extracts of *S. solfataricus* grown on several sugars as well as tryptone confirmed the presence of a constitutively expressed enzyme which is capable of hydrolyzing pNPG. The αGal was purified from

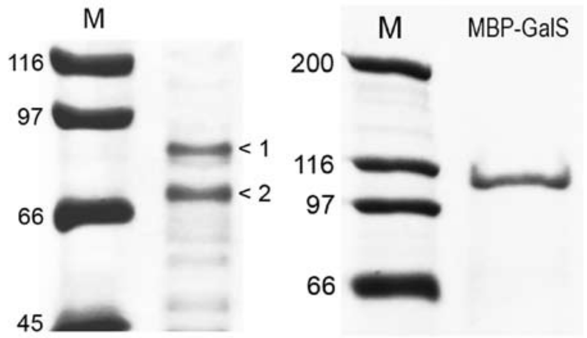


Figure 6.1
(Left) Purification of the native α-galactosidase from *S. solfataricus*. 8% SDS-PAGE gel loaded with 5 μg total protein of the α-galactosidase enriched fraction after gel filtration chromatography. The gel was stained with Coomassie Brilliant Blue G250. The sizes of the broad range protein marker are shown in kDa. Arrow 1 and 2 indicate the two proteins bands that were identified by mass-spectrometry. (Right) 6% SDS-PAGE gel loaded with 2 μg GalS fusion protein after affinity- and anion exchange chromatography.

8.8 grams of cells, which were harvested from a 3.6 L late exponential phase culture grown on raffinose. In four subsequent chromatographic steps (Q-Sepharose, Hydroxyapatite, MonoQ and Superdex) the specific activity was enriched 23-fold (Table 6.1). The enriched fractions were found to contain two major protein bands of approximately 73 and 87 kDa on a sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 6.1). Size exclusion chromatography showed that the αGal eluted in fractions corresponding to proteins of a native molecular mass of 225 ± 15 kDa, suggesting a trimeric oligomerization (data not shown). Separations of the αGal using columns based on hydrophobic interactions resulted in an almost complete loss of activity. Similarly, affinity purification using immobilized D-galactose (Pierce) proved unsuccessful, since the enzyme did not bind to the resin. Therefore, the protein could not be purified to apparent homogeneity. To identify their corresponding genes, both major protein bands were excised and treated with trypsin, after which the peptides were eluted from the gel-slice, desalted, concentrated and analyzed by ESI qQ-TOF tandem mass spectrometry. Twenty-one different peptides (40% sequence coverage, MOWSE score of 831) were found that matched band 1 with the gene-product of Sso2760 (cutA-5), an 84.9 kDa protein which is orthologous to the 80.5 kDa α-subunit of the aldehyde oxidoreductase of *S. acidocaldarius* [185]. In band two, 17 unique peptides were found that matched the gene-product of Sso3127 (Uniprot: Q97U94), a 74.6 kDa hypothetical protein that shares sequence similarity with raffinose synthases and seed imbibition proteins of GH36. The MOWSE score for this protein was 517 and a sequence coverage of 34% was achieved. This protein was likely to be responsible for the thermostable αGal activity, and was termed GalS.

Characteristics of native GalS

Substrate specificity - The enzyme was able to hydrolyze pNPG, but showed no activity towards pNP-α-substituted hexoses such as D-glucose, D-mannose, L-rhamnose or *N*-acetyl-D-galactosaminide. pNP-β-substituted

Table 6.1 - Purification of the native *S. solfataricus* αGal

Step	Protein (mg)	Activity (U)	Specific activity (U mg^{-1})	Purification factor	Recovery (%)
Cell free extract	174	31.5	0.18	1	100
Q-sepharose	49	19.5	0.4	2	62
Hydroxyapatite	7	7.8	1.16	6	25
MonoQ	3	6.4	2.31	13	20
Superdex	0.26	1.1	4.25	23	4

hexoses such as D-galactose, D-glucose and D-mannose did not support catalysis, nor did the pentose pNP-β-D-xylopyranoside. Of the tested artificial substrates, only pNP-β-L-arabinopyranoside was hydrolysed by the enzyme. However, the Michaelis-Menten constant for this substrate was very high (K_m 37.4 ± 2.2 mM) compared to that of pNPG (K_m 0.08 ± 0.01 mM). Of the natural substrates tested, the αGal was most active on the trisaccharide raffinose (2.9 U mg^{-1}), followed by the disaccharide melibiose (2.6 U mg^{-1}) and tetrasaccharide stachyose (1.2 U mg^{-1}).

Catalytic and stability properties - Activity assays in a range of pH values indicated that the enzyme is more than 50% active between pH 4.1 and 6.7, and that it has a sharp optimum at pH 5.0 (Fig. 6.2A). When the assay temperature was varied, optimal αGal activity was found at 90 °C, while the enzyme was 50% active at a more physiological temperature of 75 °C (Fig. 6.2B). GalS was completely inactive at temperatures below 50 °C. At its optimal temperature for catalysis, the enzyme showed a half-life of 30 min (Fig. 6.2C). At 70 or 80 °C, the enzyme did not show any significant decrease in activity during 2.5 h of incubation.

Inhibition of activity - The αGal was tested for the inactivation by divalent cations at a concentration of 10 mM. Most metal ions completely inhibited the activity of the enzyme (Ag^{2+}, Ca^{2+}, Cd^{2+}, Co^{2+}, Cu^{2+}, Hg^{2+}, Mn^{2+}, Ni^{2+} and Zn^{2+}), but Mg^{2+} and Mo^{2+} had no effect. Assays in the presence divalent cation chelator EDTA or reducing agent DTT did not alter the activity. Enzyme activity was also unaffected in the presence of several saccharides such as D-galactose, L-arabinose, D-fucose and sucrose up to concentrations of 20 mM.

Recombinant GalS overexpression

The *galS* gene (Sso3127, Uniprot: Q97U94) was cloned in a T7 RNA polymerase-based vector and overexpressed *E. coli* BL21(DE3)-RIL. The heat-treated soluble fraction confirmed the presence of αGal activity at 80 °C in the *galS* extract. Activity of *E. coli* transformants was also observed on selective plates containing 20 μg ml^{-1} of the chromogenic substrate 5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside (X-α-Gal) after a short incubation at 60 °C. However, the soluble expression levels were too low to conduct any detailed experimental studies. Therefore an orthologous gene from *S. tokodaii* (*galSt*, St2554, Q96XG2) was cloned and expressed, yielding an enzyme with similar characteristics and expression levels as the *S. solfataricus* αGal. Several reported strate-

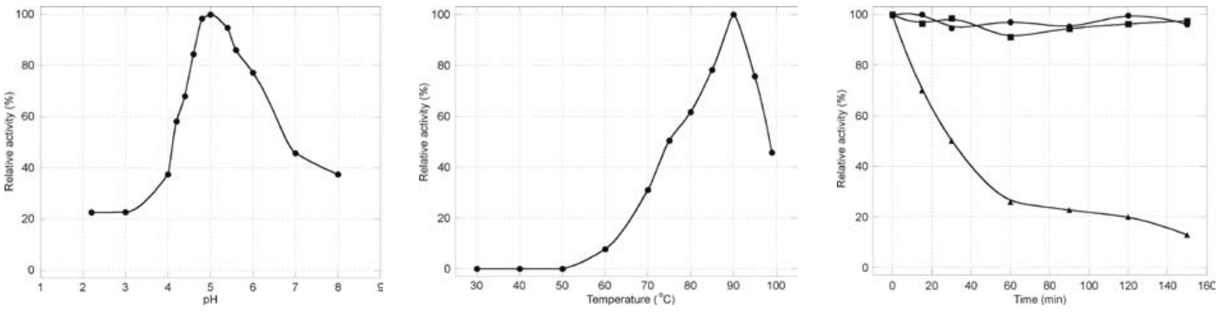


Figure 6.2
Properties of the native GalS. Relative reaction rates as a function of (left) pH and (middle) temperature. (right) Thermal inactivation curves. Residual activity is given after a pre-incubation at 70 °C (dot), 80 °C (square) and 90 °C (triangle).

gies were then employed to obtain a sufficient amount of soluble recombinant protein. These included: monitoring the protein overexpression levels in time (1, 2, 5, 8 h and overnight induction), lowering the temperature during overnight protein overexpression (15, 20, 25 °C), decreasing the IPTG concentration (0.01, 0.05, 0.1 mM), changing the cell lysis buffer, the pH and the ionic strength (HEPES-KOH, Tris-HCl, NaP_i, pH 5-10, ionic strength 0-0.5 M NaCl), but none of these strategies prevented the formation of inclusion bodies. However, since it is known that the *E. coli* maltose binding protein (MBP) promotes the solubility of proteins to which it is fused [184], this strategy was also tested. The MBP was translationally fused to the N-terminus of GalS, giving rise to a 116 kDa fusion protein. Overexpression in *E. coli* HB101 now yielded much more soluble fusion protein, which could easily be isolated using amylose affinity purification and subsequent anion exchange chromatography (Fig. 6.1).

The catalytic properties of the *E. coli*-produced GalS were comparable to those determined for the native enzyme. The Michaelis-Menten constants K_m and V_{max} for hydrolysis of pNPG were 0.085 ± 0.009 mM and 48.3 ± 1.1 $U \cdot mg^{-1}$, respectively. The catalytic efficiency of the *E. coli*-produced GalS was 703.6 $s^{-1} \cdot mM^{-1}$.

Phylogeny of αGals and prediction of their catalytic amino acids

GHs and transglycosidases are classified into families based on their sequence similarities [153]. Family 27, which mainly contains α-galactosidases (EC 3.2.1.22) and α-*N*-acetylgalactosaminidases (EC 3.2.1.49) is related to family 36, which additionally consists of stachyose synthases (EC 2.4.1.67) and raffinose synthases (EC 2.4.1.82). Both families make up GH Clan D (GH-D) and their catalytic domain adopts a common $(\beta\alpha)_8$-barrel fold [153]. This prediction was recently confirmed by the elucidation of 4 structures of members of GH27 [127,133,134,142]. In order to analyze the diversity of both families and to position the *Sulfolobus* αGal sequences in a phylogenetic tree, we aligned the catalytic domain of 43 representative members of GH-D (not shown). A phylogenetic tree was constructed from this alignment which is shown in Figure 6.3. The tree indicates that α-galactosidases of clan GH-D can be divided in at least three major types; the eukaryal type (GH27), the bacterial type (GH36b), and a type consisting of mainly plant enzymes (GH36p). The latter type additionally comprises uncharacterized sequences from Archaea, intestinal bacteria and a fungus. The two *Sulfolobus* αGal sequences form the deepest branch of this plant subfamily, and as yet, constitute the only Archaeal sequences of clan GH-D. Moreover, the *Sulfolobus* sequences are clearly distinct from other thermostable αGals that are produced by thermophilic bacteria, such as those from the genera of *Thermotoga* (Uniprot: O33835) [228], *Thermus* (Q746I3) [124] and *Geobacillus* (Q9LBD1) [123].

The catalytic domain of GH-D members comprises only 241 (amino acid 225 to 466) out of the 648 amino acids of the total GalS sequence. It is encompassed by a 26 kDa N-terminal domain and a 21 kDa C-terminal domain of unknown function. These extra domains are partly shared by members of GH36p (PF05691: raffinose synthase domain) and not

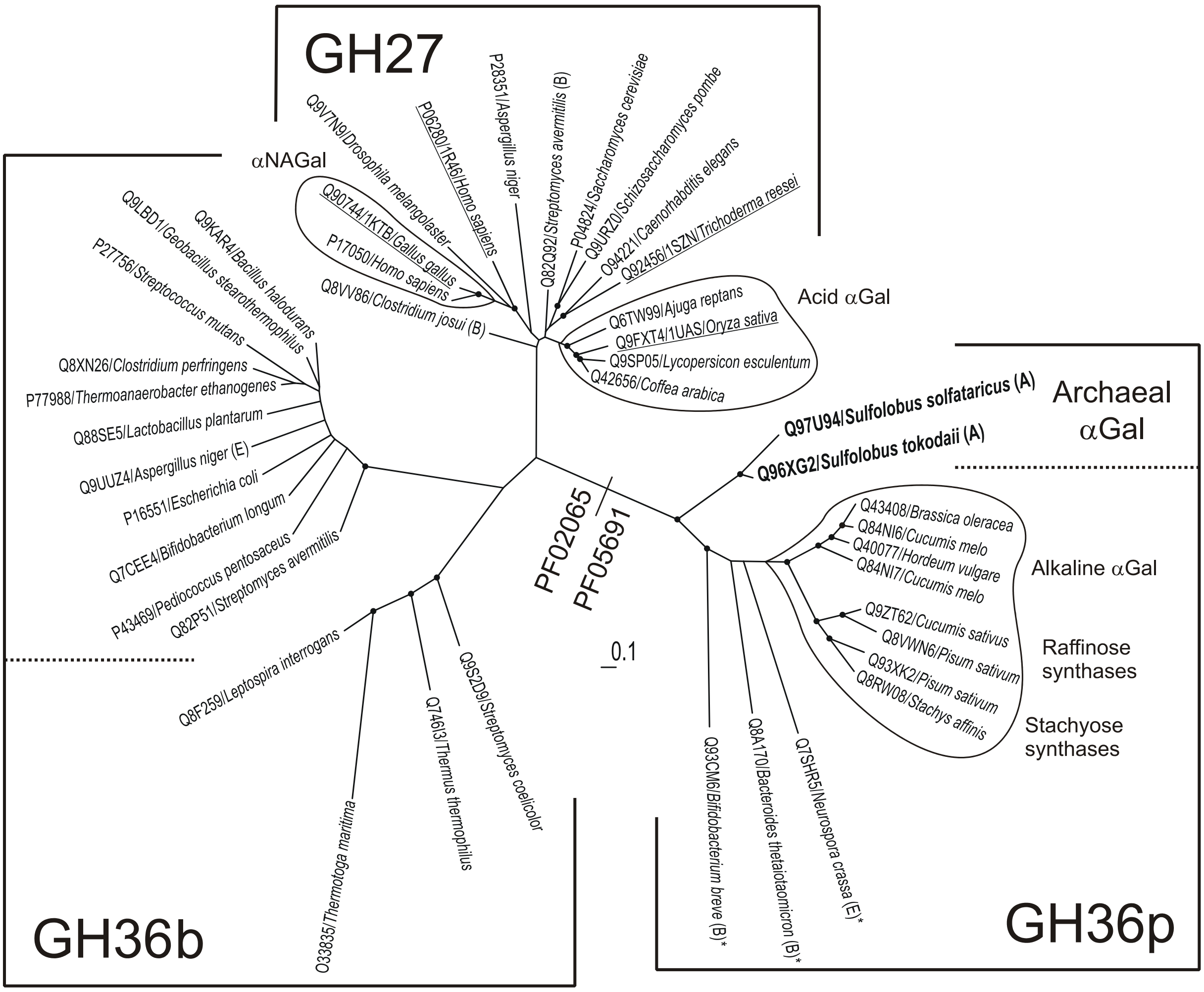

GH27
αNAGal
Q9V7N9/Drosophila melangolaster
P06280/1R46/Homo sapiens
P28351/Aspergillus niger
Q82Q92/Streptomyces avermitilis (B)
P04824/Saccharomyces cerevisiae
Q9URZ0/Schizosaccharomyces pombe
O94221/Caenorhabditis elegans
Q92456/1SZN/Trichoderma reesei
Q90744/1KTB/Gallus gallus
P17050/Homo sapiens
Q8VV86/Clostridium josui (B)
Acid αGal
Q6TW99/Ajuga reptans
Q9FXT4/1UAS/Oryza sativa
Q9SP05/Lycopersicon esculentum
Q42656/Coffea arabica
Q9LBD1/Geobacillus stearothermophilus
Q9KAR4/Bacillus halodurans
P27756/Streptococcus mutans
Q8XN26/Clostridium perfringens
P77988/Thermoanaerobacter ethanogenes
Q88SE5/Lactobacillus plantarum
Q9UUZ4/Aspergillus niger (E)
P16551/Escherichia coli
Q7CEE4/Bifidobacterium longum
P43469/Pediococcus pentosaceus
Q82P51/Streptomyces avermitilis
Q8F259/Leptospira interrogans
O33835/Thermotoga maritima
Q746I3/Thermus thermophilus
Q9S2D9/Streptomyces coelicolor
GH36b
PF02065
PF05691
0.1
Q97U94/Sulfolobus solfataricus (A)
Q96XG2/Sulfolobus tokodaii (A)
Archaeal αGal
Q43408/Brassica oleracea
Q84NI6/Cucumis melo
Q40077/Hordeum vulgare
Q84NI7/Cucumis melo
Alkaline αGal
Q9ZT62/Cucumis sativus
Q8VWN6/Pisum sativum
Q93XK2/Pisum sativum
Q8RW08/Stachys affinis
Raffinose synthases
Stachyose synthases
Q93CM6/Bifidobacterium breve (B)*
Q8A170/Bacteroides thetaiotaomicron (B)*
Q7SHR5/Neurospora crassa (E)*
GH36p

Figure 6.3
Unrooted neighbor-joining tree of the catalytic domain of a representative set of α-galactosidases belonging to GH-D. Nodes with bootstrap probability >70% are indicated by dots. The scale bar indicates an evolutionary distance of 0.1 substitutions per position. Abbreviations: αGal alpha-galactosidase, αNAGal α-N-acetylgalactosaminidases, E eukaryal sequences, B bacterial sequences, A archaeal sequences. Asterisks indicate uncharacterized sequences.

by GH27 or GH36b members (PF02065: melibiase domain).

Figure 6.4 depicts the invariant aspartic acid residues that are predicted to be involved in catalysis as either the nucleophile or the acid/base (A/B) of the double displacement reaction mechanism of α-galactosyl hydrolysis. The prediction is based on the alignment combined with experimental and structural evidence for the nucleophile and A/B which is available for GH27 [127,133,134,142,150,239]. While the sequence motif for the catalytic nucleophile is fully conserved within GH-D (K(Y/V/L/W)**D**, catalytic Asp in boldface), the motif comprising the A/B aspartic acid is much less conserved (RX$_3$**D**). The mesophilic bacterial αGal group seems to deviate most from the main A/B consensus, but appears to contain another motif (**D**X$_2$D), which is mutually shared with GH27. Yet another motif that is strongly conserved among αGals comprises two invariant aspartic acids that are involved in substrate binding of the C4- and C6-OH group of D-galactose [127,142]. These residues have long been wrongfully assumed to be involved in catalysis as the A/B residues. Interestingly, members of GH27 and GH36 can be distinguished on the bases of either a DDC or a DDG motif at this position, respectively. The cysteine of GH27 is fully conserved, since it is involved in the formation of a disulfide bond within the active site.

Mutation of predicted catalytic aspartates

The predicted nucleophilic and A/B aspartic acids of GalS were changed into glycines, yielding mutant Asp367Gly and Asp425Gly, respectively. The wild-type and both mutant enzymes were produced and purified. Activity assays using various substrate concentrations indicated that the activity of mutant Asp367Gly was below the detection limit (<1x10^{-3} of the wild-type), whereas mutant Asp425Gly showed approximately 5x10^{-3} of the activity of the wild-type (data not shown). Molar concentrations of sodium-azide or sodium-formate were unable to restore or increase the activity of the mutant enzymes (data not shown).

DISCUSSION

In the present study we have identified a novel thermostable αGal in cell-extracts of the hyperthermophilic crenarchaeon *Sulfolobus solfataricus* and studied the biochemical properties of the enzyme. Sequence analysis revealed an unusual phylogenetic position within the widely distributed class of GH-D αGals.

Properties of GalS

GalS and GalSt belong to the most thermoactive αGals known to date. Despite the fact that their primary sequence is very different from the α-galactosidases from thermophilic bacteria, their catalytic properties, such as the optimal pH of 5 and the optimal temperature of catalysis of 90 °C, are similar. Interestingly, most related enzymes belonging to GH36p have neutral or slightly alkaline pH optima's. This might be due to the fact that these enzymes contain an aspartic acid residue that is in juxtaposition to the predicted A/B residue, which may cause an increase in the pK$_a$ of the A/B. Catalysis can only proceed when the correct protonation state of the A/B is achieved, which may thus be at slightly higher pH. Since GalS contains an isoleucine residue at the corresponding position, its optimum activity may therefore be at slightly acidic pH.

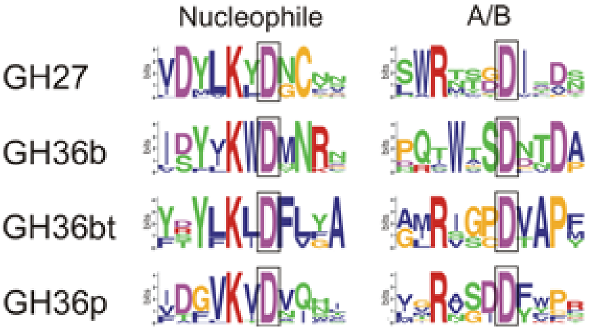


Figure 6.4
Weblogo representations of the predicted catalytic amino acid sequence motifs of the different classes of GH-D. Character size indicates the relative occurrence of amino acids in the sequences encompassing the catalytic aspartic acids (boxed). The thermophilic bacterial sequences from the *Thermus* and *Thermotoga* subgroup are indicated by GH36bt, the other groups are defined as depicted in Figure 6.3.

GalS is specific for α-linked D-galactosides and β-linked L-arabinosides. However, the Michaelis-Menten constant K$_m$, which is roughly the inverse measure of the enzyme's substrate affinity, is approximately 460 times lower for pNPG than for pNP-β-L-arabinoside, which probably excludes a physiological role of GalS as an arabinosidase. The activity on β-linked L-arabinoside is not surprising since it has an identical configuration of the hydroxyl groups at C1, C2, C3 and C4, but lacks the CH_2OH at its C5 atom in comparison with galactose [89]. Activity assays using the natural substrates melibiose, raffinose and stachyose indicated that the enzyme can release D-galactose from oligosaccharides up to at least four sugar moieties, with the maximum catalytic rates on the trisaccharide. Several related alkaline αGals were reported to prefer the longer raffinose-type oligosaccharides [53,132,223]. The enzyme showed no product inhibition, but it was severely inhibited by several divalent metal ions. This strong inhibitory effect of transition metal ions, such as Zn^{2+}, Mn^{2+} and Hg^{2+} was also observed with the raffinose synthase from *Vicia faba* seeds [224]. Similarly, the raffinose synthase from *Pisum sativum* (Uniprot: Q8VWN6) was only 30% active after 1 h of incubation with 1 mM $NiCl_2$ [296]. This finding suggests the presence of an oxidizable group, such as a free cysteine, that is required for substrate binding or catalysis.

Mechanism of catalysis

Based on sequence similarity with GH27, GH36 members are assumed to hydrolyze the glycosidic bond with retention of configuration of the released D-galactose [42]. Although this assumption has never been experimentally verified, it is highly likely that GH36 enzymes are retaining instead of inverting enzymes, since only the covalent enzyme-galactose intermediate of the retaining mechanism will allow transglycosylation reactions to occur. The raffinose and stachyose synthases of GH36p make use of this principle by coupling the galactose-moiety of galactinol to an incoming sucrose or raffinose molecule in the second stage of the reaction, respectively [296].

Mutagenesis of the proposed nucleophilic and A/B aspartic acids of GalS gave rise to similar protein yields, indicating properly folded proteins, but virtually inactive enzymes. Although many mutations may have a near lethal effect on enzymatic activity, this finding combined with multiple sequence alignment data, suggests that both substituted aspartic acids are involved in α-galactosyl hydrolysis. Unfortunately, reactivation experiments with external nucleophiles such as azide and formate failed to restore the activity of either the nucleophile or the A/B mutant enzymes. While retaining β-glycosidases can sometimes be reactivated up to wild-type levels [265], for unknown reasons this strategy seems to be far less efficient for retaining α-glycosidases [62,85].

Catalytic residues of glycosidases can also be identified by an approach that consists of covalent modification of either the nucleophile or A/B, followed by mass-spectrometric analysis of tryptic digests of the modified enzyme. This method requires specifically designed enzyme inhibitor molecules that create a stable bond between the enzyme and inhibitor. Using this strategy, the nucleophilic aspartic acid of the GH27 α-galactosidases from *Phanerochaete chrysosporium* and *Coffea arabica* (Uniprot: Q42656) was identified [150,239].

Metabolism of α-linked galactosides

The ability to use α-linked galactosides, such as melibiose and raffinose as a sole carbon and energy source implies that an organism should at least have a suitable sugar uptake system, an αGal and an efficient catabolic pathway for the liberated monosaccharides. Within the *Sulfolo-*

bus genus, the αGal described here was only found in the completed genome sequences of *S. solfataricus* [353] and *S. tokodaii* [189] and not in *S. acidocaldarius* [57]. This observation correlates well with the ability of *S. solfataricus* and *S. tokodaii*, but not *S. acidocaldarius*, to grow on melibiose and raffinose and also the monosaccharide D-galactose [144,388]. In addition, cellular growth on raffinose and D-galactose was also observed with *S. shibatae*, *S. yangmingensis* and *S. tengchongensis*, which may implicate that these species do have an αGal gene in their genome as well [144,168,437].

To date, two main catabolic pathways for D-galactose have been reported. Most ubiquitous is the Leloir pathway, in which galactose is converted to glucose-1-phosphate by the action of a sugar kinase, uridylyltransferase and an epimerase [121]. Galactose can also be degraded according to a scheme that was originally proposed by De Ley and Doudoroff [83]. In this pathway, D-galactose is oxidized to D-galactonate, subsequently dehydrated to 2-keto-3-*deoxy*-D-galactonate, phosphorylated at the C6 position and cleaved by an aldolase to yield pyruvate and D-glyceraldehyde-3-phosphate. *S. solfataricus* was recently demonstrated to use this pathway by employing a promiscuous enzyme set that can convert both D-galactose and D-glucose as well as their derivatives [213,215]. Since the aldolase can convert non-phosphorylated substrates, the pathway may also occur non-phosphorylative *in vivo*.

EXPERIMENTAL PROCEDURES

All chemicals were of analytical grade and purchased from Sigma, unless stated otherwise. Primers were obtained from MWG Biotech AG (Ebersberg, Germany). Polymerase chain reactions were performed with *Pfu* TURBO polymerase (Stratagene). The chromogenic substrate 5-Bromo-4-Chloro-3-indolyl-α-D-galactopyranoside (X-α-Gal) was purchased from Glycosynth (Warrington, UK).

Growth of *Sulfolobus* species

S. solfataricus P2 (DSM1617) and *S. tokodaii* (JCM10545) were grown aerobically at pH 3.5 in a rotary shaker at 80 °C. The medium contained: 2.5 g L^{-1} $(NH_4)_2SO_4$, 3.1 g L^{-1} KH_2PO_4, 203.3 mg L^{-1} $MgCl_2 \bullet 6\ H_2O$, 70.8 mg L^{-1} $Ca(NO_3)_2 \bullet 4\ H_2O$, 2 mg L^{-1} $FeSO_4 \bullet 7\ H_2O$, 1.8 mg L^{-1} $MnCl_2 \bullet 4\ H_2O$, 4.5 mg L^{-1} $Na_2B_4O_7 \bullet 2\ H_2O$, 0.22 mg L^{-1} $ZnSO_4 \bullet 7\ H_2O$, 0.06 mg L^{-1} $CuCl_2 \bullet 2\ H_2O$, 0.03 mg L^{-1} $Na_2MoO_4 \bullet 2\ H_2O$, 0.03 mg L^{-1} $VOSO_4 \bullet 2\ H_2O$, 0.01 mg L^{-1} $CoCl_2 \bullet 6\ H_2O$ and was supplemented with 3 g L^{-1} carbon source and Wollin vitamins.

The Wollin vitamin stock (100x) contained 2 mg L^{-1} D-Biotin, 2 mg L^{-1} Folic acid, 10 mg L^{-1} Pyridoxine-HCl, 10 mg L^{-1} Riboflavin, 5 mg L^{-1} Thiamine-HCl, 5 mg L^{-1} Nicotinic acid, 5 mg L^{-1} DL-Ca-Pantothenate, 0.1 mg L^{-1} Vitamin B12, 5 mg L^{-1} p-Aminobenzoic acid, 5 mg L^{-1} Lipoic acid.

Purification of the α-galactosidase from *S. solfataricus* P2 extracts

A 3.6 L culture of *S. solfataricus* P2 was grown on raffinose to an A_{600} of 0.8, after which the cells were centrifuged (10 min, 6,000 xg, 4 °C). The cell-pellet (8.8 g wet weight) was resuspended in 10 ml 50 mM Tris-HCl (pH 7.5) and frozen at -80 °C. After thawing, the cells were sonicated and diluted 1:1 in the same buffer that contained 1 mM Phenylmethylsulphonylfluoride (PMSF). The suspension was clarified by centrifugation (60 min, 19,000 xg, 4 °C). All subsequent chromatographic steps were performed on an ÄKTA FPLC system (Amersham Biosciences) at room temperature.

First, the supernatant was applied to a 70 ml Q-Sepharose Fast Flow (Amersham Biosciences) anion exchange column which was equilibrated with 50 mM Tris-HCl (pH 7.5). After extensive washing, the proteins were eluted by a linear gradient of buffer with 0.5 M NaCl. Active fractions were pooled and dialyzed overnight against 10 mM NaP_i buffer (pH 6.8). The dialyzed fraction was then loaded onto a Hydroxyapatite column CTH5-I (Bio-Rad) and eluted by a linear gradient of 500 mM NaP_i buffer (pH 6.8). Fractions that contained αGal activity were pooled, dialyzed overnight against 50 mM NaP_i buffer (pH 6.8), and applied to a MonoQ 5/50 GL column (Amersham Biosciences). The proteins were eluted by a linear gradient of 50 mM NaP_i buffer (pH 6.8) supplemented with 0.5 M NaCl. Subsequently, the pooled MonoQ fractions were purified by gel-filtration chromatography using a Superdex 200 HR 10/30 column (Amersham Biosciences) and 50 mM NaP_i buffer (pH 6.8) supplemented with 100 mM NaCl.

Protein identification

The protocol to identify the αGal was slightly modified from Snijders *et al.* [369]. Coomassie stained protein bands were excised with a scalpel from SDS-PAGE gel and transferred to low adhesion Eppendorf tubes. Next, they were destained twice with 200 mM NH_4HCO_3 in 40% acetonitrile (ACN) for 30 min at 37 °C. After this, the gel pieces were shrunk with ACN, dried in a vacuum centrifuge and the proteins were reduced with 10 mM DTT (30 min, 56 °C) and alkylated with 50 µl 55mM Iodoacetamide in 50 mM NH_4HCO_3 (20 min, room temperature, in the dark). After this, gel pieces were washed with 50 mM NH_4HCO_3 and shrunk with ACN. Next, the trypsin solution was added, consisting of 8 µg of trypsin in 100 µl of 9% ACN and 50 mM NH_4HCO_3 (overnight, 37 °C). The next day, peptides were extracted in four sequential extraction steps: (1) 50 µl of 50 mM NH_4HCO_3 (10 min, room temperature), (2) 75 µl of ACN (15 min, 37 °C), (3) 75 µl of 5% formic acid (FA). (4) 75 µl of ACN (15 mins, 37 °C). Extracts were pooled, dried down to completeness in a vacuum centrifuge, and peptides were redissolved in 0.1% FA and 3% ACN. The peptide mixture was separated on a PepMap C-18 RP capillary column (LC Packings, Amsterdam, Netherlands) and eluted directly onto a QStarXL electrospray ionisation quadrupole time-of-flight tandem mass spectrometer (ESI qQ-TOF; Applied Biosystems/ MDS Sciex; www.appliedbiosystems.com). Gradients and data acquisition were set up as previously described [369,370]. Database search was performed with Mascot 2.0 (www.matrixscience.com) against the Mass Spectrometry protein sequence DataBase (MSDB). The peptide tolerance was set to 1.0 Da, the MS/MS tolerance was 0.6 Da. Carbamidomethyl modification of cysteine was set as a fixed and methionine oxidation as a variable modification. A maximum of 1 missed cleavage site by trypsin was allowed.

Protein quantitation

Protein concentrations were determined by using the Bicinchoninic acid Protein Assay (Pierce) and the Bradford assay (Bio-Rad) according to the supplied protocol.

Enzyme assays

Standard activity assays were performed with 1.0 mM *para*-nitrophenol-α-D-galactopyranoside (pNPG) in a 50 mM Citric acid – Na_2HPO_4 buffer (pH 5.0) at 80 °C. The reaction was stopped by adding 0.7 volumes of cold 1 M Na_2CO_3 after which the sample was put on ice. The amount of released *para*-nitrophenol (pNP) was measured at 420 nm and calculated using an extinction coefficient of 0.0135 μM^{-1} cm^{-1} **[228]**. One unit of activity was defined as the amount of enzyme required to convert 1 µmol of pNP per minute. Thermal inactivation assays were performed by incubating 45 µg ml^{-1} of enzyme at 70, 80 and 90 °C and drawing aliquots at regular intervals during 2.5 h. The residual activity was then determined using the standard assay. The optimal pH of the enzyme was determined using the McIlvaine citrate-phosphate buffer system **[254]** in a range of pH 2.2 to 8.0. Several pNP-substituted hexoses and pentoses were tested as a substrate at a concentration of 20 mM under standard conditions.

The activity towards the disaccharide melibiose (D-galactose-α(1,6)-D-glucose), trisaccharide raffinose (D-galactose-α(1,6)-D-glucose-(α1,β2)-D-fructose) and tetrasaccharide stachyose (D-galactose-α(1,6)-raffinose) (Fluka) was tested in a continuous assay at 80 °C using 20 mM of substrate, 0.4 mM $NADP^+$ in a 50 mM NaP_i buffer (pH 7.0). The amount of liberated D-galactose was measured by adding 20 U of the *Thermoplasma acidophilum* Glucose dehydrogenase (Sigma) and calculated using an extinction coefficient of 6.22 mM^{-1} cm^{-1} for NADPH **[228]**.

The effect of different divalent metal ions, ethylenediaminetetraacetic acid (EDTA) and dithiothreitol (DTT) was tested at a concentration of 10 mM in a standard assay. Sugars such as D-galactose, sucrose, L-arabinose and D-fucose were analysed for inhibitory effects at a concentration of 5, 10 and 25 mM.

Gene cloning and mutagenesis

The genomic fragments corresponding to Sso3127 and St2554 were PCR-amplified from genomic DNA that was prepared according to the method of Pitcher **[301]**, and the amplified genes were cloned into vector pET24d or pMAL-CT, respectively, using *Escherichia coli* HB101 as a host (Table 6.2). Mutations were introduced by employing either Excite or Quikchange protocols (Stratagene). Inserts of plasmids used in this study were sequenced by Westburg Genomics (Leusden, Netherlands).

Recombinant protein overexpression and purification

Plasmids pWUR269 and pWUR270 containing the *galS* gene from *S. solfataricus* P2 and *galSt* gene *S. tokodaii*, respectively, were transformed into *E. coli* BL21(DE3)-RIL. Transformants were grown overnight at 37 °C and used to inoculate 1 L cultures which were incubated until an A_{600} of approximately 0.8. Then the cultures were placed on ice to induce cold-shock proteins which may prevent inclusion body formation [67]. After 1 h, recombinant

Table 6.2 - Strains, plasmids and primers used in this study

Strains	**Entry code / genotype**	**Reference / Supplier**
S. solfataricus P2	DSM1617 wild-type	446
S. tokodaii strain 7	JCM 10545 wild-type	388
E. coli HB101	*F- hsdS*20 (r$_B$-, m$_B$-) *ara*-14 *galK*2 *lacY*1 *leuB*6 *mcrB mtl*-1 *proA*2 *recA*13 *rpsL*20 *supE*44 *thi*-1 *xyl*-5 (Str^R)	33
E. coli BL21(DE3)-RIL	*hsdS gal* (λcI*ts*857 *ind*1*S*am7 *nin*5 *lac*UV5-T7 gene *1*)	Novagen
Plasmids	**Description**	
pET24d	T7 RNA polymerase expression system	Novagen
pMAL-CT	N-terminal fusion-tag of *E. coli* MBP	New England Biolabs
pWUR269	*galS* (*Sso3127*) in pET24d (BspHI/NcoI – XhoI)	This study
pWUR270	*galSt* (*St2554*) in pET24d (NcoI – BamHI)	This study
pWUR271	*galS* (*Sso3127*) in pMAL-CT (BamHI – PstI)	This study
pWUR272	pWUR271 with substitution D367G	This study
pWUR273	pWUR271 with substitution D425G	This study
Primers	**Sequence (5' - 3')**[a]	**Restriction site / codon change**
galS-fw (Sso3127)	CGTGATCATGATTTGGATAGAAGACGAGAATGGG	BspHI
galS-rv (Sso3127)	GCTCACTCGAGTCATTCTATAGTAACTGGGATTCC	XhoI
galSt-fw (St2554)	GGGCGCCATGGCTATTTGGATATATGATGAAAATGGG	NcoI
galSt-rv (St2554)	GCCCGGGATCCTTACTCGATACTAACTATTTCTTCAGC	BamHI
MBP-galS-fw (Sso3127)	CGCGGATCCATGATTTGGATAGAAGACGAG	BamHI
MBP-galS-rv (Sso3127)	CCGCGCTGCAGTCATTCTATAGTAACTGGGATTCC	PstI
galS-fw D367G[a]	GTAATCAATGGGTAATTCACGC	GAT > GGT
galS-rv D367G[a]	CAACCTTAACGAGATCGAAG	GAT > GGT
galS-fw D425G[b]	GAGGAATTCTATAGGCTACGTACCCTTC	GAC > GGC
galS-rv D425G[b]	GAAGGGTACGTAGCCTATAGAATTCCTC	GAC > GGC

[a] Restriction sites are underlined, and nucleotide mismatches are indicated in boldface, [b] Excite mutagenesis primer. [c] QuikChange mutagenesis primer

protein expression was induced by adding 0.5 mM of isopropyl-β-D-thiogalactopyranoside (IPTG). The cultures were allowed to grow for 4 h at room temperature, after which the cells were centrifuged.

Plasmids pWUR271, pWUR272 and pWUR273 containing wild-type and mutant αGal genes translationally fused to the *E. coli* maltose binding protein (MBP) were transformed into *E. coli* HB101 and the protein was overexpressed as described above. Frozen cell pellets were resuspended in 20 mM Tris-HCl (pH 7.5) buffer which was supplemented with 150 mM NaCl, 10% glycerol (v/v) and complete protease inhibitors (Roche). Cell suspensions were then sonified (Branson sonifier) while keeping the sample in an ice-ethanol bath of -10 °C. Next, insoluble cell-matter was removed by centrifugation (60 min, 26,500 xg, 4 °C). Clarified cell-free extracts were then loaded onto equilibrated Amylose resin (New England Biolabs). After extensive washing with the same buffer without glycerol and protease inhibitors, the fusion protein was eluted in buffer supplemented with 10 mM of maltose. To concentrate the fusion protein and to remove contaminating maltose, the sample was diluted 5 times in NaCl-free buffer and loaded onto a MonoQ 5/50 GL column (Amersham Biosciences). Pure fusion proteins were eluted by a linear gradient of the same buffer with 1 M NaCl. The maltose binding protein was removed by thrombin cleavage according to the instructions from the manufacturer (Sigma), after which the recombinant GalS was re-isolated with a MonoQ as described above.

Sequence analysis of α-galactosidases

A total of 171 αGal sequences belonging to PF0265 and PF05691 were obtained from the Pfam database. A small subset of sequences, including the ones for which a crystal structure is available and selected members of the different subgroups of GH-D, was then aligned using the TCoffee program [275]. Additional sequences were subsequently added to the alignment using the ClustalX profile alignment mode. Extensive manual alignment editing was performed with the BioEdit software package, after which highly similar sequences were discarded. Neighbor-joining trees of 43 representative sequences were then calculated and bootstrapped in ClustalX, while correcting for multiple substitutions. Trees were drawn with the program Treeview. Sequence motifs comprising the catalytic nucleophile and acid/base were made using the Weblogo server [73].

ACKNOWLEDGEMENTS

This work was supported by a grant from the European Union in the framework of the SCREEN project. APLS thanks the University of Sheffield and the EPSRC for a scholarship. PCW thanks the UK's Engineering and Physical Sciences Research Council (EPSRC) for provision of an Advanced Research Fellowship (GR/A11311/01).

7

DNA family shuffling of hyperthermostable β-glycosidases

Thijs Kaper
Stan J.J. Brouns
Ans C. M. Geerling
Willem M. de Vos
John van der Oost

Biochem J (2002) **368**: 461-470

Abstract

The structural compatibility of two hyperthermostable family 1 β-glycosidases, *Pyrococcus furiosus* CelB and *Sulfolobus solfataricus* LacS, as well as their kinetic potential were studied by construction of a library of 2048 hybrid β-glycosidases using DNA family shuffling. The hybrids were tested for their thermostability, ability to hydrolyze lactose and sensitivity towards inhibition by glucose. Three screening rounds at 70 °C led to the isolation of three high performance hybrid enzymes (hybrid 11, 18 and 20) that had 1.5-3.5 fold and 3.5–8.6 fold increased lactose hydrolysis rates compared to parental CelB and LacS, respectively. The three variants were the result of a single crossover event, which gave rise to hybrids with a LacS N-terminus and a main CelB sequence. Constructed 3D models of the hybrid enzymes visualized that the catalytic $(\beta\alpha)_8$-barrel was composed of both LacS and CelB elements. In addition, an extra intersubunit H-bond in hybrids 18 and 20 might explain their superior stability over hybrid 11. This study demonstrates that extremely thermostable enzymes with limited homology and different mechanisms of stabilization can be efficiently shuffled to stable hybrids with improved catalytic features.

INTRODUCTION

Enzyme optimization has taken off with the recent developments in laboratory evolution of proteins. Random mutagenesis combined with DNA shuffling [378], DNA family shuffling [71], iterative truncation for creation of hybrid enzymes [279] and sequence homology independent recombination [355] are several techniques that have been applied to modify or improve the performance of a wide array of biologically active proteins. A higher operational stability is one of the most desired enzyme improvements and has been found hard to achieve by rational protein design. By directed evolution, the half-life of thermal inactivation of several mesophilic proteins has been extended dramatically, often without compromising low-temperature activity [140]. Subtle structural changes were observed in the stabilized enzyme variants, which resulted in immobilization of flexible loops or optimization of long-range interactions [17,374].

Relatively few studies report on the directed evolution of extreme thermostable proteins or thermozymes. These proteins already have an extreme resistance against thermal inactivation, but often have relatively low activities at moderate temperatures. In the few known cases, single genes that encode hyperthermophilic proteins were subjected to in vitro or in vivo random mutagenesis, and enzyme variants were screened for increased activities at ambient temperatures [218,257,322]. An increase in low-temperature activity resulted from a small number of mutations, which often lead to a severe reduction in thermostability. To our knowledge, there have been no reports, yet, on the creation of enzyme hybrids by DNA family shuffling of genes from hyperthermophilic origin. This could be due to the fact that studies on the thermostability of proteins from hyperthermophiles have indicated that no general rules exist for protein thermostabilization and that even within a protein family, members may have evolved distinct mechanisms to resist extreme conditions [305]. Thus, each protein seems to have been evolved individually to withstand the high temperatures of hyperthermophilic habitats [419]. In this respect it is not unlikely that family shuffling of proteins from hyperthermophiles does yield hybrid enzymes with affected stability.

Despite these uncertainties, we have attempted to generate stable, functional hybrids by DNA shuffling of two genes coding for hyperthermostable β-glycosidases. The parental enzymes are the β-glucosidase CelB from *Pyrococcus furiosus* (optimal growth temperature 100 °C) and β-glycosidase LacS from *Sulfolobus solfataricus* (optimal growth temperature 85 °C), which are among the most thermo-active members of family 1 glycosyl hydrolases with optimal temperatures for hydrolysis of 105 and 95 °C, respectively [195,300]. Thermostable family 1 glycosidases are being recognized as promising candidates for carbohydrate engineering [118,298,299,402] and several site-directed mutagenesis studies on their substrate specificity and activity engineering have been reported [69,180,265]. CelB and LacS are 53% and 56% identical at the protein and DNA level, respectively, and are similar with respect to catalytic mechanism and substrate specificity [305]. However, the molecular basis of the high thermostability appears to be different in CelB and LacS. A biochemical comparison suggested that CelB is mainly stabilized by hydrophobic interactions, while salt-bridge interactions are crucial for the stability of LacS [305]. Indeed, analysis of the crystal structures of CelB [180] and LacS [3] revealed a higher number of ion-pairs in the less stable LacS protein [221]. Besides being model enzymes for the study of enzyme stability and high-temperature catalysis, CelB and LacS have been recognized as potentially interesting biocatalysts for the hydrolysis of the milk sugar lactose (galactose β-1,4-D-glucose) to monomeric glucose and galactose [298]. However, for CelB a significant glucose inhibition was observed, whereas LacS displayed limited stability.

In the present study, we report the generation of functional glycosidase hydrolase hybrids by shuffling of the genes coding for *P. furiosus* CelB and *S. solfataricus* LacS, and it is the first study on DNA family shuffling of extreme thermostable enzymes. In order to study catalysis at sub-optimal temperatures, a library of 2048 active variants has been screened for optimized activity at 70 °C, for which hydrolysis of lactose in combination with a reduced inhibition by glucose has been taken as a model system. After three screening rounds, three high-performance variants and one variant with reduced stability were isolated, purified, and characterized in detail on a biochemical and molecular level. Using crystal structures of CelB [180] and LacS [3], 3D models of the hybrids were constructed by homology modeling. The results show that distantly related thermozymes can be shuffled into hybrids with improved catalytic features. Moreover, we found that these enzymes can tolerate a significant number of non-wild-type residues while remaining extremely thermostable.

RESULTS

Isolation of high performance hybrids

After the DNA shuffling procedure, hybrid PCR fragments were ligated in pET9d, and transformed to *E. coli* JM109(DE3). Approximately 10,000 colonies were screened for X-β-Gal hydrolyzing activity on TY-agar plates at 37 °C. About 20% of the colonies showed a blue phenotype indicating functional β-galactosidase expression. At this level the variety in functional hybrid enzymes was visible by the differences in blue color intensity between individual colonies. The enzymes were then tested for their (i) thermostability at 70 °C, (ii) lactose-hydrolyzing capacity and

Table 7.1 - Characterization of high-performance hybrids: lactose hydrolysis at 70 °C and thermostability. Residual activity is expressed as the percentage of the initial activity after a 1.5 h incubation at 70 °C. ND not determined.

Hybrid	Specific activity on lactose (U mg^{-1})		Residual activity (%)
	100 mM	300 mM	
CelB	400	581	ND
LacS	223	263	ND
2	333	540	128
3	157	396	3
4	261	562	89
5	308	456	140
6	384	553	138
7	166	310	2
8	378	540	134
9	385	549	116
10	782	1119	107
11	554	926	117
12	604	959	128
13	545	960	130
14	804	941	139
15	588	949	125
16	619	996	114
17	624	948	121
18	728	1146	139
19	32	73	3
20	727	1089	110
21	637	902	132

Table 7.2 - Kinetic parameters of wild-type and hybrid enzymes for the hydrolysis of lactose and cellobiose at 70 °C.

	Lactose K_m (mM)	k_{cat} (s^{-1})	k_{cat} K_m^{-1} ($mM^{-1}s^{-1}$)	Cellobiose K_m (mM)	k_{cat} (s^{-1})	k_{cat} K_m^{-1} ($mM^{-1}s^{-1}$)	$K_{m, lactose}$/ $K_{m, cellobiose}$
CelB[a]	59.8 ± 12.1	286 ± 15	4.78	3.2 ± 2.6	37.6 ± 3.2	11.8	18.7
LacS	37.9 ± 17.3	121 ± 13	3.19	13.5 ± 8.4	23.2 ± 2.9	1.72	2.81
11	71.8 ± 10.2	442 ± 18	6.16	11.8 ± 3.9	42.5 ± 2.7	3.60	6.08
18	88.6 ± 8.4	1022 ± 32	11.5	4.6 ± 1.8	264 ± 13	57.4	19.3
20	59.5 ± 5.5	447 ± 11	7.51	12.1 ± 2.4	60.1 ± 2.3	4.97	4.92

[a] Typical substrate inhibition was observed with cellobiose.

(iii) inhibition by glucose. The heat incubation inactivated a significant part of the variants, indicating the presence of functional, but thermolabile hybrids. One of these instable variants, hybrid 1, was studied in more detail with respect to its stability and, after sequencing, a 3D model was constructed by homology modeling (see below). In each of the two subsequent screening rounds, 10% of the hybrids were selected for re-screening, which finally yielded 20 thermostable hybrid enzymes with improved lactose hydrolyzing activity and/or a reduced inhibition by glucose. These variants were partially purified and tested at 70 °C for their activity on lactose, glucose inhibition, and stability. This more detailed characterization revealed that 11 hybrids, indeed showed significantly higher lactose-hydrolyzing activity at 70 °C compared to the parental enzymes (Table 7.1). Remarkably, all of these hybrids were specifically selected with primers for an N-terminal LacS and a C-terminal CelB sequence. Hybrid 18 was the most active variant, while hybrid 11 and 20 represented high-performance hybrids with a relatively low (11) and average (20) activity. These three hybrids were selected for further analysis.

Kinetic parameters

To obtain more insight in their altered biochemical properties, the hybrid enzymes 11, 18 and 20 were purified to homogeneity together with the CelB and LacS parent enzymes and characterized. At 70 °C the kinetic parameters of the purified enzymes were determined for the hydrolysis of lactose, cellobiose and their respective chromogenic analogs pNP-Gal and pNP-Glc (Table 7.2 and Table 7.3). At 70 °C, CelB is a little more active on the galactosides lactose and pNP-Gal, than on pNP-Glc, whereas the efficiency of cellobiose hydrolysis is relatively low. With doubled activities on galactosides compared to glucosides, LacS displays a more pronounced β-galactosidase activity than CelB, as reported previously [305]. On all tested substrates, LacS is less active than

Table 7.3 - Kinetic parameters of wild-type and hybrid enzymes for the hydrolysis of pNP-Gal and pNP-Glc at 70 °C.

	pNP-Gal K_m (mM)	k_{cat} (s^{-1})	k_{cat} K_m^{-1} ($mM^{-1}s^{-1}$)	pNP-Glc K_m (mM)	k_{cat} (s^{-1})	k_{cat} K_m^{-1} ($mM^{-1}s^{-1}$)	Glc K_i (mM)[b]	PheImGlc K_i (nM)[b]
CelB[a]	2.60 ± 0.26	298 ± 9	115	0.80 ± 0.11	206 ± 11	258	43	9.3
LacS	1.57 ± 0.21	165 ± 6	105	0.12 ± 0.02	45 ± 2	375	46	8.4
11	3.99 ± 0.35	327 ± 9	82.0	0.71 ± 0.29	235 ± 28	331	110	5.9
18[a]	5.63 ± 0.51	743 ± 13	132	0.58 ± 0.09	440 ± 21	759	88	12
20[a]	6.58 ± 0.85	412 ± 20	62.6	0.75 ± 0.13	276 ± 17	368	80	13

[a] Typical substrate inhibition was observed with pNP-Glc for CelB (> 3 mM) and hybrids 18 and 20 (> 4 mM). Kinetic parameters have been determined using activities at lower substrate concentrations. [b] Determined with pNP-Gal as a substrate.

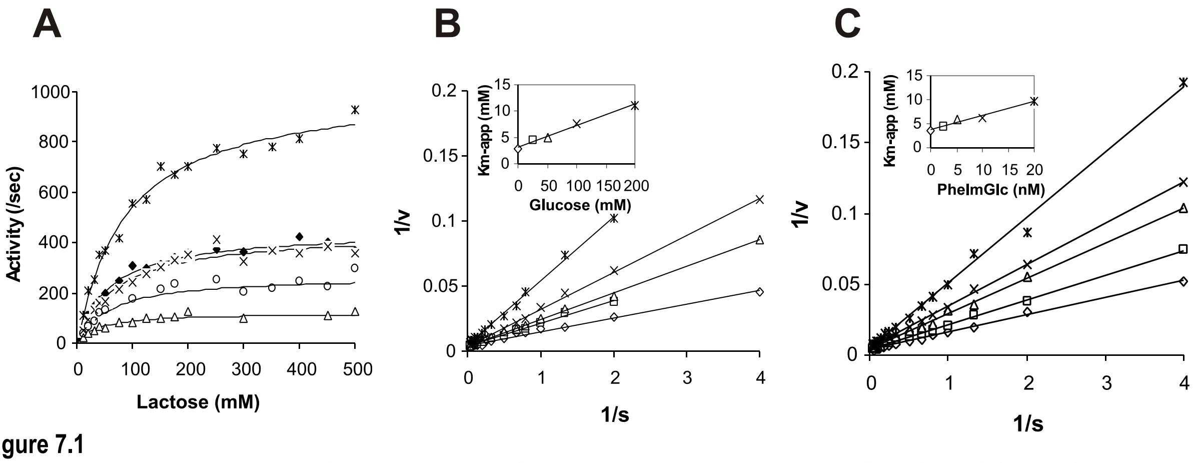


Figure 7.1
Michaelis-Menten and Dixon plots of hybrid enzymes. **A)** Michaelis-Menten kinetics of CelB (circle), LacS (triangle), hybrids 11 (cross), 18 (star) and 20 (diamond) for the hydrolysis of lactose. **B)** Inhibition of hybrid 20 by various glucose concentrations, 0 mM (diamond), 25 mM (square), 50 mM (triangle), 100 mM (cross), 200 mM (star). **C)** Inhibition of hybrid 20 by the transition state analogue PheImGlc at various concentrations, 0 nM (diamond), 2.5 nM (square), 5 nM (triangle), 10 nM (cross), 20 nM (star). Insets in **B)** and **C)** show the dependence of the K_m for glucose and PheImGlc, respectively.

the other enzymes, but has the highest affinity for all substrates except cellobiose. Compared to LacS, CelB is about twice as active on all substrates, however, with equally augmented Michaelis constants. The substrate affinities of the hybrids are similar to those of CelB. Hybrids 11, 18 and 20 have highest activity for the hydrolysis of lactose, which was the substrate used in the screening. The efficiency of lactose hydrolysis is about 1.5-3.5 and 3.6-8.6 times higher than that of CelB and LacS, respectively (Fig. 7.1A). All hybrids show increased turnover rates (k_{cat}) on cellobiose and chromogenic substrates, but only hybrid 18 has a higher efficiency for their hydrolysis ($k_{cat}\ K_m^{-1}$) compared to CelB and LacS. CelB was inhibited at higher concentrations of the substrates cellobiose and pNP-Glc. Hybrid 20 showed similar inhibition, but only with pNP-Glc. All other substrates were hydrolyzed according to Michaelis-Menten kinetics.

Inhibition by glucose and PheImGlc

The inhibition constants of glucose and the thermostable transition state analogue PheImGlc on the hydrolytic activity of CelB, LacS and hybrids 11, 18 and 20 were determined (Table 7.3). pNP-Gal was chosen as a substrate, since no substrate inhibition was observed for the hydrolysis of this compound with any of the tested enzymes. The inhibiting effect of glucose on the activity of CelB and LacS is similar; the inhibition constants for glucose are comparable, whereas the three high-performance hybrids are about 2 times less sensitive for glucose inhibition (Fig. 7.1B). This indicates that the affinity for glucose in the ground state conformation has been similarly diminished in all hybrids with respect to LacS and CelB. PheImGlc is a glucose derivative that has a flat geometry at the anomeric C1 saccharide atom and, as such, is believed to resemble the transition-state in the glycoside hydrolysis reaction in retaining β-glycosidases [286]. For all enzymes the strong inhibition by PheImGlc followed kinetics for competitive inhibition (Fig. 7.1C). The inhibition constants for CelB and LacS were similar. The K_i value determined for hybrid 11 was slightly lower than that for CelB and LacS, while the K_i-value for hybrid 18 and 20 was about 2 times the value of hybrid 11. This means that the active sites of the enzymes have subtle differences in structure, which affect the stabilization of the transition state of the reaction.

Stability

The hybrids 11, 18 and 20 were compared to CelB and LacS with respect to their thermal inactivation rate, optimum temperature for catalysis and melting temperature (Table 7.4). The different analysis methods roughly display similar trends for the enzyme variants. Hybrid 1 is the least stable variant, while the stability of hybrids 11, 18 and 20 are intermediate of that of CelB and LacS. Hybrid 20 is the most stable hybrid regarding the inactivation rate and melting temperature, while hybrid 18 has a slightly higher optimum temperature for hydrolysis (Fig. 7.2). Remarkably, LacS is readily inactivated at 92 °C, while it displays a broad optimum temperature for catalysis, with a peak around 95 °C. For CelB and hybrid 1 and 18, the determined optimal temperature for catalysis is very close to the determined melting temperature. For hybrid 11 and 20, however, the T_{opt}-value is almost 10 degrees lower than the determined T_m-value. This might suggest different inactivation patterns between the hybrids. It was not possible to determine a T_m value for LacS. This is most likely due to the fact that the unfolding of LacS does not involve two states, but rather occurs via a stable intermediate, which has been identified during guanidine hydrochloride-induced unfolding of LacS (Kaper and Van der Oost, unpublished results) [54]. In previous differential scanning calorimetric experiments with LacS, a two-state unfolding was observed, but these experiments were performed in 10 mM CAPS buffer at pH 10 [80]. Under those conditions, the melting temperature of LacS was 86.3 °C.

Table 7.4 - Thermal inactivation rate, optimal temperature of catalysis and melting temperature for the wild-type and hybrid enzymes.

Enzyme	Half-life at 92 °C (min)	T_{opt} (°C)	T_m (°C)
CelB	>> 100	104	106.0
LacS	< 3	95	ND
Hybrid 1	30[a]	80	80.7
Hybrid 11	7	85	94.0
Hybrid 18	8	96	94.5
Hybrid 20	100	93	101.8

[a] at 80 °C. ND not determined.

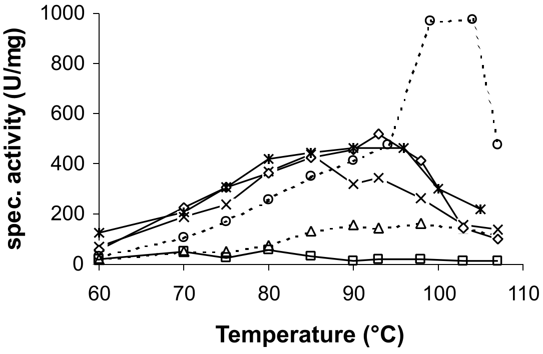


Figure 7.2
Influence of temperature on the activity of CelB, LacS and hybrids 1, 11, 18 and 20. Reactions were performed in 250 µl 20 mM pNP-Gal in 150 mM sodium citrate (pH 5.0) with 0.05 µg enzyme (CelB (circle), LacS (triangle), hybrid 1 (square), 11 (cross), 18 (star) and 20 (diamond)). Reactions were incubated for 10 min at 60 and 70 °C, 5 min at 70-85 °C, and 3 min at 90-107 °C. 1 U equals the hydrolysis of 1 µmol of pNP-Gal per min.

Structural analysis of hybrids

The primary structure of several hybrids was determined by DNA sequence analysis. All hybrids are the result of a single crossover event in different regions with 5 to 22 bases identical DNA sequence (Fig. 7.3). All sequenced hybrids contain mainly the CelB amino acid sequence in which the C-terminus or N-terminus has

A

```
CelB  387 VSHLKAVYNAMKEGADVRGYLHWSLTDNYEWAQGFRMRFGLVYVD 432
LacS  402 VSHVYQVHRAINSGADVRGYLHWSLADNYEWASGFSMRFGLLKVD 447
1     387 VSHLKAVYNAMKEGADVRGYLHWSLTDNYEWAQGFRMRFGLVYVD 432
7     387 VSHLKAVYNAMKEGADVRGYLHWSLTDNYEWAQGFRMRFGLVYVD 432
          ***   *  *   ************ ****** ** *****  **

CelB  433 FETKKRYLRPSALVFREIATQKEIPEELAHLADLKFVT--RK    472
LacS  448 YNTKRLYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRH    489
1     433 YNTKRLYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRH    474
7     433 YNTKRLYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRH    474
            **  * ****** *****   *  *  **     *   *
```

B

```
CelB  M-KFPKNFMFGYSWSGFQFEMGLPGSE-VESDWWVWVHDKENIASGLVSGDLPENGPA 56
LacS  MYSFPNSFRFGWSQAGFQSEMGTPGSEDPNTDWYKWVHDPENMAAGLVSGDLPENGPG 58
11    MYSFPNSFRFGWSQAGFQSEMGTPGSEDPNTDWYKWVHDPENMAAGLVSGDLPENGPA 58
18    MYSFPNSFRFGWSQAGFQSEMGPPGSEDPNTDWWVWVHDKENIASGLVSGDLPENGPA 58
14/20 MYSFPNSFRFGWSQAGFQSEMGTPGSEDPNTDWYKWVHDPENIASGLVSGDLPENGPA 58
         **  * ** *  *** *** ****    **  ****  * * ************
                                                ⇑ ⇑
```

Figure 7.3
N- and C-terminal amino acid sequences of CelB, LacS and hybrids. CelB sequence is underlined, whereas the corresponding regions of nucleotide sequence identity where crossovers took place, are indicated in bold and italics. **A)** C-terminal amino acid sequence of CelB, LacS, hybrid 1 and 7. **B)** N-terminal sequences of CelB, LacS and hybrids 11, 18 and 20. Differences between hybrid 11 and 20 are indicated by arrows. Sequence identity is indicated by *. The alanine residue at position 1a in the sequences is not shown for clarity reasons.

been substituted with LacS sequence (Fig. 7.3 and 7.4A). Starting at the N-terminus, hybrids 1 and 7 have 419 and 413 amino acids derived from CelB, followed by 55 and 61 residues of LacS until the C-terminus, respectively (Fig. 7.3A). Hybrid 9 was identical to wild-type CelB. Hybrid 11 and 20 are very similar. Both proteins start with a LacS sequence of about 40 residues, followed by CelB sequence until the C-terminus. The only differences between hybrid 11 and 20 are the residues at position 43 and 45 (Fig. 7.3B). Hybrid 11 has LacS residues at this position, respectively a methionine and an alanine, whereas hybrid 20, like CelB, has an isoleucine and a serine at these positions. Hybrid 14 was found to be identical to hybrid 20, while hybrid 18 was identical to LacS until residue 33, except for a point mutation that resulted in Thr23Pro, followed by 441 residues of CelB sequence (Fig. 7.3B). The gene of hybrid 19 consisted of 68 nucleotides of LacS sequence followed by the complete wild-type *celB* gene sequence.

To obtain insight in the structural characteristics of the hybrids, three-dimensional models were constructed of hybrids 1, 11, 18 and 20 by homology modeling using the available crystal structures of CelB [180] and LacS [3]. Although the resolution of the CelB and LacS structures (3.3 and 2.6 Å, respectively) do not allow for a detailed analysis of the exact positions of the amino acid side chains, they do provide a basis for the interpretation of the structural rearrangements in the chimeric enzymes. In hybrid 1, the LacS sequence replaces almost completely the CelB-CelB intersubunit contacts at the small subunit interface of the tetramer (Fig. 7.4B). The 55 amino acid residues of the LacS C-terminus are 55% identical to the C-terminus of CelB. The LacS sequence is two residues longer than in CelB and fills up the cavity that is present in the center of CelB at the tetrameric subunit interface. The C-terminus is involved in an elaborate ion-pair network in LacS, where the penultimate arginine residue reaches across and participates in a 16-residue ion-pair network that bridges the four subunits [3]. In general, ion pairs have been shown to be involved in stabilization of proteins against high temperatures [418]. In the

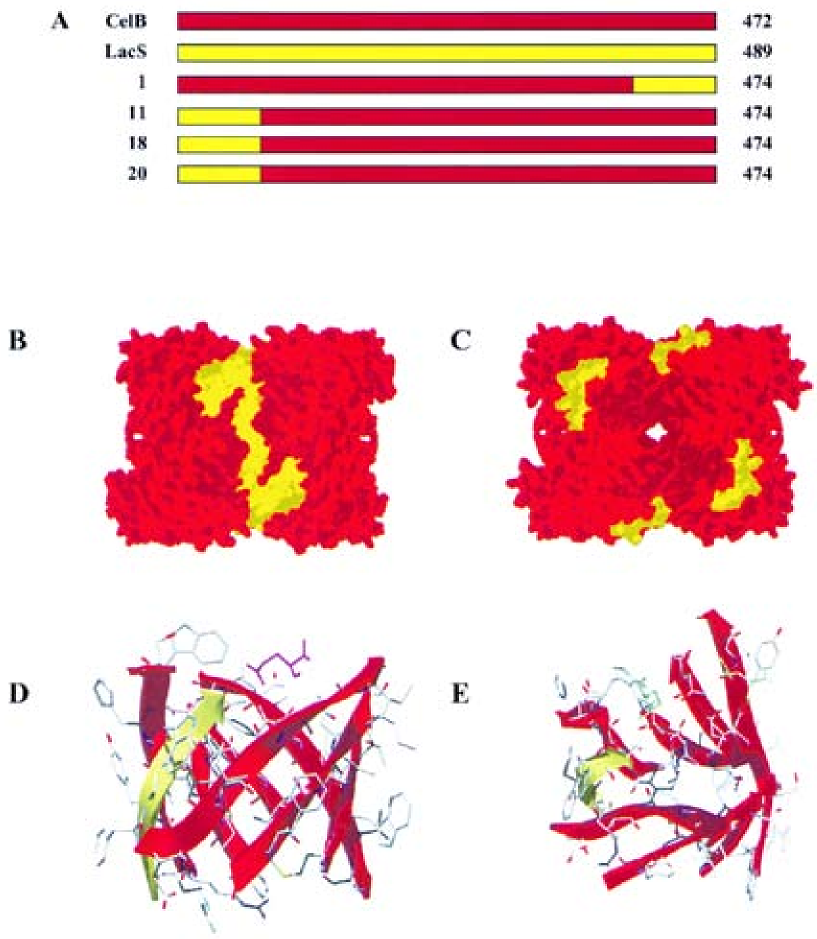


Figure 7.4 (in color on p.158) Structural representation of hybrids. **A)** Schematic representation of primary structures of CelB, LacS, hybrids 1, 11, 18 and 20. Surface plot of tetrameric **B)** hybrid 1 and **C)** hybrid 20. **D)** Side view and **E)** top view of the $(\beta\alpha)$8-barrel of hybrid 20.

C-terminus of CelB several charged residues are located that are possibly involved in an ion-pair, but align with non-charged residues in LacS and vice versa. Therefore, in hybrid 1 these possible ion pairs can not be formed. Interestingly, the ion partner of the next-to-last arginine residue in LacS, Glu345, aligns with Glu330 in CelB, which could indicate a LacS-like C-terminal organization in hybrid 1. However, the changes in hydrophobic interactions and H-bridges can not be evaluated with these models. Overall, the loss in thermostability can be interpreted as an incompatibility of LacS and CelB in the structural stabilization of the C-terminus and the subunit interface. The active site in hybrid 1 is equal to that of CelB. This could explain the catalytic profile of hybrid 1, which is similar to that of CelB (data not shown).

In contrast to hybrid 1, hybrids 11, 18 and 20 contain LacS sequences that completely traverse the subunit of the protein (Fig. 7.4C). A close up of the catalytic $(\beta\alpha)_8$-barrel visualizes that in both variants the first β-strand of the CelB barrel has been replaced by its LacS counterpart (Fig. 7.4D and Fig. 7.4E). CelB and LacS are 51% identical in the first 43 and 45 amino acid residues of the protein (Fig. 7.3B). In LacS, the N-terminus is tightly bound to the protein body by an ion-ion interaction of the α-amino group with the side chain of Asp473 [3]. Such an interaction is not possible in the hybrids, where an alanine (Ala 1a) has been introduced after the N-terminal methionine residue to enable a translational fusion of the *lacS* gene to the T7 promoter in a NcoI restriction site. In CelB, however, the α-amino group does not appear to make ionic interactions. The side chain of Lys2 could form an ion pair with Glu455 in CelB. The interactions of the charged residues in the LacS stretch in hybrid 11, 18 and 20 appear to be similar to those in wild-type LacS, which could explain the relatively high stability of the hybrids.

Hybrids 11 and 20 differ only at two positions, Met43 and Ala45 in hybrid 11 versus Ile43 and Ser45 in hybrid 20. Since we measured a significant higher stability for hybrid 20, this must solely result from the effects of the two different residues. Interestingly, both residues are located at the larger subunit interface of the glycosidase tetramer. Earlier studies have shown that quaternary interactions contribute largely to the stability of proteins [419]. These intersubunit interactions can be hydrophobic interactions, salt-bridges or hydrogen bonds. In modeling studies the interactions of Met43 and Ile43 in hybrid 11 and 20, respectively, appear to be very similar. However, an extra possible hydrogen bond was identified in hybrid 20, which involves the side chain of Ser45 and the backbone amide atom of Leu221, located at the adjacent subunit (Fig. 7.5A). Since hybrid 11 has an alanine at position 45 a similar interaction is absent in this variant (Fig. 7.5B). In hybrid 18 the residues at positions 43 and 45 are the same as in hybrid 20. Surprisingly, it is significantly less stable than hybrid 20. The point mutation in hybrid 18, which resulted in Thr23Pro, causes a small shift in the position of the N-terminal peptide chain in the 3D model, compared to that in hybrid 20. This shift could have a destabilizing effect on the interactions that 23 N-terminal residues have with the rest of the protein and this can explain the lower stability of hybrid 18.

Throughout family 1 of glycoside hydrolases, the residues that line the active site have been highly conserved [180]. Therefore, CelB and LacS share a nearly identical active site topology. However, the serine residue at position 14 in CelB is a serine in half of the known family 1 sequences, while an alanine is present in the other sequences, including LacS (Ala15). The hydroxyl group of this serine side

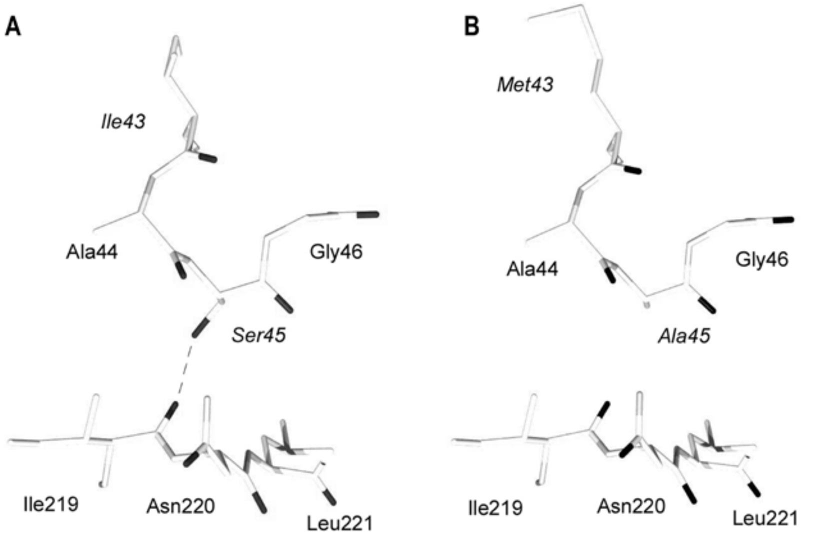


Figure 7.5
Modeled subunit interface interactions of **A)** hybrid 20 and **B)** hybrid 11 around residue 45.

chain could possibly form a hydrogen bond with the residue at the non-reducing end of a hexose substrate in the active site. Overlay studies with the structure of the 6-phospho-β-galactosidase LacG from Lactococcus lactis with a bound galactose-6-phosphate in the active site [431] showed that Ser14 might interact with the 3-hydroxyl of galactose or glucose, which have an identical orientation at this position. To investigate whether this residue was involved in substrate inhibition, Ser14 was substituted by an alanine by site-directed mutagenesis resulting in mutant CelB Ser14Ala. However, this mutant was similarly inhibited by glucose as CelB and LacS and resembled CelB in substrate specificity and activity (data not shown).

DISCUSSION

We have succeeded in generating thermostable hybrid β-glycosidases and isolation of high-performance variants. A library was constructed by shuffling of the genes encoding the hyperthermostable family 1 β-glycosidases CelB of *P. furiosus* and LacS from *S. solfataricus*. This is the first example of improving extremely thermostable enzymes by a DNA family shuffling approach. The β-glycosidase-encoding genes used in this study are 56% identical, which is relatively low compared to other shuffling studies in which 60-90% identical genes were used [56,71,197,271]. However, the combination of hybrid primer sets with extremely low annealing temperatures in the reassembly procedure did lead to the formation of a variety of hybrid genes (Fig. 7.3).

Previously, one of the parental enzymes in this study, the pyrococcal CelB, has been optimized for catalysis at room temperature by random mutagenesis in combination with DNA shuffling [218]. It was demonstrated that variants that were selected for increased activity on pNP-Glc at room temperature, were not equally more active on the β-(1,4) linked glucose disaccharide cellobiose [218]. In contrast, the three high performance hybrids that were selected in the present study for an improved hydrolysis of lactose also have increased turnover rates for hydrolysis of cellobiose, pNP-Glc and pNP-Gal (Tables 7.2 and Table 7.3). However, only hybrid 18 is more efficient than the wild-type enzymes in hydrolysis of all four tested substrates. This demonstrates the selectivity of the screening procedure: 'You get what you screen for' [440]. In the hydrolysis of pNP-Gal, the three selected hybrids were inhibited less by glucose, compared to the parental enzymes (Table 7.3), which suggests a correlation with the increased turnover rates. Judging from the increased K_m-values for the hydrolysis of pNP-Gal the affinity for the Michaelis-complex has been reduced in the hybrids with respect to the parental enzymes, which correlates with the reduced glucose inhibition. Therefore, the increase in k_{cat} might result from a less favorable binding of the ground state of the substrate, which would lower the activation energy to reach the transition state of the reaction [117]. On the other hand, the relation between the inhibition by transition-state analogue PheImGlc and the activity of the enzymes seems less clear since individual hybrids were either slightly more (hybrid 11) or less (hybrids 18 and 20) inhibited compared to the parental enzymes (Table 7.3). When the hybrids are compared, the higher affinity of 11 for PheImGlc correlates with the observed affinity for hydrolysis of pNP-Gal.

The three high-performance hybrids resulted from single crossover events near the N-terminus of the protein. Based on the determined sequences of hybrids 1, 11, 18 and 20, three-dimensional structural models were constructed using the available structures of CelB [180] and LacS [3]. The hybrids are near-identical in sequence and consequently in structure, which means that structural explanations for substrate specificity have to result from differences between residues at position 34 to 45 in the hybrids (Fig. 7.3). CelB has a more pronounced β-glucosidase character than LacS for the hydrolysis of disaccharides (Table 7.2). Interestingly, hybrid 18 has the shortest LacS sequence (residues 1-33) and resembles CelB more considering the ratio of the affinity for lactose and cellobiose (K_m), while hybrid 11 and 20 have more LacS-like specificities (Table 7.2). The peptide stretch from residue 33 to 45 is the determining factor in the substrate specificity for the hydroxyl group at C4 of the sugar residue at the non-reducing end. However, such specificity correlation can not be extended to the hydrolysis of chromogenic substrates, since these substrates have a different leaving group. For hydrolysis of disaccharides, the initial breaking of the glycosidic bond is most critical for the reaction rate due to the high pK_a-value of the leaving glucose (*i.e.* 12.4 [434]) [194]. This requires considerable distortion of the non-reducing sugar residue [81]. Since nitrophenol has a good leaving group ability (pK_a 7.2) [194], the rate-limiting step could change from glycosylation to deglycosylation of the enzyme. However, this depends on the substrate and makes it impossible to compare the kinetic data of the hydrolysis of pNP-substrates in a similar fashion as for lactose and cellobiose. Interestingly, the residues at position 33 to 45 could be best described as third or fourth shell residues and are located distantly from the active site, but nevertheless appear to affect its structure.

The three hybrids have a stability that is between that of LacS and CelB, with hybrid 20 being the most stable variant (Table 7.4, Fig. 7.2). The loss in stability compared to CelB and increase in activity could be the result of increased flexibility, although this is not an absolute requirement for improved turnover numbers at lower temperatures, as has been found in other directed evolution studies [17,140]. The structural models were analysed with respect to the hybrids' stability. The inferior stability of hybrid 1 most probably results from the loss of a relatively large number of stabilising interactions at the C-terminus. The difference in stability between hybrid 11 and 20 is intriguing since the two hybrids differ only in two residues. The 3D model shows four extra intersubunit H-bridges in the hybrid 20 tetramer, which could explain the 7 °C difference in melting temperatures between the two hybrids (Fig. 7.5). The forces that counteract conformational entropy loss and result in protein folding have a net energy gain for folded proteins that equals only a few weak intermolecular interactions [167]. Therefore, the introduction of four H-bridges might result in such a large stability increase. Mutational studies of subunit interface interactions have mainly dealt with ion-pair interactions, but similar differences in melting temperatures have been reported upon altering single residues at subunit interfaces [220,245].

This study has demonstrated that DNA family shuffling is also applicable for directed evolution of highly thermostable enzymes with different mechanisms of stabilization. We have successfully shuffled the genes coding for the extremely thermostable β-glycosidases CelB from *Pyrococcus furiosus* and LacS from *Sulfolobus solfataricus*. Hybrid enzymes were selected for thermostability and improved catalytic properties for the hydrolysis of lactose at 70°C. Composition of the hybrids showed that β-glycosidases from hyperthermophiles can tolerate significant foreign peptide stretches, while remaining very thermostable. These results have encouraged us to continue directed evolution of enzymes from extreme thermophilic origin. Laboratory evolution of enzymes remains the most promising strategy for improvement of enzymes, since the necessary structural changes for altered catalytic or stability features are too subtle to be predicted by rational design. Analysis of high-performance hybrids is likely to increase our understanding of protein stability and catalysis at elevated temperatures.

EXPERIMENTAL PROCEDURES

Library construction

For the production of *P. furiosus* β-glucosidase CelB in *E. coli* an efficient expression system has been developed by cloning the *celB* gene in the pET9d vector, resulting in pLUW511 [219]. Similarly, the *lacS* gene [74] was cloned in the pET9d vector (kan^R) (Novagen). The *lacS* gene was PCR-amplified using *Pfu* TURBO polymerase (Stratagene) and homologous primers BG745 (5'-GCGCGCCC**ATG***GCA*TACCATTTCCAGATAGCTTT-3'; introduced NcoI restriction site underlined, *lacS* start codon in bold face, introduced codon for alanine at position 1a in italics) and BG746 (5'-GCGCGCGGATCC**TTA**GTGCTTTAATGGCTTTACTG-3', introduced BamHI restriction site underlined, *lacS* stop anticodon in bold face). The PCR product was digested with NcoI and BamHI and cloned in NcoI-BamHI digested pET9d vector to form pWUR6.

Subsequently, the plasmids pLUW511 and pWUR6 were used as starting material for the DNA shuffling procedure in which *Pfu* TURBO DNA polymerase was used in all assembly and DNA amplification steps. The *celB* and *lacS* genes were amplified by PCR using the homologous primer sets BG238/BG239 [219] and BG745/BG746, respectively. Using DNaseI the genes were randomly digested as described before [234]. The digestion mixture contained 4 µg DNA and 1 µg DNaseI per 50 µl DNaseI buffer (5 mM Tris-HCl (pH 7.4), 1 mM $MnCl_2$) and was performed at room temperature in triplicate. Reactions were stopped at intervals by addition of 5 µl 0.5 M EDTA and stored on ice. DNA fragments were analyzed on a 1.5% agarose gel and fragments in the range of 50-300 bp (derived from 13-15 min incubation) were excised and purified using the QiaexII gel extraction kit (Qiagen). Performing a PCR without primers assembled the fragmented genes. The reaction mix contained 0.5 µg of each fragmented gene, 0.2 mM of each dNTP, 2.5 U *Pfu* TURBO DNA polymerase in the supplied buffer in a total volume of 50 µl. Reported assembly reactions have been carried out with annealing temperatures of 40 °C [421]. Due to the limited homology of the *celB* and *lacS* genes, however, the use of lower annealing temperatures for re-assembly was explored. Attempts with annealing temperatures of 10 °C were unsuccessful. Hybrid genes were obtained by using a $T_{gradient}$ thermocycler (Biometra, Göttingen, Germany), when the mixtures were subjected to 5 min 95 °C, followed by two temperature programs of 21 cycles. The first 21 cycles consisted of 1 min 95 °C, 1 min at 20 °C with 1 °C increase per cycle and 1 min 72 °C with 5 s increase per cycle. The second 21 cycles were 1 min 95 °C, 1 min 40 °C and 2 min 40 s at 72 °C with a 2 s increase in elongation time per cycle. To prevent amplification of wild-type genes, the shuffled constructs were enriched in PCR reactions using hybrid primer sets. The amplification mixtures contained 100 ng recombined DNA fragments, 5 pmol of each primer, 0.2 mM of each dNTP, and 2.5 U *Pfu* TURBO polymerase in the supplied buffer in a total volume of 50 µl. The temperature program was 2 min at 95 °C followed by 25 repeats of 30 s at 95 °C, 1 min at 58 °C, and 45 s at 72 °C with a 20 s increase per cycle, with a final step of 5 min at 72 °C. The PCR products were purified using the Qiaquick PCR purification kit (Qiagen) before digestion with NcoI and BamHI. Using T4 DNA ligase, hybrid genes were ligated in the NcoI/BamHI sites of pET9d vector. *E. coli* JM109(DE3) (*lacZ*$^-$ strain) electro-competent cells were transformed with the ligation mixture using a Electroporator 2510 (Bio-Rad) and plated on selective TY-agar (1% tryptone, 0.5% yeast extract, 0.5% NaCl, 30 µg ml^{-1} kanamycin, with 1.5% granulated agar) containing 1.6 µg ml^{-1} 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-β-gal). Colonies with complemented β-glycosidase activity showed a blue phenotype and were transferred to sterile flat bottom a 96-well microplates that contained selective TY-medium with 10% glycerol. Each plate contained negative (*E. coli* JM109(DE3)/pET9d) and positive controls (*E. coli* JM109(DE3)/pLUW511 or pWUR6). 2048 colonies were isolated and grown overnight at 37 °C while shaking. These glycerol stocks were stored at -80 °C.

Screening for lactose hydrolysis at 70 °C

For the preparation of cell-free extract (CFE), the glycerol stock plates were replicated in 96-well plates containing selective TY-medium and grown for 72 h at 37 °C while shaking. Induction of the expression system was not necessary, because leakage of the lacUV5 promoter results in a low level of constitutive T7 polymerase expression, as described previously [219]. A freezing-thawing step lyzed the cells and the cell-free extract was subjected to a heat incubation of 60 min at 70 °C, which denatured most of the *E. coli* host proteins. After overnight incubation at 4 °C the majority of the denatured proteins precipitated at the well bottom, leaving a clear CFE. Microplates containing either 200 µl L-buffer (100 mM lactose in citrate buffer (150 mM sodium citrate, pH 5.0)) or 200 µl LG-buffer (L-buffer containing 10 mM glucose) per well, were covered to prevent evaporation and preheated in a 70 °C waterbath for 15 min. In an insulated microplate holder, 5 µl CFE was transferred to the plates containing L-buffer and LG-buffer. Next, the plates were covered, manually inverted for mixing, and placed in a 70 °C waterbath. Hydrolysis was stopped after 60 min by placing the microplates in a water-ice mixture. The liberated glucose was measured by transferring 5 µl of the hydrolysis reaction to 200 µl GOD-PAP (Roche, Basel, Switzerland) glucose detection mixture in 96 well microplates. After 30 min incubation at room temperature, the developed color was measured at 492 nm in a Thermomax microplate reader (Molecular Devices, Sunnyvale, CA, USA). The amount of liberated glucose was calculated from a calibration curve. The accuracy of the GOD-PAP kit allowed for a reliable calculation of the liberated glucose in the presence of 10 mM glucose.

Initial characterization of high performance hybrids

For a more detailed analysis, hybrids were grown overnight in 15 ml selective TY-medium at 37 °C while shaking. At A_{600} 1.0, isopropylthio-β-D-galactopyranoside (IPTG) was added to final concentration of 0.5 mM. Cells were harvested and resuspended in 0.75 ml citrate buffer (pH 5.0). Cells were lyzed by sonication and CFE was incubated at 70 °C for 60 min. This yielded at least 90% pure hybrid β-glycosidase after centrifugation at 13,000 ×g for 30 min, as was judged by SDS-PAGE analysis. The hybrids were tested for their lactose-hydrolyzing activity and thermostability. For the lactose-hydrolyzing activity, samples of 0.25 µg enzyme were added to 0.5 ml of pre-heated 100 mM lactose and 300 mM lactose both in citrate buffer and incubated at 70°C for 15 min. Glucose was detected with the GOD-PAP kit. Thermostability of the hybrids was tested by incubating 50 µl of 50 µg ml^{-1} enzyme solution at 70 °C for 1.5 h and measuring the residual activity, which was divided by the initial activity to give the inactivation rate.

Purification high performance hybrids

E. coli JM109(DE3) clones harboring pLUW511, pWUR6 or pET9d derivatives with hybrid genes, were grown o/n in 125 ml selective TY medium at 37°C while shaking. Prior to harvesting, the cultures were induced for 4 h by the addition of IPTG to a final concentration of 0.5 mM. Cells were collected by centrifugation at 5,400 ×g for 10 min, resuspended in citrate buffer, and lyzed by sonication (Sonifier B12, Branson). *E. coli* proteins were denatured by a heat incubation at 70 °C for 1.5 h and pelleted by centrifugation at 45,000 ×g for 30 min. Ammonium sulfate was added to the supernatant to a final concentration of 1 M. After passage through a 0.2 µm filter (Schleicher & Schuell, Dassel, Germany), the supernatant was loaded on a phenyl-superose column coupled to an Äkta FPLC (Amersham-Pharmacia-Biotech, Uppsala, Sweden), equilibrated with 20 mM Tris-HCl (pH 8.0) buffer with 1 M $(NH_4)_2SO_4$. Proteins were eluted by a linearly decreasing $(NH_4)_2SO_4$ gradient and eluted from the column at approximately 0.2 M $(NH_4)_2SO_4$. Active fractions were pooled and dialyzed against 20 mM sodium phosphate (pH 7.5). Proteins were pure as judged by SDS-PAGE analysis. Absence of contaminating DNA was verified in wavelength scans in the range of 250-300 nm. Protein concentrations were determined by measuring the absorption at 280 nm, in which calculated specific ε_{280} values of 128,280

$M^{-1}cm^{-1}$ and 140,370 $M^{-1}cm^{-1}$ were used for CelB and LacS, respectively [137].

Enzyme activity assays

Enzymes were tested for the hydrolysis of lactose, cellobiose, para-nitrophenyl-β-galactopyranoside (pNP-Gal) and para-nitrophenyl-β-glucopyranoside (pNP-Glc) at 70 °C by discontinuous assays, as described before [182]. Under the used conditions, the ε_{405} of pNP was determined as 15.8 $mM^{-1}cm^{-1}$. Kinetic data were obtained by fitting the data to the Michaelis-Menten equation using the non-linear regression program Tablecurve 2D (Jandel Scientific Software, San Rafael, CA, USA). Inhibition constants were determined for glucose and the thermostable transition state analogue (5R,6R,7S,8S)-5-(hydroxymethyl)-2-phenyl-5,6,7,8-tetrahydroimidazol[1,2-a] pyridine-6,7,8-triol (PheImGlc, generously supplied by Dr. Andrea Vasella, ETH Zurich, Switzerland) [286] with pNP-Gal as a substrate, as described before [181]. However, after the apparent K_m values were determined, K_i values were determined in a Dixon plot.

Stability studies

Half-live values of thermal inactivation of parent and hybrid enzymes were determined at 92 °C in 20 mM sodium phosphate (pH 7.5) using 50 µg ml^{-1} enzyme, as described before [182]. Optimal temperatures for catalysis were determined by measuring specific activities at 60, 70, 75, 80, 85, 90, 95, 100, 105 and 110 °C for the hydrolysis of 20 mM pNP-Gal in citrate buffer. Below 90 °C, aliquots of 0.5 ml buffer with substrate were preheated for 5 min in a water bath, before addition of 5 µl enzyme solution. After 5 min the incubations were stopped by addition of 1 ml 0.5 M Na_2CO_3. At 90 °C and above, 250 µl buffer with substrate was preheated in 1 ml stoppered glass HPLC vials in a silicon oil bath for 5 min. Reactions were started by addition of 5 µl enzyme solution, and terminated by addition of 0.5 ml 0.5 M Na_2CO_3. Liberated nitrophenol was measured at 405 nm using a spectrophotometer (Hitachi). Melting temperatures of wild-type and hybrid enzymes were determined by differential scanning calorimetry in a VP-DCS micro calorimeter (MicroCal, Northampton, MA, USA), as described previously [181].

DNA sequencing and homology modeling

The DNA sequence of hybrid genes was determined using the Thermo Sequenase kit (Amersham-Pharmacia-Biotech, Uppsala, Sweden) with IR-labelled primers (MWG, Ebersberg, Germany) and subsequent analysis on an automated sequencer (LiCor, Lincoln, USA). Three-dimensional tetrameric models of the hybrid enzymes were obtained by molecular replacement using the crystal structures of CelB [180] and LacS (PDB ID 1GOW) as search models by SWISS-MODEL [145].

Construction of CelB Ser14Ala mutant

Using pLUW511 as a template, a Ser14Ala substitution was introduced in the *celB* gene via the QuikChange Site-Directed Mutagenesis Kit (Stratagene) using primers BG996 (5'-TATTCTTGGgCTGGTTTCCAG-3', sense, introduced mutation in lower case font) and BG997 (5'-CTGGAAACCAGcCCAAGAA-TA-3', antisense). The sequence with introduced mutation was verified by DNA sequence analysis (see above). CelB Ser14Ala protein was produced and purified as wild-type CelB.

ACKNOWLEDGEMENTS

This work was supported by the EU in contract FAIR CT96-1048. The authors thank Dr. Andrea Vasella (ETH, Zurich, Switzerland) for the kind gift of PheImGlc and Anton Korteweg for technical assistance with the DSC experiments.

8

Engineering a selectable marker for hyperthermophiles

Stan J.J. Brouns
Hao Wu
Jasper Akerboom
Andrew P. Turnbull
Willem M. de Vos
John van der Oost

J Biol Chem (2005) **280**: 11422-11431

Abstract

Limited thermostability of antibiotic resistance markers has restricted genetic research in the field of extremely thermophilic archaea and bacteria. In this study, we have employed directed evolution and selection in the thermophilic bacterium *Thermus thermophilus* HB27 to find thermostable variants of a bleomycin binding protein from the mesophilic bacterium, *Streptoalloteichus hindustanus*. In a single selection round we identified 8 clones bearing 5 types of double mutated genes that provided *T. thermophilus* transformants with bleomycin resistance at 77 °C, while the wild-type gene could only do so up to 65 °C. Only 6 different amino acid positions were altered, three of which were glycine residues. All variant proteins were produced in *Escherichia coli* and analyzed biochemically for thermal stability and functionality at high temperature. A synthetic mutant resistance gene with low GC content was designed which combined 4 of the observed substitutions. The encoded protein showed up to 17 °C increased thermostability and unfolded at 85 °C in the absence of bleomycin, whereas in its presence the protein unfolded at 100 °C. Despite these highly thermophilic properties, this mutant protein was still able to function normally at mesophilic temperatures *in vivo*. To understand the effect of the mutations, the mutant protein was co-crystallized with bleomycin and the structure of the binary complex was determined to a resolution of 1.5 Å. Detailed structural analysis revealed that increased thermostability and enhanced antibiotic binding was due to the introduction of an intersubunit hydrogen bond network, improved hydrophobic packing of surface indentations, reduction of loop flexibility and α-helix stabilization. The applicability of the thermostable selection marker is discussed.

INTRODUCTION

Despite the vast amount of protein sequences and structures from micro-organisms that grow optimally at temperatures above 80 °C, improving a protein's thermal stability is still a challenging task. This is mainly because the laws governing protein stability are not easily extracted, since they are highly variable and complex [15,419]. It seems generally accepted that the extreme stability of certain natural proteins results from the cumulative effect of small adaptations in protein architecture and amino acid composition. Although some of these stabilizing features, such as optimized surface ion pair networks [186], are unlikely to be engineered into a protein of interest, other strategies like α-helix capping [350], and the introduction of disulfide bonds and prolines in β-turns [412] can be applied very successfully when carefully designed on the basis of a high resolution crystal structure. However, in many cases atomic resolution three dimensional information of a protein is unavailable. Directed evolution approaches, by contrast, do not require any structural information, and commonly rely on random mutagenesis and recombination followed by screening or selection schemes [15,378]. Thermostability screens of mutant libraries are usually carried out by applying a thermal challenge at nonpermissive temperatures, after which the remaining functionality of the individual clones is tested [140,283]. To explore sufficient sequence space, requires the testing of large numbers of mutant clones, which necessitates high throughput approaches such as the use of robotics. Conversely, efficient selection procedures allow the testing of a large set of variants while reducing the effort of finding improved ones to a minimum.

A convenient selection system for finding protein variants in a library with improved thermostability is based on *in vivo* screening in a thermophilic expression host. Cloning and selection in thermophilic micro-organisms such as *Geobacillus stearothermophilus* (30 to 60 °C) or *Thermus thermophilus* (50 to 80 °C), mimics natural evolution, but is only applicable when the gene of interest encodes a protein that is of biological relevance to growth or survival of the host organism [227,393]. The selective pressure can be fine-tuned by raising the temperature of growth, enabling only hosts that bear thermo-adapted variants to grow on solid media. For instance, a combination of *in vitro* mutagenesis methods and *in vivo* selection schemes have led to a highly thermostable kanamycin nucleotidyltransferase gene that is able to function at temperatures up to 79 °C [156]. Such mutant selection markers have permitted the development of genetic tools which are very useful in the study of gene-function relationships in thermophilic bacteria [126].

In contrast to thermophiles (optimum temperature for growth 60 - 80 °C), antibiotic-based genetic systems for hyperthermophilic bacteria and archaea (optimum temperature for growth > 80 °C) are still in their infancy. This is primarily due to the absence of thermostable antibiotics and their corresponding resistance factors, since most known antibiotic producing micro-organisms are mesophilic bacteria and fungi. Often, the common antibiotics cannot be used, since many of them are unstable at high temperatures, or hyperthermophiles are simply insensitive to them [274]. The glycopeptide bleomycin is an exception, since it is a highly thermostable molecule and effective against many aerobic micro-organisms and eukaryotic cell-lines [135,385]. The bleomycin family of antibiotics, including phleomycin and tallysomycin, are DNA- and RNA-cleaving glycopeptides that are produced by the actinomycetes *Streptoalloteichus hindustanus* and *Streptomyces verticillus*. As little as a few hundred bleomycin molecules can effectively kill aerobic cells [302]. For this reason, bleomycin is currently clinically employed as an antitumor agent against squamous cell carcinomas and malignant lymphomas [360]. Resistance against bleomycin-like antibiotics is conferred by N-acetylation, deamidation and sequestration of the molecule [385]. The latter mechanism involves bleomycin binding proteins (BBPs) which have been found only in mesophilic Bacteria. Two proteins, Shble and BlmA, provide self-immunity for bleomycin-producers *St. hindustanus* and *S. verticillus*, respectively [135,387], and may be involved the transport and excretion of the molecule [92]. Two genes, *blmT* and *blmS*, are located on the *Klepsiella pneumoniae* transposon Tn*5* [136] and on the *Staphylococcus aureus* plasmid pUB110 [349], respectively. All four proteins are highly negatively-charged cytoplasmic proteins of around 14 kDa, which form homodimers that bind two positively charged antibiotic molecules at a hydrophobic subunit interface cleft [92,250,386]. The small protein size and the wide applicability of the drug have made both *shble* and *blmT* popular dominant selection markers in vector systems for lower and higher eukaryotes, bacteria and halophilic archaea [91,277,385]. This prompted us to investigate whether we could thermostabilize Shble and BlmS to allow for its application in aerobic thermophiles and hyperthermophiles.

In this study, we have performed directed evolution using selection in the thermophilic bacterium *T. thermophilus*, and obtained various mutant proteins which could operate at highly thermophilic growth conditions. Their enhanced performance at high temperature was analyzed biochemically and possible stabilizing effects were identified.

RESULTS AND DISCUSSION

Selection of stabilized Shble variants

Randomly mutated *shble* genes were introduced in the *E. coli* - *T. thermophilus* shuttle vector pMK18 under the control of the promoter of the surface layer protein A gene (*slpA*) from *Thermus thermophilus* HB8 [116]. This promoter is known to drive efficient transcription of the single selection marker in both bacteria. An error-prone library of approximately 20,000 functional clones was generated in *E. coli* HB101. Colonies appeared of similar size and there was no difference between the mutant and wild-type *shble* phenotype. The plasmid library was harvested and transformed into *Thermus thermophilus* HB27, making use of its high natural competence [208]. *Thermus* clones appeared on bleomycin-containing plates up to 65 °C after transformation with the wild-type *shble* shuttle vector, whereas wild-type *blmS* was unable to generate a resistant phenotype at either 50 or 65 °C. The transformation efficiency of the *shble* shuttle vector was approximately five times lower at 65 °C than the kanamycin-based vector pMK18 [82]. This difference might be due to the lethal effect of bleomycin and the non-catalytic nature of its elimination, which requires at least one protein molecule per bleomycin molecule.

Upon increasing the temperature of selection, a dramatic decrease in the number of colonies was observed after transformation of 8 μg of mutant library DNA. While 1200 colonies appeared at 67 °C, this number decreased to 800 at 69 °C, 600 at 70 °C, 106 at 75 °C and 8 at 77 °C. No colonies appeared at 78 and 80 °C. Plating efficiencies at these temperatures have been reported to be severely reduced, which complicates selection up to 85 °C, the maximum temperature of growth [156]. The eight *Thermus* clones found at 77 °C (termed 77-1 to 77-8) were grown overnight in selective media at 70 °C and their plasmids were isolated, transformed into *E. coli* HB101, and subsequently re-isolated and their inserts were sequenced. This revealed that all variants were double mutants bearing, in total, 6 different amino acid substitutions and 3 silent mutations (Table 8.1). Five types of double mutants could be distinguished at the protein level and two sets of double mutants were identical. Remarkably, 3 out of 6 mutations found were glycine substitutions, of which glycine 98 was replaced by either a valine or a serine. The fact that only double mutants were found, seems to be a clear indication of the high stringency that was used during selection. Interestingly, some substitutions, such as Leu63Gln, had occurred in combination with either Gly18Glu, Asp32Val or Gly98Val, which may point to the independent effects of the different mutations. A multiple sequence alignment of BBPs and the position of the mutations are shown in Figure 8.1. To assess the reason why these mutants performed better at elevated temperatures *in vivo*, we produced and purified wild-type Shble and all double mutants and studied their biochemical behavior *in vitro*. Furthermore, a synthetic quadruple mutant gene with low GC content was designed by combining mutation Gly18Glu, Asp32Val, Leu63Gln and Gly98Val. The protein, designated HTS (High Temperature Shble), was produced, purified and biochemically analyzed. The HTS protein was crystallized in complex with bleomycin A2 and its structure determined.

Table 8.1 - Nucleotide and amino acid changes of mutants.

	Residue number					
	18	31	32	40	63	98
wild-type	Gly	Arg	Asp	Gly	Leu	Gly
77-1/2/8					Gln T188A	Val G293T
77-3[a]			Val A95T		Gln T188A	
77-4/6[b]		Leu G92T		Ala G119C		
77-5[c]		Leu G92T				Ser G292A
77-7	Glu G53A				Gln T188A	

Additional silent mutations [a] C360A, [b] G63A, [c] G30A.

Thermal unfolding

Shble variants were subjected to temperature-induced equilibrium unfolding experiments in the presence and absence of bleomycin. The protein was found to unfold largely irreversible, since only 40% of the native folded signal was regained after slow cooling of the thermally unfolded protein. Therefore only apparent midpoint temperatures of unfolding (T_m) could be calculated. The results are summarized in Table 8.2.

In the absence of the antibiotic, wild-type Shble appears to be a very stable protein. This is remarkable, since *St. hindustanus* grows optimally at 28 °C [401]. It is often found, however,

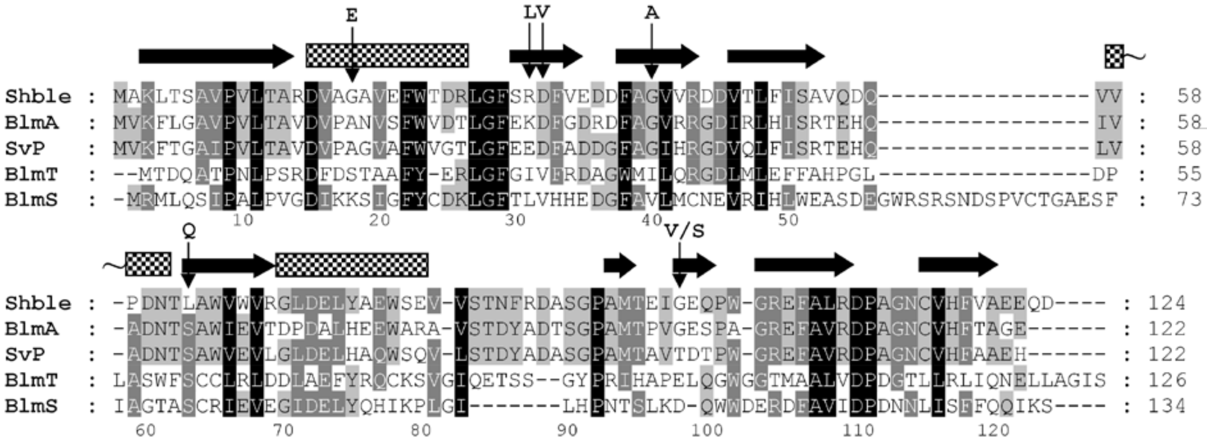

E
LV
A
Shble : MAKLTSAVPVLTARDVAGAVEFWTDRLGFSRDFVEDDFAGVVRDDVTLFISAVQDQ----------------VV : 58
BlmA : MVKFLGAVPVLTAVDVPANVSFWVDTLGFEKDFGDRDFAGVRRGDIRLHISRTEHQ----------------IV : 58
SvP : MVKFTGAIPVLTAVDVPAGVAFWVGTLGFEEDFADDGFAGIHRGDVQLFISRTEHQ----------------LV : 58
BlmT : --MTDQATPNLPSRDFDSTAAFY-ERLGFGIVFRDAGWMILQRGDLMLEFFAHPGL----------------DP : 55
BlmS : -MRMLQSIPALPVGDIKKSIGFYCDKLGFTLVHHEDGFAVLMCNEVRIHLWEASDEGWRSRSNDSPVCTGAESF : 73
10
20
30
40
50
Q
V/S
Shble : -PDNTLAWVWVRGLDELYAEWSEV-VSTNFRDASGPAMTEIGEQPW-GREFALRDPAGNCVHFVAEEQD---- : 124
BlmA : -ADNTSAWIEVTDPDALHEEWARA-VSTDYADTSGPAMTPVGESPA-GREFAVRDPAGNCVHFTAGE------ : 122
SvP : -ADNTSAWVEVLGLDELHAQWSQV-LSTDYADASGPAMTAVTDTPW-GREFAVRDPAGNCVHFAAEH------ : 122
BlmT : LASWFSCCLRLDDLAEFYRQCKSVGIQETSS--GYPRIHAPELQGWGGTMAALVDPDGTLLRLIQNELLAGIS : 126
BlmS : IAGTASCRIEVEGIDELYQHIKPLGI-------LHPNTSLKD-QWWDERDFAVIDPDNNLISFFQQIKS---- : 134
60
70
80
90
100
110
120

Figure 8.1

Structural alignment of bleomycin binding proteins from different microbial sources. Shble (PDB ID: 1BYL) from *St. hindustanus* [92], BlmA (PDB ID: 1QTO) from *S. verticillus* ATCC15003 [188], SvP from *S. verticillus* ATCC21890, BlmT (PDB ID: 1ECS) from *K. pneumoniae* transposon Tn5 [250] and BlmS from *S. aureus* plasmid pUB110. The alignment was created by backbone superimposition of the three structures and expanded with the SvP and BlmS sequences by realignment using ClustalX [398] while maintaining the original gaps. The HTS structure was used for residue numbering and topology assignment (black arrows: β-strand, checkered boxes: α-helix). Mutations are indicated by arrows.

that proteins for which low biological turnover is beneficial for a host, are prone to little local unfolding and hence are less susceptible to proteolytic attack [171]. Structurally, Shble which serves a function of self-immunity, might well be adapted to meet these criteria by its compactness, relatively high secondary structure content, high surface charge and by its embedded N and C-termini [92]. The unfolding data also clearly show the strong stabilizing effect of ligand binding on the thermostability of the BBP, since the apparent unfolding midpoint temperature increases 27.3 °C upon bleomycin binding. This effect has also been recognized in other ligand binding proteins, such as streptavidin and avidin, which become extremely thermostable in the presence of biotin [143]. In the absence of bleomycin, the stability of the various mutants is rather different. Of the double mutants, only 77-3 (Asp32Val, Leu63Gln) seems to have a marked increase in T_m as observed with circular dichroism (CD) and fluorescence spectroscopy (FS), while numbers 77-1 (Leu63Gln, Gly98Val), 77-5 (Arg31Leu, Gly98Ser) and 77-7 (Gly18Glu, Leu63Gln) remain virtually unchanged. Surprisingly, mutant 77-4 (Arg31Leu, Gly40Ala) displays significantly lower T_m values compared to the wild-type. Quadruple mutant HTS, which combines non-redundant mutations found in 77-1, 77-3 and 77-7, displays a profound increase of 13.9, 10.8 and 17.7 °C in stability in the absence of the antibiotic as found by CD, FS and differential scanning calorimetry

Table 8.2 - Apparent thermal unfolding midpoints (° C) of Shble variants.

		WT	HTS	77-1	77-3	77-4	77-5	77-7
apo	**CD**[a]	70.8 ± 0.5	84.7 ± 0.6	72.2 ± 0.5	79.5 ± 0.6	66.8 ± 0.5	71.1 ± 0.6	70.3 ± 0.5
	FS[b]	67.9 ± 0.4	78.7 ± 0.5	67.2 ± 0.4	69.1 ± 0.5	63.8 ± 0.4	64.6 ± 0.5	65.9 ± 0.4
	DSC[c]	67.4 ± 0.6	85.1 ± 0.7	ND	ND	ND	ND	ND
bleomycin	**DSC**[c]	94.7 ± 0.8	100.3 ± 0.8	ND	ND	ND	ND	ND

[a] CD circular dichroism spectroscopy (λ_{205} nm), [b] FS fluorescence spectroscopy (λ_{ex} 295 nm, λ_{em} 315 nm), [c] DSC: differential scanning calorimetry. ND not determined.

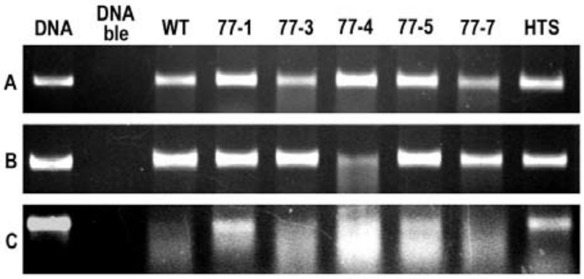


Figure 8.2
DNA protection assay. Images of 1% agarose gels showing the degree of DNA protection by Shble variants against the strand scission action of bleomycin A2. **A)** assay at 25 °C for 10 min. **B)** assay at 25 °C for 10 min after protein preincubation at 85 °C for 30 min. **C)** assay at 85 °C for 10 min.

(DSC), respectively. In its presence, the complex becomes hyperthermostable, unfolding at a temperature of just over 100 °C, 5.6 °C higher compared to the wild-type protein.

To our surprise, the double mutants 77-1, 77-5 and 77-7 had almost unchanged apparent melting temperatures compared to the wild-type. This can be understood by realizing that *in vivo*, some amino acid changes may prevent instances of local protein unfolding, and therefore may avoid further unfolding and subsequent proteolytic attack. However, this is not necessarily reflected in its *in vitro* melting temperature, which is a measure of its global stability. Only when the weakest point of a structure was compensated (Asp32Val and Leu63Gln in 77-3), an increase of its melting temperature from 70.8 to 79.5 °C with CD, and 67.9 to 69.1 °C with FS was observed. Adding mutation Gly98Val from 77-1 and Gly18Glu from 77-7 to 77-3, giving rise to HTS, further increased its melting temperature as one would expect. This observation is analogous to the findings of extensive work which has been conducted with the neutral protease from *Geobacillus stearothermophilus*, where interactions close to the N-terminus were found to be limiting the global stability [412].

Mutants improve DNA protection against bleomycin at high temperature

In vitro DNA protection assays were performed with the various Shble mutants in order to test whether the resistant phenotype of *T. thermophilus* at 77 °C was due to improved protection against the DNA degrading capability of bleomycin. The result of this is shown in Figure 8.2. At 25 °C, no significant differences in band intensities are observed. A 30 min thermal preincubation of the protein at 85 °C, however, revealed a drastic loss of function in mutant 77-4. Differences between the wild-type and mutants became pronounced when bleomycin binding capabilities were tested at 85 °C. At this temperature, the DNA was protected best by 77-1 and HTS, followed by 77-4, 77-5, 77-3, 77-7 and the wild-type. Surprisingly, mutant 77-4, which displayed a low temperature unfolding midpoint and high thermal inactivation at 85 °C, apparently bound bleomycin effectively at high temperature conditions. So although the global stability of this mutant was decreased, it had improved bleomycin binding characteristics, which in itself, stabilizes the protein dramatically as observed by DSC measurements for the wild-type. These results indicate that some of the double mutants have improved the bleomycin binding properties compared to the wild-type, which confirms the findings of the *in vivo* selection procedure in *Thermus thermophilus*. Possible structural explanations for the improved functionality at higher temperature are discussed below.

Overall structure description

The quadruple mutant HTS was crystallized in

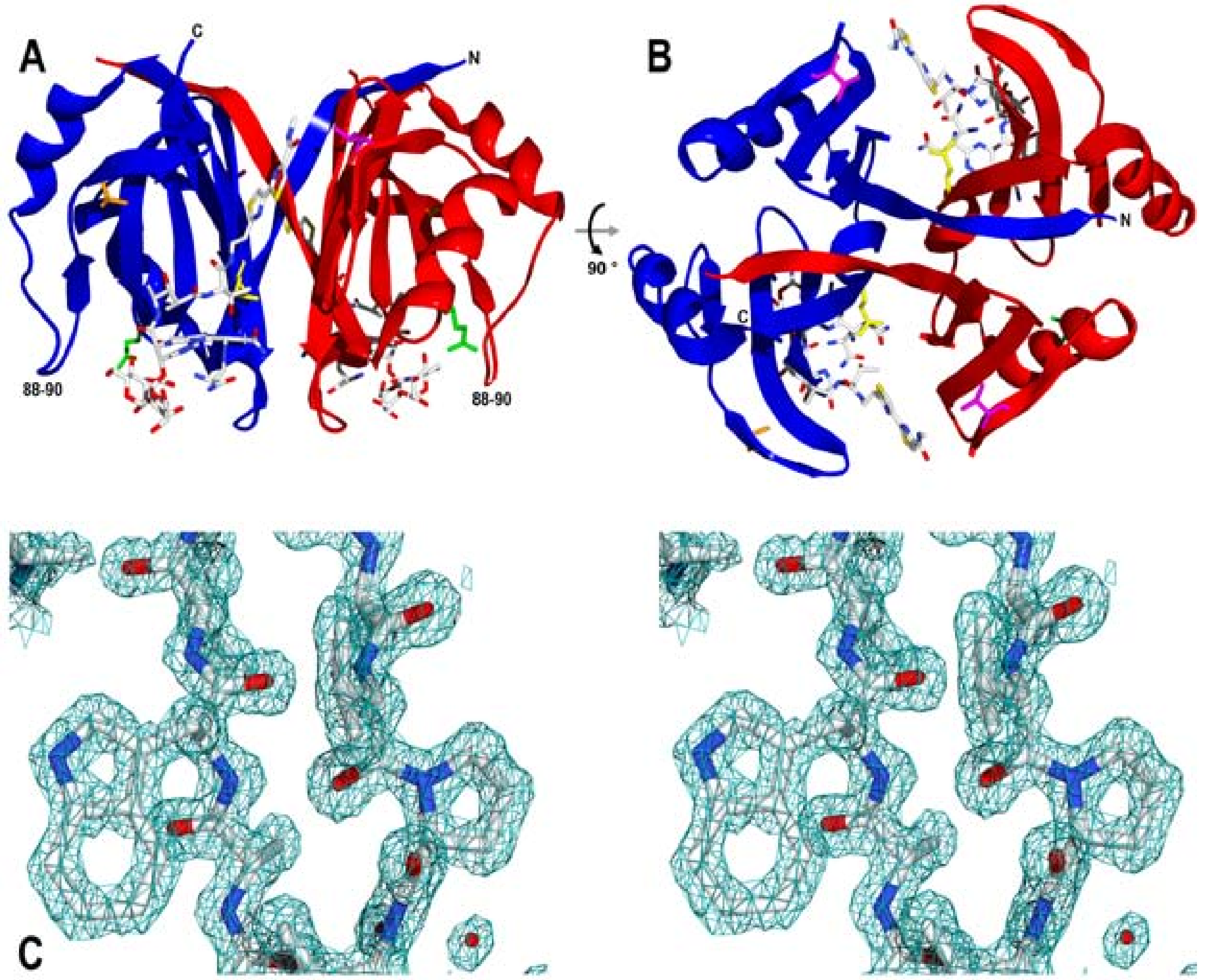


Figure 8.3 (in color on p.159)
Structure and electron density of HTS in complex with bleomycin A2. Ribbon diagram showing the dimeric structure of the fourfold mutant Shble in complex with bleomycin A2. Mutations are indicated by stick representations. Chain A in blue, chain B in red, Gly18Glu in green, Asp32Val in pink, Leu63Gln in yellow and Gly98Val in orange. **A)** side view. **B)** viewed from the N- and C-terminal side (top view) **C)** stereo view of the electron density around residue Pro9 and Trp65 contoured at 2σ. Residues are colored according to the CPK color scheme, water molecules are represented by red spheres.

the presence of bleomycin A2 and its structure was determined to 1.50 Å resolution. The crystals grown belong to space group $P2_1$ with unit cell parameters a=44.0 Å, b=66.6 Å, c=47.2 Å and β=117.4° and a dimer in the asymmetric unit (Fig. 8.3AB). Representative electron density is shown in Figure 8.3C. The structure forms a compact, homodimeric α/β protein of 121 amino acids (Met1, Gln123 and Asp124 are disordered) in which two bleomycin A2 molecules are accommodated in binding pockets at the dimer interface. These pockets consist of a hydrophilic concavity which runs into a hydrophobic intersubunit crevice. The dimer is maintained by alternate N-terminal β_1-strand hydrogen bonding between both monomers and by Van der Waals interactions at the largely hydrophobic subunit contact [92,188]. Three sulphate ions are present at the surface of the dimer of which two form ion pairs with Arg104 of both chains. The presence of a dimer in the asymmetric unit allowed the identification of certain symmetry deviations between both monomers. A backbone superimposition of both chains (RMSD: 0.38 Å, Table 8.3) only revealed large differences in a random coil region comprising residues Asp88, Ala89 and Ser90 (Fig. 8.3AB), which is spatially close to the carbamoyl group of the D-mannose moiety of bleomycin (Fig. 8.4A). Their respective C_α atoms deviate 2.0, 5.1 and 1.6 Å in position, while giving rise to almost oppositely pointing amino acid side chains. In contrast to the bleomycin bound and unbound BlmA structure, backbone B-factors in this region are only marginally higher compared to the average value, suggesting a rigid conformation [188,386]. The difference in orientation of this loop might therefore be the result of sequential binding of two bleomycin molecules. Unlike BlmA, no symmetry related differences were observed in the region between amino acids 100 to 103. The topology of the HTS protein complex and the mode of bleomycin binding are similar to other BBPs. An overview of available structures is given in Table 8.3.

The structure of the HTS mutant in complex with bleomycin completes the list of structural information of three BBPs with and without their ligands, hereby contributing to our understanding of these proteins in general. Moreover, it has revealed several molecular features, which can account for increased protein stability and improved functionality at higher temperature *in vivo* and *in vitro*.

Structural effects of mutations

Introduction of an intersubunit hydrogen bond network - The structure of the dimer shows that each of the two bleomycin A2 molecules is bound by the concerted action of 21 amino acids. Due to its intersubunit location, both binding sites are composed of residues from either subunit. These include Val32, Phe33, Glu35, Phe38, Ser51, Ala52 and Val53 of one subunit and Pro59, Asp60, Asn61, Thr62, Gln63, Trp65, Phe86, Ala89, Trp102, Ala107, Arg109, Gly113, Cys115 and His117 of the other. The crystal structure clearly reveals the central role of mutation Leu63Gln which was found in 3 out of 5 different double mutants. Gln63 is involved in an extensive hydrogen bond network at the bottom of the bleomycin binding concavity (Fig. 8.5A). It is noteworthy that the carbonyl side chains (Oε1) of both Gln63 residues in the dimer act as terminal hydrogen bond acceptors of a five-molecule water channel present at the dimer interface. A second hydrogen bond is accepted from the side chain hydroxyl group (Oγ) of Ser51 of the adjacent subunit. The amide side chain (Nε2) of Gln63 forms a hydrogen bond with one of two water molecules trapped between the bleomycin and the surface of the protein. The presence of a leucine at position 63 would

Table 8.3 - Crystal structures of bleomycin binding proteins.

Source / description	PDB ID	Form	Resolution (Å)	RMSD (Å)	Reference
Shble *St. hindustanus*	1BYL	apo	2.3	0.63[a]	92
	1XRK	bleomycin A2	1.5	0.38[b]	This study
BlmA *S. verticillus*	1QTO	apo	1.5	0.77[a]	188
	1JIE	bleomycin A2	1.8	0.73[a]	386
	1JIF	Cu^{2+}, bleomycin A2	1.6	0.75[a]	
BlmT *K. pneumoniae*	1ECS	apo	1.7	1.18[a]	250
	1EWJ	bleomycin A2	2.5	1.15[a]	

[a] Av. backbone superimposition RMSD values to HTS chain a & b, [b] backbone superimposition RMSD values of chain a to b.

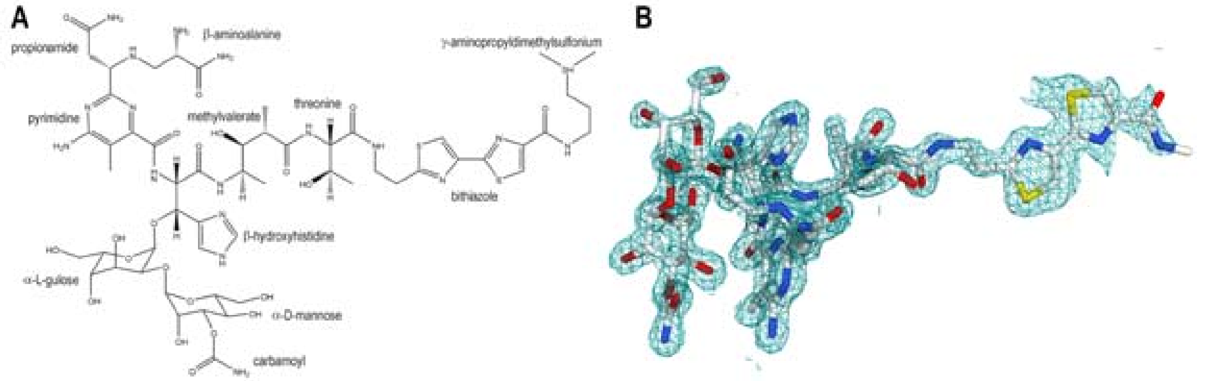


Figure 8.4 (in color on p.159)
Chemical diagram and electron density of bleomycin A2. **A)** Schematic representation of bleomycin A2. **B)** Electron density around bleomycin A2 contoured at 1.5σ. The diagram indicates the missing electron density around the γ-aminopropyldimethylsulfonium moiety suggesting a disordered conformation.

most likely not have allowed for a hydrogen-bond network of this size. The advantage of an amino acid compatible with hydrogen bonding at position 63 is also evident from the alignment, which indicates that without exception the other four BBPs have a serine at this specific site (Fig. 8.1). Although the mutant structure without bleomycin is not available, we speculate that an intersubunit hydrogen bond between Gln63 and Ser51 can persist even without the antibiotic bound, giving rise to a beneficial interaction that might stabilize the dimer at high temperatures.

Hydrophobic packing of surface indentations - In wild-type Shble and BlmA, Asp32 is located on the edge of the intersubunit binding groove for the bithiazole moiety and tail-region of bleomycin (Fig. 8.4A) [92,188,386]. In bleomycin A2, B2 and in phleomycin D1 the tail is positively charged which suggests involvement of Asp32 in electrostatic stabilization or ligand recognition. From NMR studies it has become clear that in the bound state no strong interactions occur between the protein and the positively charged tail of bleomycin [415]. These data are supported by the absence of electron density for the γ-aminopropyldimethylsulfonium moiety of bleomycin A2 in the binary complex structure of BlmA and HTS, suggesting a disordered conformation of the tail end (Fig. 8.4B) [386]. This might have allowed for an amino acid substitution to valine, which extends the hydrophobic bithiazole binding cleft at the dimer interface fitting nicely within a highly hydrophobic environment consisting of Phe33, Phe38, Val42, Thr47 and Phe49 (Fig. 8.5B). In addition, both BlmT and BlmS sequences also contain a valine at the corresponding position (Fig. 8.1). From a thermodynamic point of view, a mutation introducing surface hydrophobicity is generally believed to be unfavorable and has therefore rarely been investigated in directed mutagenesis studies. Nevertheless, some studies have reported significant improvements in protein stability by placing bulky hydrophobic amino acids at the surface of a neutral protease from *Geobacillus stearothermophilus* [410]. Recent findings using *Bacillus licheniformis* α-amylase have clearly indicated that hydrophobic surface residues can indeed be extremely stabilizing by improving hydrophobic packing of surface indentations, hereby reinforcing subsurface secondary structure elements [240]. This might explain the enhanced secondary structure preservation at high temperatures as inferred from CD of mutant 77-3. Strikingly, modeling of Gly40Ala into the wild-type structure (not shown) revealed close spatial proximity to Asp32 (5 Å between Cα and 3.5 Å between Cβ atoms), which may also underline a similar need for hydrophobicity in this part of the protein. Mutation Arg31 to Leu, which occurred in two types of double mutants came as a surprise, since it is involved in a surface ion pair with Asp25 in wild-type Shble. Apparently this electrostatic interaction does not counterweight beneficial effects of improved hydrophobic packing among residues Val20, Thr24, Val34, and Val41.

Reduction of surface loop flexibility - From previous crystallographic and NMR studies of BBPs, it has become clear that the loop following Gly98 in Shble will change its conformation upon binding of the antibiotic [188,250,386,415]. This conformational change enables the tryptophan at position 102 to stack optimally with the hydrophobic bithiazole moiety of bleomycin (Fig. 8.4A), packing both thiazole rings tightly against Phe33 and Phe38 of the adjacent subunit. In both BlmA and Shble, Gly98, located on the edge of a small β-strand leading towards the binding loop, seems to have a hinge function (Fig. 8.5C). The bending motion of the backbone is also clearly reflected in large Phi and Psi torsion angle changes of more than 20° upon the binding of bleomycin. At high temperatures, however, this flexibility might have caused problems leading to local unfolding or a decreased bleomycin binding ability. A substitution for either a valine or serine as observed, would increase the rigidity of the loop and could therefore restore the binding capacity at high temperature.

α-Helix stabilization - Mutation Gly18Glu introduces a glutamate at position N3 in the first turn of the largest α-helix of the protein. Statistical analysis as well as experimental studies have shown that glutamates are energetically highly favored over glycines at the third position in an α-helix [165,292]. This effect is most likely caused by the stabilizing effect of the negatively charged side chain on the helix macro dipole. To our surprise, chain A of the crystal structure revealed the formation of a genuine *i*, *i*+5 α-helix surface ion-pair between Glu21 and Arg26, which was absent in the wild-type structure (Fig. 8.5D). This new ion pair may have been the result of repulsion of anionic glutamate side chains of position 18 and 21, directing the latter towards the C-terminal arginine. Although *i*, *i*+5 α-helical surface ion pairs do not give rise to strong ionic interactions at ambient temperatures [365], they might be more favorable at higher temperatures. Theoretical models have indicated that the energetic cost of desolvating charged groups is much less at 100 °C, due to a drop in the dielectric constant of water [98]. This is currently the best explanation for the fact that proteins from extreme thermophiles have large ion pair networks at their surfaces, which are thought to be involved in maintaining structural integrity [186]. Additionally, a minor beneficial effect of this mutation could be the introduction of additional negative surface charge, which enhances electrostatic attraction of the cationic antibiotic under physiological conditions.

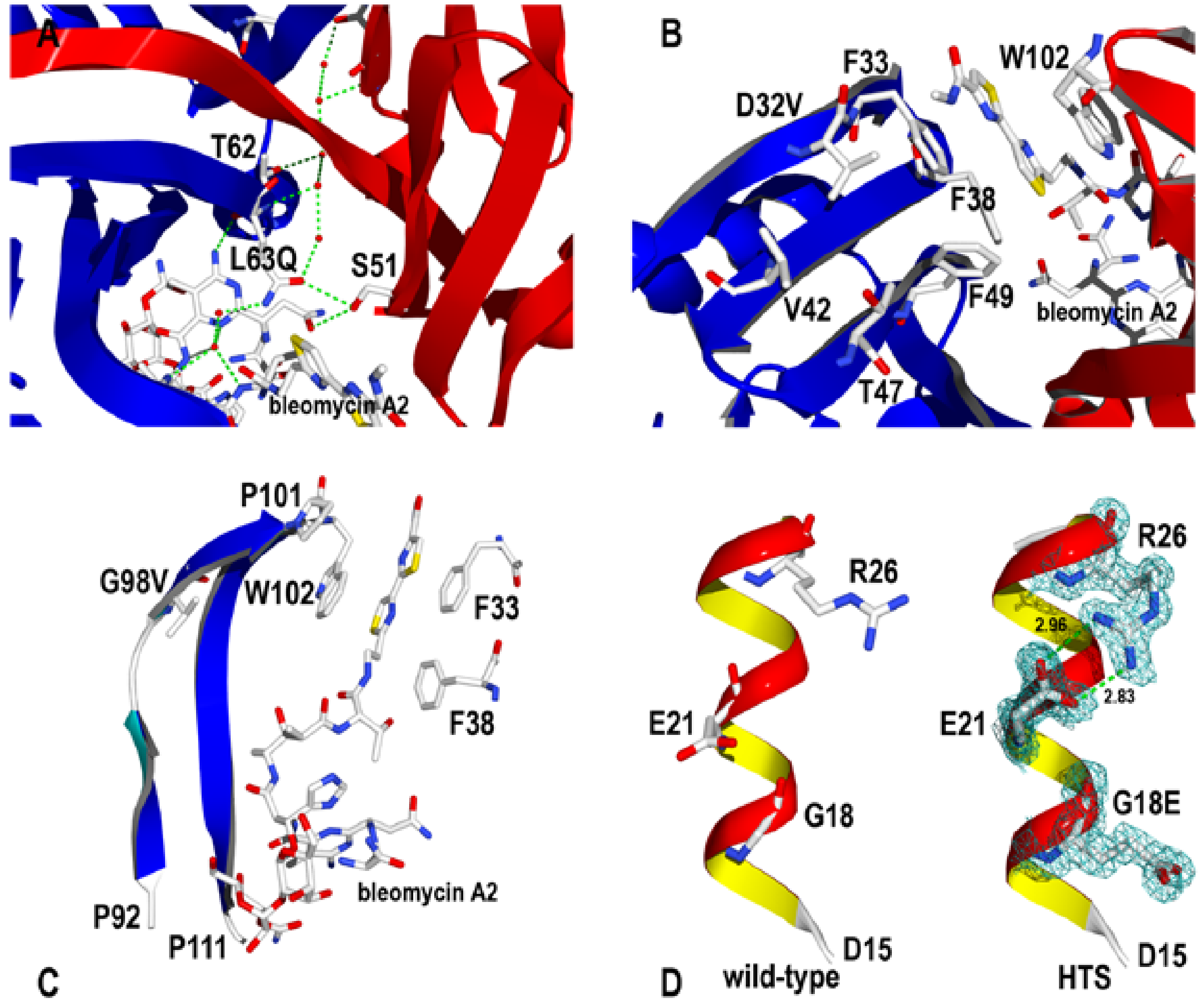


Figure 8.5 (in color on p.160)
Structural effects of the individual mutations. **A)** Leu63Gln. Ribbon diagram showing the hydrogen bond network at the dimer interface. Thr62, Gln63, Ser51 and bleomycin A2 are shown together with the intersubunit water channel. **B)** Asp32Val. Ribbon diagram showing the hydrophobic intersubunit bleomycin tail binding crevice. Val32 may be involved in improved hydrophobic packing of this surface indentation among amino acids Phe33, Phe38, Val42, Thr47 and Phe49. **C)** Gly98Val. Ribbon representation showing a loop between Pro92 and Pro111 which is involved in bleomycin binding. Val98 is located at a former hinge region which enables Trp102 to stack the bithiazole tail against Phe33 and Phe38. The electron density revealed two alternative sidechain rotamers for Val98 in chain A (not shown) and a single side chain conformation in chain B. **D)** Gly18Glu. Side-by-side comparison of α-helix one formed between Asp15 and Leu27 in the wild-type and mutant crystal structures of Shble. The existence of a surface ion-pair between Glu21 and Arg26 is visible in the electron density (contoured at 1.5σ). Interatomic distances are indicated in Å. CPK color coding was used for amino acids and bleomycin A2. Chain A is indicated in blue, chain B in red. Water molecules are represented by red spheres. Hydrogen bonds are depicted by green dotted lines.

Laboratory vs natural evolution of stability

In this study, several possible mechanisms of adaptation to high temperature were identified, such as the introduction of a hydrogen bond network, improved hydrophobic packing of surface indentations, reduction of loop flexibility and α-helix stabilization. Remarkably, half of all mutations found were glycine replacements, which could point to protein stabilization by decreasing the entropy of the unfolded state [411]. Although this could be a general strategy of stabilization, proteins from hyperthermophiles do not have a lower glycine content than their mesophilic counterparts, but rather a slightly increased one [49]. Their predicted proteomes do have an increased propensity for charged (Arg, Lys and Glu) and bulky aliphatic (Ile and Val) amino acids, which has mostly come at the cost of polar residues (Asn, Gln, Ser and Thr) [49]. This is fully in agreement with the requirements for the elevated numbers of surface salt bridges and improved hydrophobic core packing that has generally been recognized in these types of proteins. These are just two of a multitude of mechanisms that proteins from hyperthermophilic micro-organisms have employed to deal with extreme temperatures [212,419,420].

Recently, several other random mutagenesis studies have also reported large improvements in thermostability by applying directed evolution approaches. A mesophilic xylanase of family 11 was stabilized by over 35 °C by combining nine mutations found separately after extensive screening. The activity of this mutant was optimized by saturation mutagenesis of all mutated positions, yielding an enzyme variant with highly enhanced properties for high-temperature applications [283]. In another study, a highly thermostable esterase containing seven mutations was evolved in six rounds of random mutagenesis, recombination and screening [140]. The resulting enzyme was crystallized and its structure determined [374]. The structure revealed that improved stability was due to altered core packing, α-helix stabilization, the introduction of surface salt bridges and reduction of flexibility in surface loops. From these and many other directed evolution and site-directed mutagenesis studies, it has become apparent that (i) proteins can be stabilized substantially by small numbers of mutations, (ii) these mutations are often located at the protein surface and (iii) their effects are usually additive. As few as 2 out of 12 amino acid differences between a mesophilic and thermophilic cold shock protein turned out to be responsible for the difference in thermostability [293]. The remaining variation in sequence might just have occurred as a result of neutral sequence drift or specific properties required by the host, such as solubility, turnover and molecular interactions [247].

Despite the fact that only a small number of mutations are required to render a protein thermostable, finding those mutations remains a difficult task. Apart from screening vast numbers of random mutants in microtiter plates, *in vitro* and *in vivo* selection schemes offer great advantages in order to reduce the effort to encounter improved variants. Currently, two main strategies are available.

One of these thermostabilizing selection methods is called Proside (Protein stability increased by directed evolution) which is based on the empirically derived inverse correlation between protein thermostability and proteolytic susceptibility [356]. In a phage display-like procedure, a library of mutants is fused between two domains of *E. coli* filamentous phage Fd gene-3-protein (G3P) which is then subjected to proteases while inducing local unfolding of the target protein by means of temperature or chemical denaturants. The resulting instable fusion protein variants are cleaved, causing only the surviving, more stable phages to be found after infection. It was found that in an ionic denaturant, non-polar surface interactions were optimized, whereas at elevated temperature variants with improved surface electrostatics were selected [247].

Cloning and selection in a thermophile, as conducted in this study, is a second directed evolution strategy which can lead to rapid improvements in thermostability. Despite its simplicity, very few studies using this technique have been reported. This is most likely due to its limited applicability, since the gene of interest rarely confers any biologically relevant function for growth or survival of the thermophile. Nonetheless, functionally stabilized mutants have been reported up to 79 °C for a kanamycin nucleotidyltransferase [156,227] and 58 °C for a chloramphenicol acetyltransferase [403]. Auxotrophic knockout strains for leucine biosynthesis were complemented with thermolabile counterpart genes from *Bacillus subtilis* and *Saccharomyces cerevisiae* and adapted to higher temperature by serial accumulation of beneficial mutations in the *in-trans* introduced 3-isopropylmalate dehydrogenase genes [5,392]. A hybrid α-galactosidase consisting of *B. stearothermophilus* and *T. thermophilus* peptide regions was adapted to function at 67 °C by selection for growth on melibiose as the sole carbon and energy source [125].

Concluding remarks

In this study, several double mutants of a bleomycin binding protein were isolated with enhanced performance at high temperature both *in vivo* and *in vitro*. Structural analysis showed that the mutations gave rise to different means of stabilization in four parts of the protein. A combined mutant gene with a low GC content was created, which can serve as an antibiotic resistance marker for aerobic and micro-aerophilic mesophiles, thermophiles and hyperthermophiles. This may allow the development of efficient shuttle-vectors and knockout strategies for hyperthermophilic archaea and bacteria based on positive selection schemes. Moreover, the high GC content double mutant genes will now permit multigene knockout strategies in thermophiles such as *Thermus thermophilus*, allowing further exploration and exploitation of thermophilic microbial sources.

EXPERIMENTAL PROCEDURES

All chemicals were of analytical grade and purchased from Sigma. Primers were obtained from MWG Biotech AG (Ebersberg, Germany). Polymerase chain reactions were performed with *Pfu* TURBO (Stratagene) unless stated otherwise. Bleomycin A2 (Bleocin, Calbiochem) was used for all selections. *Escherichia coli* HB101 (F- hsdS20 (r_B-, m_B-) ara-14 galK2 lacY1 leuB6 mcrB mtl-1 proA2 recA13 rpsL20 supE44 thi-1 xyl-5 (Str^R)) [33] was used for cloning purposes and routinely transformed by electroporation.

Construction of bleomycin shuttle vector

Bacillus subtilis 168 8G5 carrying pUB110 was kindly provided by Dr. S. Bron (University of Groningen, Netherlands) and the plasmid was isolated by Qiagen Miniprep according to the manufacturer's instructions. The *blmS* gene was PCR amplified with primers BG1407 (sense) 5'-GGAGGTGCATATGAGAATGTTACAGTCTATCCC-3' and BG1240 (antisense) 5'-CGCGTCTAGATTAGCTTTTTATTTGTTGAAAAAAG-3' (NdeI and XbaI sites underlined). Chromosomal DNA of *Streptoalloteichus hindustanus* (ATCC31158) was prepared according to standard procedures [330] and used for PCR amplification of the *shble* gene with primers BG1410 (sense)

5’-TGAGGCATATGGCCAAGTTGACCAGTGCCG-3’ and BG1411 (antisense) 5’-GATCCTCTAGATTAGTCCTGCTCCTCGGCCACG-3’ (NdeI and XbaI sites underlined). PCR products were digested and ligated into *E. coli* - *T. thermophilus* shuttle-vector pMK18 [82] (Biotools, Madrid, Spain) thereby replacing the kanamycin nucleotidyltransferase gene. Ligation mixtures were transformed into *E. coli* HB101 and transformants were plated on 1.5% LB agar plates supplemented with 3 μg ml^{-1} bleomycin. Both *blmS* and *shble* provided resistance against the antibiotic, giving rise to the 4434 bp plasmid pWUR111 and the 4404 bp plasmid pWUR112, respectively.

Mutant library construction

Error-prone PCR was carried out using two different polymerases, namely *Taq* (Amersham) and Mutazyme (Genemorph kit, Stratagene). This approach was chosen to complement the transition and transversion bias of each enzyme to provide a more complete mutational spectrum of the PCR product. For the error-prone amplification, flanking primers BG1412 (sense) 5’-CGACCCTTAAGGAGGTGTGAGGCAT**ATG**-3’ and BG1408 (antisense) 5’-CGAGCTCGGTACCCGGGGATCCTCTAGA**TTA**-3’ (NdeI and XbaI site underlined) were designed to allow variation throughout the entire coding sequence, between the start- and stop-codon (indicated in boldface). *Taq* polymerase based PCR reactions were performed as previously described [351]. A 50 μl PCR reaction contained 5 ng of pWUR112, 5 pmol of each primer, 0.2 mM of dATP and dGTP, 1 mM of dCTP and dTTP, 5 U of polymerase, 3 mM MgCl$_2$ and three concentrations of MnCl$_2$ (0.1, 0.3 and 0.5 mM). The mixture was thermocycled as follows: 95 °C (4 min), 30 cycles of 94 °C (30 s), 55 °C (45 s) and 72 °C (25 s) post-dwelled for 4 min at 72 °C. Mutazyme PCR reactions were prepared according to the manufacturer’s instructions and thermocycled as above using an elongation time of 50 seconds.

Randomly mutated PCR products were cloned into vector pMK18 and transformed into *E. coli* HB101. A total of approximately 10,000 *Taq* and 10,000 Mutazyme derived clones were resuspended in 50 ml LB media supplemented with antibiotic and grown in 1 L media to early stationary phase. Plasmids were subsequently harvested using a Miniprep plasmid isolation kit (Qiagen).

Selection in *Thermus thermophilus*

Thermus thermophilus HB27 was kindly provided by Dr. J. Berenguer (Autonomous University of Madrid, Spain). Cells were routinely cultivated at 70 °C in a Ca^{2+} (3.9 mM) and Mg^{2+} (1.9 mM) rich media [82] containing 8 g L^{-1} tryptone, 4 g L^{-1} yeast extract and 3 g L^{-1} NaCl dissolved in Evian mineral water (pH 7.7, after autoclaving) (Evian-les-Bains, France). Transformation of *T. thermophilus* was essentially performed by the method of Koyama [208]. Frozen cell aliquots were resuspended in 25 ml of media and grown at 150 rpm to an A$_{600}$ of 0.8. The culture was then diluted 1:1 in preheated media and incubated for another h. Next, plasmids were added to 0.5 ml of culture and the mixture was incubated for 2-3 h at 70 °C with occasional shaking before being plated on 3% agar plates (Beckton Dickinson) supplemented with 30 μg ml^{-1} kanamycin or 15 µg ml^{-1} bleomycin (Calbiochem) for selection. Colonies appeared within 36 h at 60 - 70 °C. At temperatures above 70 °C, 1% Gelrite plates (Roth, Karlsruhe, Germany) were used, supplemented with 100 µg ml^{-1} kanamycin or 20 µg ml^{-1} bleomycin for selection. Colonies were grown overnight in liquid media containing 30 µg ml^{-1} kanamycin or 5 µg ml^{-1} bleomycin. *T. thermophilus* plasmid DNA was prepared using a plasmid Miniprep-kit (Qiagen) after a 2 mg ml^{-1} pre-incubation with lysozyme for 30 min at 37 °C.

Gene cloning, protein production and purification

Wild-type and double mutant *shble* genes were PCR amplified from their respective pWUR112 plasmids using primers BG1503 (sense) 5'-GATGGCCATGGCCAAGTTGACCAGTGC-3' and BG1504 (antisense) 5'-GCCGCAAGCTTAGTCCTGCTCCTCGGCC-3' (NcoI and HindIII site underlined). PCR products were cloned into vector pET26b (Novagen) and fused to an *Erwinia carotovora* pectate lyase (*pelB*) signal sequence allowing periplasmic protein overexpression in *E. coli* BL21(DE3) (Novagen). Periplasmic fractions of 1 L cultures were prepared by osmotic shock according to the manufacturers instructions and dialyzed overnight against 20 mM Tris-HCl (pH 7.5). Samples were loaded onto a MonoQ HR 5/50 connected to a FPLC system (Amersham) and eluted using a 1 M NaCl gradient. Shble containing fractions were pooled and dialyzed against a 10 mM NaP_i buffer (pH 7.0) supplemented with 50 mM NaCl and subsequently purified by size exclusion chromatography using a Superdex 200 HR 10/30 column (Amersham).

Synthetic gene construction

A synthetic mutant *shble* gene based on archaeal codon usage was constructed by oligonucleotide assembly PCR [379]. This gene contains the point mutations Gly18Glu, Asp32Val, Leu63Gln and Gly98Val, and has a GC content of 40.8% compared to 70.2% of wild-type *shble*. The synthetic gene was denominated HTS (High Temperature Shble). The sequence has been deposited to Genbank (accession number AY780486).

Assembly PCR mixtures contained 10 oligonucleotides (BG1542 - BG1451, online Suppl. Table) with an overlap of 20 bases. Both flanking primers were 40 bases in length whereas the 8 central primers consisted of 80 to 90 bases. The assembly PCR mixture contained 2.5 µM of each primer, 0.2 mM of dNTP's and 0.05 U µl^{-1} of *Pfu* polymerase. The mixture was thermocycled at 94 °C (30 s), 55 °C (30 s) and 72 °C (60 s) for 40 cycles. The PCR products were purified over a PCR purification column (Qiagen) and diluted 1:1 in fresh PCR mix containing only both flanking primers BG1542 and BG1551 at 0.1 µM concentration and were thermocycled according to standard procedures. PCR products of the expected size were isolated from agarose gel using Qiaex II gel extraction kit (Qiagen), digested with NdeI and BglII and cloned into vector pET26b. This allowed for efficient cytoplasmic protein overproduction in *E. coli* BL21(DE3)-RIL (Novagen). Positive clones were picked from LB agar plates containing 3 µg ml^{-1} bleomycin and 50 µg ml^{-1} chloramphenicol. A 4 L culture was grown at 37 °C until A_{600} 0.5, induced with

0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and was incubated for another 5 h. Cells were harvested, resuspended in 20 mM Tris-HCl (pH 7.5) and sonicated. Cell extracts were clarified by centrifugation (30 min, 26,500 xg, 4 °C) and applied to a 70 ml Q-Sepharose Fast Flow (Amersham) anion exchange column. Proteins were eluted by a 1 M NaCl gradient, High Temperature Shble (HTS) containing fractions were pooled and concentrated over a YM10 filter (Amicon) and further purified by size exclusion chromatography as described above.

DNA sequencing

Inserts of plasmids used in this study were sequenced by Westburg Genomics (Wageningen, Netherlands).

Protein quantitation

Protein concentrations were determined by using a Bradford assay [34] (Bio-Rad). Purified proteins were quantified from A_{280} measurements using a protein extinction coefficient of 29,000 $M^{-1}cm^{-1}$ [320].

DNA protection assay

Protein functionality assays were essentially performed as described elsewhere [135], employing a 10-fold excess molar concentration of protein over bleomycin A2 (Calbiochem). In each assay, 0.2 μg of PstI linearized plasmid, pUC19, was used. Assays were performed by first incubating DNA and protein shortly at 85 °C, after which bleomycin A2, dithiothreitol (DTT) and Fe^{2+} were sequentially added to the reaction mixture.

Circular dichroism spectroscopy

CD experiments were performed on a Jasco J-715 spectropolarimeter (Jasco, Tokyo, Japan) equipped with a PTC-348WI Peltier temperature control system. Far-UV CD measurements were conducted with Suprasil quartz cuvettes (Hellma Benelux, Rijswijk, Netherlands) with a 1 mm cell length. During all experiments, the sample cell chamber was purged by dry N_2 gas at a flow rate of 10 L min^{-1}. In temperature-induced unfolding experiments, the cuvette containing 1.7 μM of protein sample in degassed 10 mM NaP_i (pH 7.0) and 50 mM NaCl, was heated from 25 to 95 °C at 0.4 °C min^{-1} and subsequently cooled to 25 °C at the same rate. The ellipticity at 205 nm was measured every 0.5 °C with a 2 s response time to monitor the loss of β-sheet and β-turn secondary structure elements. The bandwidth of the measurement was set to 1.0 nm and the sensitivity to 100 mdeg. Data were corrected for the temperature dependent ellipticity of a blanc without protein. Averaged data of two independent scans were fit according to a two state model of unfolding and the apparent temperature unfolding midpoint (T_m) was derived from van 't Hoff plots.

Fluorescence spectroscopy

Fluorescence experiments were performed on a Varian Cary Eclipse fluorimeter (Varian, Middelburg, Netherlands) equipped with a four-cuvette Peltier multicell holder and PCB-150 waterbath. All measurements were performed in 3 ml Suprasil quartz cuvettes (Hellma Benelux, Rijswijk, Netherlands) with a 1 cm pathlength. A magnetic stirring bar ensured a homogeneous sample temperature.

The temperature of the sample was recorded by a temperature probe inside one of the four samples. Spectra and thermal unfolding curves were recorded of 1.7 μM protein solutions in degassed 10 mM NaP_i (pH 7.0) buffer supplemented with 50 mM NaCl. Tryptophans were excited at 295 nm and fluorescence was recorded from 300 to 550 nm with both the excitation and the emission slits set to 5 nm. During temperature-induced unfolding and refolding studies, fluorescence emission intensities were monitored at 315 nm from 27 to 92 °C at a heating and cooling rate of 0.4 °C min^{-1}. Data were corrected for the fluorescence emission of corresponding blanc solutions. Data of two independent scans were treated and fit as described above.

Differential scanning calorimetry (DSC)

DSC measurements were performed on a Microcal III system (Setaram, Caluire, France). Degassed protein samples of 0.28 mg ml^{-1} (20 μM) in 10 mM NaPi (pH 7.0) and 50 mM NaCl in the presence and absence of an 8-fold molar excess bleomycin A2, were heated from 20 to 120 °C at 0.5 °C min^{-1}. Midpoint temperatures of unfolding were determined by curved baseline analysis from two independent scans.

Protein crystallization, data collection and processing

The HTS protein was extensively dialyzed against 10 mM NaP_i, pH 7.0 and was subsequently crystallized by the sitting drop method of vapor diffusion at 20 °C and a protein concentration of 3.3 mg ml^{-1} in the presence of a 10 fold molar excess of bleomycin A2 - HCl (Calbiochem). Crystals grew in 2.0 M Ammonium-Sulfate as the precipitant in 0.1 M Sodium-Acetate buffer, pH 4.6. Data were collected from a single flash-frozen native crystal (100K) to 1.5 Å resolution using a MAR345 imaging plate at the Protein Structure Factory beamline BL14.2 of the Free University of Berlin at the BESSY synchrotron source (Berlin, Germany) (Table 8.4). All data were reduced with DENZO and SCALEPACK [280]. The crystal used for data collection had unit cell parameters *a*=44.0 Å, *b*=66.6 Å and *c*=47.2 Å and *β*=117.4° and belonged to space group $P2_1$ with a dimer in the asymmetric unit.

Table 8.4 - Data collection and refinement. Values in parentheses refer to the highest resolution shell.

X-ray data collection statistics	
Wavelength (Å)	0.90830
Resolution (Å)	30 – 1.5 (1.53 – 1.50)
Total observations	87740 (3602)
Unique observations	36444 (1714)
Completeness (%)	94.4 (88.6)
$I / \sigma(I)$	20.3 (3.9)
$R_{sym}{}^{a}$	0.043 (0.189)
Refinement statistics	
Resolution (Å)	30.0 - 1.5
$R_{work}{}^{b}$	0.174
$R_{free}{}^{c}$	0.194
RMSD bond distances (Å)	0.01
RMSD bond angles (°)	1.664
Total number of non-H atoms	1883
Av. protein B value (Å^2)	13.3
Solvent molecules	288
Av. solvent B value (Å^2)	27.0
Bleomycin atoms	182
Av. ligand B value (Å^2)	22.8
Sulphate ion atoms	15
Av. sulphate ion B value (Å^2)	31.4

[a] $R_{sym} = \Sigma_{hkl} \Sigma_i |I-<I_i>| / \Sigma_{hkl} \Sigma_i I_i$, where I_i is the intensity of a given measurement and the sums are over all measurements and reflections. Values in parentheses refer to the highest resolution shell.

[b] $R_{work} = \Sigma ||F(obs)|-|F(calc)|| / \Sigma |F(obs)|$ for the 95% of the reflection data used in refinement.

[c] $R_{free} = \Sigma ||F(obs)|-|F(calc)|| / \Sigma |F(obs)|$ for the remaining 5%.

The structure of HTS was determined by molecular replacement using the program MOLREP [408] and the *Streptomyces verticillus* BlmA dimer (PDB ID: 1JIE) [386] as the search model. The initial phases were improved using the free-atom refinement method together with automatic model tracing in ARP/wARP [294]. TLS parameters were determined and TLS-restrained refinement was performed using REFMAC [269]. Several rounds of iterative model building and refinement followed and water molecules were added using ARP/wARP [294]. The final model (comprising 241 amino acids, 288 water molecules, 2 bleomycin molecules and 3 sulphate ions), refined using data between 30 and 1.5 Å resolution, has an R- and free R-factor of 17.4 and 19.4%, respectively, with good geometry (Table 8.4). Residues Met1, Glu122, Gln123 and Asp124 in chain A and Met1, Gln123 and Asp124 in chain B are not visible in the electron density map and therefore have been excluded from the model. Additionally, the side chains of two residues in chain A (Asp36 and Arg87), four side chains in chain B (Glu21, Asp36, Arg87 and Glu122), and the γ-aminopropyldimethylsulfonium moiety of bleomycin have been truncated in the final model. The stereochemical quality of the model and the model fit to the diffraction data were analyzed with the programs PROCHECK [216] and SFCHECK [409].

The coordinates and experimental structure factors have been deposited in the Protein Data Bank with accession number 1XRK. Figures were prepared with Swiss-PDBviewer v3.7 SP5 [146] and rendered with POV-Ray v3.6.

ACKNOWLEDGEMENTS

The authors thank Dr. J. Berenguer for helpful suggestions and Anton Korteweg for technical assistance with DSC. This work was supported by a grant from the European Union in the framework of the SCREEN project.

9

Summary and concluding remarks

The work presented in this thesis describes a study of various aspects of the sugar metabolism of the hyperthermophilic archaeon *Sulfolobus solfataricus*. During these studies a diverse set of molecular techniques was employed, among which DNA microarray analysis, quantitative proteomics, classical genetics and biochemistry, X-ray crystallography and protein engineering. The wealth of genetic information available in the databases was explored using bioinformatics. The integrated approach that was taken to discover new metabolic pathways, and to analyze the activities and structures of the enzymes in these pathways, has led to a better general understanding of the lifestyle that these microbes have developed to thrive in hot acid environments.

Chapter one gives a general overview of thermophiles, their genetics and provides an introduction to the hyperthermophilic model archaeon *Sulfolobus solfataricus*, its vector developments and glycoside hydrolases. Directed evolution approaches are also briefly discussed.

Chapter two describes the results of a reconstruction of the central carbon metabolism (glycolysis, gluconeogenesis and tricarboxylic acid cycle) of *S. solfataricus* on the basis of genome, proteome, transcriptome and biochemical data. The expression level of the genes involved in these pathways was compared at the mRNA and protein level during growth on peptides and sugars. Most of the observed differences in expression between the two conditions were surprisingly small. Despite this, some key enzymes in glycolysis (non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase), and gluconeogenesis (phospho*enol*pyruvate synthase, phosphoglycerate kinase) were differentially expressed. Since these pathways in *Sulfolobus* generally appear not to be regulated by the abundance of the respective enzymes, it is proposed that the activity of some key enzymes may be regulated by allosteric control and protein post-translational modifications instead.

Chapter three presents the results of an integrative genomics approach to elucidate the D-arabinose metabolism of *S. solfataricus*. Key enzymes of the pathway were identified using DNA microarray and quantitative proteomics analyses, while their activities were studied by classic biochemistry. This revealed that the pentose was converted into 2-oxoglutarate by the action of four previously uncharacterized enzymes: a D-arabinose dehydrogenase, a D-arabinonate dehydratase, a 2-keto-3-*deoxy*-D-arabinonate dehydratase and a 2,5-dioxopentanoate dehydrogenase. Although this pathway has been shown by Weimberg and Doudoroff in the late 50's and early 60's in *Pseudomonas saccharophila* and *P. fragi*, it has taken nearly 50 years for the responsible genes to be found. Bioinformatics was employed to make use of the wealth of microbial genome information to establish the occurrence of the pathway components in other micro-organisms. Interesting evolutionary observations were extracted from the similarities of the pentose oxidation pathway with degradation schemes for hexaric acids and L-hydroxyproline, which support a pathway genesis scheme by enzyme recruitment events between pathways with shared intermediates.

Chapter four gives a detailed view

of the three dimensional structure and biochemical properties of the enzyme responsible for first step of the above mentioned pentose oxidation pathway: the D-arabinose dehydrogenase. As most nicotinamide-adenine-dinucleotide utilizing enzymes from *Sulfolobus* spp, the AraDH enzyme showed a clear preference for the electron acceptor $NADP^+$ over NAD^+, which could be explained by the presence of an adenyl-phosphate binding pocket in the structure. Molecular understanding of the substrate preference for D-arabinose and L-fucose was obtained by ligand docking simulations. The information from this study exemplifies once more the diverse functions that are carried out by members of the zinc-dependent medium-chain dehydrogenase/reductase protein superfamily, and can be instrumental for protein engineering and genome annotation purposes.

Chapter five describes the crystal structure of a novel enzyme of the *Sulfolobus* D-arabinose degradation pathway: the 2-keto-3-*deoxy*-D-arabinonate dehydratase. This enzyme is responsible for the third step of the pentose oxidation pathway, which concerns the water elimination of 2-keto-3-*deoxy*-D-arabinonate leading to the aldehyde 2,5-dioxopentanoate. The structure uncovered an unusual tetrameric ring of two-domain subunits, resembling an elongated donut. Insight into substrate binding and catalysis was obtained from co-crystal structures with a substrate analog and the enzyme product bound. A catalytic mechanism for base catalyzed dehydration was proposed based on these observations.

Chapter six describes the purification and characterization of an unusual intracellular α-galactosidase from the hyperthermophilic crenarchaeon *S. solfataricus*. This class of enzymes liberates galactose from the non-reducing end of sugars, and can be commonly found in Bacteria and in Eukaryotes such as yeasts, fungi, plants and mammals. Within glycosyl hydrolase family 36, the primary sequence of the archaeal enzyme variant was shown to cluster with mostly alkaline alpha-galactosidases, and raffinose and stachyose synthases from plants, and appeared more distantly related to the thermostable bacterial α-galactosidases. Catalytic aspartate candidates for the nucleophile and acid/base of the reaction were predicted based on conserved sequence motifs, and their involvement in catalysis was verified by site directed mutagenesis.

Chapter seven reports on the results of a project to engineer the properties of a β-glycosidase by DNA family shuffling. A hybrid enzyme was constructed with improved catalytic rates at 70 °C, which had retained the desired properties of the parental enzymes CelB (high thermostability) from *Pyrococcus furiosus* and LacS (low product inhibition) from *Sulfolobus solfataricus*. The result came as a surprise, since it is commonly accepted that amino acid side chain interactions in proteins from hyperthermophiles are highly optimized to achieve extreme thermostability. This study however shows that it is possible to exchange a large portion of the central $(\beta\alpha)_8$-barrel of the enzyme, in effect by substituting seventeen amino acids in CelB and introducing a single insertion, without dramatically affecting the structural integrity of the enzyme.

Chapter eight describes the directed evolution of a bleomycin binding protein towards increased thermal tolerance. An *in vivo* selection system was set up, by making use of the thermophilic bacterium *Thermus thermophilus* HB27, capable of growing at temperatures ranging from approximately 50 to 80 °C. This system allowed for the simple identification of mutants of the bleomycin binding protein that were operative up to 77 °C within a library of 20,000 functional clones. Strikingly, while the wild-type bleomycin binding protein provided antibiotic resistance up to 65 °C, only eight double mutants could do so at 77 °C, indicating a high stringency of the selection procedure. A total of six amino acid positions were altered in the protein, of which three were glycine residues. A quadruple mutant was constructed which showed that additional mutations give rise to an increase in thermostability, suggesting independent and cumulative effects of some of the mutations. The molecular basis of increased thermostability was deduced from the crystal structure of the quadruple mutant, which displayed additional hydrogen bonds, improved hydrophobic packing of surface indentations, reduced flexibility of a surface loop, and a stabilized α-helix.

References

1. **Aagaard, C., Leviev, I., Aravalli, R.N., Forterre, P., Prieur, D. & Garrett, R.A. (1996)** General vectors for archaeal hyperthermophiles: strategies based on a mobile intron and a plasmid. *FEMS Microbiol Rev* **18:**93-104.
2. **Adams, M.W., Holden, J.F., Menon, A.L., Schut, G.J., Grunden, A.M., Hou, C., Hutchins, A.M., Jenney, F.E., Jr., Kim, C., Ma, K., Pan, G., Roy, R., Sapra, R., Story, S.V. & Verhagen, M.F. (2001)** Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* **183:**716-724.
3. **Aguilar, C.F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M. & Pearl, L.H. (1997)** Crystal structure of the beta-glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. *J Mol Biol* **271:**789-802.
4. **Ahmed, H., Ettema, T.J.G., Tjaden, B., Geerling, A.C.M., Van der Oost, J. & Siebers, B. (2005)** The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochem J* **390:**529-540.
5. **Akanuma, S., Yamagishi, A., Tanaka, N. & Oshima, T. (1998)** Serial increase in the thermal stability of 3-isopropylmalate dehydrogenase from *Bacillus subtilis* by experimental evolution. *Protein Sci* **7:**698-705.
6. **Akerboom, J., Turnbull, A.P., Hargreaves, D., Fisher, M., De Geus, D., Sedelnikova, S.E., Berrisford, J.M., Baker, P.J., Verhees, C.H., Van der Oost, J. & Rice, D.W. (2003)** Purification, crystallization and preliminary crystallographic analysis of phosphoglucose isomerase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Acta Crystallogr D Biol Crystallogr* **59:**1822-1823.
7. **Albers, S.V. & Driessen, A.J. (2007)** Conditions for gene disruption by homologous recombination of exogenous DNA into the *Sulfolobus solfataricus* genome. *Archaea* **2:**i-v.
8. **Albers, S.V., Jonuscheit, M., Dinkelaker, S., Urich, T., Kletzin, A., Tampe, R., Driessen, A.J. & Schleper, C. (2006)** Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Appl Environ Microbiol* **72:**102-111.
9. **Allers, T. & Mevarech, M. (2005)** Archaeal genetics - the third way. *Nat Rev Genet* **6:**58-73.
10. **Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997)** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:**3389-3402.
11. **Amako, K., Fujita, K., Shimohata, T.A., Hasegawa, E., Kishimoto, R. & Goda, K. (2006)** NAD+-specific D-arabinose dehydrogenase and its contribution to erythroascorbic acid production in *Saccharomyces cerevisiae*. *FEBS Lett* **580:**6428-6434.
12. **Andersson, A., Bernander, R. & Nilsson, P. (2005)** Dual-genome primer design for construction of DNA microarrays. *Bioinformatics* **21:**325-332.
13. **Aravalli, R.N. & Garrett, R.A. (1997)** Shuttle vectors for hyperthermophilic archaea. *Extremophiles* **1:**183-191.
14. **Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V. & Notredame, C. (2006)** Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res* **34:**W604-608.
15. **Arnold, F.H., Wintrode, P.L., Miyazaki, K. & Gershenson, A. (2001)** How enzymes adapt: lessons from directed evolution. *Trends Biochem Sci* **26:**100-106.
16. **Arnold, H.P., She, Q., Phan, H., Stedman, K., Prangishvili, D., Holz, I., Kristjansson, J.K., Garrett, R. & Zillig, W. (1999)** The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. *Mol Microbiol* **34:**217-226.
17. **Arrizubieta, M.J. & Polaina, J. (2000)** Increased thermal resistance and modification of the catalytic properties of a beta-glucosidase by random mutagenesis and in vitro recombination. *J Biol Chem* **275:**28843-28848.
18. **Aucelli, T., Contursi, P., Girfoglio, M., Rossi, M. & Cannio, R. (2006)** A spreadable, non-integrative and high copy number shuttle vector for *Sulfolobus solfataricus* based on the genetic element pSSVx from *Sulfolobus islandicus*. *Nucleic Acids Res* **34:**e114.
19. **Babbitt, P.C., Mrachko, G.T., Hasson, M.S., Huisman, G.W., Kolter, R., Ringe, D., Petsko, G.A., Kenyon, G.L. & Gerlt, J.A. (1995)** A functionally diverse enzyme superfamily that abstracts the alpha protons of carboxylic acids. *Science* **267:**1159-1161.
20. **Baker, P.J., Britton, K.L., Rice, D.W., Rob, A. & Stillman, T.J. (1992)** Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold. Implications for nucleotide specificity. *J Mol Biol* **228:**662-671.
21. **Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J., Chen, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y., Ling, L., Tan, H., Chen, R., Wang, J., Yu, J. & Yang, H. (2002)** A complete sequence of the *T. tengcongensis* genome. *Genome Res* **12:**689-700.
22. **Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. & Horvath, P. (2007)** CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315:**1709-1712.
23. **Barry, E.R. & Bell, S.D. (2006)** DNA replication in the archaea. *Microbiol Mol Biol Rev* **70:**876-887.
24. **Bartolucci, S., Rella, R., Guagliardi, A., Raia, C.A., Gambacorta, A., De Rosa, M. & Rossi, M. (1987)** Malic enzyme from archaebacterium *Sulfolobus solfataricus*. Purification, structure, and kinetic properties. *J Biol Chem* **262:**7725-7731.
25. **Bartolucci, S., Rossi, M. & Cannio, R. (2003)** Characterisation and functional complementation of a nonlethal deletion in the chromosome of a Beta-glycosidase mutants of *Sulfolobus solfataricus*. *J Bacteriol*

185:3948-3957.

26. **Bateman, R.L., Ashworth, J., Witte, J.F., Baker, L.J., Bhanumoorthy, P., Timm, D.E., Hurley, T.D., Grompe, M. & McClard, R.W. (2006)** Slow-onset inhibition of fumarylacetoacetate hydrolase by phosphinate mimics of the tetrahedral intermediate: kinetics, crystal structure, and pharmacokinetics. *Biochem J.*

27. **Bateman, R.L., Bhanumoorthy, P., Witte, J.F., McClard, R.W., Grompe, M. & Timm, D.E. (2001)** Mechanistic inferences from the crystal structure of fumarylacetoacetate hydrolase with a bound phosphorus-based inhibitor. *J Biol Chem* **276:**15284-15291.

28. **Bergeron, A., D'Astous, M., Timm, D.E. & Tanguay, R.M. (2001)** Structural and functional analysis of missense mutations in fumarylacetoacetate hydrolase, the gene deficient in hereditary tyrosinemia type 1. *J Biol Chem* **276:**15225-15231.

29. **Berkner, S., Grogan, D., Albers, S.V. & Lipps, G. (2007)** Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus*, model organisms of the (cren-)archaea. *Nucleic Acids Res.*

30. **Berkowitz, D.B. & Benner, S.A. (1987)** Anomeric specificity of L-fucose dehydrogenase: a stereochemical imperative in aldopyranose dehydrogenases? *Biochemistry* **26:**2606-2611.

31. **Blochl, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H.W. & 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.

32. **Bomati, E.K. & Noel, J.P. (2005)** Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. *Plant Cell* **17:**1598-1611.

33. **Boyer, H.W. & Rouland-Dussoix, R. (1969)** A complementation analysis of the restriction and modification of DNA in *Escherichia coli. J Mol Biol* **41:**459-472.

34. **Bradford, M.M. (1976)** A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72:**248-254.

35. **Brinkman, A.B., Bell, S.D., Lebbink, R.J., De Vos, W.M. & Van der Oost, J. (2002)** The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J Biol Chem* **277:**29537-29549.

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

37. **Brock, T.D. & Freeze, H. (1969)** *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *J Bacteriol* **98:**289-297.

38. **Brouns, S.J.J., Smits, N., Wu, H., Snijders, A.P.L., Wright, P.C., De Vos, W.M. & Van der Oost, J. (2006)** Identification of a Novel alpha-Galactosidase from the Hyperthermophilic Archaeon *Sulfolobus solfataricus. J Bacteriol* **188:**2392-2399.

39. **Brouns, S.J.J., Walther, J., Snijders, A.P.L., Van de Werken, H.J.G., Willemen, H.L.D.M., Worm, P., De Vos, M.G.J., Andersson, A., Lundgren, M., Mazon, H.F.M., Van den Heuvel, R.H.H., Nilsson, P., Salmon, L., De Vos, W.M., Wright, P.C., Bernander, R. & Van der Oost, J. (2006)** Identification of the missing links in prokaryotic pentose oxidation pathways: evidence for enzyme recruitment. *J Biol Chem* **281:**27378-27388.

40. **Brugger, K., Chen, L., Stark, M., Zibat, A., Redder, P., Ruepp, A., Awayez, M., She, Q., Garrett, R.A. & Klenk, H.P. (2006)** The genome of *Hyperthermus butylicus*: a sulfur-reducing, peptide fermenting, neutrophilic Crenarchaeote growing up to 108 degrees C. *Archaea* **2:**127-135.

41. **Brugger, K., Torarinsson, E., Redder, P., Chen, L. & Garrett, R.A. (2004)** Shuffling of *Sulfolobus* genomes by autonomous and non-autonomous mobile elements. *Biochem Soc Trans* **32:**179-183.

42. **Brumer, H., 3rd, Sims, P.F. & Sinnott, M.L. (1999)** Lignocellulose degradation by *Phanerochaete chrysosporium*: purification and characterization of the main alpha-galactosidase. *Biochem J* **339 (Pt 1):**43-53.

43. **Brünger, A.T. (1991)** Simulated annealing in crystallography. *Annu Rev Phys Chem* **42:**197-223.

44. **Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. & Warren, G.L. (1998)** Crystallography & NMR system (CNS): A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54:**905-921.

45. **Brunner, N.A., Brinkmann, H., Siebers, B. & Hensel, R. (1998)** NAD+-dependent glyceraldehyde-3 -phosphate dehydrogenase from *Thermoproteus tenax*. The first identified archaeal member of the aldehyde dehydrogenase superfamily is a glycolytic enzyme with unusual regulatory properties. *J Biol Chem* **273:**6149-6156.

46. **Buchanan, C.L., Connaris, H., Danson, M.J., Reeve, C.D. & Hough, D.W. (1999)** An extremely thermostable aldolase from *Sulfolobus solfataricus* with specificity for non-phosphorylated substrates. *Biochem J* **343 Pt 3:**563-570.

47. **Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S. & Venter, J.C. (1996)** Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science* **273:**1058-1073.

48. **Camacho, M.L., Brown, R.A., Bonete, M.J., Danson, M.J. & Hough, D.W. (1995)** Isocitrate dehydrogenases from *Haloferax volcanii* and *Sulfolobus solfataricus*: enzyme purification, characterisation and N-terminal sequence. *FEMS Microbiol Lett* **134:**85-90.

49. **Cambillau, C. & Claverie, J.M. (2000)** Structural and genomic correlates of hyperthermostability. *J*

Biol Chem **275:**32383-32386.

50. **Cannio, R., Contursi, P., Rossi, M. & Bartolucci, S. (1998)** An autonomously replicating transforming vector for *Sulfolobus solfataricus*. *J Bacteriol* **180:**3237-3240.

51. **Cannio, R., Contursi, P., Rossi, M. & Bartolucci, S. (2001)** Thermoadaptation of a mesophilic hygromycin B phosphotransferase by directed evolution in hyperthermophilic Archaea: selection of a stable genetic marker for DNA transfer into *Sulfolobus solfataricus*. *Extremophiles* **5:**153-159.

52. **Cardona, S., Remonsellez, F., Guiliani, N. & Jerez, C.A. (2001)** The glycogen-bound polyphosphate kinase from *Sulfolobus acidocaldarius* is actually a glycogen synthase. *Appl Environ Microbiol* **67:**4773-4780.

53. **Carmi, N., Zhang, G., Petreikov, M., Gao, Z., Eyal, Y., Granot, D. & Schaffer, A.A. (2003)** Cloning and functional expression of alkaline alpha-galactosidase from melon fruit: similarity to plant SIP proteins uncovers a novel family of plant glycosyl hydrolases. *Plant J* **33:**97-106.

54. **Catanzano, F., Graziano, G., De Paola, B., Barone, G., D'Auria, S., Rossi, M. & Nucci, R. (1998)** Guanidine-induced denaturation of beta-glycosidase from *Sulfolobus solfataricus* expressed in *Escherichia coli*. *Biochemistry* **37:**14484-14490.

55. **Ceccarelli, C., Liang, Z.X., Strickler, M., Prehna, G., Goldstein, B.M., Klinman, J.P. & Bahnson, B.J. (2004)** Crystal structure and amide H/D exchange of binary complexes of alcohol dehydrogenase from *Bacillus stearothermophilus*: insight into thermostability and cofactor binding. *Biochemistry* **43:**5266-5277.

56. **Chang, C.C., Chen, T.T., Cox, B.W., Dawes, G.N., Stemmer, W.P., Punnonen, J. & Patten, P.A. (1999)** Evolution of a cytokine using DNA family shuffling. *Nat Biotechnol* **17:**793-797.

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

58. **Chiang, C. & Knight, S.G. (1960)** Metabolism of D-xylose by moulds. *Nature* **188:**79-81.

59. **Clarke, A.R. & Daffron, T.R.** 1998. Nicotinamide cofactor-dependent enzymes, p. 1-75, Comprehensive biological catalysis: a mechanistic reference, vol. 3. Academic Press, San Diego.

60. **Cline, A.L. & Hu, A.S. (1965)** Enzymatic characterization and comparison of three sugar dehydrogenases from a pseudomonad. *J Biol Chem* **240:**4493-4497.

61. **Cobucci-Ponzano, B., Trincone, A., Giordano, A., Rossi, M. & Moracci, M. (2003)** Identification of an archaeal alpha-L-fucosidase encoded by an interrupted gene. Production of a functional enzyme by mutations mimicking programmed -1 frameshifting. *J Biol Chem* **278:**14622-14631.

62. **Cobucci-Ponzano, B., Trincone, A., Giordano, A., Rossi, M. & Moracci, M. (2003)** Identification of the catalytic nucleophile of the family 29 alpha-L-fucosidase from *Sulfolobus solfataricus* via chemical rescue of an inactive mutant. *Biochemistry* **42:**9525-9531.

63. **Cohen, G.N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Querellou, J., Ripp, R., Thierry, J.C., Van der Oost, J., Weissenbach, J., Zivanovic, Y. & Forterre, P. (2003)** An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol Microbiol* **47:**1495-1512.

64. **Colombo, S., Grisa, M., Tortora, P. & Vanoni, M. (1994)** Molecular cloning, nucleotide sequence and expression of a *Sulfolobus solfataricus* gene encoding a class II fumarase. *FEBS Lett* **337:**93-98.

65. **Colonna-Cesari, F., Perahia, D., Karplus, M., Eklund, H., Braden, C.I. & Tapia, O. (1986)** Interdomain motion in liver alcohol dehydrogenase. Structural and energetic analysis of the hinge bending mode. *J Biol Chem* **261:**15273-15280.

66. **Consalvi, V., Chiaraluce, R., Politi, L., Gambacorta, A., De Rosa, M. & Scandurra, R. (1991)** Glutamate dehydrogenase from the thermoacidophilic archaebacterium *Sulfolobus solfataricus*. *Eur J Biochem* **196:**459-467.

67. **Constantinesco, F., Forterre, P. & Elie, C. (2002)** NurA, a novel 5'-3' nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea. *EMBO Rep* **3:**537-542.

68. **Contursi, P., Cannio, R., Prato, S., Fiorentino, G., Rossi, M. & Bartolucci, S. (2003)** Development of a genetic system for hyperthermophilic Archaea: expression of a moderate thermophilic bacterial alcohol dehydrogenase gene in *Sulfolobus solfataricus*. *FEMS Microbiol Lett* **218:**115-120.

69. **Corbett, K., Fordham-Skelton, A.P., Gatehouse, J.A. & Davis, B.G. (2001)** Tailoring the substrate specificity of the beta-glycosidase from the thermophilic archaeon *Sulfolobus solfataricus*. *FEBS Lett* **509:**355-360.

70. **Crameri, A., Dawes, G., Rodriguez, E., Jr., Silver, S. & Stemmer, W.P. (1997)** Molecular evolution of an arsenate detoxification pathway by DNA shuffling. *Nat Biotechnol* **15:**436-438.

71. **Crameri, A., Raillard, S.A., Bermudez, E. & Stemmer, W.P. (1998)** DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391:**288-291.

72. **Crestia, D., Guerard, C., Bolte, J. & Demuynck, C. (2001)** Rabbit muscle aldolase (RAMA) as a catalyst in a new approach for the synthesis of 3-*deoxy*-D-manno-2-octulosonic acid and analogues. *J Mol Catal B: Enzym* **11:**207-212.

73. **Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. (2004)** WebLogo: a sequence logo generator. *Genome Res* **14:**1188-1190.

74. **Cubellis, M.V., Rozzo, C., Montecucchi, P. & Rossi, M. (1990)** Isolation and sequencing of a new beta-galactosidase-encoding archaebacterial gene. *Gene* **94:**89-94.

75. **Dagley, S. & Trudgill, P.W. (1965)** The Metabolism of Galactarate, D-Glucarate and Various Pentoses

by Species of *Pseudomonas*. *Biochem J* **95:**48-58.

76. **Dahms, A.S. & Anderson, R.L. (1969)** 2-keto-3-*deoxy*-L-arabonate aldolase and its role in a new pathway of L-arabinose degradation. *Biochem Biophys Res Commun* **36:**809-814.

77. **Danson, M.J. (1988)** Archaebacteria: the comparative enzymology of their central metabolic pathways. *Adv Microb Physiol* **29:**165-231.

78. **Danson, M.J., Black, S.C., Woodland, D.L. & Wood, P.A. (1985)** Citric acid cycle enzymes of the archaebacteria: citrate synthase and succinate thiokinase. *FEBS* **179:**120-124.

79. **Darwin, C. (1859)** On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

80. **D'Auria, S., Barone, R., Rossi, M., Nucci, R., Barone, G., Fessas, D., Bertoli, E. & Tanfani, F. (1997)** Effects of temperature and SDS on the structure of beta-glycosidase from the thermophilic archaeon *Sulfolobus solfataricus*. *Biochem J* **323 (Pt 3):**833-840.

81. **Davies, G.J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A.M., Schulein, M. & Withers, S.G. (1998)** Snapshots along an enzymatic reaction coordinate: analysis of a retaining beta-glycoside hydrolase. *Biochemistry* **37:**11707-11713.

82. **de Grado, M., Castan, P. & Berenguer, J. (1999)** A high-transformation-efficiency cloning vector for *Thermus thermophilus*. *Plasmid* **42:**241-245.

83. **De Ley, J. & Doudoroff, M. (1957)** The metabolism of D-galactose in *Pseudomonas saccharophila*. *J Biol Chem* **227:**745-757.

84. **De Rosa, M., Gambacorta, A., Nicolaus, B., Giardina, P., Poerio, E. & Buonocore, V. (1984)** Glucose metabolism in the extreme thermoacidophilic archaebacterium *Sulfolobus solfataricus*. *Biochem J* **224:**407-414.

85. **Debeche, T., Bliard, C., Debeire, P. & O'Donohue, M.J. (2002)** Probing the catalytically essential residues of the alpha-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *Protein Eng* **15:**21-28.

86. **Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J. & Swanson, R.V. (1998)** The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392:**353-358.

87. **DeLano, W.L.** 2002. The PyMol Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.

88. **Deshpande, A.D., Baheti, K.G. & Chatterjee, N.R. (2004)** Degradation of beta-lactam antibiotics. *Curr Sci* **87:**1684-1695.

89. **Dey, P.M. & Pridham, J.B. (1972)** Biochemistry of alpha-galactosidases. *Adv Enzymol Relat Areas Mol Biol* **36:**91-130.

90. **Diaz, E., Ferrandez, A., Prieto, M.A. & Garcia, J.L. (2001)** Biodegradation of aromatic compounds by *Escherichia coli*. *Microbiol Mol Biol Rev* **65:**523-569.

91. **Drocourt, D., Calmels, T., Reynes, J.P., Baron, M. & Tiraby, G. (1990)** Cassettes of the *Streptoalloteichus hindustanus* ble gene for transformation of lower and higher eukaryotes to phleomycin resistance. *Nucleic Acids Res* **18:**4009.

92. **Dumas, P., Bergdoll, M., Cagnon, C. & Masson, J.M. (1994)** Crystal structure and site-directed mutagenesis of a bleomycin resistance protein and their significance for drug sequestering. *Embo J* **13:**2483-2492.

93. **Duncan, M.J. (1979)** L-arabinose metabolism in rhizobia. *J Gen Microbiol* **113:**177-179.

94. **Duncan, M.J. & Fraenkel, D.G. (1979)** alpha-Ketoglutarate dehydrogenase mutant of *Rhizobium meliloti*. *J Bacteriol* **137:**415-419.

95. **Egan, S.E., Fliege, R., Tong, S., Shibata, A., Wolf, R.E., Jr. & Conway, T. (1992)** Molecular characterization of the Entner-Doudoroff pathway in *Escherichia coli*: sequence analysis and localization of promoters for the edd-eda operon. *J Bacteriol* **174:**4638-4646.

96. **Ehrig, T., Hurley, T.D., Edenberg, H.J. & Bosron, W.F. (1991)** General base catalysis in a glutamine for histidine mutant at position 51 of human liver alcohol dehydrogenase. *Biochemistry* **30:**1062-1068.

97. **Eklund, H., Samma, J.P., Wallen, L., Branden, C.I., Akeson, A. & Jones, T.A. (1981)** Structure of a triclinic ternary complex of horse liver alcohol dehydrogenase at 2.9 A resolution. *J Mol Biol* **146:**561-587.

98. **Elcock, A.H. (1998)** The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J Mol Biol* **284:**489-502.

99. **Elferink, M.G., Albers, S.V., Konings, W.N. & 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.

100. **Elferink, M.G., Schleper, C. & Zillig, W. (1996)** Transformation of the extremely thermoacidophilic archaeon *Sulfolobus solfataricus* via a self-spreading vector. *FEMS Microbiol Lett* **137:**31-35.

101. **Elsinghorst, E.A. & Mortlock, R.P. (1994)** Molecular cloning of the *Escherichia coli* B L-fucose-D-arabinose gene cluster. *J Bacteriol* **176:**7223-7232.

102. **Emsley, P. & Cowtan, K. (2004)** Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:**2126-2132.

103. **Entner, N. & Doudoroff, M. (1952)** Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*. *J Biol Chem* **196:**853-862.

104. **Erauso, G., Marsin, S., Benbouzid-Rollet, N., Baucher, M.F., Barbeyron, T., Zivanovic, Y., Prieur, D. & Forterre, P. (1996)** Sequence of plasmid pGT5 from the archaeon *Pyrococcus abyssi*: evidence for rolling-circle replication in a hyperthermophile. *J Bacteriol* **178:**3232-3237.

105. **Erauso, G., Stedman, K.M., Van de Werken, H.J.G., Zillig, W. & Van der Oost, J. (2006)** Two novel conjugative plasmids from a single strain of *Sulfolobus*. *Microbiology* **152:**1951-1968.

106. **Ernst, H.A., Lo Leggio, L., Willemoes, M.,**

Leonard, G., Blum, P. & Larsen, S. (2006) Structure of the *Sulfolobus solfataricus* alpha-glucosidase: implications for domain conservation and substrate recognition in GH31. *J Mol Biol* **358:**1106-1124.

107. **Esposito, L., Bruno, I., Sica, F., Raia, C.A., Giordano, A., Rossi, M., Mazzarella, L. & Zagari, A. (2003)** Crystal structure of a ternary complex of the alcohol dehydrogenase from *Sulfolobus solfataricus. Biochemistry* **42:**14397-14407.

108. **Esposito, L., Sica, F., Raia, C.A., Giordano, A., Rossi, M., Mazzarella, L. & Zagari, A. (2002)** Crystal structure of the alcohol dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus* at 1.85 A resolution. *J Mol Biol* **318:**463-477.

109. **Eswaramoorthy, S., Kumaran, D., Daniels, B., Studier, F.W. & Swaminathan, S. (2003)** 1NKQ: Crystal Structure of Yeast Hypothetical Protein YNQ8_YEAST. *www.pdb.org.*

110. **Ettema, T., Van der Oost, J. & Huynen, M. (2001)** Modularity in the gain and loss of genes: applications for function prediction. *Trends Genet* **17:**485-487.

111. **Ettema, T.J.G., De Vos, W.M. & Van der Oost, J. (2005)** Discovering novel biology by in silico archaeology. *Nat Rev Microbiol* **3:**859-869.

112. **Ettema, T.J.G., Makarova, K.S., Jellema, G.L., Gierman, H.J., Koonin, E.V., Huynen, M.A., De Vos, W.M. & Van der Oost, J. (2004)** Identification and functional verification of archaeal-type phosphoenolpyruvate carboxylase, a missing link in archaeal central carbohydrate metabolism. *J Bacteriol* **186:**7754-7762.

113. **Fabry, J. (1898)** Ein Beitrag Zur Kenntnis der Purpura haemorrhagica nodularis (Purpura papulosa hemorrhagica Hebrae). *Arch Derm Syph* **43:**187-200.

114. **Fedorov, A.A., Fedorov, E.V., Sharp, A. & Almo, S.C. (2004)** 1SV6: Crystal structure of 2-hydroxypentadienoic acid hydratase from *Escherichia coli. www.pdb.org.*

115. **Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., Tang, Y., Liu, X., Han, W., Peng, X., Liu, R. & Wang, L. (2007)** Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proc Natl Acad Sci U S A* **104:**5602-5607.

116. **Fernandez-Herrero, L.A., Olabarria, G. & Berenguer, J. (1997)** Surface proteins and a novel transcription factor regulate the expression of the S-layer gene in *Thermus thermophilus* HB8. *Mol Microbiol* **24:**61-72.

117. **Fersht, A.R. (1999)** Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. Freeman, New York.

118. **Fischer, L., Bromann, R., Kengen, S.W., De Vos, W.M. & Wagner, F. (1996)** Catalytical potency of beta-glucosidase from the extremophile *Pyrococcus furiosus* in glucoconjugate synthesis. *Biotechnology (N Y)* **14:**88-91.

119. **Fitz-Gibbon, S.T., Ladner, H., Kim, U.J., Stetter, K.O., Simon, M.I. & Miller, J.H. (2002)** Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum. Proc Natl Acad Sci U S A* **99:**984-989.

120. **Fossitt, D., Mortlock, R.P., Anderson, R.L. & Wood, W.A. (1964)** Pathways of L-Arabitol and Xylitol Metabolism in *Aerobacter aerogenes. J Biol Chem* **239:**2110-2115.

121. **Frey, P.A. (1996)** The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *Faseb J* **10:**461-470.

122. **Frickey, T. & Lupas, A.N. (2004)** PhyloGenie: automated phylome generation and analysis. *Nucleic Acids Res* **32:**5231-5238.

123. **Fridjonsson, O., Watzlawick, H., Gehweiler, A. & Mattes, R. (1999)** Thermostable alpha-galactosidase from *Bacillus stearothermophilus* NUB3621: cloning, sequencing and characterization. *FEMS Microbiol Lett* **176:**147-153.

124. **Fridjonsson, O., Watzlawick, H., Gehweiler, A., Rohrhirsch, T. & Mattes, R. (1999)** Cloning of the gene encoding a novel thermostable alpha-galactosidase from *Thermus brockianus* ITI360. *Appl Environ Microbiol* **65:**3955-3963.

125. **Fridjonsson, O., Watzlawick, H. & Mattes, R. (2002)** Thermoadaptation of alpha-galactosidase AgaB1 in *Thermus thermophilus. J Bacteriol* **184:**3385-3391.

126. **Friedrich, A., Prust, C., Hartsch, T., Henne, A. & Averhoff, B. (2002)** Molecular analyses of the natural transformation machinery and identification of pilus structures in the extremely thermophilic bacterium *Thermus thermophilus* strain HB27. *Appl Environ Microbiol* **68:**745-755.

127. **Fujimoto, Z., Kaneko, S., Momma, M., Kobayashi, H. & Mizuno, H. (2003)** Crystal structure of rice alpha-galactosidase complexed with D-galactose. *J Biol Chem* **278:**20313-20318.

128. **Fukuda, E. & Wakagi, T. (2002)** Substrate recognition by 2-oxoacid:ferredoxin oxidoreductase from *Sulfolobus* sp. strain 7. *Biochim Biophys Acta* **1597:**74-80.

129. **Fukuda, W., Fukui, T., Atomi, H. & Imanaka, T. (2004)** First characterization of an archaeal GTP-dependent phosphoenolpyruvate carboxykinase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J Bacteriol* **186:**4620-4627.

130. **Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S. & Imanaka, T. (2005)** Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res* **15:**352-363.

131. **Futterer, O., Angelov, A., Liesegang, H., Gottschalk, G., Schleper, C., Schepers, B., Dock, C., Antranikian, G. & Liebl, W. (2004)** Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proc Natl Acad Sci U S A* **101:**9091-9096.

132. **Gao, Z. & Schaffer, A.A. (1999)** A novel alkaline alpha-galactosidase from melon fruit with a substrate preference for raffinose. *Plant Physiol* **119:**979-988.

133. **Garman, S.C. & Garboczi, D.N. (2004)** The

molecular defect leading to Fabry disease: structure of human alpha-galactosidase. *J Mol Biol* **337:**319-335.

134. **Garman, S.C., Hannick, L., Zhu, A. & Garboczi, D.N. (2002)** The 1.9 A structure of alpha-N-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases. *Structure (Camb)* **10:**425-434.

135. **Gatignol, A., Durand, H. & Tiraby, G. (1988)** Bleomycin resistance conferred by a drug-binding protein. *FEBS Lett* **230:**171-175.

136. **Genilloud, O., Garrido, M.C. & Moreno, F. (1984)** The transposon Tn5 carries a bleomycin-resistance determinant. *Gene* **32:**225-233.

137. **Gill, S.C. & von Hippel, P.H. (1989)** Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182:**319-326.

138. **Giometti, C.S., Reich, C., Tollaksen, S., Babnigg, G., Lim, H., Zhu, W., Yates, J. & Olsen, G. (2002)** Global analysis of a "simple" proteome: *Methanococcus jannaschii. J Chromatogr B Analyt Technol Biomed Life Sci* **782:**227-243.

139. **Giordano, A., Cannio, R., La Cara, F., Bartolucci, S., Rossi, M. & Raia, C.A. (1999)** Asn249Tyr substitution at the coenzyme binding domain activates *Sulfolobus solfataricus* alcohol dehydrogenase and increases its thermal stability. *Biochemistry* **38:**3043-3054.

140. **Giver, L., Gershenson, A., Freskgard, P.O. & Arnold, F.H. (1998)** Directed evolution of a thermostable esterase. *Proc Natl Acad Sci U S A* **95:**12809-12813.

141. **Gleissner, M., Kaiser, U., Antonopoulos, E. & Schafer, G. (1997)** The archaeal SoxABCD complex is a proton pump in *Sulfolobus acidocaldarius. J Biol Chem* **272:**8417-8426.

142. **Golubev, A.M., Nagem, R.A., Brandao Neto, J.R., Neustroev, K.N., Eneyskaya, E.V., Kulminskaya, A.A., Shabalin, K.A., Savel'ev, A.N. & Polikarpov, I. (2004)** Crystal structure of alpha-galactosidase from *Trichoderma reesei* and its complex with galactose: implications for catalytic mechanism. *J Mol Biol* **339:**413-422.

143. **Gonzalez, M., Argarana, C.E. & Fidelio, G.D. (1999)** Extremely high thermal stability of streptavidin and avidin upon biotin binding. *Biomol Eng* **16:**67-72.

144. **Grogan, D.W. (1989)** Phenotypic characterization of the archaebacterial genus *Sulfolobus*: comparison of five wild-type strains. *J Bacteriol* **171:**6710-6719.

145. **Guex, N., Diemand, A. & Peitsch, M.C. (1999)** Protein modelling for all. *Trends Biochem Sci* **24:**364-367.

146. **Guex, N. & Peitsch, M.C. (1997)** SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**.

147. **Guy, J.E., Isupov, M.N. & Littlechild, J.A. (2003)** The structure of an alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix. J Mol Biol* **331:**1041-1051.

148. **Hammes-Schiffer, S. & Benkovic, S.J. (2006)** Relating protein motion to catalysis. *Annu Rev Biochem* **75:**519-541.

149. **Hansen, T., Wendorff, D. & Schonheit, P. (2004)** Bifunctional phosphoglucose/phosphomannose isomerases from the Archaea *Aeropyrum pernix* and *Thermoplasma acidophilum* constitute a novel enzyme family within the phosphoglucose isomerase superfamily. *J Biol Chem* **279:**2262-2272.

150. **Hart, D.O., He, S., Chany, C.J., 2nd, Withers, S.G., Sims, P.F., Sinnott, M.L. & Brumer, H., 3rd (2000)** Identification of Asp-130 as the catalytic nucleophile in the main alpha-galactosidase from *Phanerochaete chrysosporium*, a family 27 glycosyl hydrolase. *Biochemistry* **39:**9826-9836.

151. **Hartl, T., Grossebuter, W., Gorisch, H. & Stezowski, J.J. (1987)** Crystalline NAD/NADP-dependent malate dehydrogenase; the enzyme from the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius. Biol Chem Hoppe Seyler* **368:**259-267.

152. **Henne, A., Bruggemann, H., Raasch, C., Wiezer, A., Hartsch, T., Liesegang, H., Johann, A., Lienard, T., Gohl, O., Martinez-Arias, R., Jacobi, C., Starkuviene, V., Schlenczeck, S., Dencker, S., Huber, R., Klenk, H.P., Kramer, W., Merkl, R., Gottschalk, G. & Fritz, H.J. (2004)** The genome sequence of the extreme thermophile *Thermus thermophilus. Nat Biotechnol* **22:**547-553.

153. **Henrissat, B. (1991)** A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **280 (Pt 2):**309-316.

154. **Hess, D., Kruger, K., Knappik, A., Palm, P. & Hensel, R. (1995)** Dimeric 3-phosphoglycerate kinases from hyperthermophilic Archaea. Cloning, sequencing and expression of the 3-phosphoglycerate kinase gene of *Pyrococcus woesei* in *Escherichia coli* and characterization of the protein. Structural and functional comparison with the 3-phosphoglycerate kinase of *Methanothermus fervidus. Eur J Biochem* **233:**227-237.

155. **Holm, L. & Sander, C. (1996)** Mapping the protein universe. *Science* **273:**595-603.

156. **Hoseki, J., Yano, T., Koyama, Y., Kuramitsu, S. & Kagamiyama, H. (1999)** Directed evolution of thermostable kanamycin-resistance gene: a convenient selection marker for *Thermus thermophilus. J Biochem (Tokyo)* **126:**951-956.

157. **Hosoya, S., Yamane, K., Takeuchi, M. & Sato, T. (2002)** Identification and characterization of the *Bacillus subtilis* D-glucarate/galactarate utilization operon ycbCDEFGHJ. *FEMS Microbiol Lett* **210:**193-199.

158. **Hottes, A.K., Meewan, M., Yang, D., Arana, N., Romero, P., McAdams, H.H. & Stephens, C. (2004)** Transcriptional profiling of *Caulobacter crescentus* during growth on complex and minimal media. *J Bacteriol* **186:**1448-1461.

159. **Huang, Y., Krauss, G., Cottaz, S., Driguez, H. & Lipps, G. (2005)** A highly acid-stable and thermostable endo-beta-glucanase from the thermoacidophilic archaeon *Sulfolobus solfataricus. Biochem J* **385:**581-588.

160. **Huber, R., Huber, H. & Stetter, K.O. (2000)** Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. *FEMS Microbiol Rev* **24:**615-623.

161. **Huh, W.K., Lee, B.H., Kim, S.T., Kim, Y.R., Rhie, G.E., Baek, Y.W., Hwang, C.S., Lee, J.S. & Kang, S.O. (1998)** D-Erythroascorbic acid is an important antioxidant molecule in *Saccharomyces cerevisiae. Mol Microbiol* **30:**895-903.

162. **Hutchins, A.M., Holden, J.F. & Adams, M.W. (2001)** Phosphoenolpyruvate synthetase from the hyperthermophilic archaeon *Pyrococcus furiosus. J Bacteriol* **183:**709-715.

163. **Huynen, M.A., Dandekar, T. & Bork, P. (1999)** Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol* **7:**281-291.

164. **Imanaka, T., Ohta, T., Sakoda, H., Widhyastuti, N. & Matsuoka, M. (1993)** Cloning, nucleotide sequence, and efficient expression of the gene coding for thermostable aldehyde dehydrogenase from *Bacillus stearothermophilus*, and characterization of the enzyme. *J Ferm Bioeng* **76:**161-167.

165. **Iqbalsyah, T.M. & Doig, A.J. (2004)** Effect of the N3 residue on the stability of the alpha-helix. *Protein Sci* **13:**32-39.

166. **Ishikawa, K., Higashi, N., Nakamura, T., Matsuura, T. & Nakagawa, A. (2007)** The First Crystal Structure of L-Threonine Dehydrogenase. *J Mol Biol* **366:**857-867.

167. **Jaenicke, R. (2000)** Stability and stabilization of globular proteins in solution. *J Biotechnol* **79:**193-203.

168. **Jan, R.L., Wu, J., Chaw, S.M., Tsai, C.W. & Tsen, S.D. (1999)** A novel species of thermoacidophilic archaeon, *Sulfolobus yangmingensis* sp. nov. *Int J Syst Bacteriol* **49 Pt 4:**1809-1816.

169. **Jansen, R., Embden, J.D., Gaastra, W. & Schouls, L.M. (2002)** Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43:**1565-1575.

170. **Janssen, S., Schafer, G., Anemuller, S. & Moll, R. (1997)** A succinate dehydrogenase with novel structure and properties from the hyperthermophilic archaeon *Sulfolobus acidocaldarius*: genetic and biophysical characterization. *J Bacteriol* **179:**5560-5569.

171. **Jaswal, S.S., Sohl, J.L., Davis, J.H. & Agard, D.A. (2002)** Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature* **415:**343-346.

172. **Jeffcoat, R., Hassall, H. & Dagley, S. (1969)** Purification and properties of D-4-*deoxy*-5-oxoglucarate hydro-lyase (decarboxylating). *Biochem J* **115:**977-983.

173. **Jensen, R.A. (1976)** Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* **30:**409-425.

174. **Johnsen, U. & Schonheit, P. (2004)** Novel xylose dehydrogenase in the halophilic archaeon *Haloarcula marismortui. J Bacteriol* **186:**6198-6207.

175. **Jonuscheit, M., Martusewitsch, E., Stedman, K.M. & Schleper, C. (2003)** A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. *Mol Microbiol* **48:**1241-1252.

176. **Kabsch, W. (1993)** Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* **26:**795-800.

177. **Kabsch, W. & Sander, C. (1983)** Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22:**2577-2637.

178. **Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. (2004)** The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32:**D277-280.

179. **Kaper, T., Brouns, S.J.J., Geerling, A.C.M., De Vos, W.M. & Van der Oost, J. (2002)** DNA family shuffling of hyperthermostable beta-glycosidases. *Biochem J* **368:**461-470.

180. **Kaper, T., Lebbink, J.H., Pouwels, J., Kopp, J., Schulz, G.E., Van der Oost, J. & De Vos, W.M. (2000)** Comparative structural analysis and substrate specificity engineering of the hyperthermostable beta-glucosidase CelB from *Pyrococcus furiosus. Biochemistry* **39:**4963-4970.

181. **Kaper, T., Van Heusden, H.H., Van Loo, B., Vasella, A., Van der Oost, J. & De Vos, W.M. (2002)** Substrate specificity engineering of beta-mannosidase and beta-glucosidase from *Pyrococcus* by exchange of unique active site residues. *Biochemistry* **41:**4147-4155.

182. **Kaper, T., Verhees, C.H., Lebbink, J.H., Van Lieshout, J.F.T., Kluskens, L.D., Ward, D.E., Kengen, S.W., Beerthuyzen, M.M., De Vos, W.M. & Van der Oost, J. (2001)** Characterization of beta-glycosylhydrolases from *Pyrococcus furiosus. Methods Enzymol* **330:**329-346.

183. **Kappler, U., Sly, L.I. & McEwan, A.G. (2005)** Respiratory gene clusters of *Metallosphaera sedula* - differential expression and transcriptional organization. *Microbiology* **151:**35-43.

184. **Kapust, R.B. & Waugh, D.S. (1999)** *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* **8:**1668-1674.

185. **Kardinahl, S., Schmidt, C.L., Hansen, T., Anemuller, S., Petersen, A. & Schafer, G. (1999)** The strict molybdate-dependence of glucose-degradation by the thermoacidophile *Sulfolobus acidocaldarius* reveals the first crenarchaeotic molybdenum containing enzyme--an aldehyde oxidoreductase. *Eur J Biochem* **260:**540-548.

186. **Karshikoff, A. & Ladenstein, R. (2001)** Ion pairs and the thermotolerance of proteins from hyperthermophiles: a “traffic rule” for hot roads. *Trends Biochem Sci* **26:**550-556.

187. **Kashefi, K. & Lovley, D.R. (2003)** Extending the upper temperature limit for life. *Science* **301:**934.

188. **Kawano, Y., Kumagai, T., Muta, K., Matoba, Y., Davies, J. & Sugiyama, M. (2000)** The 1.5 Å crystal structure of a bleomycin resistance determinant from bleomycin-producing *Streptomyces verticillus. J Mol Biol* **295:**915-925.

189. **Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai,**

A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M., Kato, Y., Yoshizawa, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Masuda, S., Yanagii, M., Nishimura, M., Yamagishi, A., Oshima, T. & Kikuchi, H. (2001) Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. *DNA Res* **8:**123-140.

190. **Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H., Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A. & Kikuchi, H. (1999)** Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res* **6:**83-101, 145-152.

191. **Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K. & Kikuchi, H. (1998)** Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res* **5:**55-76.

192. **Kawashima, T., Amano, N., Koike, H., Makino, S., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T., Nunoshiba, T., Yamamoto, Y., Aramaki, H., Makino, K. & Suzuki, M. (2000)** Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proc Natl Acad Sci U S A* **97:**14257-14262.

193. **Keller, F. & Pharr, M.** 1996. Metabolism of carbohydrates in sinks and sources: galactosyl-sucrose oligosaccharides, p. 157-183. *In* E. Zamski and A. A. Schafer (ed.), Photoassimilate Distribution in Plants and Crops. Marcel Dekker, New York.

194. **Kempton, J.B. & Withers, S.G. (1992)** Mechanism of *Agrobacterium* beta-glucosidase: kinetic studies. *Biochemistry* **31:**9961-9969.

195. **Kengen, S.W., Luesink, E.J., Stams, A.J. & Zehnder, A.J. (1993)** Purification and characterization of an extremely thermostable beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Eur J Biochem* **213:**305-312.

196. **Kerscher, L., Nowitzki, S. & Oesterhelt, D. (1982)** Thermoacidophilic archaebacteria contain bacterial-type ferredoxins acting as electron acceptors of 2-oxoacid:ferredoxin oxidoreductases. *Eur J Biochem* **128:**223-230.

197. **Kikuchi, M., Ohnishi, K. & Harayama, S. (1999)** Novel family shuffling methods for the in vitro evolution of enzymes. *Gene* **236:**159-167.

198. **Kim, M.S., Park, J.T., Kim, Y.W., Lee, H.S., Nyawira, R., Shin, H.S., Park, C.S., Yoo, S.H., Kim, Y.R., Moon, T.W. & Park, K.H. (2004)** Properties of a novel thermostable glucoamylase from the hyperthermophilic archaeon *Sulfolobus solfataricus* in relation to starch processing. *Appl Environ Microbiol* **70:**3933-3940.

199. **Kim, S. & Lee, S.B. (2006)** Catalytic Promiscuity in Dihydroxy-Acid Dehydratase from the Thermoacidophilic Archaeon *Sulfolobus solfataricus*. *J Biochem (Tokyo)* **139:**591-596.

200. **Kim, S. & Lee, S.B. (2005)** Identification and characterization of *Sulfolobus solfataricus* D-gluconate dehydratase: a key enzyme in the non-phosphorylated Entner-Doudoroff pathway. *Biochem J* **387:**271-280.

201. **Kim, Y., Joachimiak, A., Edwards, A., Skarina, T. & Savchenko, A. (2003)** 1NR9: Crystal Structure of *Escherichia coli* Putative Isomerase EC1262 (APC5008). *www.pdb.org*

202. **Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Woese, C.R. & Venter, J.C. (1997)** The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390:**364-370.

203. **Kohen, A., Cannio, R., Bartolucci, S. & Klinman, J.P. (1999)** Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase. *Nature* **399:**496-499.

204. **Kohlhoff, M., Dahm, A. & Hensel, R. (1996)** Tetrameric triosephosphate isomerase from hyperthermophilic Archaea. *FEBS Lett* **383:**245-250.

205. **Komorowski, L., Verheyen, W. & Schafer, G. (2002)** The archaeal respiratory supercomplex SoxM from *S. acidocaldarius* combines features of quinole and cytochrome c oxidases. *Biol Chem* **383:**1791-1799.

206. **Koonin, E.V., Makarova, K.S. & Aravind, L. (2001)** Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55:**709-742.

207. **Korkhin, Y., Kalb, A.J., Peretz, M., Bogin, O., Burstein, Y. & Frolow, F. (1998)** NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of *Clostridium beijerinckii* and *Thermoanaerobacter brockii*. *J Mol Biol* **278:**967-981.

208. **Koyama, Y., Hoshino, T., Tomizuka, N. & Furukawa, K. (1986)** Genetic transformation of the extreme thermophile *Thermus thermophilus* and of other *Thermus* spp. *J Bacteriol* **166:**338-340.

209. **Kraus, G.A. & Landgrebe, K. (1984)** A direct synthesis of omega-Bromo-1-alkenes. *Synthesis* **1984:**885.

210. **Krissinel, E. & Henrick, K.** 2005. Detection

of protein assemblies in crystals, p. 163-174. *In* M. R. Berthold (ed.), CompLife 2005, vol. 3695. Springer-Verlag, Berlin Heidelberg.

211. **Kuchner, O. & Arnold, F.H. (1997)** Directed evolution of enzyme catalysts. *Trends Biotechnol* **15:**523-530.

212. **Kumar, S. & Nussinov, R. (2001)** How do thermophilic proteins deal with heat? *Cell Mol Life Sci* **58:**1216-1233.

213. **Lamble, H.J., Heyer, N.I., Bull, S.D., Hough, D.W. & Danson, M.J. (2003)** Metabolic pathway promiscuity in the archaeon *Sulfolobus solfataricus* revealed by studies on glucose dehydrogenase and 2-keto-3-*deoxy*gluconate aldolase. *J Biol Chem* **278:**34066-34072.

214. **Lamble, H.J., Milburn, C.C., Taylor, G.L., Hough, D.W. & Danson, M.J. (2004)** Gluconate dehydratase from the promiscuous Entner-Doudoroff pathway in *Sulfolobus solfataricus*. *FEBS Lett* **576:**133-136.

215. **Lamble, H.J., Theodossis, A., Milburn, C.C., Taylor, G.L., Bull, S.D., Hough, D.W. & Danson, M.J. (2005)** Promiscuity in the part-phosphorylative Entner-Doudoroff pathway of the archaeon *Sulfolobus solfataricus*. *FEBS Lett* **579:**6865-6869.

216. **Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993)** PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst* **26:**283-291.

217. **Lebbink, J.H., Eggen, R.I., Geerling, A.C., Consalvi, V., Chiaraluce, R., Scandurra, R. & De Vos, W.M. (1995)** Exchange of domains of glutamate dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus* and the mesophilic bacterium *Clostridium difficile*: effects on catalysis, thermoactivity and stability. *Protein Eng* **8:**1287-1294.

218. **Lebbink, J.H., Kaper, T., Bron, P., Van der Oost, J. & De Vos, W.M. (2000)** Improving low-temperature catalysis in the hyperthermostable *Pyrococcus furiosus* beta-glucosidase CelB by directed evolution. *Biochemistry* **39:**3656-3665.

219. **Lebbink, J.H., Kaper, T., Kengen, S.W., Van der Oost, J. & De Vos, W.M. (2001)** beta-Glucosidase CelB from *Pyrococcus furiosus*: production by *Escherichia coli*, purification, and in vitro evolution. *Methods Enzymol* **330:**364-379.

220. **Lebbink, J.H., Knapp, S., Van der Oost, J., Rice, D., Ladenstein, R. & De Vos, W.M. (1999)** Engineering activity and stability of *Thermotoga maritima* glutamate dehydrogenase. II: construction of a 16-residue ion-pair network at the subunit interface. *J Mol Biol* **289:**357-369.

221. **Lebbink, J.H.G. (1999)** PhD Thesis: Molecular characterisation of the thermostability and catalytic properties of enzymes from hyperthermophiles.

222. **Lee, N., Gielow, W., Martin, R., Hamilton, E. & Fowler, A. (1986)** The organization of the *araBAD* operon of *Escherichia coli*. *Gene* **47:**231-244.

223. **Lee, R.H., Lin, M.C. & Chen, S.C. (2004)** A novel alkaline alpha-galactosidase gene is involved in rice leaf senescence. *Plant Mol Biol* **55:**281-295.

224. **Lehle, L. & Tanner, W. (1973)** The function of myo-inositol in the biosynthesis of raffinose. Purification and characterization of galactinol:sucrose 6-galactosyltransferase from *Vicia faba* seeds. *Eur J Biochem* **38:**103-110.

225. **Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., Prosser, J.I., Schuster, S.C. & Schleper, C. (2006)** Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442:**806-809.

226. **Levin, I., Meiri, G., Peretz, M., Burstein, Y. & Frolow, F. (2004)** The ternary complex of *Pseudomonas* aeruginosa alcohol dehydrogenase with NADH and ethylene glycol. *Protein Sci* **13:**1547-1556.

227. **Liao, H., McKenzie, T. & Hageman, R. (1986)** Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. *Proc Natl Acad Sci U S A* **83:**576-580.

228. **Liebl, W., Wagner, B. & Schellhase, J. (1998)** Properties of an alpha-galactosidase, and structure of its gene galA, within an alpha-and beta-galactoside utilization gene cluster of the hyperthermophilic bacterium *Thermotoga maritima*. *Syst Appl Microbiol* **21:**1-11.

229. **Lieshout, J.F.T., Verhees, C.H., Ettema, T.J.G., Van der Sar, S., Imamura, H., Matsuzawa, H., Van der Oost, J. & De Vos, W.M. (2003)** Identification and Molecular Characterization of a Novel Type of alpha-galactosidase from *Pyrococcus furiosus*. *Biocatal Biotransfor* **21:**243-252.

230. **Lillestol, R.K., Redder, P., Garrett, R.A. & Brugger, K. (2006)** A putative viral defence mechanism in archaeal cells. *Archaea* **2:**59-72.

231. **Limauro, D., Cannio, R., Fiorentino, G., Rossi, M. & Bartolucci, S. (2001)** Identification and molecular characterization of an endoglucanase gene, *celS*, from the extremely thermophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* **5:**213-219.

232. **Lohlein-Werhahn, G., Goepfert, P. & Eggerer, H. (1988)** Purification and properties of an archaebacterial enzyme: citrate synthase from *Sulfolobus solfataricus*. *Biol Chem Hoppe Seyler* **369:**109-113.

233. **Londei, P. (2005)** Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol Rev* **29:**185-200.

234. **Lorimer, I.A. & Pastan, I. (1995)** Random recombination of antibody single chain Fv sequences after fragmentation with DNaseI in the presence of Mn^{2+}. *Nucleic Acids Res* **23:**3067-3068.

235. **Lubelska, J.M., Jonuscheit, M., Schleper, C., Albers, S.V. & Driessen, A.J. (2006)** Regulation of expression of the arabinose and glucose transporter genes in the thermophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* **10:**383-391.

236. **Lundgren, M., Andersson, A., Chen, L., Nilsson, P. & Bernander, R. (2004)** Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A* **101:**7046-7051.

237. **Lundgren, M. & Bernander, R. (2005)** Archaeal cell cycle progress. *Curr Opin Microbiol* **8:**662-668.

238. **Lusty, C.J. (1999)** A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. *J Appl Cryst* **32:**106-112.

239. **Ly, H.D., Howard, S., Shum, K., He, S., Zhu, A. & Withers, S.G. (2000)** The synthesis, testing and use of 5-fluoro-alpha-D-galactosyl fluoride to trap an intermediate on green coffee bean alpha-galactosidase and identify the catalytic nucleophile. *Carbohydr Res* **329:**539-547.

240. **Machius, M., Declerck, N., Huber, R. & Wiegand, G. (2003)** Kinetic stabilization of *Bacillus licheniformis* alpha-amylase through introduction of hydrophobic residues at the surface. *J Biol Chem* **278:**11546-11553.

241. **Macritchie, J.A., Silcock, A. & Willis, C.L. (1997)** Enantioselective synthesis of unsaturated alpha-hydroxy acids. *Tetrahedron: Asymmetry* **8:**3895-3902.

242. **Magonet, E., Hayen, P., Delforge, D., Delaive, E. & Remacle, J. (1992)** Importance of the structural zinc atom for the stability of yeast alcohol dehydrogenase. *Biochem J* **287 (Pt 2):**361-365.

243. **Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I. & Koonin, E.V. (2006)** A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1:**7.

244. **Makarova, K.S. & Koonin, E.V. (2003)** Comparative genomics of Archaea: how much have we learned in six years, and what's next? *Genome Biol* **4:**115.

245. **Mandelman, D., Schwarz, F.P., Li, H. & Poulos, T.L. (1998)** The role of quaternary interactions on the stability and activity of ascorbate peroxidase. *Protein Sci* **7:**2089-2098.

246. **Manjasetty, B.A., Niesen, F.H., Delbruck, H., Gotz, F., Sievert, V., Bussow, K., Behlke, J. & Heinemann, U. (2004)** X-ray structure of fumarylacetoacetate hydrolase family member *Homo sapiens* FLJ36880. *Biol Chem* **385:**935-942.

247. **Martin, A., Sieber, V. & Schmid, F.X. (2001)** In-vitro selection of highly stabilized protein variants with optimized surface. *J Mol Biol* **309:**717-726.

248. **Martusewitsch, E., Sensen, C.W. & Schleper, C. (2000)** High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements. *J Bacteriol* **182:**2574-2581.

249. **Maruta, K., Mitsuzumi, H., Nakada, T., Kubota, M., Chaen, H., Fukuda, S., Sugimoto, T. & Kurimoto, M. (1996)** Cloning and sequencing of a cluster of genes encoding novel enzymes of trehalose biosynthesis from thermophilic archaebacterium *Sulfolobus acidocaldarius*. *Biochim Biophys Acta* **1291:**177-181.

250. **Maruyama, M., Kumagai, T., Matoba, Y., Hayashida, M., Fujii, T., Hata, Y. & Sugiyama, M. (2001)** Crystal structures of the transposon Tn5-carried bleomycin resistance determinant uncomplexed and complexed with bleomycin. *J Biol Chem* **276:**9992-9999.

251. **Mathias, A.L., Rigo, L.U., Funayama, S. & Pedrosa, F.O. (1989)** L-arabinose metabolism in *Herbaspirillum seropedicae*. *J Bacteriol* **171:**5206-5209.

252. **Matsumi, R., Manabe, K., Fukui, T., Atomi, H. & Imanaka, T. (2007)** Disruption of a sugar transporter gene cluster in a hyperthermophilic archaeon using a host-marker system based on antibiotic resistance. *J Bacteriol* **189:**2683-2691.

253. **Matthews, B.W. (1968)** Solvent content of protein crystals. *J Mol Biol* **33:**491-497.

254. **McIlvaine, T.C. (1921)** A buffer solution for colorimetric comparison. *J Biol Chem* **49:**183-186.

255. **McRee, D.E. (1999)** XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* **125:**156-165.

256. **Meisenzahl, A.C., Shapiro, L. & Jenal, U. (1997)** Isolation and characterization of a xylose-dependent promoter from *Caulobacter crescentus*. *J Bacteriol* **179:**592-600.

257. **Merz, A., Yee, M.C., Szadkowski, H., Pappenberger, G., Crameri, A., Stemmer, W.P., Yanofsky, C. & Kirschner, K. (2000)** Improving the catalytic activity of a thermophilic enzyme at low temperatures. *Biochemistry* **39:**880-889.

258. **Milburn, C.C., Lamble, H.J., Theodossis, A., Bull, S.D., Hough, D.W., Danson, M.J. & Taylor, G.L. (2006)** The structural basis of substrate promiscuity in glucose dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J Biol Chem* **281:**14796-14804.

259. **Mizutani, H. & Kunishima, M. (2006)** 1WZO: Crystal structure of HpcE from *Thermus thermophilus* HB8. *www.pdb.org*.

260. **Mizutani, H. & Kunishima, M. (2006)** 2DFU: Crystal structure of the 2-hydroxyhepta-2,4-diene-1-,7-dioate isomerase from *Thermus thermophilus* HB8. *www.pdb.org*.

261. **Mobley, P.W., Metzger, R.P. & Wick, A.N. (1975)** L-fucose dehydrogenase from sheep liver. *Methods Enzymol* **41:**173-177.

262. **Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. (2005)** Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60:**174-182.

263. **Moore, R.A., Reckseidler-Zenteno, S., Kim, H., Nierman, W., Yu, Y., Tuanyok, A., Warawa, J., DeShazer, D. & Woods, D.E. (2004)** Contribution of gene loss to the pathogenic evolution of *Burkholderia pseudomallei* and *Burkholderia mallei*. *Infect Immun* **72:**4172-4187.

264. **Moracci, M., Cobucci Ponzano, B., Trincone, A., Fusco, S., De Rosa, M., Van der Oost, J., Sensen, C.W., Charlebois, R.L. & Rossi, M. (2000)** Identification and molecular characterization of the first alpha -xylosidase from an archaeon. *J Biol Chem* **275:**22082-22089.

265. **Moracci, M., Trincone, A., Perugino, G., Ciaramella, M. & Rossi, M. (1998)** Restoration of the activity of active-site mutants of the hyperthermophilic beta-glycosidase from *Sulfolobus solfataricus*: depen-

dence of the mechanism on the action of external nucleophiles. *Biochemistry* **37:**17262-17270.

266. **Morana, A., Paris, O., Maurelli, L., Rossi, M. & Cannio, R. (2007)** Gene cloning and expression in *Escherichia coli* of a bi-functional beta-D-xylosidase/alpha-L-arabinosidase from *Sulfolobus solfataricus* involved in xylan degradation. *Extremophiles* **11:**123-132.

267. **Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchell, A., Nikolskaya, A.N., Orchard, S., Pagni, M., Ponting, C.P., Quevillon, E., Selengut, J., Sigrist, C.J., Silventoinen, V., Studholme, D.J., Vaughan, R. & Wu, C.H. (2005)** InterPro, progress and status in 2005. *Nucleic Acids Res* **33:**D201-205.

268. **Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997)** Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:**240-255.

269. **Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997)** Refinement of macromolecular structures by the maximum-likelyhood method. *Acta Crystallogr D Biol Crystallogr* **53:**240-255.

270. **Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. & Fraser, C.M. (1999)** Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399:**323-329.

271. **Ness, J.E., Welch, M., Giver, L., Bueno, M., Cherry, J.R., Borchert, T.V., Stemmer, W.P. & Minshull, J. (1999)** DNA shuffling of subgenomic sequences of subtilisin. *Nat Biotechnol* **17:**893-896.

272. **Nishimasu, H., Fushinobu, S., Shoun, H. & Wakagi, T. (2004)** The first crystal structure of the novel class of fructose-1,6-bisphosphatase present in thermophilic archaea. *Structure (Camb)* **12:**949-959.

273. **Nishimasu, H., Fushinobu, S., Shoun, H. & Wakagi, T. (2006)** Identification and characterization of an ATP-dependent hexokinase with broad substrate specificity from the hyperthermophilic archaeon *Sulfolobus tokodaii*. *J Bacteriol* **188:**2014-2019.

274. **Noll, K.M. & Vargas, M. (1997)** Recent advances in genetic analyses of hyperthermophilic archaea and bacteria. *Arch Microbiol* **168:**73-80.

275. **Notredame, C., Higgins, D.G. & Heringa, J. (2000)** T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302:**205-217.

276. **Novick, N.J. & Tyler, M.E. (1982)** L-arabinose metabolism in *Azospirillum brasiliense*. *J Bacteriol* **149:**364-367.

277. **Nuttall, S.D., Deutschel, S.E., Irving, R.A., Serrano-Gomicia, J.A. & Dyall-Smith, M.L. (2000)** The ShBle resistance determinant from *Streptoalloteichus hindustanus* is expressed in *Haloferax volcanii* and confers resistance to bleomycin. *Biochem J* **346 Pt 2:**251-254.

278. **Osterman, A. & Overbeek, R. (2003)** Missing genes in metabolic pathways: a comparative genomics approach. *Curr Opin Chem Biol* **7:**238-251.

279. **Ostermeier, M., Nixon, A.E., Shim, J.H. & Benkovic, S.J. (1999)** Combinatorial protein engineering by incremental truncation. *Proc Natl Acad Sci U S A* **96:**3562-3567.

280. **Otwinowski, A. & Minor, W. (1997)** Processing of X-ray diffraction data collected in oscillation mode. *Meth Enzymol* **276:**307-326.

281. **Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E., Jr., Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I., Bhattacharyya, A., Burd, H., Gardner, W., Hanke, P., Kapatral, V., Mikhailova, N., Vasieva, O., Osterman, A., Vonstein, V., Fonstein, M., Ivanova, N. & Kyrpides, N. (2003)** The ERGO genome analysis and discovery system. *Nucleic Acids Res* **31:**164-171.

282. **Pace, N.R. (1997)** A molecular view of microbial diversity and the biosphere. *Science* **276:**734-740.

283. **Palackal, N., Brennan, Y., Callen, W.N., Dupree, P., Frey, G., Goubet, F., Hazlewood, G.P., Healey, S., Kang, Y.E., Kretz, K.A., Lee, E., Tan, X., Tomlinson, G.L., Verruto, J., Wong, V.W., Mathur, E.J., Short, J.M., Robertson, D.E. & Steer, B.A. (2004)** An evolutionary route to xylanase process fitness. *Protein Sci* **13:**494-503.

284. **Palleroni, N.J. & Doudoroff, M. (1956)** Characterization and properties of 2-keto-3-*deoxy*-D-arabonic acid. *J Biol Chem* **223:**499-508.

285. **Palleroni, N.J. & Doudoroff, M. (1957)** Metabolism of carbohydrates by *Pseudomonas saccharophila*. III. Oxidation of D-arabinose. *J Bacteriol* **74:**180-185.

286. **Panday, N., Canac, Y. & Vasella, A. (2000)** Very strong inhibition of glucosidases by C2-substituted tetrahydroimidazopyridines. *Helv Chim Acta* **83:**58-79.

287. **Pappin, D.J., Hojrup, P. & Bleasby, A.J. (1993)** Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **3:**327-332.

288. **Park, S.H., Pastuszak, I., Drake, R. & Elbein, A.D. (1998)** Purification to apparent homogeneity and properties of pig kidney L-fucose kinase. *J Biol Chem* **273:**5685-5691.

289. **Pauly, T.A., Ekstrom, J.L., Beebe, D.A., Chrunyk, B., Cunningham, D., Griffor, M., Kamath, A., Lee, S.E., Madura, R., McGuire, D., Subashi, T., Wasilko, D., Watts, P., Mylari, B.L., Oates, P.J., Adams, P.D. & Rath, V.L. (2003)** X-ray crystallographic and kinetic studies of human sorbitol dehydrogenase. *Structure* **11:**1071-1085.

290. **Peak, M.J., Peak, J.G., Stevens, F.J., Blamey, J., Mai, X., Zhou, Z.H. & Adams, M.W. (1994)** The hyperthermophilic glycolytic enzyme enolase in the

archaeon, *Pyrococcus furiosus*: comparison with mesophilic enolases. *Arch Biochem Biophys* **313:**280-286.

291. **Pedrosa, F.O. & Zancan, G.T. (1974)** L-Arabinose metabolism in *Rhizobium japonicum*. *J Bacteriol* **119:**336-338.

292. **Penel, S., Hughes, E. & Doig, A.J. (1999)** Side-chain structures in the first turn of the alpha-helix. *J Mol Biol* **287:**127-143.

293. **Perl, D., Mueller, U., Heinemann, U. & Schmid, F.X. (2000)** Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat Struct Biol* **7:**380-383.

294. **Perrakis, A., Morris, R. & Lamzin, V.S. (1999)** Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6:**458-463.

295. **Perrakis, A., Sixma, T.K., Wilson, K.S. & Lamzin, V.S. (1997)** wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr D Biol Crystallogr* **53:**448-455.

296. **Peterbauer, T., Mach, L., Mucha, J. & Richter, A. (2002)** Functional expression of a cDNA encoding pea (*Pisum sativum* L.) raffinose synthase, partial purification of the enzyme from maturing seeds, and steady-state kinetic analysis of raffinose synthesis. *Planta* **215:**839-846.

297. **Peterson, J.D., Umayam, L.A., Dickinson, T., Hickey, E.K. & White, O. (2001)** The Comprehensive Microbial Resource. *Nucleic Acids Res* **29:**123-125.

298. **Petzelbauer, I., Nidetzky, B., Haltrich, D. & Kulbe, K.D. (1999)** Development of an ultra-high-temperature process for the enzymatic hydrolysis of lactose. I. The properties of two thermostable beta-glycosidases. *Biotechnol Bioeng* **64:**322-332.

299. **Petzelbauer, I., Splechtna, B. & Nidetzky, B. (2002)** Development of an ultrahigh-temperature process for the enzymatic hydrolysis of lactose. III. Utilization of two thermostable beta-glycosidases in a continuous ultrafiltration membrane reactor and galacto-oligosaccharide formation under steady-state conditions. *Biotechnol Bioeng* **77:**394-404.

300. **Pisani, F.M., Rella, R., Raia, C.A., Rozzo, C., Nucci, R., Gambacorta, A., De Rosa, M. & Rossi, M. (1990)** Thermostable beta-galactosidase from the archaebacterium *Sulfolobus solfataricus*. Purification and properties. *Eur J Biochem* **187:**321-328.

301. **Pitcher, D.G., Saunders, N.A. & Owen, R.J. (1989)** Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett Appl Microbiol* **8:**151-156.

302. **Poddevin, B., Orlowski, S., Belehradek, J., Jr. & Mir, L.M. (1991)** Very high cytotoxicity of bleomycin introduced into the cytosol of cells in culture. *Biochem Pharmacol* **42 Suppl:**S67-75.

303. **Pollard, J.R. & Bugg, T.D. (1998)** Purification, characterisation and reaction mechanism of monofunctional 2-hydroxypentadienoic acid hydratase from *Escherichia coli*. *Eur J Biochem* **251:**98-106.

304. **Portsmouth, D., Stoolmiller, A.C. & Abeles, R.H. (1967)** Studies on the mechanism of action of 2-keto-3-*deoxy*-L-arabonate dehydratase. The participation of an enzyme-substrate Schiff base in a dehydration. *J Biol Chem* **242:**2751-2759.

305. **Pouwels, J., Moracci, M., Cobucci-Ponzano, B., Perugino, G., Van der Oost, J., Kaper, T., Lebbink, J.H., De Vos, W.M., Ciaramella, M. & Rossi, M. (2000)** Activity and stability of hyperthermophilic enzymes: a comparative study on two archaeal beta-glycosidases. *Extremophiles* **4:**157-164.

306. **Prangishvili, D., Albers, S.V., Holz, I., Arnold, H.P., Stedman, K., Klein, T., Singh, H., Hiort, J., Schweier, A., Kristjansson, J.K. & Zillig, W. (1998)** Conjugation in archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*. *Plasmid* **40:**190-202.

307. **Puchegger, S., Redl, B. & Stoffler, G. (1990)** Purification and properties of a thermostable fumarate hydratase from the archaeobacterium *Sulfolobus solfataricus*. *J Gen Microbiol* **136:**1537-1541.

308. **Ramaswamy, S., el Ahmad, M., Danielsson, O., Jornvall, H. & Eklund, H. (1996)** Crystal structure of cod liver class I alcohol dehydrogenase: substrate pocket and structurally variable segments. *Protein Sci* **5:**663-671.

309. **Ramaswamy, S., Kratzer, D.A., Hershey, A.D., Rogers, P.H., Arnone, A., Eklund, H. & Plapp, B.V. (1994)** Crystallization and preliminary crystallographic studies of *Saccharomyces cerevisiae* alcohol dehydrogenase I. *J Mol Biol* **235:**777-779.

310. **Ramaswamy, S., Park, D.H. & Plapp, B.V. (1999)** Substitutions in a flexible loop of horse liver alcohol dehydrogenase hinder the conformational change and unmask hydrogen transfer. *Biochemistry* **38:**13951-13959.

311. **Ramaswamy, S.G. (1984)** Hydroxyproline 2-epimerase of *Pseudomonas*. Subunit structure and active site studies. *J Biol Chem* **259:**249-254.

312. **Rao, S.T. & Rossmann, M.G. (1973)** Comparison of super-secondary structures in proteins. *J Mol Biol* **76:**241-256.

313. **Read, R.J. (1986)** Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallograph Sect A Found Crystallogr* **42:**140-149.

314. **Reher, M., Bott, M. & Schonheit, P. (2006)** Characterization of glycerate kinase (2-phosphoglycerate forming), a key enzyme of the nonphosphorylative Entner-Doudoroff pathway, from the thermoacidophilic euryarchaeon *Picrophilus torridus*. *FEMS Microbiol Lett* **259:**113-119.

315. **Reilly, M.S. & Grogan, D.W. (2001)** Characterization of intragenic recombination in a hyperthermophilic archaeon via conjugational DNA exchange. *J Bacteriol* **183:**2943-2946.

316. **Reiter, W., Palm, P., Yeats, S. & Zillig, W. (1987)** Gene expression in archaebacteria: Physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. *Mol Gen Genet* **209:**270-275.

317. **Rivera, M.C. & Lake, J.A. (2004)** The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431:**152-155.

318. **Robb, F.T., Maeder, D.L., Brown, J.R., DiRuggiero, J., Stump, M.D., Yeh, R.K., Weiss, R.B. & Dunn, D.M. (2001)** Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. *Methods Enzymol* **330:**134-157.

319. **Rolfsmeier, M. & Blum, P. (1995)** Purification and characterization of a maltase from the extremely thermophilic crenarchaeote *Sulfolobus solfataricus. J Bacteriol* **177:**482-485.

320. **Rondeau, J.M., Cagnon, C., Moras, D. & Masson, J.M. (1989)** Crystallization and preliminary X-ray data of a phleomycin-binding protein from *Streptoalloteichus hindustanus. J Mol Biol* **207:**645-646.

321. **Ronimus, R.S. & Morgan, H.W. (2003)** Distribution and phylogenies of enzymes of the Embden-Meyerhof-Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. *Archaea* **1:**199-221.

322. **Roovers, M., Sanchez, R., Legrain, C. & Glansdorff, N. (2001)** Experimental evolution of enzyme temperature activity profile: selection in vivo and characterization of low-temperature-adapted mutants of *Pyrococcus furiosus* ornithine carbamoyltransferase. *J Bacteriol* **183:**1101-1105.

323. **Rosenfeld, S.A., Stevis, P.E. & Ho, N.W. (1984)** Cloning and characterization of the *xyl* genes from *Escherichia coli. Mol Gen Genet* **194:**410-415.

324. **Rothschild, L.J. & Mancinelli, R.L. (2001)** Life in extreme environments. *Nature* **409:**1092-1101.

325. **Ruepp, A., Graml, W., Santos-Martinez, M.L., Koretke, K.K., Volker, C., Mewes, H.W., Frishman, D., Stocker, S., Lupas, A.N. & Baumeister, W. (2000)** The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum. Nature* **407:**508-513.

326. **Russo, A.D., Rullo, R., Masullo, M., Ianniciello, G., Arcari, P. & Bocchini, V. (1995)** Glyceraldehyde-3-phosphate dehydrogenase in the hyperthermophilic archaeon *Sulfolobus solfataricus*: characterization and significance in glucose metabolism. *Biochem Mol Biol Int* **36:**123-135.

327. **Sabin, C., Mitchell, E.P., Pokorna, M., Gautier, C., Utille, J.P., Wimmerova, M. & Imberty, A. (2006)** Binding of different monosaccharides by lectin PA-IIL from *Pseudomonas aeruginosa*: thermodynamics data correlated with X-ray structures. *FEBS Lett* **580:**982-987.

328. **Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. (1988)** Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:**487-491.

329. **Sako, Y., Takai, K., Uchida, A. & Ishida, Y. (1996)** Purification and characterization of phosphoenolpyruvate carboxylase from the hyperthermophilic archaeon *Methanothermus sociabilis. FEBS Lett* **392:**148-152.

330. **Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989)** Molecular Cloning: A Laboratory Manual, 2nd Ed. ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

331. **Santos, H. & da Costa, M.S. (2001)** Organic solutes from thermophiles and hyperthermophiles. *Methods Enzymol* **334:**302-315.

332. **Sato, T., Fukui, T., Atomi, H. & Imanaka, T. (2003)** Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J Bacteriol* **185:**210-220.

333. **Satory, M., Furlinger, M., Haltrich, D., Kulbe, K.D., Pittner, F. & Nidetzky, B. (1997)** Continuous enzymatic production of lactobionic acid using glucose-fructose oxidoreductase in an ultrafiltration membrane bioreactor. *Biotech Lett* **19:**1205-1208.

334. **Schafer, G. (1996)** Bioenergetics of the archaebacterium *Sulfolobus. Biochim Biophys Acta* **1277:**163-200.

335. **Schelert, J., Dixit, V., Hoang, V., Simbahan, J., Drozda, M. & Blum, P. (2004)** Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon *Sulfolobus solfataricus* by use of gene disruption. *J Bacteriol* **186:**427-437.

336. **Schleper, C., Holz, I., Janekovic, D., Murphy, J. & Zillig, W. (1995)** A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating. *J Bacteriol* **177:**4417-4426.

337. **Schleper, C., Kubo, K. & Zillig, W. (1992)** The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. *Proc Natl Acad Sci U S A* **89:**7645-7649.

338. **Schleper, C., Roder, R., Singer, T. & Zillig, W. (1994)** An insertion element of the extremely thermophilic archaeon *Sulfolobus solfataricus* transposes into the endogenous beta-galactosidase gene. *Mol Gen Genet* **243:**91-96.

339. **Schmidt, A., Bahr, U. & Karas, M. (2001)** Influence of pressure in the first pumping stage on analyte desolvation and fragmentation in nano-ESI MS. *Anal Chem* **73:**6040-6046.

340. **Schmidt, K.J., Beck, K.E. & Grogan, D.W. (1999)** UV stimulation of chromosomal marker exchange in *Sulfolobus acidocaldarius*: implications for DNA repair, conjugation and homologous recombination at extremely high temperatures. *Genetics* **152:**1407-1415.

341. **Schmidt, S., Sunyaev, S., Bork, P. & Dandekar, T. (2003)** Metabolites: a helping hand for pathway evolution? *Trends Biochem Sci* **28:**336-341.

342. **Schneider, T.R. & Sheldrick, G.M. (2002)** Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58:**1772-1779.

343. **Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. & Schomburg, D. (2004)** BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* **32:**D431-433.

344. **Schonheit, P. & Schafer, G. (1995)** Metabolism of hyperthermophiles. *World J Microbiol Biotechnol* **11:**26-54.

345. **Schramm, A., Siebers, B., Tjaden, B., Brinkmann, H. & Hensel, R. (2000)** Pyruvate kinase of the hyperthermophilic crenarchaeote *Thermoproteus tenax*: physiological role and phylogenetic aspects. *J Bacteriol* **182:**2001-2009.

346. **Schut, G.J., Brehm, S.D., Datta, S. & Adams, M.W. (2003)** Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides. *J Bacteriol* **185:**3935-3947.

347. **Seemann, J.E. & Schulz, G.E. (1997)** Structure and mechanism of L-fucose isomerase from *Escherichia coli*. *J Mol Biol* **273:**256-268.

348. **Selig, M., Xavier, K.B., Santos, H. & Schonheit, P. (1997)** Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga*. *Arch Microbiol* **167:**217-232.

349. **Semon, D., Movva, N.R., Smith, T.F., el Alama, M. & Davies, J. (1987)** Plasmid-determined bleomycin resistance in *Staphylococcus aureus*. *Plasmid* **17:**46-53.

350. **Serrano, L. & Fersht, A.R. (1989)** Capping and alpha-helix stability. *Nature* **342:**296-299.

351. **Shafikhani, S., Siegel, R.A., Ferrari, E. & Schellenberger, V. (1997)** Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. *Biotechniques* **23:**304-310.

352. **Sharma, B.S. & Blumenthal, H.J. (1973)** Catabolism of D-glucaric acid to alpha-ketoglutarate in *Bacillus megaterium*. *J Bacteriol* **116:**1346-1354.

353. **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., Erauso, G., Fletcher, C., Gordon, P.M., Heikamp-de Jong, I., Jeffries, A.C., Kozera, C.J., Medina, N., Peng, X., Thi-Ngoc, H.P., Redder, P., Schenk, M.E., Theriault, C., Tolstrup, N., Charlebois, R.L., Doolittle, W.F., Duguet, M., Gaasterland, T., Garrett, R.A., Ragan, M.A., Sensen, C.W. & Van der Oost, J. (2001)** The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* **98:**7835-7840.

354. **Shivvers, D.W. & Brock, T.D. (1973)** Oxidation of elemental sulfur by *Sulfolobus acidocaldarius*. *J Bacteriol* **114:**706-710.

355. **Sieber, V., Martinez, C.A. & Arnold, F.H. (2001)** Libraries of hybrid proteins from distantly related sequences. *Nat Biotechnol* **19:**456-460.

356. **Sieber, V., Pluckthun, A. & Schmid, F.X. (1998)** Selecting proteins with improved stability by a phage-based method. *Nat Biotechnol* **16:**955-960.

357. **Siebers, B., Brinkmann, H., Dorr, C., Tjaden, B., Lilie, H., Van der Oost, J. & 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.

358. **Siebers, B. & Schonheit, P. (2005)** Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. *Curr Opin Microbiol* **8:**695-705.

359. **Siebers, B., Tjaden, B., Michalke, K., Dorr, C., Ahmed, H., Zaparty, M., Gordon, P., Sensen, C.W., Zibat, A., Klenk, H.P., Schuster, S.C. & Hensel, R. (2004)** Reconstruction of the central carbohydrate metabolism of *Thermoproteus tenax* by use of genomic and biochemical data. *J Bacteriol* **186:**2179-2194.

360. **Sikic, B.I. (1986)** Biochemical and cellular determinants of bleomycin cytotoxicity. *Cancer Surv* **5:**81-91.

361. **Singh, R.M. & Adams, E. (1965)** Enzymatic deamination of delta-1-pyrroline-4-hydroxy-2-carboxylate to 2,5-dioxovalerate (alpha-ketoglutaric semialdehyde). *J Biol Chem* **240:**4344-4351.

362. **Skorko, R., Osipiuk, J. & Stetter, K.O. (1989)** Glycogen-bound polyphosphate kinase from the archaebacterium *Sulfolobus acidocaldarius*. *J Bacteriol* **171:**5162-5164.

363. **Slesarev, A.I., Mezhevaya, K.V., Makarova, K.S., Polushin, N.N., Shcherbinina, O.V., Shakhova, V.V., Belova, G.I., Aravind, L., Natale, D.A., Rogozin, I.B., Tatusov, R.L., Wolf, Y.I., Stetter, K.O., Malykh, A.G., Koonin, E.V. & Kozyavkin, S.A. (2002)** The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc Natl Acad Sci U S A* **99:**4644-4649.

364. **Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J.N. & et al. (1997)** Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol* **179:**7135-7155.

365. **Smith, J.S. & Scholtz, J.M. (1998)** Energetics of polar side-chain interactions in helical peptides: salt effects on ion pairs and hydrogen bonds. *Biochemistry* **37:**33-40.

366. **Smith, K.S. & Ingram-Smith, C. (2007)** *Methanosaeta*, the forgotten methanogen? *Trends Microbiol* **15:**150-155.

367. **Smith, L.D., Stevenson, K.J., Hough, D.W. & Danson, M.J. (1987)** Citrate synthase from the thermophilic archaebacteria *Thermoplasma acidophilum* and *Sulfolobus acidocaldarius*. *FEBS Lett* **225:**277-281.

368. **Snijders, A.P.L., De Vos, M.G., De Koning, B. & Wright, P.C. (2005)** A fast method for quantitative proteomics based on a combination between two-dimensional electrophoresis and 15N-metabolic labelling. *Electrophoresis* **26:**3191-3199.

369. **Snijders, A.P.L., De Vos, M.G., De Koning, B. & Wright, P.C. (2005)** A fast method for quantitative proteomics based on a combination between two-dimensional electrophoresis and (15)N-metabolic labelling. *Electrophoresis* **26:**3191-3199.

371. **Snijders, A.P.L., Walther, J., Peter, S., Kinnman, I., De Vos, M.G.J., Van de Werken, H.J.G., Brouns, S.J.J., Van der Oost, J. & Wright, P.C. (2006)** Reconstruction of central carbon metabolism in *Sulfolobus solfataricus* using a two-dimensional gel electropho-

resis map, stable isotope labelling and DNA microarray analysis. *Proteomics* **6:**1518-1529.

372. **Solow, B., Bischoff, K.M., Zylka, M.J. & Kennelly, P.J. (1998)** Archael phosphoproteins. Identification of a hexosephosphate mutase and the alpha-subunit of succinyl-CoA synthetase in the extreme acidothermophile *Sulfolobus solfataricus*. *Protein Sci* **7:**105-111.

373. **Sperber, N., Zaugg, H.E. & Sandstrom, W.M. (1947)** The controlled sodium amalgam reduction of aldonolactones and their esters to aldoses and an improved synthesis of D-arabinose. *J Am Chem Soc* **69:**915-920.

374. **Spiller, B., Gershenson, A., Arnold, F.H. & Stevens, R.C. (1999)** A structural view of evolutionary divergence. *Proc Natl Acad Sci U S A* **96:**12305-12310.

375. **Stedman, K.M., Schleper, C., Rumpf, E. & Zillig, W. (1999)** Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. *Genetics* **152:**1397-1405.

376. **Stedman, K.M., She, Q., Phan, H., Holz, I., Singh, H., Prangishvili, D., Garrett, R. & Zillig, W. (2000)** pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota. *J Bacteriol* **182:**7014-7020.

377. **Stemmer, W.P. (1994)** DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A* **91:**10747-10751.

378. **Stemmer, W.P. (1994)** Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370:**389-391.

379. **Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M. & Heyneker, H.L. (1995)** Single-step assembly of a gene and entire plasmid from large numbers of oligo*deoxy*ribonucleotides. *Gene* **164:**49-53.

380. **Stephens, C., Christen, B., Fuchs, T., Sundaram, V., Watanabe, K. & Jenal, U. (2007)** Genetic analysis of a novel pathway for D-xylose metabolism in *Caulobacter crescentus*. *J Bacteriol* **189:**2181-2185.

381. **Stetter, K.O. (1999)** Extremophiles and their adaptation to hot environments. *FEBS Lett* **452:**22-25.

382. **Stetter, K.O. (1996)** Hyperthermophilic prokaryotes. *FEMS Microbiol Rev* **18:**149-158.

383. **St-Louis, M. & Tanguay, R.M. (1997)** Mutations in the fumarylacetoacetate hydrolase gene causing hereditary tyrosinemia type I: overview. *Hum Mutat* **9:**291-299.

384. **Stoolmiller, A.C. & Abeles, R.H. (1966)** Formation of alpha-ketoglutaric semialdehyde from L-2-keto-3-*deoxy*arabonic acid and isolation of L-2-keto-3-*deoxy*arabonate dehydratase from *Pseudomonas saccharophila*. *J Biol Chem* **241:**5764-5771.

385. **Sugiyama, M. & Kumagai, T. (2002)** Molecular and structural biology of bleomycin and its resistance determinants. *J Biosci Bioeng* **93:**105-116.

386. **Sugiyama, M., Kumagai, T., Hayashida, M., Maruyama, M. & Matoba, Y. (2002)** The 1.6-A crystal structure of the copper(II)-bound bleomycin complexed with the bleomycin-binding protein from bleomycin-producing *Streptomyces verticillus*. *J Biol Chem* **277:**2311-2320.

387. **Sugiyama, M., Thompson, C.J., Kumagai, T., Suzuki, K., Deblaere, R., Villarroel, R. & Davies, J. (1994)** Characterisation by molecular cloning of two genes from *Streptomyces verticillus* encoding resistance to bleomycin. *Gene* **151:**11-16.

388. **Suzuki, T., Iwasaki, T., Uzawa, T., Hara, K., Nemoto, N., Kon, T., Ueki, T., Yamagishi, A. & Oshima, T. (2002)** *Sulfolobus tokodaii* sp. nov. (f. *Sulfolobus* sp. strain 7), a new member of the genus *Sulfolobus* isolated from Beppu Hot Springs, Japan. *Extremophiles* **6:**39-44.

389. **Svensson, S., Hoog, J.O., Schneider, G. & Sandalova, T. (2000)** Crystal structures of mouse class II alcohol dehydrogenase reveal determinants of substrate specificity and catalytic efficiency. *J Mol Biol* **302:**441-453.

390. **Tahallah, N., Pinkse, M., Maier, C.S. & Heck, A.J. (2001)** The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Commun Mass Spectrom* **15:**596-601.

391. **Takami, H., Takaki, Y., Chee, G.J., Nishi, S., Shimamura, S., Suzuki, H., Matsui, S. & Uchiyama, I. (2004)** Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*. *Nucleic Acids Res* **32:**6292-6303.

392. **Tamakoshi, M., Nakano, Y., Kakizawa, S., Yamagishi, A. & Oshima, T. (2001)** Selection of stabilized 3-isopropylmalate dehydrogenase of *Saccharomyces cerevisiae* using the host-vector system of an extreme thermophile, *Thermus thermophilus*. *Extremophiles* **5:**17-22.

393. **Tamakoshi, M., Yamagishi, A. & Oshima, T. (1995)** Screening of stable proteins in an extreme thermophile, *Thermus thermophilus*. *Mol Microbiol* **16:**1031-1036.

394. **Tame, J.R., Namba, K., Dodson, E.J. & Roper, D.I. (2002)** The crystal structure of HpcE, a bifunctional decarboxylase/isomerase with a multifunctional fold. *Biochemistry* **41:**2982-2989.

395. **Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J. & Natale, D.A. (2003)** The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:**41.

396. **Tatusov, R.L., Koonin, E.V. & Lipman, D.J. (1997)** A genomic perspective on protein families. *Science* **278:**631-637.

397. **Terwilliger, T.C. (2003)** SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol* **374:**22-37.

398. **Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997)** The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25:**4876-4882.

399. **Thompson, W., Rouchka, E.C. & Lawrence, C.E. (2003)** Gibbs Recursive Sampler: finding transcription factor binding sites. *Nucleic Acids Res* **31:**3580-3585.

400. **Timm, D.E., Mueller, H.A., Bhanumoorthy, P., Harp, J.M. & Bunick, G.J. (1999)** Crystal structure and mechanism of a carbon-carbon bond hydrolase. *Structure* **7:**1023-1033.

401. **Tomita, K., Nakakita, Y., Hoshino, Y., Numata, K. & Kawaguchi, H. (1987)** New genus of the Actinomycetales: *Streptoalloteichus hindustanus* gen. nov., nom. rev.; sp. nov., nom. rev. *Int J Syst Bacteriol* **37:**211-213.

402. **Trincone, A., Perugino, G., Rossi, M. & Moracci, M. (2000)** A novel thermophilic glycosynthase that effects branching glycosylation. *Bioorg Med Chem Lett* **10:**365-368.

403. **Turner, S.L., Ford, G.C., Mountain, A. & Moir, A. (1992)** Selection of a thermostable variant of chloramphenicol acetyltransferase (Cat-86). *Protein Eng* **5:**535-541.

404. **Tusher, V.G., Tibshirani, R. & Chu, G. (2001)** Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98:**5116-5121.

405. **Ueda, K., Yamashita, A., Ishikawa, J., Shimada, M., Watsuji, T.O., Morimura, K., Ikeda, H., Hattori, M. & Beppu, T. (2004)** Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism. *Nucleic Acids Res* **32:**4937-4944.

406. **Uhrigshardt, H., Walden, M., John, H. & Anemuller, S. (2001)** Purification and characterization of the first archaeal aconitase from the thermoacidophilic *Sulfolobus acidocaldarius*. *Eur J Biochem* **268:**1760-1771.

407. **Uhrigshardt, H., Walden, M., John, H., Petersen, A. & Anemuller, S. (2002)** Evidence for an operative glyoxylate cycle in the thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius*. *FEBS Lett* **513:**223-229.

408. **Vagin, A.A. & Teplyakov, A. (1997)** MOLREP: an automated program for molecular replacement. *J Appl Cryst* **30:**1022-1025.

409. **Vaguine, A.A., Richelle, J. & 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.

410. **Van den Burg, B., Dijkstra, B.W., Vriend, G., Van der Vinne, B., Venema, G. & Eijsink, V.G. (1994)** Protein stabilization by hydrophobic interactions at the surface. *Eur J Biochem* **220:**981-985.

411. **Van den Burg, B. & Eijsink, V.G. (2002)** Selection of mutations for increased protein stability. *Curr Opin Biotechnol* **13:**333-337.

412. **Van den Burg, B., Vriend, G., Veltman, O.R., Venema, G. & Eijsink, V.G. (1998)** Engineering an enzyme to resist boiling. *Proc Natl Acad Sci U S A* **95:**2056-2060.

413. **Van den Heuvel, R.H.H., Van Duijn, E., Mazon, H., Synowsky, S.A., Lorenzen, K., Versluis, C., Brouns, S.J.J., Langridge, D., Van der Oost, J., Hoyes, J. & Heck, A.J. (2006)** Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* **78:**7473-7483.

414. **Van der Oost, J., Huynen, M.A. & Verhees, C.H. (2002)** Molecular characterization of phosphoglycerate mutase in archaea. *FEMS Microbiol Lett* **212:**111-120.

415. **Vanbelle, C., Brutscher, B., Blackledge, M., Muhle-Goll, C., Remy, M.H., Masson, J.M. & Marion, D. (2003)** NMR study of the interaction between Zn(II) ligated bleomycin and *Streptoalloteichus hindustanus* bleomycin resistance protein. *Biochemistry* **42:**651-663.

416. **Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W. & Taylor, R.D. (2003)** Improved protein-ligand docking using GOLD. *Proteins* **52:**609-623.

417. **Verhees, C.H., Kengen, S.W., Tuininga, J.E., Schut, G.J., Adams, M.W., De Vos, W.M. & Van der Oost, J. (2003)** The unique features of glycolytic pathways in Archaea. *Biochem J* **375:**231-246.

418. **Vetriani, C., Maeder, D.L., Tolliday, N., Yip, K.S., Stillman, T.J., Britton, K.L., Rice, D.W., Klump, H.H. & Robb, F.T. (1998)** Protein thermostability above 100 degreesC: a key role for ionic interactions. *Proc Natl Acad Sci U S A* **95:**12300-12305.

419. **Vieille, C. & Zeikus, G.J. (2001)** Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* **65:**1-43.

420. **Vogt, G., Woell, S. & Argos, P. (1997)** Protein thermal stability, hydrogen bonds, and ion pairs. *J Mol Biol* **269:**631-643.

421. **Volkov, A.A. & Arnold, F.H. (2000)** Methods for in vitro DNA recombination and random chimeragenesis. *Methods Enzymol* **328:**447-456.

422. **Warren, L. (1960)** Thiobarbituric acid spray reaction for *deoxy* sugars and sialic acids. *Nature* **186:**237.

423. **Watanabe, S., Kodak, T. & Makino, K. (2006)** Cloning, expression, and characterization of bacterial L-arabinose 1-dehydrogenase involved in an alternative pathway of L-arabinose metabolism. *J Biol Chem* **281:**2612-2623.

424. **Watanabe, S., Kodaki, T. & Makino, K. (2005)** Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc. *J Biol Chem* **280:**10340-10349.

425. **Watanabe, S., Kodaki, T. & Makino, K. (2006)** A novel alpha-ketoglutaric semialdehyde dehydrogenase: evolutionary insight into an alternative pathway of bacterial L-arabinose metabolism. *J Biol Chem* **281:**28876-28888.

426. **Watanabe, S., Shimada, N., Tajima, K., Kodaki, T. & Makino, K. (2006)** Identification and characterization of L-arabonate dehydratase, L-2-keto-3-*deoxy*arabonate dehydratase, and L-arabinolactonase involved in an alternative pathway of L-arabinose me-

tabolism. Novel evolutionary insight into sugar metabolism. *J Biol Chem* **281:**33521-33536.

427. **Watanabe, S., Yamada, M., Ohtsu, I. & Makino, K. (2007)** alpha-Ketoglutaric Semialdehyde Dehydrogenase Isozymes Involved in Metabolic Pathways of D-Glucarate, D-Galactarate, and Hydroxy-L-proline: molecular and metabolic convergent evolution. *J Biol Chem* **282:**6685-6695.

428. **Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G.G., Simon, M., Soll, D., Stetter, K.O., Short, J.M. & Noordewier, M. (2003)** The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci U S A* **100:**12984-12988.

429. **Weimberg, R. (1961)** Pentose oxidation by *Pseudomonas fragi*. *J Biol Chem* **236:**629-635.

430. **Weimberg, R. & Doudoroff, M. (1955)** The oxidation of L-arabinose by *Pseudomonas saccharophila*. *J Biol Chem* **217:**607-624.

431. **Wiesmann, C., Hengstenberg, W. & Schulz, G.E. (1997)** Crystal structures and mechanism of 6-phospho-beta-galactosidase from *Lactococcus lactis*. *J Mol Biol* **269:**851-860.

432. **Woese, C.R. (1987)** Bacterial evolution. *Microbiol Rev* **51:**221-271.

433. **Wojtkiewicz, B., Szmidzinski, R., Jezierska, A. & Cocito, C. (1988)** Identification of a salvage pathway for D-arabinose in *Mycobacterium smegmatis*. *Eur J Biochem* **172:**197-203.

434. **Wolfenden, R., Lu, X. & Young, G. (1998)** Spontaneous hydrolysis of glycosides. *J Am Chem Soc* **120:**6814-6815.

435. **Worthington, P., Hoang, V., Perez-Pomares, F. & Blum, P. (2003)** Targeted Disruption of the alpha-Amylase Gene in the Hyperthermophilic Archaeon *Sulfolobus solfataricus*. *J Bacteriol* **185:**482-488.

436. **Wu, M., Ren, Q., Durkin, A.S., Daugherty, S.C., Brinkac, L.M., Dodson, R.J., Madupu, R., Sullivan, S.A., Kolonay, J.F., Haft, D.H., Nelson, W.C., Tallon, L.J., Jones, K.M., Ulrich, L.E., Gonzalez, J.M., Zhulin, I.B., Robb, F.T. & Eisen, J.A. (2005)** Life in hot carbon monoxide: the complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet* **1:**e65.

437. **Xiang, X., Dong, X. & Huang, L. (2003)** *Sulfolobus tengchongensis* sp. nov., a novel thermoacidophilic archaeon isolated from a hot spring in Tengchong, China. *Extremophiles* **7:**493-498.

438. **Yeats, S., McWilliam, P. & Zillig, W. (1982)** A plasmid in the archaebacterium *Sulfolobus acidocaldarius*. *Embo J* **1:**1035-1038.

439. **Yoneya, T. & Adams, E. (1961)** Hydroxyproline metabolism. V. Inducible allohydroxy-D-proline oxidase of *Pseudomonas*. *J Biol Chem* **236:**3272-3279.

440. **You, L. & Arnold, F.H. (1996)** Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng* **9:**77-83.

441. **Zaigler, A., Schuster, S.C. & Soppa, J. (2003)** Construction and usage of a onefold-coverage shotgun DNA microarray to characterize the metabolism of the archaeon *Haloferax volcanii*. *Mol Microbiol* **48:**1089-1105.

442. **Zhang, J.H., Dawes, G. & Stemmer, W.P. (1997)** Directed evolution of alpha-fucosidase from alpha-galactosidase by DNA shuffling and screening. *Proc Natl Acad Sci U S A* **94:**4504-4509.

443. **Zhang, Q., Iwasaki, T., Wakagi, T. & Oshima, T. (1996)** 2-oxoacid:ferredoxin oxidoreductase from the thermoacidophilic archaeon, *Sulfolobus* sp. strain 7. *J Biochem (Tokyo)* **120:**587-599.

444. **Zillig, W., Arnold, H.P., Holz, I., Prangishvili, D., Schweier, A., Stedman, K., She, Q., Phan, H., Garrett, R. & Kristjansson, J.K. (1998)** Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. *Extremophiles* **2:**131-140.

445. **Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janecovik, D., Hain, J., Lanzendorfer, M. & Kristjansson, J. (1994)** Screening for Sulfolobales, their plasmids and viruses in Islandic solfataras. *Syst Appl Microbiol* **16:**609-628.

446. **Zillig, W., Stetter, K.O., Wunderl, S., Schulz, W., Priess, H. & Scholz, I. (1980)** The *Sulfolobus*-"Caldariella" group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch Mikrobiol* **125:**259-269.

Appendix I: Color figures

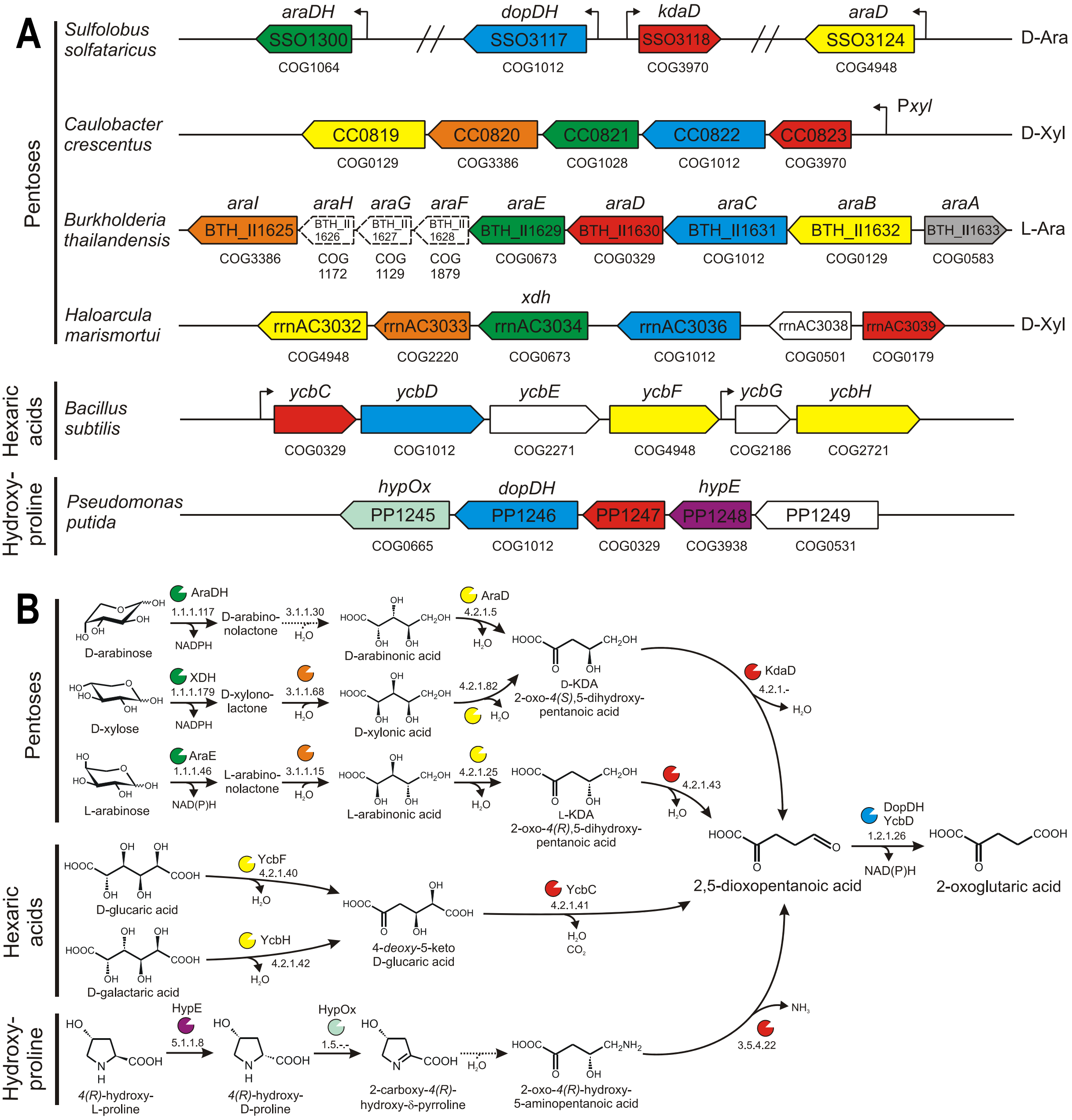

A
Pentoses
Sulfolobus solfataricus
araDH
SSO1300
COG1064
dopDH
SSO3117
COG1012
kdaD
SSO3118
COG3970
araD
SSO3124
COG4948
D-Ara
Caulobacter crescentus
CC0819
COG0129
CC0820
COG3386
CC0821
COG1028
CC0822
COG1012
CC0823
COG3970
Pxyl
D-Xyl
Burkholderia thailandensis
araI
BTH_II1625
COG3386
araH
BTH_II 1626
COG 1172
araG
BTH_II 1627
COG 1129
araF
BTH_II 1628
COG 1879
araE
BTH_II1629
COG0673
araD
BTH_II1630
COG0329
araC
BTH_II1631
COG1012
araB
BTH_II1632
COG0129
araA
BTH_II1633
COG0583
L-Ara
Haloarcula marismortui
rrnAC3032
COG4948
rrnAC3033
COG2220
xdh
rrnAC3034
COG0673
rrnAC3036
COG1012
rrnAC3038
COG0501
rrnAC3039
COG0179
D-Xyl
Hexaric acids
Bacillus subtilis
ycbC
COG0329
ycbD
COG1012
ycbE
COG2271
ycbF
COG4948
ycbG
COG2186
ycbH
COG2721
Hydroxy-proline
Pseudomonas putida
hypOx
PP1245
COG0665
dopDH
PP1246
COG1012
PP1247
COG0329
hypE
PP1248
COG3938
PP1249
COG0531
B
Pentoses
D-arabinose
AraDH
1.1.1.117
NADPH
D-arabino-nolactone
3.1.1.30
H2O
D-arabinonic acid
AraD
4.2.1.5
H2O
D-KDA
2-oxo-4(S),5-dihydroxy-pentanoic acid
KdaD
4.2.1.-
H2O
D-xylose
XDH
1.1.1.179
NADPH
D-xylono-lactone
3.1.1.68
H2O
D-xylonic acid
4.2.1.82
H2O
L-arabinose
AraE
1.1.1.46
NAD(P)H
L-arabino-nolactone
3.1.1.15
H2O
L-arabinonic acid
4.2.1.25
H2O
L-KDA
2-oxo-4(R),5-dihydroxy-pentanoic acid
4.2.1.43
H2O
2,5-dioxopentanoic acid
DopDH
YcbD
1.2.1.26
NAD(P)H
2-oxoglutaric acid
Hexaric acids
D-glucaric acid
YcbF
4.2.1.40
H2O
D-galactaric acid
YcbH
4.2.1.42
H2O
4-deoxy-5-keto D-glucaric acid
YcbC
4.2.1.41
H2O
CO2
Hydroxy-proline
4(R)-hydroxy-L-proline
HypE
5.1.1.8
4(R)-hydroxy-D-proline
HypOx
1.5.-.-
2-carboxy-4(R)-hydroxy-δ-pyrroline
H2O
2-oxo-4(R)-hydroxy-5-aminopentanoic acid
NH3
3.5.4.22

Figure 3.7

A) Cartoon of the organization of conserved gene clusters involved in the pentose, hexaric acid and hydroxyproline degradation. Proposed analogous gene functions are indicated in the same color (green: pentose dehydrogenase, orange: pentonolactonase, yellow: aldonic acid dehydratase, red: 2-keto-3-*deoxy*-aldonic acid dehydratase, blue: 2,5-dioxopentanoate dehydrogenase. Dashed genes are displayed smaller than their relative size. Protein family numbers are displayed below each gene according to Clusters of Orthologous Groups of proteins classification system (COG) [396]. The genes indicated in white or gray encode the following putative functions: araA: transcriptional regulator, *araF-araH*: L-Ara ABC transporter (periplasmic L-Ara binding protein, ATP binding protein, permease), *rrnAC3038*: heat shock protein X, *ycbE*: glucarate/galactarate permease, *ycbG*: transcriptional regulator, *PP1249*: hydroxyproline permease. **B)** Schematic representation of the convergence of catabolic pathways for pentoses, hexaric acids [75,157,172,352] and hydroxyproline [311,361,439] at the level of 2,5-dioxopentanoate. Enzymatic activities are indicated by their EC number. Dashed lines indicate proposed spontaneous reactions.

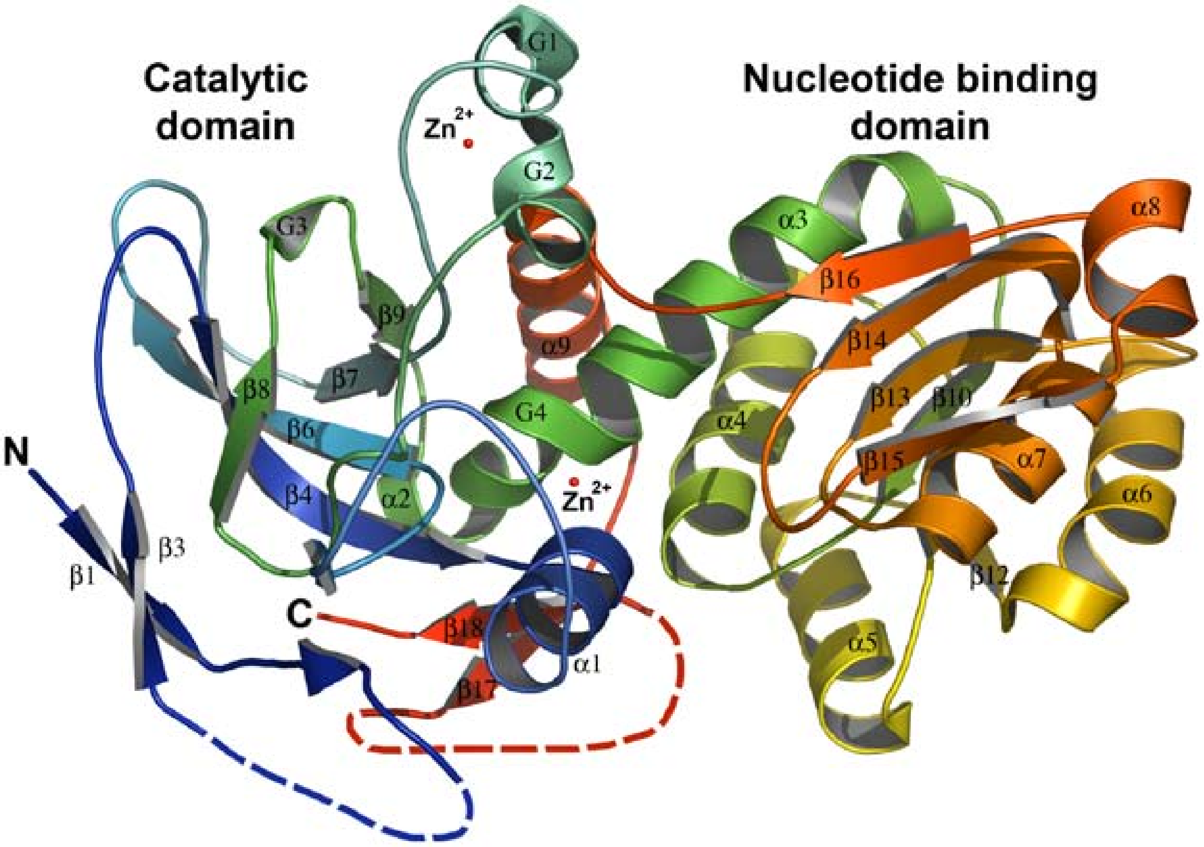

Catalytic domain
Nucleotide binding domain
G1
Zn2+
G2
G3
β9
α3
β16
α8
α9
β14
β7
β8
β13
β10
G4
α4
β6
N
β15
α7
α2
β4
Zn2+
α6
β3
β1
β12
C
β18
α1
α5
β17

Figure 4.1
Schematic representation of the *Sulfolobus solfataricus* AraDH structural model of the monomer showing the catalytic and nucleotide binding domain. The N-terminus of the polypeptide chain (N, in blue) follows a color gradient towards the C-terminus (C, in red). Interruptions in the structure due to missing residues corresponding to regions of weak electron density are indicated by dashed lines. The structural (top) and the catalytic (bottom) coordinated zinc ions are displayed as gray spheres.

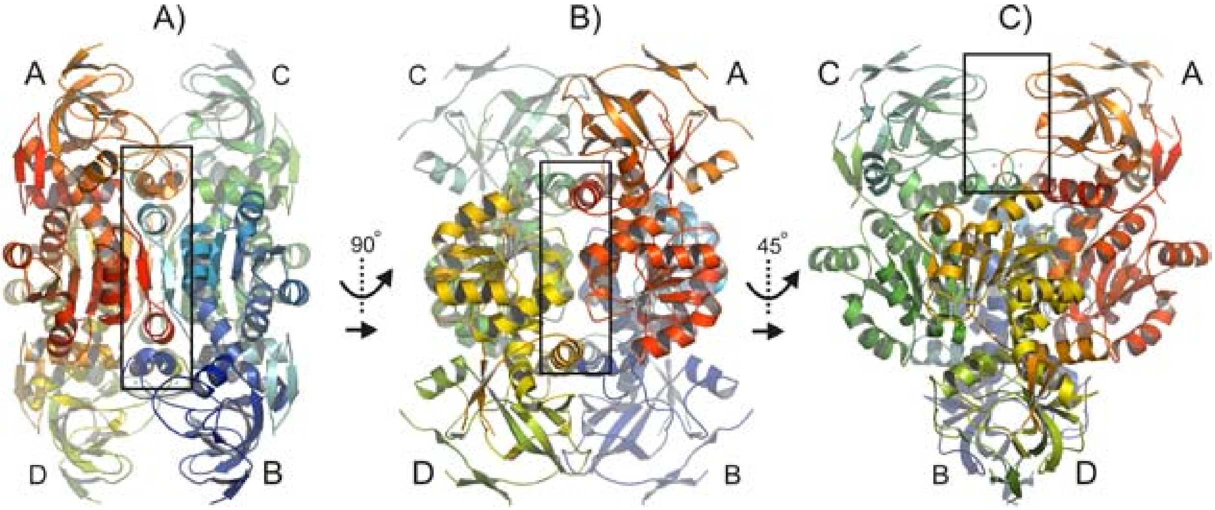

A)
B)
C)
A
C
D
B
90°
C
A
D
B
45°
C
A
B
D

Figure 4.2
Schematic diagrams of the *S. solfataricus* AraDH homo-tetramer generated by crystallographic symmetry operations. View perpendicular to **A)** the AB, **B)** the AD, and **C)** the AC dimer interface showing the intersubunit contacts (boxed).

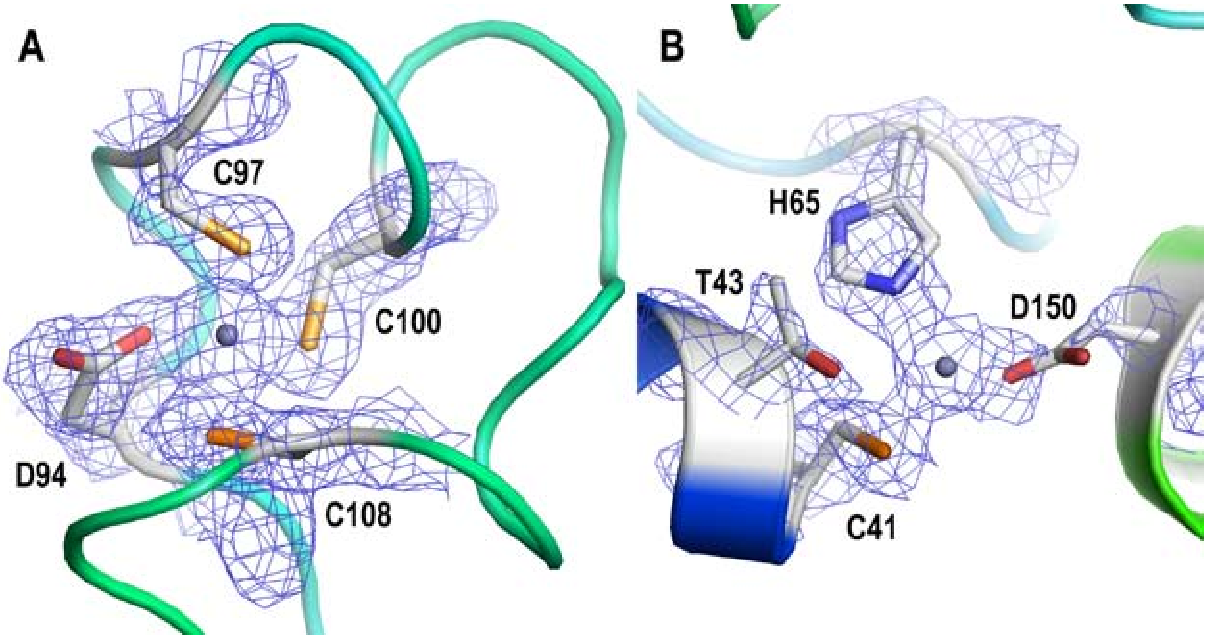

A
C97
C100
D94
C108
B
H65
T43
D150
C41

Figure 4.3
View of the composite omit map calculated using the program SFCHECK 409 contoured at 1.0σ around **A)** the structural zinc ion and **B)** the active site zinc ion. **A)** The structural zinc ion is tetrahedrally coordinated by Asp94 Oδ1 at a distance of 1.84 Å, Cys97 Sγ at 2.30 Å, Cys100 Sγ at 2.30 Å and Cys108 Sγ at 2.34 Å. **B)** The active site zinc ion is coordinated by Cys41 Sγ at a distance of 2.36 Å, His65 Nε2 at 2.35 Å and Asp150 Oδ2 at 1.84 Å. A water molecule is located 4.55 Å from the active site zinc ion (not shown).The threonine residue (Thr43) involved in the proton relay mechanism is also shown.

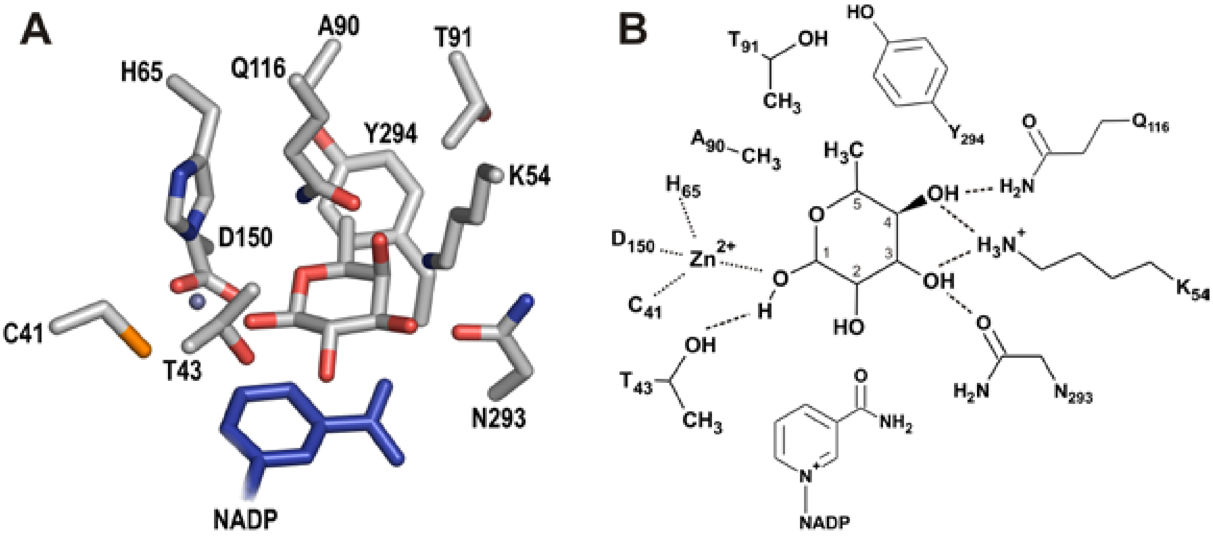

A
A90
T91
H65
Q116
Y294
K54
D150
C41
T43
N293
NADP
B
HO
T91
OH
CH3
Y294
O
Q116
A90–CH3
H3C
H65
OH
H2N
5
4
D150
Zn2+
1
3
H3N+
C41
O
2
OH
K54
H
HO
OH
T43
O
NH2
H2N
N293
CH3
N+
NADP

Figure 4.7
Docking solution of the Michaelis complex for α-L-fucopyranoside (L-fucose) and the nicotinamide cofactor in **A)** stick and **B)** schematic representation. The C1O of the substrate forms the fourth ligand of the tetrahedrally coordinated zinc ion (gray sphere). The C6 methyl group of L-fucose that distinguishes it from D-arabinose is stacked against the phenyl ring of Tyr294 and forms additional hydrophobic contacts with the aliphatic side chains of Ala90, Thr91 and Ile113 (latter not shown). Hydroxyl groups at C3 and C4 form extensive hydrogen bonds with Lys54, Gln116 and Asn293. The C1H is positioned correctly to allow hydride transfer to the nicotinamide cofactor.

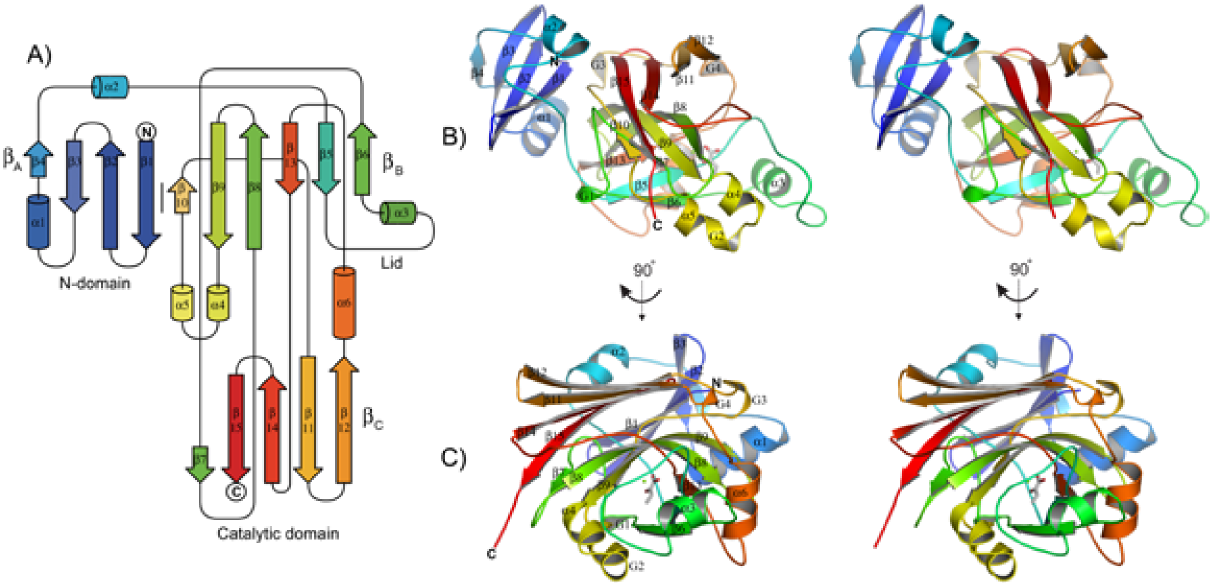
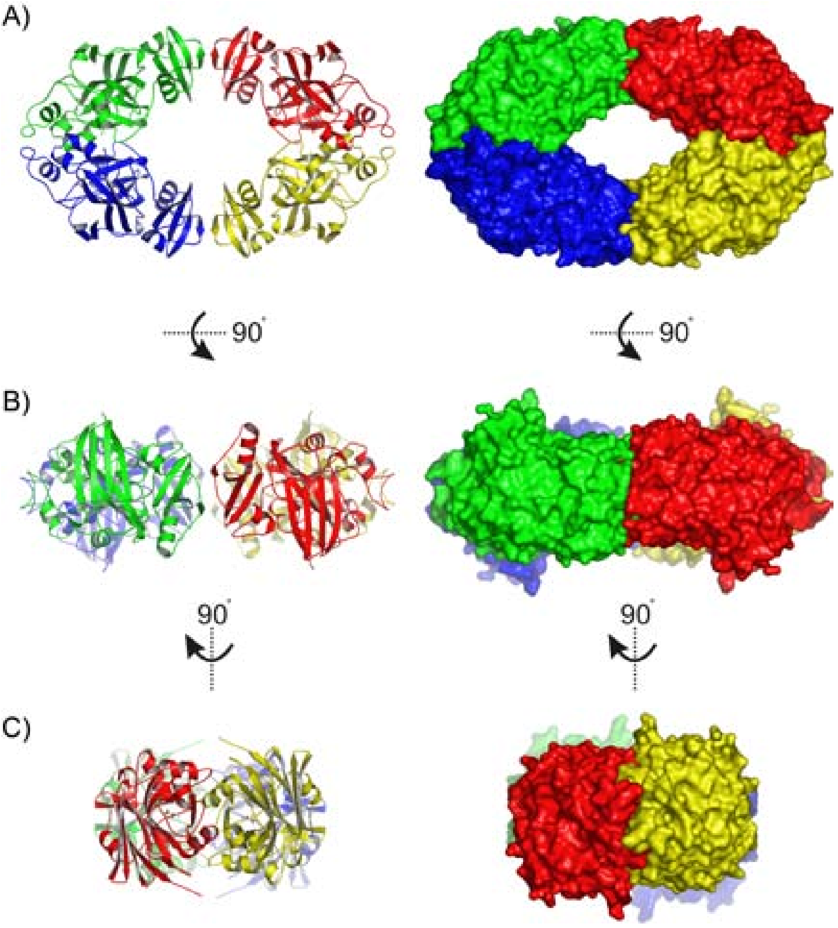


Figure 5.2
A) Topology diagram showing the connectivity of secondary structure elements and domain organization of the *S. solfataricus* KdaD monomer. The N-terminus of the polypeptide chain (N, in blue) follows a color gradient towards the C-terminus (C, in red). **B)** and **C)** Stereo ribbon diagrams at two viewing angles of the KdaD monomer with Mg^{2+} and 2-oxobutyrate bound.

Figure 5.3
Schematic ribbon and surface diagrams are shown of the *S. solfataricus* KdaD homo-tetramer generated by crystallographic symmetry operations. View **A)** perpendicular to the tetrameric ring, **B)** along the N-domain dimer-dimer interface, and **C)** along the catalytic domain dimer interface.



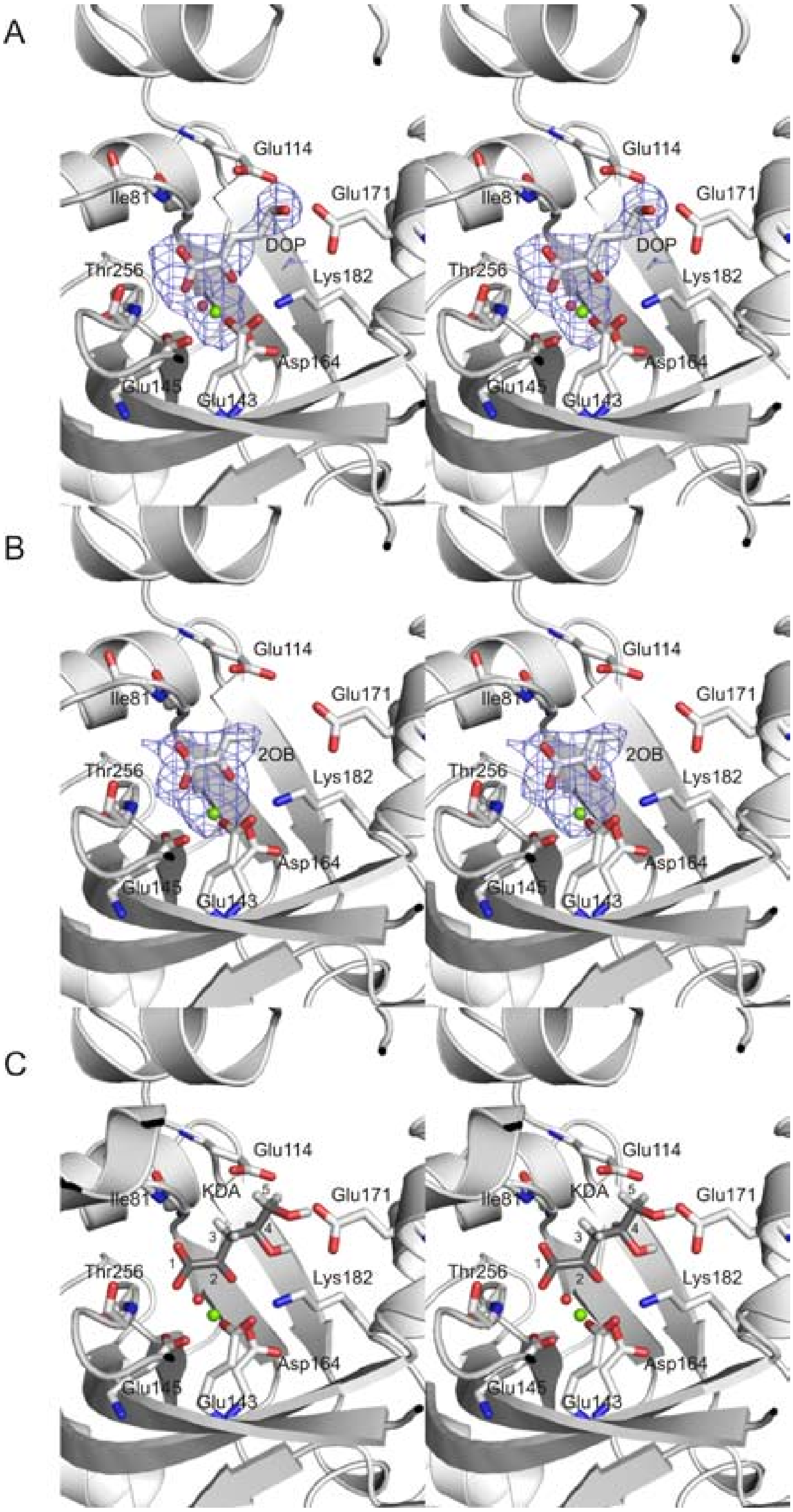

A
Glu114
Ile81
Glu171
DOP
Thr256
Lys182
Asp164
Glu145
Glu143
B
Glu114
Ile81
Glu171
2OB
Thr256
Lys182
Asp164
Glu145
Glu143
C
Glu114
Ile81
KDA
5
Glu171
3
4
1
2
Thr256
Lys182
Asp164
Glu145
Glu143

Figure 5.4
Stereo diagrams are shown of the KdaD active site with Mg^{2+} (green spheres), water molecules (red spheres) and **A)** 2,5-dioxopentanoate or **B)** 2-oxobutyrate bound. Simulated annealing F_o-*D*F_c omit electron density maps for the ligand and the metal ion are displayed at a 3σ contour level. **C)** Stereo image of a model of the Michaelis complex of KdaD with its substrate 2-keto-3-*deoxy*-D-arabinonate (D-KDA) bound. Hydrogen atoms for D-KDA are shown as well as the carbon atom numbers.

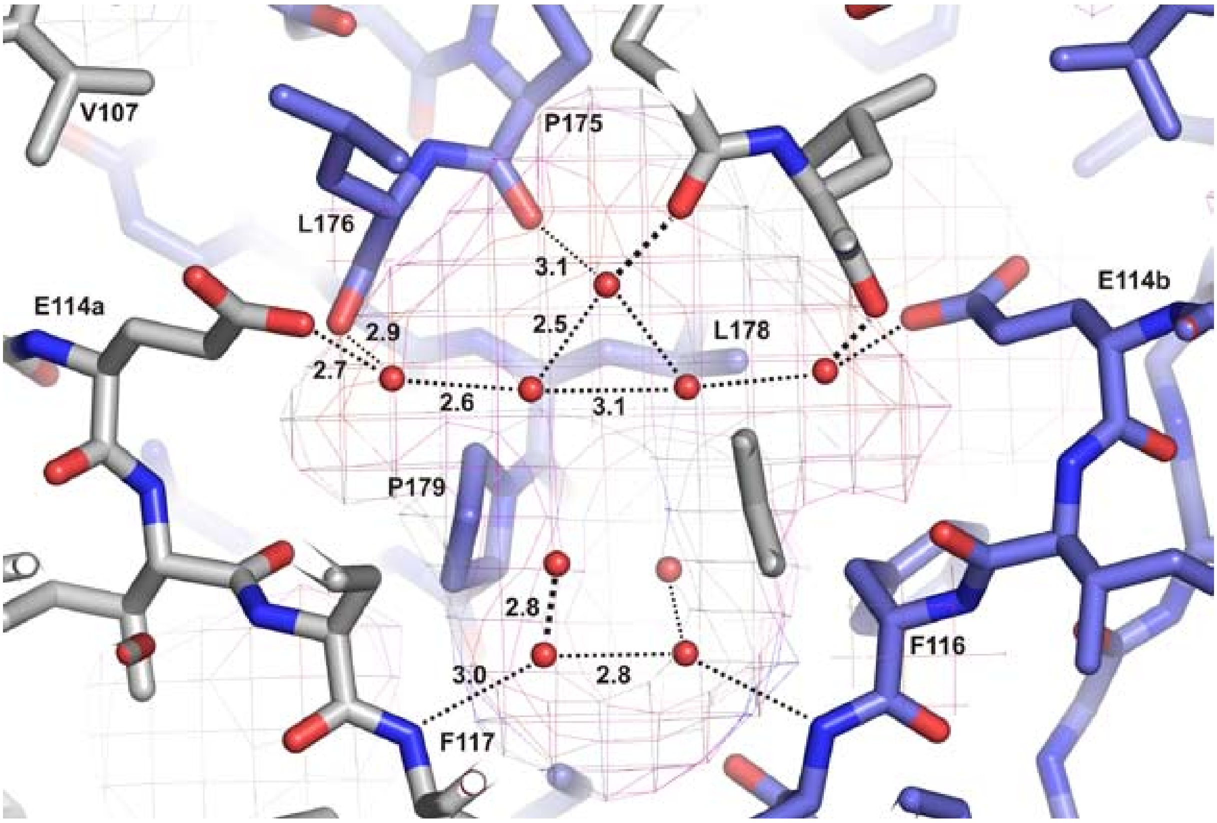

V107
P175
L176
E114a
E114b
L178
P179
F116
F117
3.1
2.5
2.9
2.7
2.6
3.1
2.8
3.0
2.8

Figure 5.6
A schematic representation of the catalytic dimer interface is shown, indicating the hydrogen bond connectivity between Glu114a and Glu114b residues of both active sites through a mostly hydrophobic water tunnel. Interatomic distances are displayed in Å.

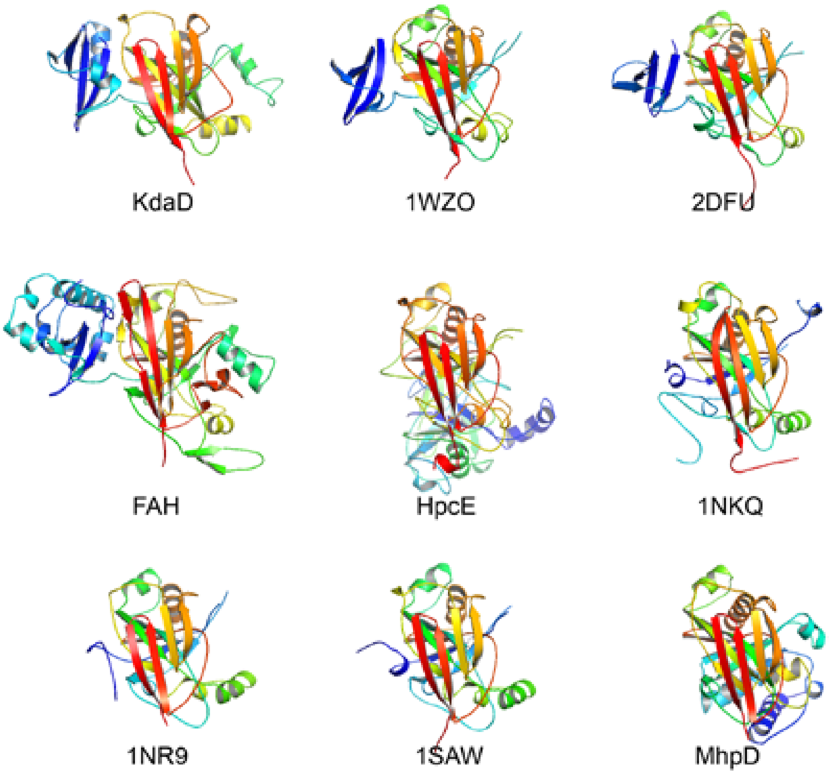

KdaD
1WZO
2DFU
FAH
HpcE
1NKQ
1NR9
1SAW
MhpD

Figure 5.9
A ribbon representation is shown of members of the FAH superfamily as monomers. See Table 5.1 and Fig. 5.8 for details.

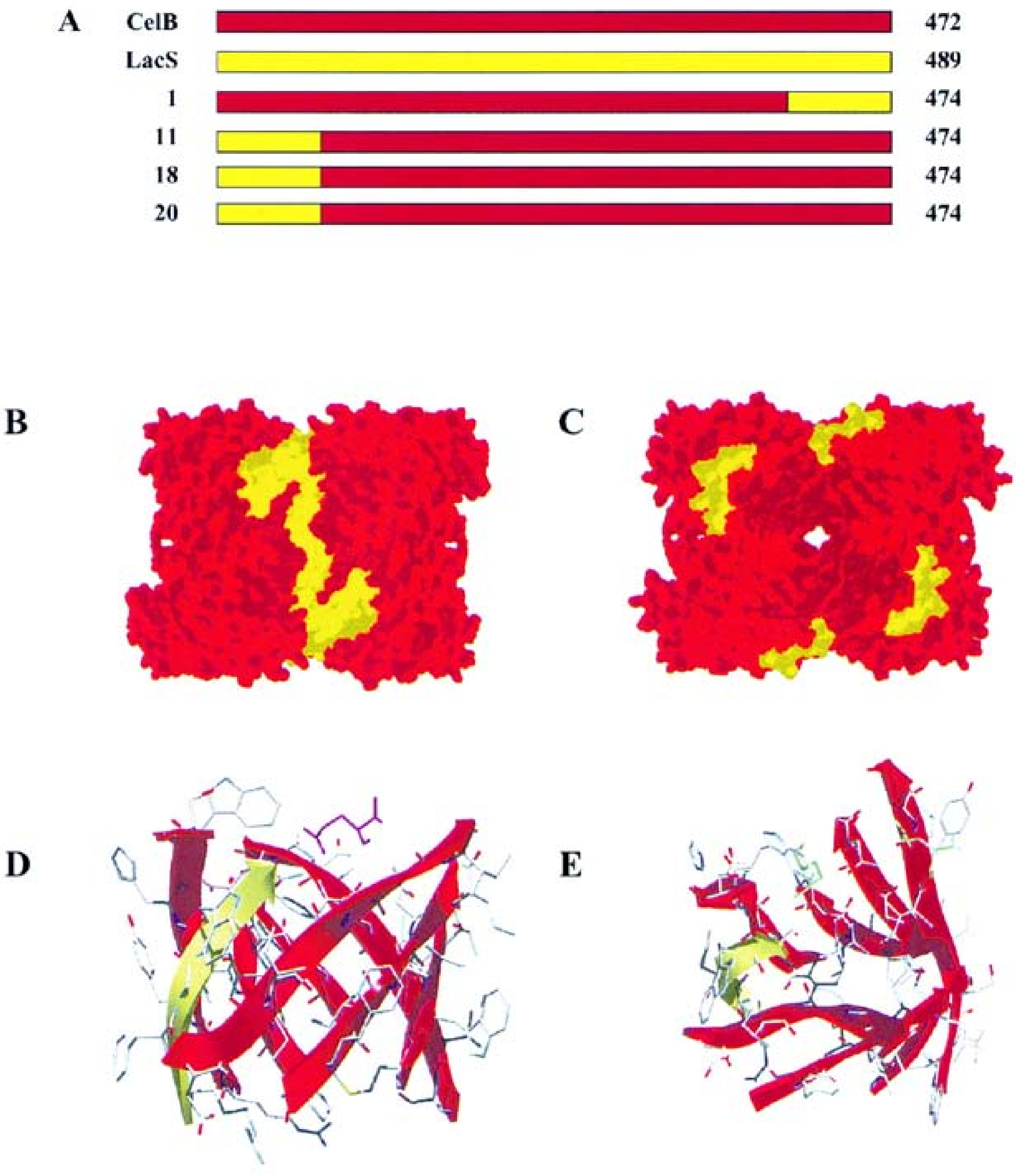

A
CelB
472
LacS
489
1
474
11
474
18
474
20
474
B
C
D
E

Figure 7.4
Structural representation of hybrids. **A)** Schematic representation of primary structures of CelB, LacS, hybrids 1, 11, 18 and 20. Surface plot of tetrameric **B)** hybrid 1 and **C)** hybrid 20. **D)** Side view and **E)** top view of the $(\beta\alpha)8$-barrel of hybrid 20.

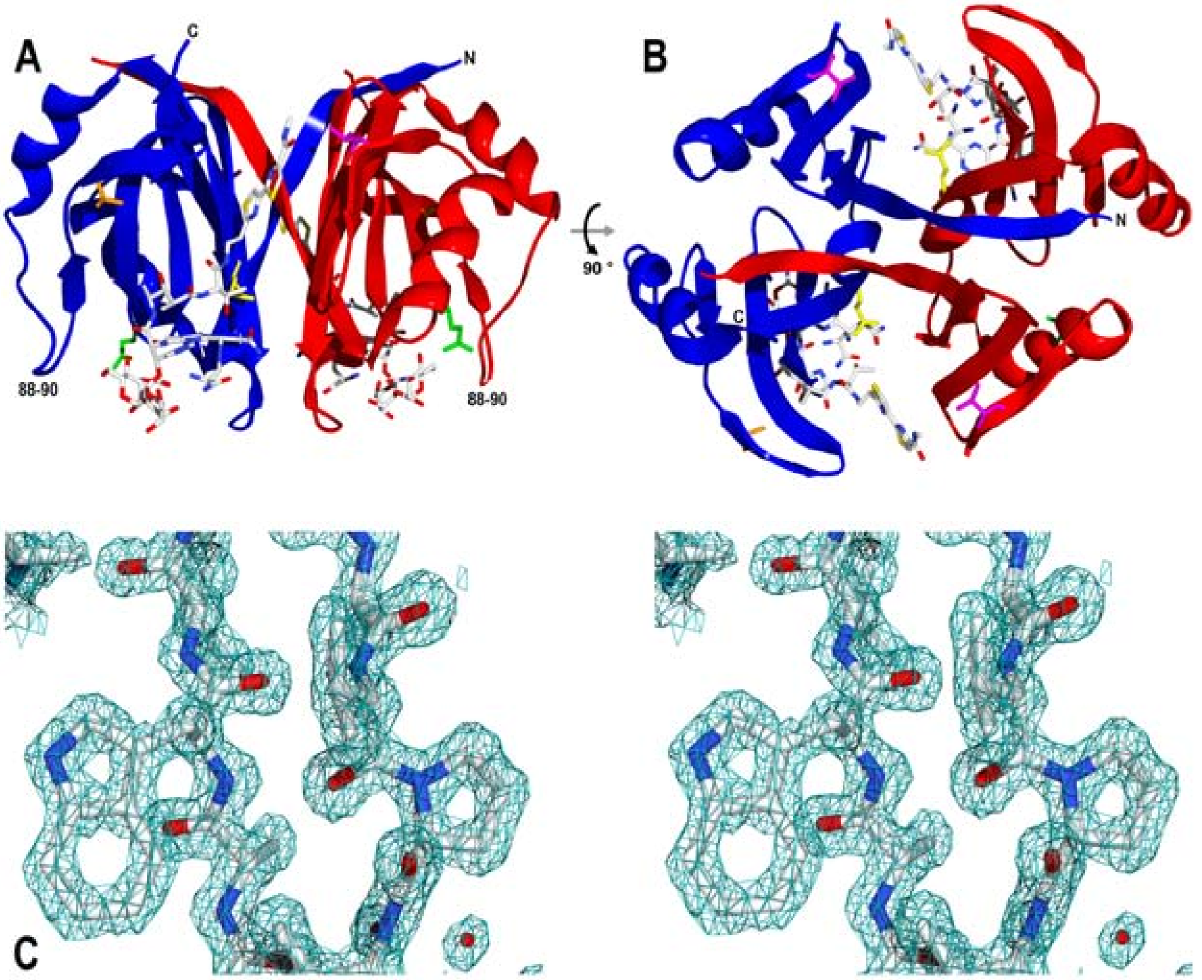

A
C
N
88-90
88-90
B
90 °
N
C
C

Figure 8.3
Structure and electron density of HTS in complex with bleomycin A2. Ribbon diagram showing the dimeric structure of the fourfold mutant Shble in complex with bleomycin A2. Mutations are indicated by stick representations. Chain A in blue, chain B in red, Gly18Glu in green, Asp32Val in pink, Leu63Gln in yellow and Gly98Val in orange. **A)** side view. **B)** viewed from the N- and C-terminal side (top view) **C)** stereo view of the electron density around residue Pro9 and Trp65 contoured at 2σ. Residues are colored according to the CPK color scheme, water molecules are represented by red spheres.

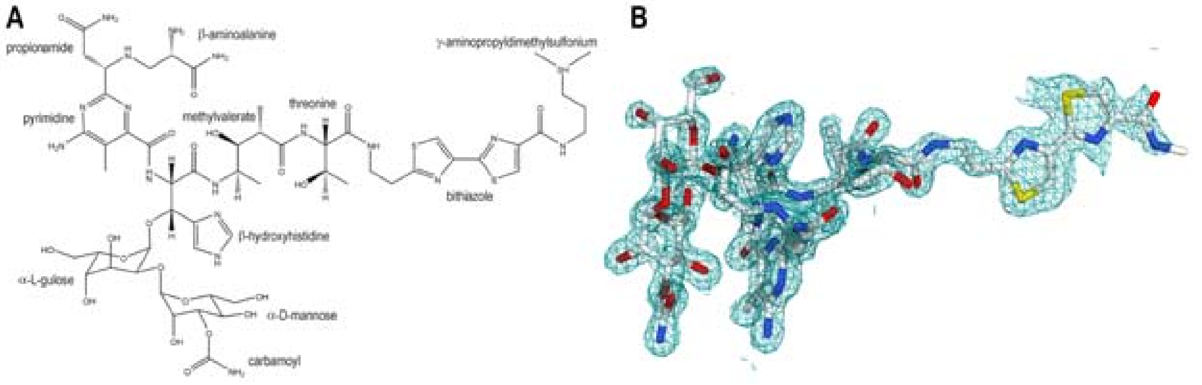

A
propionamide
β-aminoalanine
pyrimidine
methylvalerate
threonine
γ-aminopropyldimethylsulfonium
bithiazole
β-hydroxyhistidine
α-L-gulose
α-D-mannose
carbamoyl
B

Figure 8.4
Chemical diagram and electron density of bleomycin A2. **A)** Schematic representation of bleomycin A2. **B)** Electron density around bleomycin A2 contoured at 1.5σ. The diagram indicates the missing electron density around the γ-aminopropyldimethylsulfonium moiety suggesting a disordered conformation.

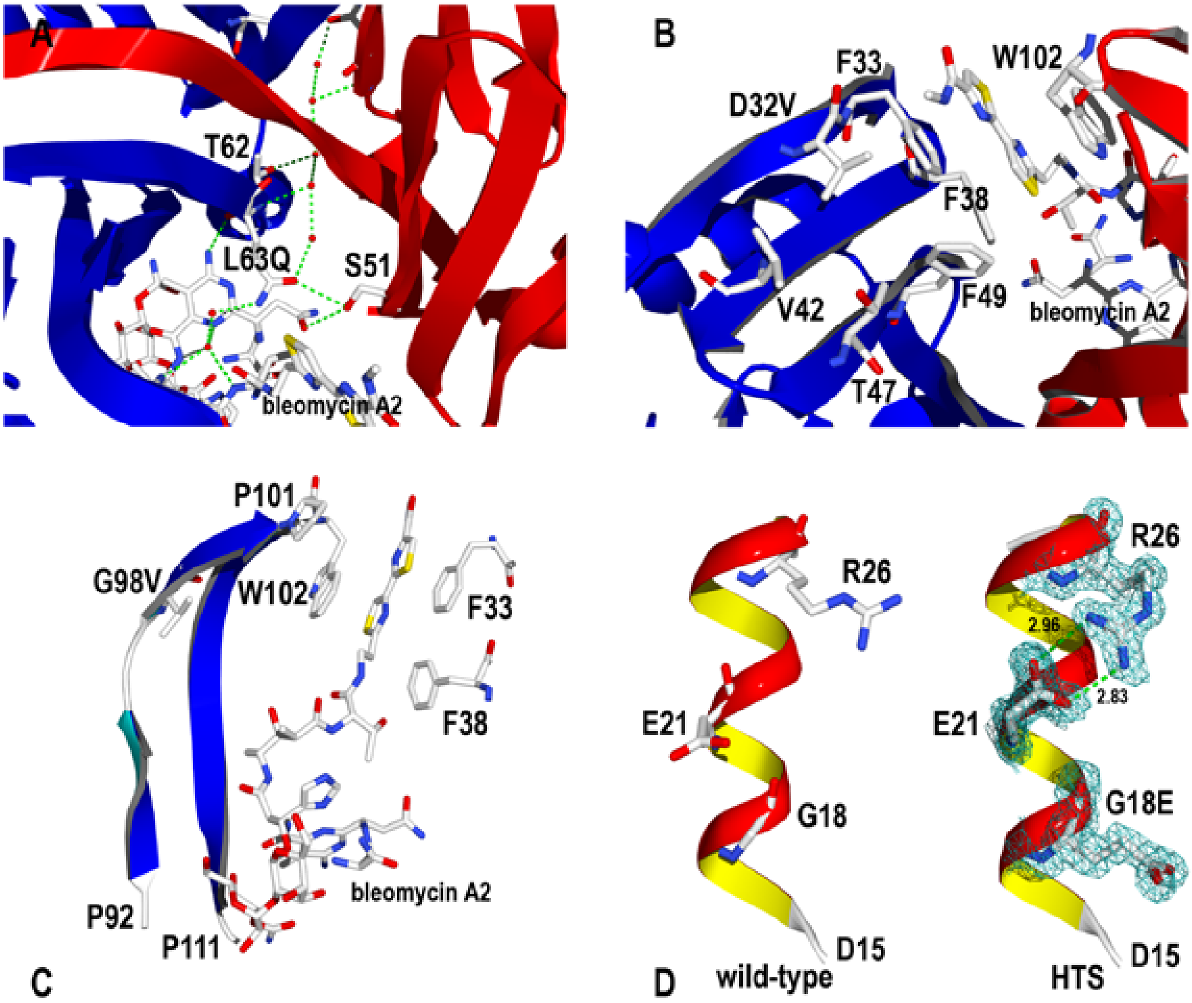

A
T62
L63Q
S51
bleomycin A2
B
D32V
F33
W102
F38
V42
F49
bleomycin A2
T47
P101
G98V
W102
F33
F38
bleomycin A2
P92
P111
C
R26
E21
G18
D15
D
wild-type
R26
2.96
2.83
E21
G18E
D15
HTS

Figure 8.5
Structural effects of the individual mutations. **A)** Leu63Gln. Ribbon diagram showing the hydrogen bond network at the dimer interface. Thr62, Gln63, Ser51 and bleomycin A2 are shown together with the intersubunit water channel. **B)** Asp32Val. Ribbon diagram showing the hydrophobic intersubunit bleomycin tail binding crevice. Val32 may be involved in improved hydrophobic packing of this surface indentation among amino acids Phe33, Phe38, Val42, Thr47 and Phe49. **C)** Gly98Val. Ribbon representation showing a loop between Pro92 and Pro111 which is involved in bleomycin binding. Val98 is located at a former hinge region which enables Trp102 to stack the bithiazole tail against Phe33 and Phe38. The electron density revealed two alternative sidechain rotamers for Val98 in chain A (not shown) and a single side chain conformation in chain B. **D)** Gly18Glu. Side-by-side comparison of α-helix one formed between Asp15 and Leu27 in the wild-type and mutant crystal structures of Shble. The existence of a surface ion-pair between Glu21 and Arg26 is visible in the electron density (contoured at 1.5σ). Interatomic distances are indicated in Å. CPK color coding was used for amino acids and bleomycin A2. Chain A is indicated in blue, chain B in red. Water molecules are represented by red spheres. Hydrogen bonds are depicted by green dotted lines.

APPENDIX II: CO-AUTHOR AFFILIATIONS

Jasper Walther, Harmen G.J. van de Werken, Jasper Akerboom[a]*, Thijs J.G. Ettema*[b]*, Hao Wu, Nicole Smits, Hanneke L.D.M. Willemen, Petra Worm, Thijs Kaper*[c]*, Ans C.M. Geerling, Hauke Smidt, Willem M. de Vos, John van der Oost* - Laboratory of Microbiology, Dpt. of Agrotechnology and Food Sciences, Wageningen University, Dreienplein 10, 6703 HB Wageningen, Netherlands

Ambrosius P.L. Snijders, Stefan Peter, Iris Kinnman, Marjon G.J. de Vos, Phillip C. Wright - Biological and Environmental Systems Group, Dpt. of Chemical and Process Engineering, University of Sheffield. Mappin Street, Sheffield S1 3JD, United Kingdom

Kenneth M. Stedman - Biology Dpt., Portland State University, Portland, OR 97207, United States

Mark J. Young - Thermal Biology Institute, Montana State University, Bozeman, MT 59717, United States

Robert H.H. van den Heuvel[d]*, Hortense F.M. Mazon* - Dpt. of Biomolecular Mass spectrometry, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, Netherlands

Magnus Lundgren[e]*, Rolf Bernander* - Dpt. of Molecular Evolution, Evolutionary Biology Center, Uppsala University, Norbyvägen 18C, SE-752 36 Uppsala, Sweden

Anders Andersson[f]*, Peter Nilsson* - Dpt. of Biotechnology, Royal Institute of Technology (KTH), SE-106 91, Stockholm, Sweden

Laurent Salmon - Laboratoire de Chimie Bioorganique et Bioinorganique, CNRS-UMR 8182, Institut de Chimie Moléculaire et des Matériaux d'Orsay (ICMMO), Université de Paris-Sud, Bâtiment 420, F-91405, Orsay, France

Thomas R.M. Barends - Dpt. of Biomolecular Mechanisms, Max-Planck Institute for Medical Research, Jahnstrasse 29, D-69120 Heidelberg, Germany

Andrew P. Turnbull - Proteinstrukturfabrik c/o BESSY GmbH, Albert Einstein Strasse 15, D-12489 Berlin, Germany

present address:
[a] Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA 19700, United States
[b] Dpt. of Molecular Evolution, Evolutionary Biology Center, Uppsala University, Norbyvägen 18C, SE-752 36 Uppsala, Sweden
[c] Genencor, Palo Alto (CA), United States
[d] Organon, Oss, Netherlands
[e] Laboratory of Microbiology, Dpt. of Agrotechnology and Food Sciences, Wageningen University, Dreienplein 10, 6703 HB Wageningen, Netherlands
[f] University of California, Berkeley, United States

NEDERLANDSE SAMENVATTING

Dit proefschrift beschrijft diverse aspecten van het suikermetabolisme van het hyperthermofiele archaeon *Sulfolobus solfataricus*. Bij het onderzoek is gebruik gemaakt van verschillende biomoleculaire technieken, zoals DNA microarray analyse, kwantitatieve proteomics, klassieke genetica en biochemie, eiwit kristallografie en mutagenese. Bioinformatica werd gebruikt om de grote hoeveelheid genoom informatie die in databanken beschikbaar is te doorzoeken. De geïntegreerde benadering die ten grondslag lag aan de ontdekking van de nieuwe metabole routes en de betrokken enzymen, heeft onder andere tot betere inzichten geleid in de levenswijze van extremofiele micro-organismen. Daarnaast zijn er in het onderzoek twee eiwitten geschikt gemaakt voor toepassingen in suikeromzettingen en genetische studies aan thermofiele micro-organismen.

Hoofdstuk één geeft een historisch overzicht van de ontdekking van thermofiele micro-organismen en hun genetica. Het presenteert het hyperthermofiele archaeon *Sulfolobus solfataricus* als model organisme.

In het **tweede hoofdstuk** staan de resultaten van een studie beschreven die gericht was op het analyseren van de expressie van genen die betrokken zijn bij centrale metabole routes, zoals de glycolyse, gluconeogenese en de citroenzuurcyclus. Tot onze verrassing bleken de verschillen in expressie op zowel mRNA- als eiwitniveau klein te zijn, wanneer *S. solfataricus* werd gecultiveerd op suikers of peptiden. Slechts de expressie van drie enzymen die betrokken zijn bij de glycolyse en gluconeogenese was significant verschillend. Mogelijk worden de meeste enzymen in deze metabole routes gereguleerd door allostere effecten en post-translationele modificatie.

Het **derde hoofdstuk** beschrijft de opheldering van de metabole route waarmee de pentose suiker D-arabinose wordt afgebroken. Wederom gebruikmakend van een geïntegreerde onderzoeksstrategie werden eerst de verantwoordelijke enzymen geïdentificeerd. Vervolgens werd van deze enzymen op een biochemische wijze vastgesteld welke activiteit zij hadden. De suiker bleek door vier opeenvolgende enzymen te worden omgezet in 2-oxoglutaraat: een centrale metaboliet uit de citroenzuurcyclus. Met behulp van bioinformatica kon worden voorspeld dat andere bacteriën en archaea ook deze oxidatieve metabole pathway bezaten. Ook konden evolutionaire verbanden worden gelegd tussen de metabole routes voor pentoses, hexaarzuren en hydroxyproline, die een interessant licht werpen op het ontstaan van metabole routes.

In **hoofdstuk vier** worden de kristal structuur en de biochemische eigenschappen van het D-arabinose dehydrogenase gepresenteerd. Net als de meerderheid van de bekende nicotinamide-adenine-dinucleotide afhankelijke enzymen van *S. solfataricus* geeft dit enzym de voorkeur aan de cofactor $NADP^+$ boven NAD^+. Op moleculair niveau kon die worden verklaard door een aminozuur sequentie die een pocket vormt voor de extra fosfaatgroep van $NADP^+$. Ook de substraatvoorkeur voor

D-arabinose en L-fucose kon worden begrepen door middel van substraat-docking studies. De informatie verkregen in dit hoofdstuk illustreert de diverse functies die uitgevoerd worden door enzymen die behoren tot de MDR superfamilie. Ook is de informatie nuttig bij genoom annotaties en eiwit mutagenese studies.

Hoofdstuk vijf beschrijft de kristal structuur van het 2-keto-3-*deoxy*-D-arabinonate dehydratase, het derde enzym uit de pentose oxidatie route van *S. solfataricus*. De structuur van dit enzym bleek een ongebruikelijke ringvormige tetrameer te zijn, waarvan iedere subunit uit twee domeinen bestond. Kristal structuren van het enzym met het product 2,5-dioxopentanoaat en substraat analoog 2-oxobutyraat lieten zien hoe substraten in de active site zijn gebonden. Op grond hiervan worden speculaties gedaan over het katalytisch mechanisme van dit nieuwe enzym. Ook werd dit enzym uitgebreid vergeleken met homologe structuren waarvan nog geen functie bekend is.

In **hoofdstuk zes** staat de zuivering en biochemische karakterisatie van een nieuw α-galactosidase van *S. solfataricus* beschreven. Dit enzym haalt galactose van niet-reducerende uiteinden van suikers af. Het archaeale type bleek phylogenetisch niet te behoren tot het bacteriële thermostabiele type α-galactosidase van glycosyl hydrolase familie 36, maar tot een subklasse die veel eiwitsequenties van planten bevat. De betrokkenheid van voorspelde katalytische aspartaten werd geverifieerd door middel van plaatsgerichte mutagenese.

In **hoofdstuk zeven** wordt de modificatie van een β-glycosidase door middel van DNA family shuffling beschreven. Om de hydrolyse van afvalproduct lactose optimaal bij 70 °C te laten verlopen, werden de goede eigenschappen van twee β-glycosidases uit de hyperthermofiele Archaea *Pyrococcus furiosus* (hoge thermostabiliteit) en *S. solfataricus* (lage productinhibitie) gecombineerd in een hybride enzym. Naast verbeterde catalytische eigenschappen van het hybride enzym was het resultaat zeer verrassend omdat het aantoont dat het mogelijk is om netto 17 aminozuren tegelijkertijd te veranderen, zonder desastreuze effecten op de hyperthermostabiliteit van het enzym te veroorzaken.

In **hoofdstuk acht** is een selectiemarker ontwikkeld voor thermofiele micro-organismen, waardoor genetisch onderzoek aan deze klasse organismen makkelijker zal worden. Thermostabiele random mutanten van een bleomycine bindend eiwit werden geselecteerd in de thermofiele bacterie *Thermus thermophilus*. In een enkele selectieronde bij 77 °C werden acht dubbelmutanten geïdentificeerd in een populatie van 20.000. Deze mutanten werden biochemisch en structureel bestudeerd, waardoor de mechanismen die ten grondslag lagen aan de verbeterde thermostabiliteit konden worden opgehelderd. Het eiwit bleek op vier onafhankelijke manieren gestabiliseerd te zijn, waardoor het functioneel kon blijven bij aanzienlijk hogere temperaturen dat het wild-type eiwit.

EDUCATION STATEMENT OF GRADUATE SCHOOL VLAG

The graduate school hereby declares that the PhD candidate has complied with the educational requirements of the graduate school VLAG by following the discipline specific activities listed below.

Discipline specific activities

Courses	**ECTS**
3 months training, Thermal Biology Institute, Bozeman, USA	6.0
Protein Engineering (oral presentation), Wageningen, 2001 & 2003	1.5
Bio-informatics 2, Wageningen, 2002	2.8
Radiation expert 5B, WU & Larenstein, Wageningen/Velp, 2003	1.4
High Troughput Screening, RU, Groningen, 2002	1.4
Biomolecular crystallography, BESSY, Berlin, 2005	1.4
X-ray structure elucidation, SGC, Oxford, 2005	1.4
Meetings	
Annual EU SCREEN project meetings, Germany, Italy, UK, Sweden, 2001-2004	2.3
Annual Dutch Molecular Genetics meeting (poster and oral presentation), Lunteren, 2001-2007	4.4
Annual Dutch Protein Study Group (poster and oral presentation), Lunteren, 2003-2006	3.3
Annual Dutch National Biotechnological Conference, Ede, 2002-2003	0.6
Extremofiles conference (poster presentation), Napels, Italy, 2002	1.9
Gordon research conference Archaea (poster presentation), Boston, USA, 2003	1.9
Gordon research conference Archaea (poster and oral presentation), Oxford, UK, 2005	2.4
Extremofiles conference (oral presentation), Brest, France, 2006	1.9
General courses	
Scientific writing, Wageningen University, 2005	1.7
Optionals	
Preparation PhD research proposal	6.0
Bacterial Genetics weekly group meeting, 2001-2007	3.0
Microbiology biweekly PhD & postdoc meeting, 2001-2007	3.0
VLAG PhD trip Japan, 2005	2.5
Total	**50.8**

1 ECTS (European Credit Transfer System) represents the normative study load of 28 hours

CURRICULUM VITAE

Stan Johan Jozef Brouns werd op 10 maart 1978 geboren te ‘s-Hertogenbosch. In 1996 behaalde zijn VWO diploma aan het Jeroen Bosch college te ‘s-Hertogenbosch. Stan verliet vervolgens het ouderlijk huis om in Wageningen Moleculaire Wetenschappen te gaan studeren. Al snel werd duidelijk dat de biologische kant van de studie hem meer interesseerde dan de fysische, wat in 1999 resulteerde in een afstudeervak bij de vakgroep Microbiologie. Onder begeleiding van Thijs Kaper en dr. John van der Oost werd met behulp van nieuwe DNA shuffling technieken een hybride enzym gemaakt dat geschikt was om het afvalproduct lactose om te zetten in bruikbare suikers. Voor dit project ontving hij in 2000 de Unilever Research prijs.

Na dit succes werd bij de vakgroep Biochemie onder begeleiding van Yves Bollen en dr. Carlo van Mierlo de stabiliteit van een esterase onderzocht met behulp van spectroscopische technieken. Een 6 maanden durende stage naar het Thermal Biology Institute in Bozeman (Verenigde Staten) volgde om ervaring op te doen met het werken met thermofiele micro-organismen. Onder supervisie van dr. Ken Stedman en prof. dr. Mark Young werd de productie van een cellulase van het thermofiele archaeon *Sulfolobus solfataricus* met behulp van virale vectoren geoptimaliseerd.

Eenmaal terug in Nederland studeerde hij in juni 2001 af, waarna hij aan de slag ging als AIO bij het Laboratorium voor Microbiologie onder leiding van prof. dr. Willem M. de Vos in de Bacteriële Genetica groep van dr. John van der Oost. De resultaten van het promotieonderzoek die hij in de daarop volgende 4 jaar behaalde, zijn in dit proefschrift beschreven.

Na zijn promotieonderzoek bleef hij als Postdoc werkzaam binnen de Bacteriële genetica groep. Daar hield hij zich bezig met regulatienetwerken in Archaea en meer recentelijk met een nieuw immuun systeem in Bacteriën en Archaea.

LIST OF PUBLICATIONS

Kaper, T., **S.J.J. Brouns**, A.C. Geerling, W.M. de Vos, and J. van der Oost (2002) DNA family shuffling of hyperthermostable β-glycosidases. *Biochem J* **368**:461-470.

Berrisford, J.M., J. Akerboom, **S.J.J. Brouns**, S.E. Sedelnikova, A.P. Turnbull, J. van der Oost, L. Salmon, R. Hardre, I.A. Murray, G.M. Blackburn, D.W. Rice, and P. J. Baker (2004) The structures of inhibitor complexes of *Pyrococcus furiosus* phosphoglucose isomerase provide insights into substrate binding and catalysis. *J Mol Biol* **343**:649-657.

Brouns, S.J.J., T.J.G. Ettema, K.M. Stedman, J. Walther, H. Smidt, A.P.L. Snijders, M.J. Young, R. Bernander, P.C. Wright, B. Siebers, and J. van der Oost (2005) The hyperthermophilic archaeon *Sulfolobus*: from exploration to exploitation, p. 261-276. In W.P. Inskeep and T.R. McDermott (ed.), *Geothermal biology and geochemistry in Yellowstone national park*. Thermal Biology Institute, Montana State University, Bozeman.

Gödde, C., K. Sahm, **S.J.J. Brouns**, L.D. Kluskens, J. van der Oost, W.M. de Vos, and G. Antranikian (2005) Cloning and expression of islandisin, a new thermostable subtilisin from *Fervidobacterium islandicum*, in *Escherichia coli*. *Appl Environ Microbiol* **71**:3951-3958.

Brouns, S.J.J., H. Wu, J. Akerboom, A.P. Turnbull, W.M. de Vos, and J. van der Oost (2005) Engineering a selectable marker for hyperthermophiles. *J Biol Chem* **280**:11422-11431.

Snijders, A.P.L., J. Walther, S. Peter, I. Kinnman, M.G.J. de Vos, H.J.G. van de Werken, **S.J.J. Brouns**, J. van der Oost, and P.C. Wright (2006) Reconstruction of central carbon metabolism in *Sulfolobus solfataricus* using a two-dimensional gel electrophoresis map, stable isotope labeling and DNA microarray analysis. *Proteomics* **6**:1518-1529.

Berrisford, J.M., A.M. Hounslow, J. Akerboom, W.R. Hagen, **S.J.J. Brouns**, J. van der Oost, I.A. Murray, G.M. Blackburn, J.P. Waltho, D.W. Rice, and P.J. Baker (2006) Evidence supporting a *cis-ene*diol-based mechanism for *Pyrococcus furiosus* phosphoglucose isomerase. *J Mol Biol* **358**:1353-1366.

Brouns, S.J.J., N. Smits, H. Wu, A.P.L. Snijders, P.C. Wright, W.M. de Vos, and J. van der Oost (2006) Identification of a novel alpha-galactosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J Bacteriol* **188**:2392-2399.

Van der Oost, J., J. Walther, **S.J.J. Brouns**, H.J.G. van de Werken, A.P.L. Snijders, P.C. Wright, A. Andersson, R. Bernander, and W.M. de Vos (2006) Functional genomics of the thermo-acidophilic archaeon *Sulfolobus solfataricus*, p. 201-232. In F. Rainey and A. Oren (ed.), Extremophiles, *Methods in Microbiology*, vol. 35. Elsevier.

Brouns, S.J.J., J. Walther, A.P.L. Snijders, H.J.G. van de Werken, H.L.D.M. Willemen, P. Worm, M.G.J. de Vos, A. Andersson, M. Lundgren, H.F.M. Mazon, R.H.H. van den Heuvel, P. Nilsson, L. Salmon, W.M. de Vos, P.C. Wright, R. Bernander, and J. van der Oost (2006) Identification of the missing links in prokaryotic pentose oxidation pathways: evidence for enzyme recruitment. *J Biol Chem* **281**:27378-27388.

Van den Heuvel, R.H.H., E. van Duijn, H.F.M. Mazon, S.A. Synowsky, K. Lorenzen, C. Versluis, **S.J.J. Brouns**, D. Langridge, J. van der Oost, J. Hoyes, and A.J. Heck (2006) Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* **78**:7473-7483.

Wu, H., L. Sun, **S.J.J. Brouns**, S. Fu, J. Akerboom, X. Li, and J. van der Oost (2007) Purification, crystallization and preliminary crystallographic analysis of a GTP-binding protein from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Acta Crystallograph Sect F Struct Biol Cryst Commun* **63**:239-241.

Van de Werken, H.J.G., **S.J.J. Brouns**, and J. van der Oost (2007) Pentose metabolism in Archaea. *In press.*

Brouns, S.J.J., A.P. Turnbull, H.L.D.M. Willemen, J. Akerboom, and J. van der Oost (2007) Crystal structure and biochemical properties of the D-arabinose dehydrogenase from *Sulfolobus solfataricus*. *J Mol Biol* **371**: 1249-1260.

Brouns, S.J.J., T.R.M. Barends, P. Worm, J. Akerboom, A.P. Turnbull, L. Salmon, and J. van der Oost (2007) Structural insight into substrate binding and catalysis of a novel 2-keto-3-*deoxy*-D-arabinonate dehydratase illustrates common mechanistic features of the FAH superfamily. *Submitted for publication.*

Machielsen, R., Leferink, N., Hendriks, A., **Brouns, S.J.J.**, Henneman, H., Daussman, T., Van der Oost, J. (2007) Laboratory evolution of *Pyrococcus furiosus* Alcohol Dehydrogenase to improve the production of (2S, 5S)-hexanediol at moderate temperatures. *Submitted for publication.*