Structural evolution of function and stability in muconate lactonizing enzymes

Tommi Kajander

Department of Biosciences, Division of Biochemistry and
Institute of Biotechnology, Research Program in Structural Biology and Biophysics,
Faculty of Science, University of Helsinki,
Helsinki, Finland

Academic Dissertation

To be presented for public criticism, with the permission of the Faculty of Science, University of Helsinki, in the auditorium 2041 at Viikki Biocenter, Viikinkaari 5, Helsinki, on February 21st, at 12 o'clock noon.

Supervised by: Professor Adrian Goldman
Institute of Biotechnology
Research Program in Structural Biology and Biophysics
University of Helsinki
Helsinki, Finland

Reviewed by: Docent Sarah Butcher
Institute of Biotechnology
Research Program in Structural Biology and Biophysics and
Deparment of Biosciences
University of Helsinki
Helsinki, Finland

Professor Juha Rouvinen
Department of Chemistry
University of Joensuu
Joensuu, Finland

Official Opponent: Professor Günter Schneider
Division of Structural Biology
Department of Medical Biochemistry and Biophysics
Karolinska Institute
Stockholm, Sweden

ISBN 952-10-0338-3
ISBN 952-10-0339-1 (PDF)
http://ethesis.helsinki.fi
Yliopistopaino
Helsinki 2003

Table of Contents

Original publications

I Schell U, Helin S, Kajander T, Schlömann M and Goldman A (1999). Structural basis for the activity of two muconate lactonizing enzyme variant toward substituted muconates. Proteins 34:125-136.

II Kajander T, Kahn PC, Passila SH, Cohen DM, Lehtiö L, Adolfsen W, Warwicker J, Schell U and Goldman A (2000). Buried charged surface in proteins. Structure Fold Des 8:1203-1214.

III Merckel MC, Kajander T, Deacon AM, Thompson A, Grossman JG, Kalkkinen N and Goldman A (2002). 3-carboxy-*cis,cis*-muconate lactonizing enzyme from *Neurospora crassa*: MAD phasing with 80 seleniums. Acta Crystallogr D58:727-734.

IV Kajander T, Merckel MC, Thompson A, Deacon AM, Mazur P, Kozarich JW and Goldman A (2002). The stucture of *Neurospora crassa* 3-carboxy-*cis,cis*-muconate lactonizing enzyme, a β propeller cycloisomerase. Structure 10:483-492.

V Kajander T*, Lehtiö L* and Goldman A (2003). The structure of *Pseudomonas* P51 Cl-muconate lactonizing enzyme – co-evolution of the hydrophobic core structure and dynamics with the dehalogenation function (Manuscript).

*These authors contributed equally to this work.

Abbreviations

2-PGA	2-phosphoglycerate
ASA	accessible surface area
CD	circular dichroism
Cl-MLE	chloromuconate lactonizing enzyme
CMLE	3-carboxy-*cis,cis*-muconate lactonizing enzyme
FPLC	fast performance liquid chromatography
GalD	galactose dehydratase
GlucD	glucarate dehydratase
GluDH	glutamate dehydrogenase
MAL	methylaspartate ammonia lyase
MAD	multiple wavelength anomalous diffraction
MLE	muconate lactonizing enzyme
MPD	2,5-methylpentanediol
MR	mandelate racemase
NAAAR	N-acylamino acid racemase
NCS	noncrystallographic symmetry
NMR	nuclear magnetic resonance
ORF	open reading frame
OSBS	*o*-succinylbenzoate synthase
P51 Cl-MLE	*Pseudomanas* P51 chloromuconate lactonizing enzyme
PAGE	polyacrylamide gel electroforesis
PCB	polychlorinated biphenyl
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PEP	phospoenolpyruvate
PIPES	piperazine-1,4-bis(2-ethanesulphonic acid)
PpMLE	*Pseudomonas putida cis,cis*-muconate lactonizing enzyme
SeMet	selenomethionine
SSHB	short strong hydrogen bond
ReCl-MLE	*Ralstonia euthropa* chloromuconate lactonizing enzyme
RhamD	rhamnose dehydratase
TIM	triose phosphate isomerase

Abstract

Muconate lactonizing enzymes (MLEs) convert *cis,cis*-muconates to muconolactones in soil microbes as part of the β-ketoadipate pathway that funnels aromatic breakdown products, mainly from lignin, into citric acid cycle intermediates. There are three distinct evolutionarily-unrelated classes of MLEs from eucaryotes and bacteria, all of which catalyze the same chemical reaction, the cycloisomerisation of *cis,cis*-muconates to muconolactones. The reaction mechanism is the reverse of a β-elimination reaction by a general base driven proton abstraction from the α-carbon of a carboxylate, a reaction that produces an unstable enolate intermediate. It is a matter of some debate how catalysis of proton abstraction from the relatively inert α-carbon is achieved. Furthermore, the bacterial Cl-MLEs have also evolved to catalyze a unique dehalogenation reaction, but with the same catalytic machinery as MLEs. This has required subtle changes in the active site, and perhaps elsewhere in the structure.

The work described here has concentrated on the structural analysis of the stability and function of MLEs. The structural basis of function has been probed by structure determination and site-directed mutagenesis within two of the three MLE groups with distinct structural or functional characteristics. First, for the bacterial MLEs, both structures of variants of *Pseudomonas* MLE (Ile54Val, Phe329Ile, Asp178Asn) and the wild type high resolution structure of *Pseudomonas* Cl-MLE have been solved. Secondly, the structure of an enzyme with a completely different fold and sequence, *Neurospora crassa* CMLE, has been solved and its mechanism studied. Finally, it was known that *Pseudomonas* MLE has a highly charged active site, and large amount of charged surface in total was buried (Helin *et al.* 1995). It was therefore of interest to examine the MLE hydrophobic core in terms of the individual buried charged residues and their distribution. The studies were further extended to a representative set of structures from the Protein Data Bank. It was found that MLE is no exception for a protein of its size: large proteins bury tens of charged residues. While solvation of charges by proteins is of general importance for understanding function and folding of proteins, it appears also that Cl-MLEs have perhaps utilized changes in their core structure related to stability to evolve a new function, dehalogenation. These structural data can provide a better understanding of MLE function and evolution.

1. Introduction

The muconate lactonizing enzymes (MLEs) are an evolutionarily diverse group of enzymes, comprised of three unrelated classes of enzymes. They are present in soil micro-organisms, both procaryotes and fungi (*e.g. Neurospora* and *Trichosporon*). The MLEs are enzymes of the β-ketoadipate pathway (Figure 1) and catalyze the conversion of *cis,cis*-muconic acid (and some of its substituted forms) to muconolactone as a step in the breakdown of catechols and protocatechuate to citric acid cycle intermediates. The three different classes of MLEs are the bacterial MLEs, the bacterial 3-carboxy-*cis,cis*-muconate lactonizing enzymes (CMLEs), and the eucaryotic MLEs and CMLEs. They are all unrelated to each other at the sequence and also presumably at the structural level. This non-homologous evolutionary origin is accompanied by mechanistic differences.

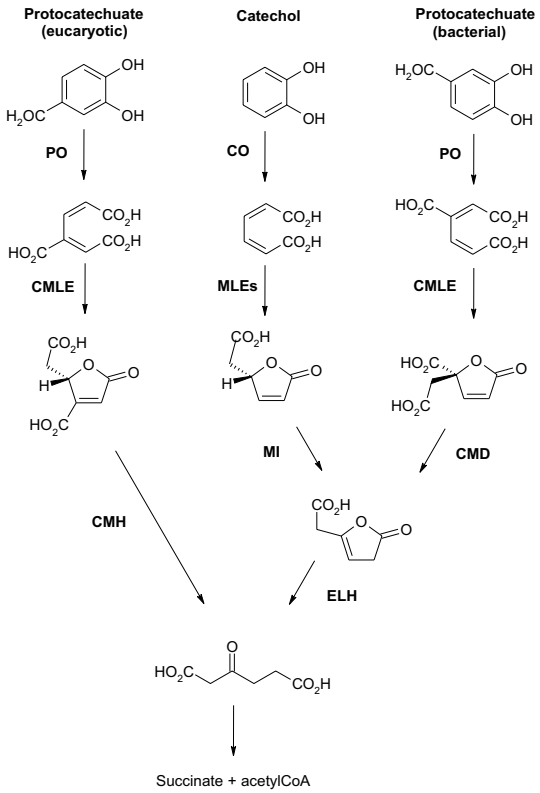

Figure 1. The three branches of the β-ketoadipate pathway. PO = protocatechuate-3,4-dioxygenase, CO = catechol-1,2-dioxygenase, MI = muconolactone isomerase, CMD = 4-carboxy-muconolactone decarboxylase, ELH = enol lactone hydrolase, CMH = 3-carboxymuconolactone hydrolase; MLE and CMLE are explained in the text.

The bacterial enzymes are homo-octameric α/β-barrel enzymes with an N-terminal α+β-domain (Figure 2) belonging to the enolase superfamily (Babbitt *et al*. 1995). The bacterial CMLEs are related by sequence homology to the class II fumarase superfamily of proteins, including fumarase, adenylosuccinate lyase and δ-crystallin (Simpson *et al.* 1994, Weaver *et al.* 1995, Toth and Yeates

2000). The structures of several proteins from this superfamily are known and they have an all-α-helical fold, with an elongated tetrameric structure (Figure 2), each active site peculiarly being constituted of three subunits (Simpson *et al.* 1994). Finally, in the eucaryotic MLEs, the CMLE and MLEs seem to be related by divergent evolution rather than having evolved separately, as in procaryotes. They are not related to any other known protein family. The bacterial MLEs catalyze a *syn*-addition and require a metal-cofactor, whereas the bacterial CMLEs catalyze *anti*-addition without a metal cofactor. The eucaryotic enzymes do not require a metal cofactor, but catalyze a *syn*-addition (Table 1).

Table 1. The classes of MLEs.

Enzyme Class	**Stereo chemistry**	**Metal requirement**	**Sequence similarity**
Bacterial MLEs	*syn*	divalent cation	enolase superfamily
Bacterial CMLEs	*anti*	-	class II fumarases
Eucaryotic MLEs and CMLEs	*syn*	-	none

We thus concluded that solving the structure of *Neurospora* CMLE might provide us with important information regarding the essential features of active site architecture for 1,2-addition-elimination reactions on carboxylic acids, as a natural example of how differing protein structures may serve as templates for the same reaction. The MLE catalyzed addition-elimination reaction of proton at the α-carbon of the carboxylate presents a theoretical problem of how catalysis might be achieved, and has been discussed in the literature recently (*e.g.* Gerlt *et al.* 1991, Gerlt and Gassman 1992, 1993 and Guthrie and Kluger 1993). This is because the α-carbon has a high pK_a of 25-30 and a carbanion intermediate must be produced. However, the existence of carbanion-intermediates has been questioned previously (Jencks 1980), and therefore the reaction might proceed differently (Gerlt *et al.* 1991).

In addition, the procaryotic MLEs have diverged to dehalogenate xenobiotic haloaromatic break-down products, mainly the chloro-substituted forms of muconates (see section 2.3.2). Also some fluorobenzoates are not catabolized fully by dehalogenating microbes: for example 2-fluoro-muconate is not converted by MLEs or Cl-MLEs (Boersma *et al.* 1998). In general, fluorinated organic compounds are considered more inert than other halogenated organics, while they may be biologically active (Key *et al.* 1997). The fate of such fluorinated aromatic and aliphatic compounds in the environment is poorly understood, but some are highly persistent (Key *et al.* 1997).

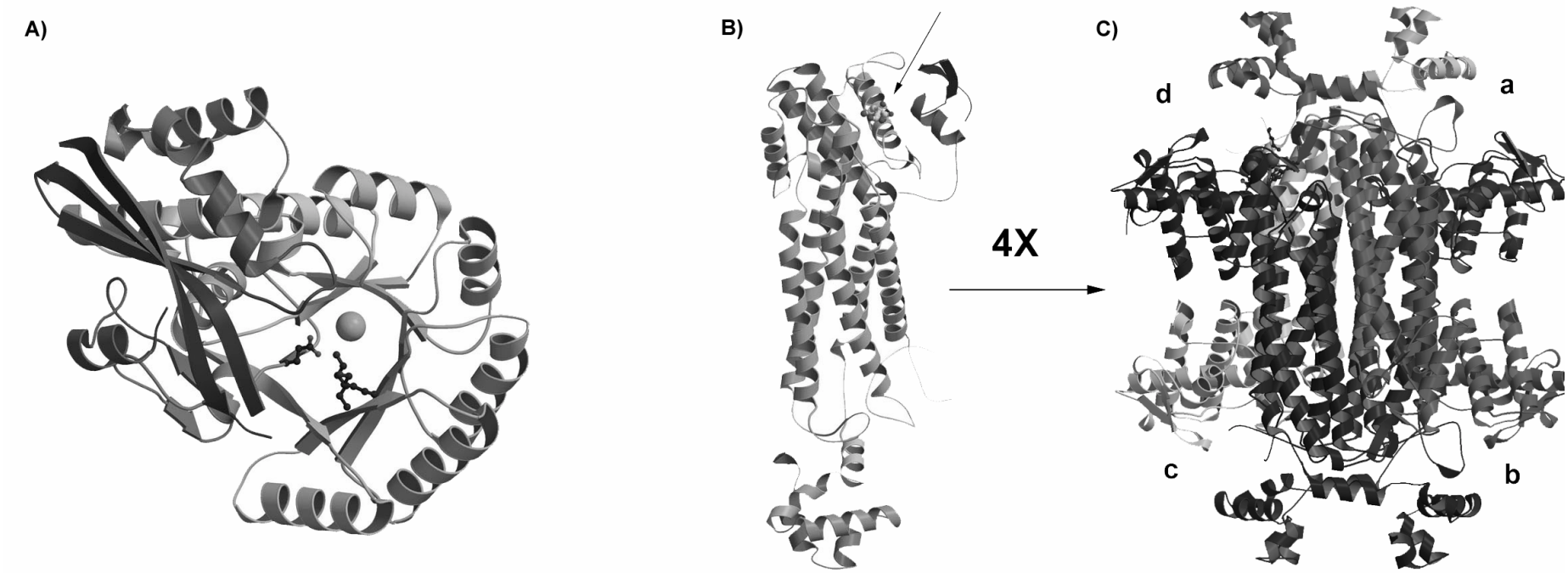

Figure 2. The structures of the two types of MLEs. A) bacterial MLE B) bacterial CMLE (represented by the fumarase monomer structure; arrow indicates active site, with bound citrate). C) The fumarase tetramer, where each of the four active sites (a-d) contains residues from three different monomers (Weaver *et al.* 1995).

One of the goals here, for the above reasons, is to understand the catalytic mechanism and requirements for dehalogenation by MLEs. In this work, the structural basis of substrate specificity was studied by examining the structures of two MLE active site variants and by modelling the binding of substituted substrates. Furthermore, the structure of *Pseudomonas* P51 Cl-MLE was solved to high resolution (1.95 Å) to study the differences between the structures of MLEs and Cl-MLEs in more detail. Comparison of the P51 Cl-MLE structure to the pre-existing low resolution (3 Å) structure of *Ralstonia euthropa* Cl-MLE (ReCl-MLE) also confirms interesting conserved structural features between Cl-MLEs, but diverging from MLEs.

Finally, the refined structure of the bacterial *Pseudomonas putida* MLE (PpMLE) revealed interesting features in terms of its charge distribution. A number of charged residues were buried in MLE, as was a high percentage of the accessible surface area of the charged atoms. This led to studies of this phenomenon in more depth, as the role of buried charges with regards to stability has remained unclear and in particular the extent to which buried charges and salt-bridges are present in protein structures in general. A representative collection of non-homologous structures from the Protein Data Bank (PDB) was analyzed with respect to the distribution of charged polar and other types of surface, while a small subset was analyzed in detail for their buried charged residues. It was confirmed that a large number of buried charges are found in larger proteins, indicating that protein interiors can be rather polar, as has been previously suggested based on electrostatics calculations and observations on some individual structures (Warshel 1978, Quiocho *et al.* 1987). However, our study showed that buried charges also occur frequently outside the active sites of proteins. It was also shown that buried charges are not always optimized for stability; in MLE, the Asp178Asn variant was more stable than wild type.

2. Review of the literature

2.1. Degradation of aromatics by microbes

Many of the catabolic activities in microbes involving breakdown of aromatic species exist in microbes associated with plants, as plants produce many aromatic compounds, including the lignin polymer of woody plants. Lignin, the major structural constituent of wood and the second most abundant polymer in the biosphere after cellulose (25% of the land-based biomass, Harwood and Parales 1996), is metabolized to relatively simple aromatic compounds that are further utilized by soil microbes. In addition, man-made aromatic pollutants, including halogenated species, are processed by soil microbes. The recycling of aromatics by microbes is thus an important step in the Earth's carbon cycle. Both aerobic and anaerobic microbes degrade aromatic compounds. In general, the aerobic pathways have been better studied. The aerobic degradation of aromatics occurs in several steps which can be divided into preparative ring modifying steps, the ring fission to produce linear hydrocarbons, and finally, processing of the products to give tricarboxylic acid cycle intermediates (Harwood and Parales 1996). The muconate lactonizing enzymes studied here are part of the aerobic *ortho*-fission degradation pathway, catalyzing the conversion of *cis,cis*-muconate, the ring fission product of catechol (Figure 1).

Pathways for the aerobic fission of aromatic rings and conversion to simpler compounds are the *ortho*-, *meta* and *gentisate* cleavage pathways. The best studied are the *ortho* and *meta*-cleavage pathways (Harwood and Parales 1996). The *meta*-pathway includes the breakdown of methyl-substituted aromatic rings, such as toluene and xylene, while the *ortho* pathway has been modified in some bacteria for processing of Cl-catechols (produced *e.g.* from chlorobenzenes and chlorobenzoates). Based on an analysis of the funneling routes of mono- and dimeric aromatic model compounds of lignin breakdown, the *ortho*-pathway is involved in the breakdown of numerous aromatic intermediates (Cain 1980).

It may be that the major aerobic pathways for catechol and protocatechuate breakdown evolved along with the appearance of woody land plants some hundreds of millions of years ago. However, other small aromatic compounds of biological origin, such as flavonoids exist in the environment, and probably existed prior to land plants. Some fungi synthesize aromatic compounds (Packter and Ward 1972), and so it is possible that this pathway is older than woody plants. The muconate lactonizing enzymes discussed here are a component of the *ortho*-pathway, also known as the β-ketoadipate pathway, and importantly they catalyze the final step in the complete dehalogenation of chlorinated catechols.

2.2. Catalysis by MLEs and related enzymes

2.2.1. The reaction catalyzed by MLE

The MLEs catalyze a 1,2-addition-elimination reaction on *cis,cis*-muconate resulting in cycloisomerisation to (4*S*)-muconolactone, by the reverse of a β-elimination reaction (Figure 3). In MLEs this is either a *syn*- or an *anti*-addition, as the stereochemistry depends on the type of enzyme. Several other enzymes structurally related to the bacterial MLEs or CMLEs catalyze similar β-elimination reactions on carboxylic acid substrates. Probably the best studied are mandelate racemase and enolase, of the enolase superfamily, to which the bacterial MLEs also belong. Therefore, I concentrate here on the description of the reaction by the bacterial MLEs that catalyze a *syn*-addition with a Mn^{2+} cofactor (Table 1).

The bacterial MLE crystal structure (Goldman *et al.* 1987, Helin *et al.* 1995) revealed the location of the active site and possible catalytic residues. Through homology to the catalytic residues of the related enzyme mandelate racemase (MR), a putative general base and acid have been identified (Neidhart *et al.* 1990, Helin *et al.* 1995) (Figure 4). All ten enzymes in the family known to date also bind a Mn^{+2} or Mg^{+2}-metal cofactor in the active site (Gerlt and Babbitt 2001, Levy *et al.* 2002). Furthermore, even without a structure of MLE complexed with substrate, the binding is still strictly defined relative to catalytic groups by the stereochemistry of the reaction: it is known that the intramolecular nucleophilic attack by the substrate's other carboxylate yields (4*S*)-muconolactone, and that the proton added to the 5-carbon (the α-carbon) is added on the same side as the nucleophile oxygen. It follows that the reaction proceeds by *syn*-addition (Figure 3, Ngai *et al.* 1983), and that the *syn*-addition occurs on one defined side of the substrate double bond plane.

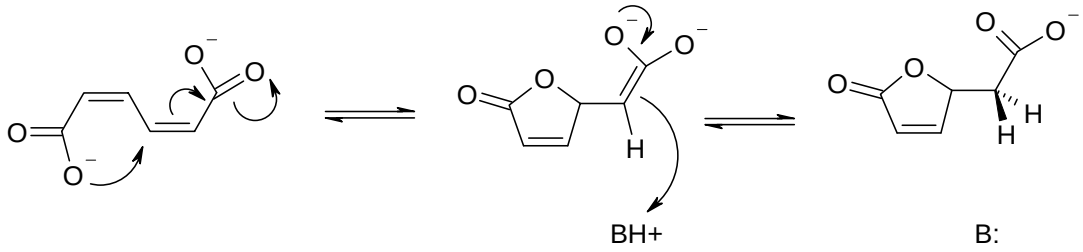

Figure 3. The MLE catalyzed cycloisomerisation with *syn*-addition of *cis,cis*-muconate through an enolate carbanion intermediate by an enzyme general base.

The metal dependence in this superfamily has led some researchers to suggest that enzyme catalysis of α-proton abstraction from enolate anions is achieved by the use of metal-cofactors (Richard and

Amyes 2002). This is not always the case: in MLEs there are two classes of enzymes that do not utilize metal cofactors (Mazur *et al.* 1994, Williams *et al.* 1992).

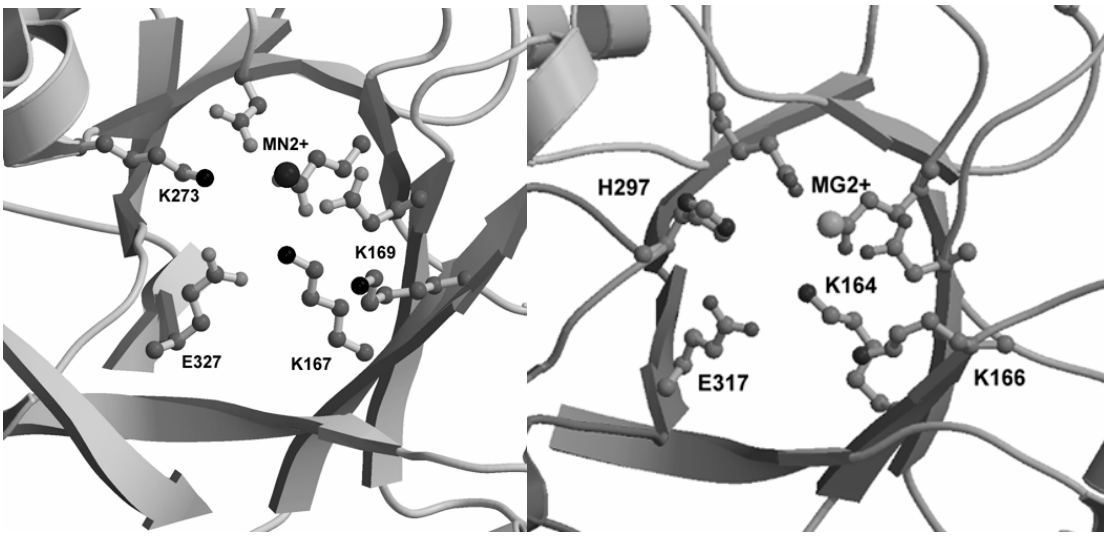

Figure 4. The active sites of MLE (left) and MR (right).

The reaction of bacterial MLEs proceeds in two steps. Firstly, an enolic or enolate intermediate is formed from the carboxylate by cycloisomerisation to lactone, and secondly this is tautomerized to the keto-form with addition of hydrogen α to the enolate to produce the muconolactone product (Figure 3). Production of the enolate as an intermediate would result from step-wise E1-elimination. However, some researches have argued such enolate carbanions are too unstable to exist for the time of one bond vibration (Jencks 1980). If so, this could not be an intermediate of the reaction. Consequently, the reaction might occur by *enforced* concerted general acid-base catalysis instead. This would result in E2-elimination, to protonate the enolate intermediate or even to generate a short strong hydrogen bond (SSHB) with an enolic, not enolate, intermediate (see below) (Jencks 1980, Gerlt and Gassman 1993). However, others suggest electrostatic stabilization of enolate does not require concerted acid-base catalysis (Gluthrie and Kluger 1993). The catalytic mechanisms of several other enzymes have also been suggested to involve the formation of enolate intermediates, including glyoxalases (Creighton and Hamilton 2001) dehydroquinases (Harris *et al.* 1996) and enolase (Stubbe and Abeles 1980, Anderson *et al.* 1994). Further, the enolate intermediate can be produced in water by general acid catalysis of an intramolecular reaction (Amyes and Kirby 1988). In this case, the reaction was intramolecular cycloisomerisation with phenolate oxygen as the nucleophile adding to the carboxylate activated double bond (the α-proton then being *added* by the conjugate base of the general acid catalyst). Thus, the reaction is actually analogous to that catalyzed by MLE (Figure 5). Thus, it is not clear whether the reaction might occur step-wise or in

concert, and whether the enolate dianion intermediate is actually formed.

2.2.2. The enolase superfamily of enzymes

Since the initial finding that MLE and MR structures are homologous despite catalyzing distinct reactions, several related enzymes have been studied both by enzymological and crystallographical methods (Gerlt and Babbitt 2001). The MR-MLE family, which was established as a superfamily of enzymes catalyzing abstraction of proton in the α-position to a carboxyl group (Babbitt *et al.* 1995), was named the enolase superfamily after the assumed progenitor of both enzymes with a related structure [enolase was originally suggested to be related to MLE by Lebioda and Stec (1988)]. Babbitt *et al.* (1995, 1996) have divided the superfamily into three subclasses: enolase, MLE and mandelate racemase subgroups, according to the conservation of key functional residues. Below I will go through the members of this superfamily to examine the common minimal requirements for the reaction.

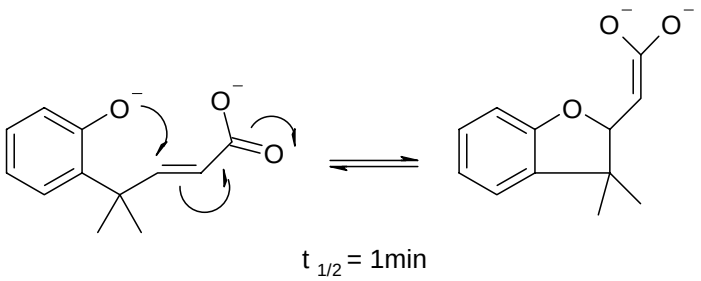

Figure 5. Intramolecular addition to double bond by phenolate oxygen in 4-methyl-4-(2-hydroxyphenyl)pent-2-enoic acid (Amyes and Kirby 1988). This reaction produces a detectable enolate intermediate; note the analogy to *cis,cis*-muconate.

2.2.3. Mandelate racemase

Mandelate racemase (MR) catalyzes the interconversion of (*R*)- and (*S*)-enantiomers of mandelic acid (Hegeman *et al.* 1970). The reaction proceeds by abstraction of the acid's α-proton and reprotonation from either face of the acid (Figure 6). The same reaction at the same carbon occurs in both directions, deprotonation always being the first step and protonation the second. In many other related enzymes, such as MLE, there is a second reaction step in addition to the enol/keto tautomerisation of the carboxylic acid; in MLE it is the closing of the lactone ring. For MR, it has been argued based on pK_a-differences of mandelic acid tautomers and the protein catalytic groups, that a step-wise (E1-elimination) reaction mechanism with a carbanion enolate formation cannot explain the observed reaction rate of $k_{cat} = 10^3\ s^{-1}$ for MR (Gerlt and Gassman 1992). The reaction requires the abstraction of the α-proton of the carboxyl group, which is relatively slow and creates an enolate intermediate that is considered to be very unstable, perhaps too unstable to exist long

enough for the reaction to happen (see above). Consequently, a concerted mechanism would be more likely (Jencks 1980, Gerlt *et al.* 1991).

Gerlt *et al.* (1991) suggested that a general acid in MR is used to protonate the carboxyl group of the carbon acid undergoing the reaction to yield a protonated species (Figure 6) which would then more readily also donate the α-proton, as protonation would be expected to lower the very high pK_a of >25 of the α-proton by 15-20 units to values similar to those observed for catalytic groups in enzyme active sites. The protonation by the general acid residue would make the first half of the reaction a concerted E2-elimination to generate the enolic carboxylic acid intermediate. In this case, different second steps such as reprotonation in MR or cycloisomerisation in MLE could occur, and the overall reaction mechanism would be step-wise (Gerlt and Gassman 1992). Others have suggested that the catalytic problem for MR is essentially thermodynamic because of the instability of the enol-tautomer of the carboxylic acid. Consequently, strong charge interactions due to the buried metal cofactor and lowered polarity of the active site compared to solution could be the main stabilizing factor for the enolate (Guthrie and Kluger 1993). However, this model is likely to be in part erroneous, as the active site cannot be a low polarity environment, and ion pairs in hydrophobic environments are unstable, despite strong attractive forces. It is still probable that electrostatic interactions stabilize the enolate intermediate considerably, but rather by optimized polar interactions by the groups on the protein (Warshel 1978, Warhsel *et al.* 1989).

Structures of MR in complex with substrate-analogues and inhibitors, and the kinetic effects of mutations are known from a wealth of data (Mitra *et al.* 1995; Landro *et al.* 1991, 1994; Kallarakal *et al.* 1995). Residues His279 and Lys166 on either side of the substrate perform the racemization by protonation reactions at the α-carbon (Powers *et al.* 1991, Neidhart *et al.* 1991, Landro *et al.* 1994) (Figure 7).

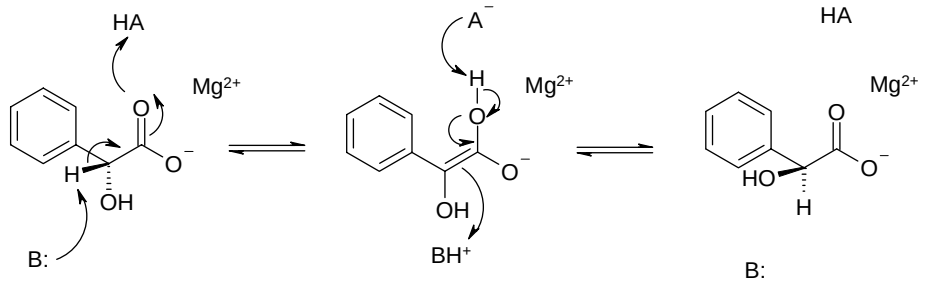

Figure 6. Racemization of mandelate by MR. BH^+ is the general base (Lys166) and HA the general acid (Glu327).

Furthermore, Lys164 and the essential Mg^{2+}-ion provide charge neutralization of the substrate carboxylate, and have been postulated to stabilize the enolate reaction intermediate. Structures of

substrate analogue complexes suggest that the substrate is bidentately coordinated to the Mg^{2+} cofactor (Figure 7) with one carboxyl oxygen, and through its α-hydroxyl-group (Neidhart *et al.* 1991, Landro *et al.* 1994). The same substrate carboxyl-oxygen that was bonded to the Mg^{2+} -ion is also hydrogen bonded to Lys164.

Probably the key interaction, however, is between Glu317 and the substrate carboxylate group. Glu317 is hydrogen bonded through its side chain carboxylate to peptide backbone and to Ser139 on opposite sides of the substrate, thus positioning it very precisely. This is reflected in the fact that the effective concentration of Glu317 is higher than that of any other catalytic residue (Bearne & Wolfenden, 2000; see below). Glu317 forms a charged hydrogen bond to the substrate carboxylate group (Figure 7), suggesting that it is protonated. Consequently, it has been proposed that concerted acid-base catalysis takes place: Glu317 protonates the enolate intermediate to stabilize it and lower the pK_a of the α-proton. Indeed, protonation of the *neutral* acid carboxyl group (to $-C=OHOH^+$) could lower the pK_a of the α-proton to ~7 (Gerlt *et al.* 1991). Consequently, Gerlt *et al.* (1991) suggested that deprotonation of the α-carbon is concerted with protonation of the nascent enolate species by Glu317, and that the enolate is further stabilized by hydrogen bonds and Mg^{2+} (Figure 7). It has also been suggested that there may be an SSHB between Glu317 and the carboxylate, with a matched pK_a (Gerlt and Gassman 1993), but the Glu317-carboxylate distances in all the structures are 2.7-2.8 Å, and there is no convincing evidence for an SSHB in this or any other enzyme active site (Warshel and Papazyan 1996).

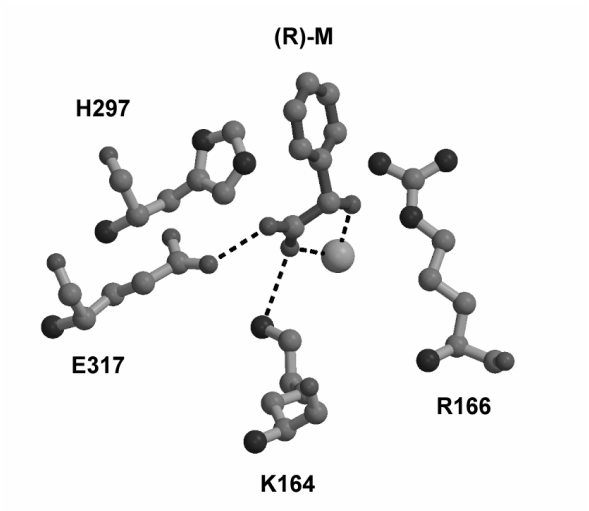

Figure 7. The substrate complex of MR. The structure is that of the inactive Lys166Arg mutant (Kallarakal *et al.* 1995), therefore Lys166 is replaced with Arg166. (*R*)-M is the (*R*)-stereo-isomer of mandelic acid.

The Glu317Gln-mutant has a 10^4-fold reduced k_{cat}, consistent with its role as a general acid, while no structural alterations were observed in the mutant enzyme active site (Mitra *et al.* 1995). The mechanism of the half-reaction is thus likely to be a concerted E2-elimination. Further evidence that Glu317 protonates the enolate is that it forms the only hydrogen bond to the enolate oxygen in the substrate analogue complexes (Kallarakal 1995, Mitra *et al.* 1995, Landro *et al.* 1994) and the enolate anion is unlikely to be more acidic than the protonated Glu317.

The MR-type reaction has recently been studied in solution with mandelic acid by mimicking the enzyme active sites groups His297, Lys166 and Glu317 with imidazole, methylamine and acetate ions as catalysts (Bearne and Wolfenden 2000). The results were used to derive effective concentrations for the catalytic groups by relating them to the specificity constants (k_{cat}/K_m) of wild type and the corresponding mutant enzymes where these groups were missing (His297Asn, Lys166Arg, and Glu317Gln), as measured by Gerlt and co-workers (Kallarakal 1995, Mitra *et al.* 1995, Landro *et al.* 1991). The effective molarities were calculated as follows. The transition state affinity constants on the enzyme were estimated as $K_{tx} = k_{non} / (k_{cat}/K_m)$ for wild-type and the variants, and the forward rate constant in solution for mandelic acid and a specific side chain mimic (imidazole, methylamine and acetate) was used to give a value for K_{tx}^{sc} ($K_{tx} = k_{non} / k_{sc}$), the transition state affinity constant for the side chain mimic. The effective molarities for the catalytic side chains are then given by $K_{tx}^{variant} \times K_{tx}^{sc} / K_{tx}^{wt}$ or $(k_{cat}/K_m)^{wt} / (k_{cat}/K_m)^{var} \times k_{non} / k_{sc}$.

The results indicated that Glu317 has the highest effective concentration (3×10^5 M), compared with 622 M for His 297 and 3×10^3 M for Lys166. This suggests that the orientation of Glu317 is the most restricted, in accordance with its proposed role. The authors measured an activation free energy for transition state formation of 34.6 kcal/mol at 25 °C for the non-catalyzed reaction. The calculated rate constant from this was $k_{non} = 3 \pm 10^{-13}$ s^{-1}, and so the results indicate that MR is an extremely efficient enzyme with a rate enhancement of 10^{15} and a $K_{tx} = 2 \times 10^{-19}$ M. Interestingly, Bearne and Wolfenden (2000) also measured a catalytic effect for magnesium in solution catalysis of mandelate racemization which was not much smaller than the contribution of methylamine.

In summary, the major contributions to MR catalysis are due to the general acid and base, and the precise positioning of substrate and enzyme catalytic groups. Additional contributions are due to hydrogen bonding and increased electrostatic stabilization of the transition state versus solution. The most important component of electrostatic catalysis must be the Mg^{2+}-cofactor. It is interesting that both electrostatic stabilization by the divalent metal-ion and protonation by the general acid may be required for efficient catalysis in MR, although some related enzymes, *e.g.* OSBS, GlucD and enolase, lack the general acid completely (Thompson *et al.* 2000, Gulick *et al.* 2000, Larsen *et*

al. 1996, Zhang *et al.* 1997). Consequently, understanding the MR mechanism is helpful for considering other similar reactions, in particular those for the enolase superfamily. I discuss below various cases of the enolase superfamily, and analyze how the requirements of these reactions fit with the hypothesis presented above.

2.2.4. Enolase

Enolase too, catalyzes the abstraction of the carboxylate α-proton from its substrate, 2-phosphoglycerate (2-PGA), to generate phosphoenolpyruvate (PEP) (Figure 8). In the first step of the reaction, the α-proton is removed by general base catalysis, and in the second reaction step, the hydroxyl from the hydroxymethyl group of 2-PGA is eliminated, resulting in the elimination of water to give the reaction: 2-PGA → PEP + H_2O. The structure of enolase was solved in 1988 (Lebioda and Stec 1988), but only recently have the major ambiguities concerning catalysis been resolved, with the solution of a substrate-product complex (mixture) with both Mg^{2+}-ions bound, and by site-directed mutagenesis and kinetics studies (Poyner *et al.* 1996, Larsen *et al.* 1996, Zhang *et al.* 1997). Enolase has been suggested to be the probable ancestor of MLE and MR (Babbitt *et al.* 1995, 1996), as it is the most widespread of the enzymes and has a central cellular function as a part of the glycolytic pathway, hence the name *enolase* superfamily (Babbitt *et al.* 1995).

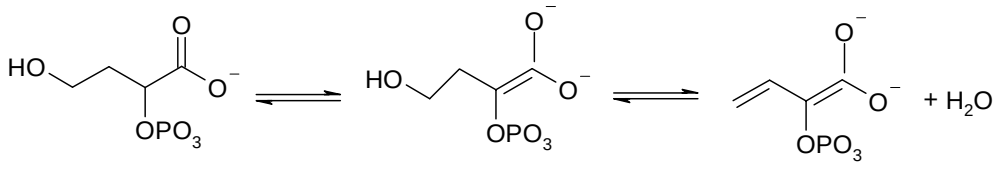

Figure 8. The enolase catalyzed reaction of conversion of 2-phosphoglycerate to phospoenolpyruvate.

Enolase is also the most divergent member of the superfamily: the sequence identity with MLE is 15 % (sequence alignment without structure superposition would not reveal homology), while the sequence identity between MLE and MR is 27 %. Enolase differs from the other members in having two divalent metal-cofactors (Mg^{2+}). The enolase α/β-barrel also differs from MLE and MR in that the barrel has $\beta\beta\alpha\alpha(\alpha\beta)_6$ topology (Lebioda *et al.* 1989; Figure 9), while in MLE and MR the last helix of the barrel is replaced by the C-terminal subdomain, resulting in $(\beta\alpha)_7\beta$ topology. There are also considerable differences in the loops of the N-terminal capping domain, and the catalytic Glu211 residue of enolase resides on a long loop which does not exist in MLE or MR (Goldman *et al.* 1987, Lebioda and Stec 1988, Neidhart *et al.* 1990).

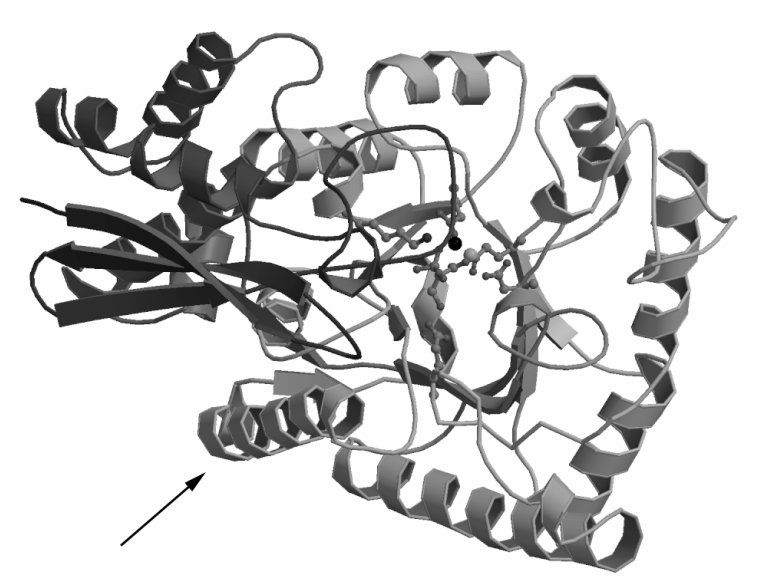

Figure 9. The fold of enolase. The arrow indicates the α/β-barrel helix missing in MLE and MR.

Studies on the mechanism of enolase support a step-wise mechanism for enolization (first half reaction to generate the enolate; Figure 8), in contrast to MR (Stubbe and Abeles 1980, Anderson *et al.* 1994, Poyner *et al.* 1996). The step-wise mechanism intimately involves two Mg^{2+}-ions. The second Mg^{2+}-ion (the activating Mg-ion) binds after the substrate and is released before the product (Faller *et al.* 1977, Poyner *et al.* 2001). Structures show that the high affinity Mg^{2+}-ion is bound to both of the substrate/product carboxylate oxygens and the three active site carboxylate groups, while the low affinity Mg^{2+}-ion is bound to one of the substrate carboxylate oxygens, the phosphate-group of the substrate, two water molecules and Ser39 (Larsen *et al.* 1996, Zhang *et al.* 1997; Figure 10). Presumably the high negative charge on the transition state (enolate + phosphate groups; -4 altogether) requires the presence of two Mg^{2+}-ions. However, it is possible that during catalysis conformational changes result in an "overneutralized phosphate" protonated by His159, because the phosphate group also binds to Arg374 and the second Mg^{2+}, thus giving a formal charge of +1 on the system (Zhang *et al.* 1997). Indeed His159 moves 3.3 Å closer to the phosphate moiety, and the interaction of Ser39 with low-affinity Mg^{2+} results in loop closure over the active site upon substrate binding (Wedekind *et al.* 1994).

In enolase Lys345 is the general base that abstracts the α-proton and Glu211 the general acid that binds to the leaving group -OH (Poyner *et al.* 1996; Figure 10). Lys345, interestingly, is the homologue of Lys273 in the MLE active site. This clearly links enolase and MLE at another conserved position, in addition to the metal binding residues of the superfamily. In MR, Asp270, hydrogen bonded to the other catalytic base of MR, His297, occupies the same position. It is clear that in enolase there is no equivalent of Glu317, the electrophilic/general acid catalyst suggested to

form an SSHB in MR, and it would seem that in enolase the two metal-ions stabilize the carbanion-intermediate of the carboxylate electrostatically (Larsen *et al.* 1996). Rate enhancement is thus largely due to electrostatic stabilization of the phosphate moiety and the enolate intermediate producing the "over-neutralization" of phosphate during binding by protonation by His159 and coordination of the enolate lowering the pKa for the α-proton to allow attack by Lys345 (Zhang *et al.* 1997, Larsen *et al.* 1996, Wedekind *et al.* 1994, Hilal *et al.* 1995), without concerted general acid-base catalysis. Consequently, both step-wise and concerted mechanisms occur in different enzymes of the same superfamily.

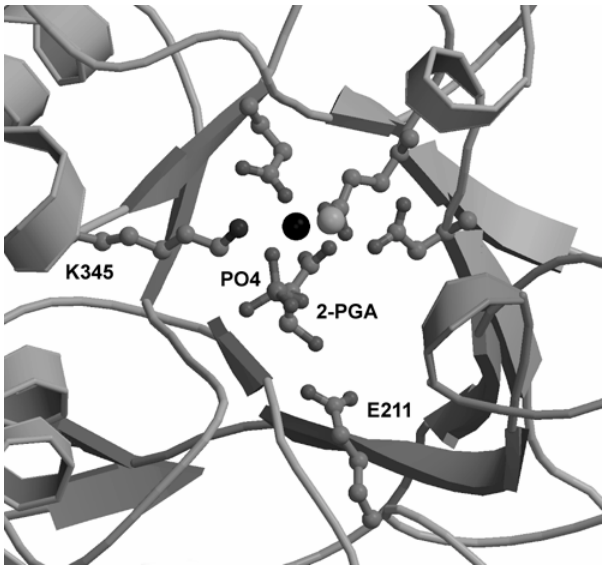

Figure 10. The enolase active site with substrate 2-PGA and two divalent metal ions bound. Lys345 is the general base and Glu211 the general acid. The metal ligands for the conserved metal binding site are shown. The high affinity metal cofactor is shown as a light grey sphere, while the second, low affinity metal ligand is shown as a black sphere.

2.2.5. Muconate lactonizing enzymes

Here I concentrate on the bacterial MLEs because they are members of the enolase superfamily. Ngai *et al.* (1983) determined the K_m and k_{cat}-values for *Pseudomonas putida* MLE. For *cis,cis*-muconate they obtained the following values: k_{cat} = 14 s^{-1} and K_m = 90 µM, but more recent data suggest that actually k_{cat} = 210 s^{-1} and K_m = 40 µM for the same *Pseudomonas* enzyme (Vollmer *et al.* 1998). The dissociation constant measured for Mn^{2+} by Ngai *et al.* (1983) was ~4 µM, and the activity with various divalent cations were: Mn (100%) >> Ni (53%) > Co (47%) > Mg (40%), but with all the above metal ligands the K_m-values for substrate were about the same (~100 µM). Other evidence for the functional groups and mechanism is largely indirectly derived from studies of the related MR reaction (see above).

The refined high resolution structure of MLE (Helin *et al.* 1995), and the low resolution structure of *Ralstonia eutropha* Cl-muconate lactonizing enzyme (ReCl-MLE) have been determined (Hoier *et al.* 1994, Kleywegt *et al.* 1996). More recently mutagenesis data on the substrate specificity determining residues has become available (Vollmer *et al.* 1998, see Section 2.4) and Lys169 has been confirmed as the general base for the catalytic reaction (Kaulmann *et al.* 2001). Direct experimental evidence for the MLE reaction mechanism in the form of structures of substrate or inhibitor complexes has been lacking, as it has not been possible to grow crystals of MLE with substrate nor soak substrate into the crystals. Undoubtedly, the MLE active site resembles that of MR (Neidhart *et al.* 1990). In particular the active site lysines 167 and 169 and the Mx^{2+}-ligands (Asp, Asp, Glu) are conserved between the two systems, as well as the general acid, Glu327 in MLE. Thus, it is evident that similar catalytic machinery is used in both enzymes (Figure 4).

The reaction for MLE, unlike for MR, is the reverse of β-elimination, and so a proton is added to the α-carbon of the enolate (Figure 3), which is the thermodynamically favorable direction for this reaction step. Nonetheless, the formation of the enolate remains the central catalytic problem. In the reverse direction from product to substrate proton-abstraction would still be the rate-limiting step. The MLE reaction in forward direction is likely to be limited by the nucleophilicity of the attacking carboxylate, and by the stability of the enolate formed (Figure 3).

Overall, in MLE, as in MR, production of the enolate intermediate is the rate limiting step. The *cis,cis*-muconate substrate carboxylate group can act as a nucleophile, but the conjugated double bonds of *cis,cis*-muconate further delocalize the charge on the carboxylate, making it a worse nucleophile than if it were not conjugated. MLE may facilitate catalysis by binding and desolvation of the substrate to increase the nucleophilicity of the attacking carboxylate. Protonation by the Glu327 general acid in MLE is beneficial as in MR because it stabilizes the enolate. In MLE protonation of enolate by the general acid (Glu327) can occur in concert with the attack by the carboxylate to produce the lactone. This is then followed by concerted acid-base catalysis for the tautomerization step from enolate to (4*S*)-muconolactone product (Helin *et al.* 1995, Figure 11).

A subgroup of bacterial MLEs has also diverged as a result of the adaptation of some bacteria to grow on Cl-aromatics. In those bacteria the enzymes catalyze a second, apparently general base catalyzed step of proton abstraction with Cl^- as the leaving group (HCl is eliminated) from 5-Cl-muconolactone or 4-Cl-muconolactone to form *cis*- or *trans*-dienelactone (Blasco *et al.* 1995, Vollmer *et al.* 1994, Vollmer *et al.* 1995). Here, perhaps the charge on the leaving Cl-anion needs to be stabilized by the active site somehow (5-Cl-muconolactone may be also disubstituted, but only

one Cl-anion is abstracted). Thus, the sequence of events in the forward direction for Cl-MLEs would be *e.g.*:

2-Cl-muconate + Base-**H**$^+$ → 5-Cl-muconolactone-**H** + Base: → *trans*-dienelactone + Base-**H**$^+$

which makes it possible for the same base (Lys169) to act as the general base in both steps. Substrate specificity and dehalogenation by MLEs have been studied mostly by Schlömann and co-workers (*e.g.* Vollmer *et al.* 1998, Kaulmann *et al.* 2001; see also Paper I); dehalogenation by MLEs will be considered in a separate section below.

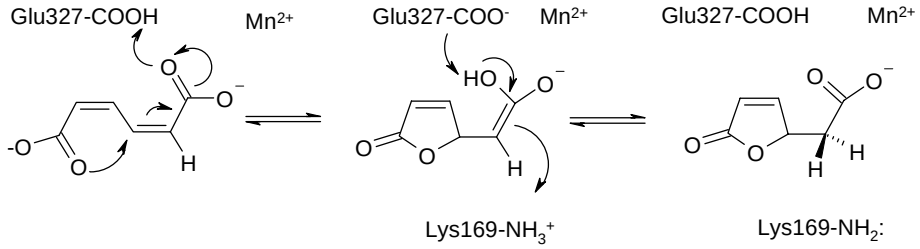

Figure 11. A possible scheme for concerted acid-base catalysis by bacterial MLEs.

2.2.6. The other enolase superfamily enzymes

Besides MLE, MR and enolase, a number of other related microbial enzymes have been characterized biochemically. Altogether the superfamily consists of at least ten enzymes of known function (Gerlt and Babbitt 2001, Babbitt *et al.* 1995, 1996). Babbitt *et al.* (1995) grouped the enzymes into three subgroups based on conservation of active site residues in each subgroup, in addition to the conserved metal ion protein-ligands in the superfamily. In the mandelate racemase subgroup the catalytic base His297 and the MR motif Lys 164-X-Lys166 are conserved, while in the MLE group, the same Lys-X-Lys-motif and the active site base Lys273 (of MLE) are conserved. Interestingly, the Lys273 position is also conserved in enolases as the general base Lys345. In the enolase group, which contains only one enzyme in addition to enolase, the general base Lys345 and His373, which hydrogen bonds the general acid Glu211 in enolase, are the key conserved residues in addition to the metal ligands (Babbitt *et al.* 1996).

The MR subgroup

The mandelate racemase subgroup includes glucarate dehydratase (GlucD), galactose dehydratase (GalD) and rhamnose dehydratase (RhamD), enzymes catalyzing the dehydration of related acidic carbohydrates (Gerlt and Babbitt 2001). The *E. coli* GlucD catalyzes both the dehydratation of D-glucarate and L-idarate and their epimerisation (Palmer *et al.* 1998). The epimerisation activity seems to follow from the fact that GlucD possesses an active site similar to MR, with two bases, (*R*) and (*S*)-specific (Palmer *et al.* 1998). Based on the structure of the *E. coli* enzyme, however, GlucD does not seem to have a general acid residue equivalent to MR Glu317 (Gulick *et al.* 2000). The equivalent residue, Asp366, is further away and there is a water molecule between it and the carboxylate of the substrate, which is bidentately coordinated to the metal cofactor. The substrate is directly coordinated to the metal ion. Presumably the enzyme achieves the required stabilization through electrostatic interactions of the substrate with the metal cofactor. Consequently, the stabilization of the enolic intermediate by a general base is not a general strategy nor is it absolutely required for the catalysis. Interestingly, unlike for all the other enzymes of the superfamily, one of the GlucD metal ligands is an uncharged asparagine (Gulick *et al.* 2000).

The MLE subgroup

The MLE-subgroup includes the enzymes *o*-succinylbenzoate synthase (OSBS) from *E. coli*, the L-Ala-D/L-Glu epimerases and methylaspartate ammonia lyase (MAL) (Babbitt *et al.* 1995). Of these enzymes, only MLE catalyzes an addition reaction; the rest catalyze proton abstraction. OSBS catalyzes the β-elimination of water from a six-carbon ring of its substrate, and makes it aromatic *via* elimination of the α-hydroxyl of a carboxylate on the ring. This produces *o*-succinyl-benzoate. The reaction is highly exothermic and the substrate is unstable at elevated temperatures and pH deviating from neutral (Gerlt and Babbitt 2001). OSBSs from several bacterial species share very low sequence similarity. In the case of *B. subtilis* the sequence identity was 15% to previously identified sequences – which apparently form a more coherent core-group (Gerlt and Babbitt 2001). Indeed, another OSBS from *Amycolaptosis* was originally classified as N-acylamino acid racemase (NAAAR), as it has low racemase activity (Palmer *et al*. 1999)! Thus, it must be capable of using the homologues of Lys273 and Lys169 of MLE in racemization, as does MR (with the homologous His297 and Lys166). The OSBS structure diverges from the canonical α/β-barrel structure with a "split barrel", where α-helix 5 of the MLE-type $(\beta/\alpha)_7\beta$-barrel is missing (Thompson *et al.* 2000). Further, as for GlucD, there is no general acid to protonate the enolate (Thompson *et al.* 2000). OSBS is also a small protein; it is only 320 amino acids long and monomeric, while MLE and MR contain 373 and 359 residues and are octameric, and enolase is 436 amino acids long and dimeric.

The L-Ala-D/L-Glu dipeptide epimerases perform MR-type 1,1-proton transfer, but are classified as members of the MLE subgroup, like OSBS, and similarly utilize two lysines as catalytic bases for racemization. They catalyze the racemase reaction on the glutamic acid residue of the dipeptide substrate. The structures of the L-Ala-D/L-Glu epimerases from *Bacillus subtilis* and *Escherichia coli* were determined recently (Gulick *et al.* 2001). The two epimerases differ in length considerably (321 versus 366 residues), with the larger being an octamer and the smaller a dimer. They also share only 31 % sequence identity, and so are almost as far apart from each other as *e.g.* MLE and MR (sequence identity 27 %). In addition, as for OSBS, the short enzyme does not align well with the longer enzymes of the subgroup, whereas MR and MLE from the different subgroups align rather well (r.m.s.d. for Cα-atoms = 1.7 Å).

The so-called MLE-subgroup is thus highly heterogeneous, as is confirmed by a closer look at the last identified member enzyme, MAL. MAL catalyzes the reversible β-elimination of ammonia from 3-methyl aspartic acid. Unlike other enzymes of the enolase superfamily, it requires not only divalent Mg^{2+}-ions for activity, but also monovalent ions, *e.g.* K^{+} (Kato and Asano 1995). The sequence identity between MAL and MLE is negligible; merely the functional residues are conserved. MAL aligns best with enolase (330 aligned Cαs with an r.m.s.d. of 1.89 Å per Cα), and like enolase, it has a long chain with 413 amino acids, and is a dimer. The α/β-barrel of MAL has all eight helices, instead of the C-terminal subdomain of MLE and MR (Levy *et al.* 2002). Furthermore, the conserved Lys-X-Lys sequence of the MLE subgroup (Babbitt *et al.* 1996) does not have any significance in the MAL-family because it is not a conserved motif (Levy *et al.* 2002). Levy *et al.* (2002) suggest Lys331, which is homologous to enolase Lys 345, as the general base and His194 as the electrophile, analogous to Lys167 of MLE. However, His194 might be the equivalent of the general acid Glu317 in MR that donates the proton to the enolate intermediate. Thus, the classification of MAL into the MLE-subgroup based on this sequence motif is not correct; it may lie somewhere between enolase and MLE, but is clearly structurally closer to enolase than to MLE! Similarly OSBS may not be closely related to MLE.

The enolase subgroup

A gene suggested to encode carboxyphoshonoenolpyruvate decarboxylase activity (Lee *et al.* 1995) was identified as an enolase-related sequence. Interestingly, the functionally related enzyme phoshonopyruvate decarboxylase was recently found to be related to the alkaline phosphatase superfamily (Galperin *et al.* 1998). Otherwise no close relative enzymes of enolase were found (Babbitt *et al.* 1995), or have been reported since, although in my opinion MAL should be included in this group, based on its structural homology, rather than in the MLE subgroup.

2.2.7. Summary of the enolase superfamily

Altogether, the data that have accumulated on the structures in the enolase superfamily would seem to suggest a need to modify some of the original proposals (Babbitt *et al.* 1994, 1995; Gerlt and Gassmann 1992, 1993). Firstly, the mechanism is similar in all cases but is not always general acid driven "electrophilic catalysis". It would seem to be an exception as at least OSBS, GlucD, and enolase appear not to use it. Enolase utilizes clearly differing catalytic machinery, with essentially only the first Mg^{2+}-ligands conserved with the rest of the superfamily, besides Lys345 of enolase being conserved in MLE as Lys273 and in MAL as Lys331. Further, unlike the others, enolase binds a second "activating" Mg^{2+}-ion with the substrate. Thus, the feature that is conserved in the family is the metal-aided stabilization of the enolate intermediate (Gerlt and Babbitt 2001), not the hydrogen bonding (SSHB) or protonation by general acid to stabilize the enolate, as originally suggested (Babbitt *et al.* 1995, 1996; Gerlt *et al.* 1991, Gerlt and Gassman 1992, 1993).

Secondly, the subgrouping to MLE, MR and enolase subgroups is not convincing. Enolase is distantly related to both MLE and MR (with sequence identity 15% to MLE); it therefore may form a defined group of proteins more distantly related to other members of the superfamily (see above). Also the enolase sequences have two major insertions relative to MLE and MR (Babbitt *et al.* 1995). The MLE-subgroup as a definition is unconvincing, as it appears to be a collection of enzymes rather loosely related by evolution, with properties that could link them more closely to MR (for OSBS and the dipeptide epimerases) and enolase (MAL) than to MLE. On the other hand MR is the closest relative to MLE among the proteins discussed.

Otherwise the MR-group appears as the least diverse, but the enzymes in the MLE group seem to be related to MR, and therefore this division to MLE and MR does not seem useful. It is clear that enolase is the most distantly related enzyme and that MAL might be a kind of a "missing link" between enolase and the rest of the superfamily. This, together with the fact that Lys345 of enolase is homologous to MLE Lys273, clearly link enolase with the rest of the enzymes. The MLE, MR and enolase "subgroups" have also somewhat similar active sites and are the only enzymes among all the TIM-barrel enzymes that have this kind of metal-coordination and domain structure. They are also very distantly related to other enzymes with the α/β-barrel fold (Copley and Bork 2000). It is thus justified to group them as a superfamily, although the relations in the superfamily are far from clear.

2.3. Dehalogenation by MLEs

2.3.1. Organohalogenic compounds and the environment

The introduction of xenobiotic pollutants into the environment, such as haloalkanes and haloaromatics, has resulted in the evolution of catalytic activities in soil microbes, particularly in bacteria, to metabolize these novel compounds. Mineralization in natural environments is often very slow, and it is therefore important to try to understand the catalytic mechanisms responsible and the pathways involved in the mineralization processes (Abraham *et al.* 2002). On the other hand, organohalogenic compounds have existed in the environment prior to their production by the chemical industry. Natural sources of halogenated organic compounds range from forest fires and volcanic eruptions (abiotic sources) to biosynthesis by corals and many other marine animals, lichens, insects, as well as higher animals and humans. For instance, the chloroaromatic diazepam (valium) has been isolated from human brains that were preserved before diazepam was synthesized in the laboratory. Organochlorines found in pristine natural forests and lakes are formed by chloroperoxidase action on humic acid, Cl^- and hydrogen peroxide (Gribble 1998).

Natural production of halo-organics can be larger than release by industrial sources: for instance about 28 x 10^6 tons/year of monochloromethane is produced from forest fires, agricultural burning and by marine organisms, while the industrial production is only *ca.* 0.6 x 10^6 tons (Fetzner 1998)! Chlorinated phenyls are produced by fungi in amounts comparable to industrial production (de Jong *et al.* 1994). Most naturally occurring organohalogens are chloro or bromo-compounds, with iodo-compounds less observed, and fluoro-compounds rarest (Fetzner 1998). The biosynthesized organohalogens are largely compounds used in chemical defense by the organism producing it (pesticides, antimicrobial agents) (Fetzner 1998).

The production of xenobiotic halogenated organic compounds by industry and release to the environment, however, is the major source for many, often toxic, halogenated compounds in nature. For instance, 1,2-dichloroethane and monochloroethene were the fourth and fifth most produced bulk organic chemicals in the USA in 1992, after ethene, propene and urea (Fetzner 1998). Pollution of the environment by organic solvents is substantial: for tetrachloroethene and trichloroethene the entire yearly production is needed to replace leakage into the environment, and halogenated aliphatic compounds are found everywhere (surface and ground water, food, urban air) (Fetzner 1998). Around 15 million tons/year of 1,2-dichloroethane and monochloroethene were produced yearly in the 1980s, and the entire production of aromatic halogenated compounds was probably close to 1 million tons per year (Fetzner 1998).

Most problematic and toxic of these compounds are those that tend to accumulate, usually high in the food-chain, in living organisms due to their fat-solubility and low rate of biodegradation, such as polychlorinated biphenyls (PCB). About 1.2 million tons of PCBs have been produced in the world since 1929 and 30 % of that is estimated to be widely spread and persisting in the environment, despite the fall in production because of restrictions (Fetzner 1998). The production of millions of tons of these chemicals for industrial purposes may override the natural capability to metabolize and mineralize these often quite toxic compounds, and local contaminations from industrial leakages can easily effectively poison the ground and water over large areas. The microbial catabolism and dehalogenation of these compounds is thus of major interest for bioremedation.

2.3.2. Dehalogenating enzymes

Several classes of enzymes with dehalogenation activity (dehalogenases) exist, and both aerobic and anaerobic pathways have been identified. Dehalogenases use a variety of catalytic mechanisms including oxidation, reduction, proton removal, and substitution by hydrolysis, thiolysis or intramolecular substitution. In some anaerobic bacteria dehalogenation occurs by methyl-transfer (Janssen *et al.* 2001, Fetzner 1998). The dehalogenating bacterial MLEs are dehydrodehalogenases.

In aerobic bacteria dehalogenation proceeds by oxidative, substitutive and dehydrodehalogenation. Aerobic dehalogenating enzymes acting on aromatic compounds catalyze dehalogenation before or during the cleavage of the aromatic ring, whereas in some catabolic pathways enzymes have been recruited that catalyze the dehalogenation step after ring cleavage (*e.g.* Cl-MLEs in the modified ortho-pathway). A number of mono- and dioxygenases cleave both haloaromatic and haloaliphatic compounds. Oxygenases are the major class of enzymes that perform dehalogenation before aromatic ring cleavage, and this is important particularly for polyhalogenated species, such as tetrachlorobenzene (Janssen *et al.* 2001). Oxygenases are further utilized to cleave aromatic ring structures (of *e.g.* catechol). Interestingly, dioxygenolytic dehalogenation of polyhalogenated aromatics is thought to proceed through formation of *cis*-diols which spontaneously reorganize to give a (Cl-)catechol with simultaneous chloride elimination (Fetzner 1998). Thus a continuous pathway of dehalogenation exists from oxidative dehalogenation by dioxygenases to the continued dehydrodehalogenation by the MLEs in the modified *ortho*-pathway.

Hydrolytic dehalogenases are the most common dehalogenation enzymes. Aliphatic halogenated substrates are processed by the hydrolytic haloalkane dehalogenases which are enzymes with an α/β-hydrolase fold domain combined with a capping domain (Ollis *et al.* 1992, Verschueren *et al.* 1993). As in other α/β-hydrolase type enzymes the active site contains a catalytic triad, here with an

aspartate as a nucleophile that abstracts the halogen from the substrate. Catalytic residues are located on the α/β-hydrolase domain, while the capping domain contains a conserved tryptophan which binds the halogen anion (Verschueren *et al.* 1993, Newman *et al.* 1999). Hydrolytic dehalogenation takes place *via* nucleophilic displacement and formation of a covalent enzyme-intermediate with the substrate alkane (Verschueren *et al.* 1993).

Under anaerobic conditions dehalogenation occurs either via reductive processes, alkyl transfer or by substitutive dehalogenation (thiolysis). Reductive dehalogenation can also be linked to carbon-metabolism or respiration (halorespiration). In halorespiration usually H_2 or formate functions as the electron donor, the halogen-compound as the electron acceptor and dehalogenation is coupled to production of ATP by a chemiosmotic mechanism (a proton gradient across the membrane) (Wohlfarth and Diekert 1997). Cobalamin-containing membrane-associated proteins and iron-sulphur cluster proteins have been associated with respiratory dehalogenation of chlorophenols, tetrachloroethene and trichloroethene. For polychlorinated aromatics, an initial reductive anaerobic cleavage of halogen-substituents is required (*e.g.* polychlorinated dibenzodioxins and dibenzofurans and PCBs) (Tiedje *et al.* 1993, Fetzner 1998, Wittich 1998); for PCB the higher Cl-substituted forms are good substrates for anaerobic degradation and this probably occurs by halorespiration (Abraham *et al.* 2002).

In *Sphingomonas chlorophenolica* tetrachlorohydroquinone dehalogenase catalyzes the reductive cleavage of chlorine from the aromatic ring structure and is involved in pentachlorophenol metabolism (Anandarajah *et al.* 2000). In addition to dehalogenation, tetrachlorohydroquinone dehalogenase also catalyzes the gluthathione-dependent isomerization of the double bond in maleylacetone. Also, a gluthathione dependent reductive dehalogenase (LinD) that acts on chlorohydroquinone has been found in *Shingomonas paucimobilis* UT26, and thiolytic enzymes from facultative methylotrophic bacteria that dehalogenate dichloromethane have been shown to be inducible gluthathione S-transferases. Thus, gluthathione S-transferase-type enzymes are involved in a number of cases of dehalogenating activity found in microbes (Fetzner 1998).

Dehydrodehalogenation results in elimination of HCl from the substrate. *S. paucimobilis* UT26 has an enzyme for the degradation of γ-hexachlorocyclohexane (*i.e.* lindane); the LinA-dehalogenase of this bacterium is an example of a dehydrodehalogenating enzyme. Complete dehalogenation of lindane actually involves four different dehalogenases: LinA, the above mentioned LinD, as well as a hydrolase and an NADH-dependent dehydrogenase (Miyauchi *et al.* 1998).

2.3.3. Dehalogenation by bacterial Cl-muconate lactonizing enzymes

The Cl-MLEs provide an example of a dehydrodehalogenation reaction. The Cl-muconate lactonizing enzymes perform a key step in the modified *ortho*-pathway in bacteria. This results in dehalogenation of Cl-catechol by removing the Cl^--ion(s) from Cl-muconolactones by elimination of Cl^- from the muconolactone product of the MLE-cycloisomerisation reaction to yield *trans*- and *cis*-dienelactones and removal of HCl (Schlömann 1994, Harwood and Parales 1996). The Cl-catechols are major intermediates of aerobic catabolism of chloroaromatics, such as the herbicide 2,4-dichlorophenoxyacetate and Cl-benzoates, which are breakdown products of PCBs (Abraham *et al.* 2002, Harwood and Parales 1996, Schlömann 1994).

Some of the substituted muconolactone products, such as 4-fluoromuconolactone, are unstable and undergo dehalogenation under physiological conditions without catalysis (Schlömann *et al.* 1990). However, it has been shown that the products 5- and 2-Cl-muconolactone from 2-Cl-muconate are stable, and Cl-muconate lactonizing enzyme is required for the decomposition of 2-Cl-muconate to *trans*-dienelactone from 5-Cl-muconolactone (Vollmer *et al.* 1994), while 2-Cl-muconolactone is less produced and not converted (Figure 12). Furthermore, 3-Cl-muconate and 2,4-diCl-muconate are also processed by Cl-MLEs to dehalogenated *cis*-dienelactone and, probably, to 2-Cl-*cis*-dienelactone, respectively (Blasco *et al.* 1995, Kaulmann *et al.* 2001). The MLEs convert 3-Cl-muconate to the bacteriotoxic protoanemonin (Blasco *et al.* 1995) by decarboxylation and associated dehalogenation (Figure 12), which results in bacterial cell death, while the MLEs are in general higly inefficient in converting any of these chlorinated muconates. It has been also noted that the toxicity of protoanemonin, produced by MLEs in certain bacteria, may be a problem associated with slow degradation of PCBs by natural microbial communities (Abraham *et al.* 2002).

There are at least three requirements for dehalogenation in Cl-MLEs. Firstly, 2-Cl-muconate needs to be bound in only one orientation so that only 3,6-cycloisomerisation occurs. This way only 5-Cl-muconolactone, not 2-Cl-muconolactone, is produced, which then can be converted to *trans*-dienelactone. Secondly, Cl-MLEs must avoid the formation of the toxic protoanemonin from 3-Cl-muconate in order for the bacteria to survive on the particular Cl-aromatic in question (Figure 12). Thirdly, the dehalogenation in Cl-MLEs occurs on the molecular level *via* elimination of Cl^- and production of dienelactone (*cis* or *trans*) by removal of a proton by an active site base. This results in a second protonation step after formation of the corresponding muconolactone. Therefore a base is required to perform this second protonation reaction.

It has been suggested that enzymes might gain novel but related substrates by broadening their substrate specificity (Schmidt and Knackmuss 1980). In the case of bacterial MLEs, however, the

Cl-MLEs have not broadened their substrate specificity, but have actually shifted it clearly towards certain types of substituted muconates (Vollmer *et al.* 1998). Further, differences in substrate specificity on Cl-muconates also exists for Cl-MLEs: 2,4-diCl-muconate is clearly the preferred substrate for most Cl-MLEs, while 2-Cl-muconate is in general poorly converted (Vollmer *et al.* 1999).

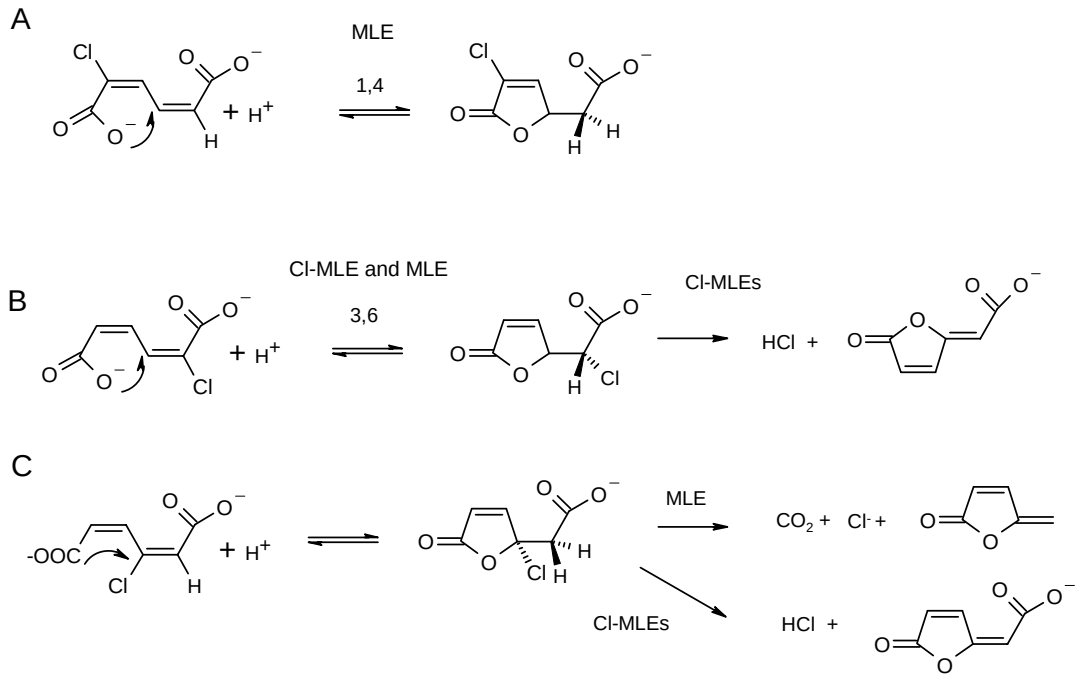

Figure 12. Reactions of MLEs and Cl-MLEs on Cl-substituted *cis,cis*-muconates. MLEs mostly catalyze the cycloisomerisation as in (A) (1,4 = 1,4-cycloisomerisation), and to lesser extent as in (B) (3,6 = 3.6-cycloisomerisation). C) MLEs produce the bacteriotoxic protoanemonin and CO_2 and Cl from 4-Cl-muconolactone derived from 3-Cl-muconate, whereas Cl-MLEs avoid protoanemonin formation.

Recently, an effort was made to determine which residues in ReCl-MLE are responsible for substrate specificity and for orienting 2-Cl-muconate correctly in the substrate binding pocket (Vollmer *et al*. 1998). Residues in the PpMLE active site were mutated to the corresponding residues in ReCl-MLE (Hoier *et al.* 1994). The mutations made were: Ile54 to Val, Ser271 to Ala, Lys276 to Asn, Tyr59 to Trp, Phe329 to Ile and Lee333 to Val, and various combinations thereof (Figure 13; Vollmer *et al.* 1998). The effect on substrate specificity and product formation was then analyzed. Some point mutations had a favorable effect on substrate turnover and specificity for 3-Cl-muconate and 2,4-Cl-muconate; in particular variants Ser271Ala and Ile54Val had over 20-fold increased activity for 3-Cl-muconate, but none of the variants displayed significant dehalogenation activity (>10%) and none of the mutations improved the turnover of 2-Cl-muconate. The mutations affected the direction of cycloisomerisation in a surprising manner: Phe329Ile and Tyr59Trp had marked preference for formation of 2-Cl-muconolactone *via* 1,4-cycloisomerisation, while Ile54Val

and Ile54Val-Tyr59Trp formed 5-Cl-muconolactone *via* 3,6-cycloisomerisation (Vollmer *et al.* 1998).

Intriguingly, it appears that the modified *ortho*-pathway for Cl-catechol degradation has evolved separately in Gram-positive bacteria and Gram-negative bacteria (Solyanikova *et al.* 1995, Eulberg *et al.* 1998). In the Gram-positive *Rhodococcus opacus* CP1, the Cl-MLE is more closely related to the *Rhodococcus* MLE than to Cl-MLEs of Gram-negative bacteria. However, the *Rhodococcus* Cl-MLE is unable to dehalogenate Cl-muconates and substrate specificity for *Rhodococcus* MLE is different from the other bacterial MLEs as it preferentially produces 5-Cl-muconolactone (Solyanikova *et al.* 1995; Figure 12). For the other types of MLEs, *i.e.* the eucaryotic MLEs and CMLEs and bacterial CMLEs, no dehalogenating activity has been reported.

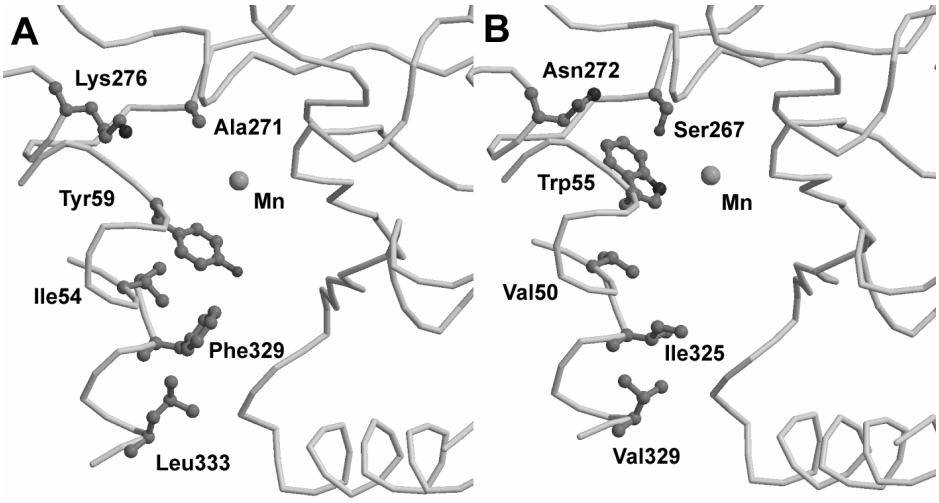

Figure 13. The differences in residues lining the active site in MLE vs Cl-MLE. A) MLE active site B) Cl-MLE active site.

2.4. The effect of charged residues on protein stability

2.4.1. Principles and theory of protein stability

Burial of hydrophobic side chains in the core of the protein away from the aqueous environment (the hydrophobic effect) is important for protein folding (Kauzmann 1959, Tanford 1970, Dill 1990), while at the same time the formation of secondary structure occurs by more specific interactions, contributing to both the stability and specificity of the native structure. Polar interactions are at least as important as the hydrophobic effect in determining precise folding. It has previously been suggested that these two types of interactions might have differing roles, the hydrophobic effect in stabilizing the fold, and polar interactions in providing strain or flexibility for

function (Shoichet *et al.* 1995). The case, however is not as simple as this, and results are often contradictory. It seems that how stability is achieved depends much on the individual protein; although general driving forces can be assigned, specific interactions can have important roles in any particular case.

The folding of a protein from unfolded (U) to native form (N) is a thermodynamic equilibrium process:

$$U \Leftrightarrow N$$

and has an equilibrium constant K_{fold},

$$K_{fold} = [N] / [U]$$

and thus is a process, for which the change in free energy, ΔG, can be given by

$$\Delta G_{fold} = \Delta H_{fold} - T\,\Delta S_{fold} = -RT\,\ln K_{fold}$$

(where ΔH_{fold} is the enthalpy change on folding and ΔS_{fold} the entropy change, R is the gas constant and T temperature). ΔG_{fold} is a measure of the stability of the protein (when $\Delta G_{fold} < 0$, the process will go towards N and when >0, towards U).

The hydrophobic effect, *i.e.* the fact that protein non-polar groups interact preferably with each other, and not with the polar solvent, is taken to be the main factor promoting protein folding (Kauzmann 1959, Tanford 1970, Dill 1990). The hydrophobic effect is really entropic in nature. At 25 °C, the ΔG of desolvation of non-polar surfaces consists of a large favorable entropy term and an unfavorable enthalpy term; this is because of a large entropy gain for the solvent on folding due to the release of water (Khecinashvili *et al.* 1995). However, at low temperatures water molecules tend to form clathrate-like (caged) structures around non-polar molecules because of the ordering of water itself, thus solubilizing the non-polar groups, mainly through the favorable enthalpy contribution from hydrogen bonding between the water molecules around the non-polar solute. Eventually with decrease in temperature this leads to cold-denaturation, unless freezing occurs first (Khechinashvili *et al.* 1995, Creighton 1993). Khechinashvili *et al.* (1995) concluded that the folding of proteins is mainly due to favorable entropy contributions from release of water from buried non-polar surfaces and the enthalpy of polar interactions. These compensate the unfavorable conformational entropy loss. Conversely, the desolvation of charges becomes more favorable as the

temperature increases (Elcock 1998; see below).

On the other hand, Creighton (1993) and Pace (2001) have suggested that hydrogen bonding may contribute more to ΔG_{fold} than does the hydrophobic effect. While it is not completely agreed upon which is most important, it is clear that hydrogen bonding and the hydrophobic effect are the components that stabilize proteins, and that most proteins are only marginally stable (10-20 kcal/mol), resulting from a small difference between large favorable and unfavorable energy terms.

2.4.2. Effects of polar interactions on stability

The protein interior was believed initially to contain mainly hydrophobic and aromatic amino acid residues, with charged and polar residues mostly on the surface (*e.g.* Tanford 1970). Although the protein core does consist largely of hydrophobic residues, numerous polar groups and charged residues are now known to be buried within proteins as they fold. Firstly, the peptide chain itself is polar and contains amide and carbonyl groups with hydrogen bond donors and acceptors, which need to be satisfied to stabilize them upon folding. This is achieved principally by the formation of secondary structure, the α-helices and β-sheets, which form hydrogen bonds between the carbonyl and amide groups of the peptide main chain. An isolated hydrogen bond is rather weak but as they form cooperative networks, the interactions becomes stronger than the expected sum of isolated hydrogen bonds. This results in the cooperativity observed in protein folding (Creighton 1993). Secondly, a number of polar and charged residues are partially or fully buried upon folding. This is almost unavoidable as the volume increases much faster than the surface area of a spherical object (Chothia 1976). The protein interior is clearly heterogeneous in terms of polarity (Warshel 1981, Quiocho *et al.* 1987, Hwang and Warshel 1988), and so ordered hydrogen-bonded interactions have an important role in defining protein structure.

It is still unclear, however, whether buried polar residues contribute to stability. Salt-bridges have been suggested to be a stabilizing factor in folding, but their burial has been reasoned to be unfavorable, based on the free energies of transfer of small compounds to organic solvents and on statistical information from experimentally determined structures (Lee and Richards 1971, Chothia 1976, Rashin and Honig 1984, Miller *et al.* 1987). However, mixed results exist on the role of buried charges (Waldburger *et al.* 1995, Hendsch and Tidor 1994, Stites *et al.* 1991, Dao-pin *et al.* 1991a, Giletto and Pace 1999). Some have found surface ion pairs or charges to be unimportant (Serrano *et al.* 1990, Dao-pin *et al.* 1991b, Horovitz *et al.* 1990), but others find them stabilizing, and even to determine the stability of some small proteins and domains (Perl *et al.* 2000, Spector *et al.* 2000). For the designed four-helix protein α_2D, a surface charge, Glu 7, was found to be

essential for the native structure, not by directly stabilizing it, but by disfavoring alternative non-native conformations (Hill and DeGrado 2000). Surface charge reversal has been suggested recently as a general strategy for increasing stability (Grimsley *et al.* 1999). For buried charges this would not work, because the whole environment needs to be preorganized to incorporate the charge (Hwang and Warshel 1988). In general, the role of surface charges is not any clearer than that of buried charges. It is clear, however, that in thermophilic proteins two major stabilizing factors are the surface salt-bridges and hydrogen bond networks, which seem more abundant than in the same proteins of mesophilic organisms (Cambillau and Claverie 2000).

The traditional view that proteins must have a hydrophobic core is incorrect. At least one naturally occurring β-sheet, that of OspA, does not contain a hydrophobic core (Pham *et al.* 1998). In addition, Cohen *et al.* (2002) recently showed that the interior of B1 domain of staphylococcal protein G is more polar than expected. This was done with a synthetic amino acid probe, "Aladan", an alanine-derivative of the fluorophore 6-dimethylamino-2-acylnapthalene which is a fluorescent amino acid sensitive to electrostatic changes.

Myers and Pace (1996) have estimated that hydrogen bonds stabilize proteins by 2.2 kcal/mol per hydrogen bond after corrections for side chain entropy and hydrophobicity. Hebert *et al.* (2001) found hydrogen bonds stabilizing by 1.5-2.7 kcal/mol in ribonucleases, when a conserved Asn residue was replaced by Ser or Ala. Similar results to these have also been obtained for tyrosines (Pace *et al.* 2001). In general, mutagenesis studies indicate a stabilizing role for hydrogen bonding groups; *e.g.* Byrne *et al.* (1995) analyzed 56 staphylococcal nuclease mutants (Val to Ser or Thr; Thr to Val, Ser or Cys and Tyr to Phe or Leu) and found that wild type hydrogen bonds stabilized the protein by 1.5 to 4 kcal/mol. Using a non-natural esterized backbone structure gave a result of 0.7-0.9 kcal/mol destabilization for each peptide hydrogen bond (Koh *et al.* 1997).

Nonetheless, it is not clear how the effect of hydrogen bonding on stability should be calculated: Karplus (1997) argues that desolvation energy estimates for polar groups depend very much on what the organic solvent chosen as a reference state is. This is important as estimates for the stabilizing effects of hydrogen bonds are constantly interpreted as the cost of transfer of a hydrogen bond from water to an organic solvent like octanol, interpreted to be analogous to the protein interior (*e.g.* Pace 2001). Hydrogen bonds are actually localized, specific interactions in the folded state, defined by the exact conformational state and the microenvironment, and thus cannot be stated in units of kcal/Å^2, because there is no simple relation to change in solvent exposed surface area upon folding, and therefore it is unclear what the "organic" reference state should be.

In addition, estimates of transfer energies for polar residues in proteins are based on small model compounds and assume structure-additivity; this is probably wrong, and the hydrophilicity of polar side chains is actually decreased by the flanking peptide groups (Roseman 1988). This is supported by the findings of Wimley *et al.* (1996a) using a pentapeptide host-guest system to evaluate the amino acid transfer energies. They found a favorable ΔG for transfer of a polar residue in a pentapeptide to octanol. In any case, it seems reasonable to assume that hydrogen bonded networks, in particular main chain hydrogen bonds, are stabilizing, especially considering the highly variable internal polarity of proteins (Warshel *et al.* 1984, Quiocho *et al.* 1987, Cohen *et al.* 2002), and the cooperativity of the system. However, charged residues remain substantially different.

2.4.3. Ionic interactions in thermophilic proteins

Many studies have compared the structures of homologous proteins from mesophiles and thermophiles or hyperthermophiles to elucidate the structural basis of thermostability. Results seem to indicate a number of means to increase thermostability, but an increase in salt bridges and hydrogen bonded networks on the surface of proteins appear to be perhaps the most general adaptations. Other features often found are shorter loops, more efficient packing and burial of non-polar surface (e.g. Argos *et al.* 1979, Goldman 1995, Karshikoff and Ladenstein 2001). The importance of ionic interactions in protein thermostability is not a new concept: it was first suggested by Perutz and Raidt (1975) almost three decades ago.

Calculations of the electrostatic contributions to stability for mesophilic *versus* thermophilic proteins (Xiao and Honig 1999) show that for all cases of the four protein families studied, electrostatic contributions to stability were more favorable for the thermophilic proteins. However, they did not depend on the number of ion pairs. In the CheY and Ferredoxin families there are fewer ion pairs in the thermophiles. The key difference was found to be the desolvation penalty for the ion pair (Xiao and Honig 1999). The authors also pointed out, that optimization of interactions is more important than the number of interactions.

In the triose phosphate isomerase (TIM) family, when 10 TIM structures from psychrophilic to thermophilic organisms were compared, increasing thermostability was associated with increased burial of non-polar accessible surface area (ASA), further connected with an increase in the number of salt bridges. No increase was observed in the number of hydrogen bonded residues (Maes *et al.* 1999). The increase in both buried hydrophobic surface area and in the number of salt-bridges is achieved at least in part through a switch to higher oligomerization states (from dimer to tetramer) (Maes *et al.* 1999). However, in *Pyrococcus furiosus* glutamate dehydrogenase (GluDH) ionic

networks were identified as the key difference between mesophile and thermophile proteins (Yip *et al.* 1995). Indeed, the GluDH from *Pyrococcus* has extensive ionic networks on the surface and at domain and subunit interfaces, the largest involving 20 residues at a subunit interface. The *Pyrococcus* enzyme has a half life of 12 h at 100 °C, while the mesophile enzyme from *Clostridium* has a half life of 20 min at 52 °C.

There are many points that suggest that ionic interactions in hyperthermophilic proteins are usually stabilizing. Firstly, there are in general more ionic residues and ion pairs on the surface of thermophilic proteins; the largest such increase has been observed in the lumazine synthase capsid, where the hyperthermophilic protein has >90% increase in salt bridge content compared to the mesophilic protein, mainly through replacing exposed polar uncharged residues with charged residues (Zhang *et al.* 2001). However, there are cases that contradict this (see above). Thus, increasing the number of ionic interactions is not sufficient for increasing thermostability. Secondly, salt-bridges can stabilize proteins through formation of extensive networks, as in the case of *Pyrococcus* GluDH; this way individual ion pairs can be stabilized. These kinds of cooperativity effects have been experimentally observed for the stabilization of ionic interactions in proteins such as barnase (Horovitz *et al.* 1990). The cooperative effect has also been studied by protein engineering in *Thermococcus litolaris* GluDH to mimic the more extensive ion-networks in *Pyrococcus* GluDH by a double-mutant (Vetriani *et al.* 1998), which regenerated a six-residue ionic network. This engineering resulted in an increase in half-life at 104 °C by four-fold, while both single mutants were destabilizing.

Also, interestingly calculations with a continuum electrostatics model on the magnitude of ionic interactions by Elcock (1998) suggests that the unfavorable effect of desolvation for the formation of salt bridges at lower temperatures (Hendsch and Tidor 1994; see below), does not apply at higher temperatures. Thus, in the higher temperature environment the salt-bridge networks will be more stabilizing than at lower temperatures, and there is an increasing energy barrier for breaking a salt bridge at higher temperatures. This, together with some experimental findings, suggests that salt-bridges, similar to covalent disulphide bridges, may provide kinetic stabilization towards unfolding at higher temperatures (Vetriani *et al.* 1998, Cavagnero *et al.* 1998).

In addition, as proper solvation or hydrogen bonding of charges in proteins (Warshel 1981, Quiocho *et al.* 1987) is important for stabilizing them, Spassov *et al.* (1995) have parameterized the optimization of charge interactions in proteins by comparing native to random arrangements of charges. They concluded that in thermophiles charge-interactions are more optimized than in mesophiles; this was also clearly linked to increased burial of hydrophobic surface.

2.4.4. Polar groups and specificity

Polar groups in proteins are also important for formation of correct native structure. Van der Waals interactions between hydrophobic residues are weak and therefore may not alone be sufficient to define the correct structure. The importance of polar groups for conformational specificity has been shown with *de novo* designed proteins. In the first experiments α-helical peptides forming four-helix bundles were created, such as the α4-family of helical proteins (Ho and DeGrado 1987, Betz *et al.* 1993). α-helical peptides of sufficient length (12-16) that self-assembled into tetramers were designed, and loops were inserted to connect the helices into two-helix and four-helix polypeptides. These were by design completely symmetrical molecules. The hydrophobic interior consisted of leucines at appropriate intervals, and Lys and Arg were placed at the surface positions to make favorable polar interactions (Regan & DeGrado 1988, DeGrado *et al.* 1989).

However, these and many other attempted designs (Betz *et al.* 1995), did not fold properly and often exhibited molten globule-like properties. The main reason was the high symmetry and the lack of constraints in the hydrophobic interactions between the Leu side chains, which can have several alternative equally favorable conformations (Betz and DeGrado 1993). The α4-proteins were refined to contain more specific internal interactions by introducing complementary side chain packing with inserted aromatic and β-branched amino acids and also metal binding sites (Betz *et al.* 1993). Nevertheless, these proteins still exhibited some non-native characteristics (Betz *et al.* 1993).

The reason became clear with the crystal structure of GCN4-activator (O'Shea *et al.* 1991). This contained a dimeric coiled-coil with a single buried asparagine that makes an inter-helical hydrogen bond to an equivalent residue. This was thought to be important for preference of a single native dimeric structure at the expense of other *e.g.* trimeric forms, which were observed in another design, "coil-Ser", together with monomeric and dimeric forms (Betz *et al.* 1995). Subsequently, native-like structures and specificity were achieved in designed proteins, but with the loss of stability, by introducing buried polar residues that disfavored alternative structures.

Mutagenesis studies have also shown that polar interactions are very important for the correct native structure, typically by disfavoring alternative near-native conformations. For example, Gonzales *et al.* (1996) showed that the conserved buried Asn16 of GCN4 was required to stabilize the dimer instead of the trimer. The only other residue at this position that favored dimers exclusively was Lys, as even the Asn16Gln variant produced trimers. The norleucine (Asn16Nle) mutant (the hydrophobic isostere to lysine) was much more stable (T_m > 100 ºC) than Asn16Lys or any other

variant at this position but formed both dimers and trimers. Lys16 as opposed to Nle16 confers specificity of oligomerization by destabilizing all oligomers, but destabilizing dimers least.

Similarly, the buried polar residues in *E. coli* thioredoxin are required to form the correct structure: The five-fold mutant (IleAlaAlaLeuVal), where buried residues Asp26, Cys32, Cys35, Thr66 and Thr77 were mutated, had a number of closely related low-energy conformations, and thus had no specific native state (Bolon and Mayo, 2001). Cordes *et al.* (2000) showed that the Asn11Leu variant of arc repressor has two alternative conformations around residue 11, β-sheet and α-helical. The exposed Asn11 residue acts as a "negative selection element" to favor the β-sheet conformation; in the α-helical conformation, position 11 is buried. This indicates the complexity and subtlety of the problem; to stabilize the native fold, other competing folds have to be destabilized specifically.

2.4.5. Buried charges and stability

Whether atoms are on the surface or buried inside a protein is usually described by the accessible surface area, as defined by Lee and Richards (1971): a sphere representing the solvent (water) molecule is rolled over the atomic surface of protein and a surface is drawn through its center, which is called the solvent accessible surface. An atom in the protein will be solvent accessible only if the sphere can reach it.

Chothia (1974) related the change in accessible surface area (ΔASA) upon folding to protein stability (to the ΔG of transfer of hydrophobic parts of the surface to water) and moreover he showed that accessible surface area is directly related to molecular weight (Chothia 1975). The ΔH and ΔG of transfer (gas to water, organic solvent to water, or organic crystal to water) for small aliphatic molecules including amino acid side chains depend linearly on the ASA (Chothia 1974) thus:

$$ASA = n\,\Delta H_{transfer}$$
$$ASA = m\,\Delta G_{transfer}$$

where *n* and *m* are proportionality coefficients with units in energy per $Å^2$ (like surface tension; for a review, see Janin 1997). A similar relationship can be assumed for polar group hydration or transfer. Further, the desolvation component for protein folding can be assumed to follow the same relationship (Chothia 1974). For protein folding ΔG_{fold} can be given as:

$$\Delta G_{fold} = \Delta G_{desolvation} + \Delta G_{interaction}$$

where

$$\Delta G_{desolvation} = \Delta G_{transfer} = m\,\Delta ASA.$$

A problem in coupling protein folding thermodynamics with structure comes from estimation of ΔASA on folding. Usually the unfolded state is approximated by either tripeptides of types Gly-X-Gly or Ala-X-Ala (where X is the residue being measured) or with an extended polypeptide. However, the nature of the unfolded state is likely to be more compact in most cases, and thus the ΔASA-values will be overestimates. This would probably have the largest effect on the buried residues in the folded state (most of which are the aliphatic and aromatic residues). An additional weakness of the approach is its static nature, which might be improved by taking into consideration residue depth, as proposed by Chakravarty and Varadarajan (1999). Nevertheless the approach can give reasonable estimates for the ΔG of solvation and is a rational basis for relating structure to the energetics of protein folding.

Others have specifically studied the distribution of charge within proteins. Barlow and Thornton (1983) found that about one third of charges are involved in ion pairs and that 17 % of ion pairs are buried, while Rashin and Honig (1984) found that there were on average two completely buried charged groups in the proteins. These early studies involved only rather small proteins. Miller *et al.* (1987) used a larger dataset, but the molecular weight range in their study was still only 4 to 35 kDa. They found that 39 % of the buried surface area is polar and 4% charged and concluded that the proportions of various surface types stay about the same within the protein core regardless of size. They also concluded that most of the buried non-polar surface contribution comes from the peptide groups. However, the statistical analysis so far has mostly concentrated on observations of mass distribution of the various atoms and groups, while it should be noted that proteins are only marginally stable and therefore even a single unfavorable charge can have deleterious effects on structure.

Hendsch and Tidor (1994) performed an important computational study on the stability of salt-bridges in various protein structures. They compared the stability of selected salt-bridges with the stability of isosteric hydrophobic residue pairs in a data set of 21 salt bridges in 9 proteins. They reported the difference in free energy of folding (ΔΔG) between an ion pair and an isosteric hydrophobic pair as a sum of three components: ΔΔG of solvation, ΔΔG of the pairwise interaction

and $\Delta\Delta G$ of interaction with the rest of the protein environment ($\Delta\Delta G_{protein}$). They thus estimated the ΔG for desolvation from the aqueous surroundings and introduction of the salt-bridge into the hydrophobic environment relative to an equivalently packed hydrophobic pair. Their reference state differed markedly from previous studies which typically included measurement of pK_a-shifts relative to the unfolded state, or conservative mutation to homologous polar uncharged residues. These kinds of studies always measure the strength of a salt-bridge relative to a charged or neutral hydrogen bond.

Hendsch and Tidor (1994) concluded that most salt-bridges were destabilizing. Their predictions have been tested experimentally and are reasonably accurate, at least in the direction of the effect on free energy (Hendsch *et al.* 1996, Spector *et al.* 2000), although the calculations tend to overestimate the strength of electrostatic effects in protein interiors (Hendsch and Tidor 1994, Hendsch *et al.* 1996, Schutz and Warshel 2001).

Buried ion pairs are calculated to be less stable than buried isosteric hydrophobic groups because the calculations assume the proteins solvate charges less well than water does. Some studies, however, indicate that protein can solvate charges better than water (Warshel *et al.* 1989, Tissot *et al.* 1996, Giletto and Pace 2000). In such cases, the protein must be pre-organized to accept the polar or charged groups; *i.e.* solvation happens through hydrogen bonding to other polar groups within the protein that are restrained both in position and orientation (Warshel 1981, Warshel *et al.* 1984, Hwang and Warshel 1988). One problem with the theoretical calculations is that the models are partly “macroscopic”, assuming bulk models in particular for the solvent and also in part for the protein system, while the true system is of course fully “microscopic”. The continuum methods (Honig and Nicholls 1995, Warshel and Papazyan 1998), used also by Hendsch and Tidor (1994), apparently model solvent relatively well (Schutz and Warshel 2001). However, the calculations of Hendsch and Tidor (1994) did not take into account the reordering or relaxation of protein permanent dipoles on addition of the charge to the protein environment, although the calculations did include the protein groups, as embedded in low dielectric medium, in their term $\Delta G_{protein}$. However, this kind of approximation, with the description of the protein as a low polarity medium, with a dielectric constant (ε), typically of $\varepsilon = 4$ (whereas for vacuum $\varepsilon=1$, octanol $\varepsilon=8$, and water $\varepsilon=80$) overestimate the energies of ion-pairs, as the reordering of the orientation of the protein dipoles around the charge can be a major contributor to stabilization, and increase local polarity (Schutz and Warshel 2001). The results of Hendsch and Tidor (1994) could be refined by including the relaxation of the protein dipoles into the calculations.

The transfer energies for a salt-bridge from water to non-polar solvents have been studied experimentally by Wimley *et al.* (1996b), and they came to similar conclusions with Hendsch and Tidor (1994). They measured the transfer energies of the pentapeptide acetyl-TrpLeu-X-LeuLeu from water to octanol (where X denotes either Lys or Arg). A salt bridge was formed between the C-terminal carboxyl group of the peptide and the residue X in octanol. The ΔG of salt-bridge formation in octanol is about -4 kcal/mol, while the cost of transfer to octanol of a pair of individual charges in the context of the peptide is about equal, so that the formation of the salt bridge approximately cancels the penalty of desolvation from water. However, relative to the transfer of a pair of leucines the transfer of any salt bridge forming charged amino acid pair is unfavorable by *ca.* 8 kcal/mol (Wimley *et al.* 1996a, 1996b). Consequently, it seems the hydrophobic interaction is more stabilizing in octanol and probably also in the protein interior than the charge interaction.

Protein engineering studies on the effects of burying charged residues have given contradictory results, undoubtedly partly as a result of the specificity and local nature of the interactions. Waldburger *et al.* (1995) mutated a buried salt-bridge triplet in arc repressor by means of creating a hydrophobic mutant library from which they selected *in vivo* for stable active mutants. They found various hydrophobic combinations, in particular with a ”central” Tyr (X-Tyr-Z) that were fully functional replacements for the charge cluster of Arg 31-Glu 36-Arg 40, respectively. The mutant Met-Tyr-Leu (MYL) was crystallographically analyzed, and found to be identical to the wild type arc-repressor structurally. This mutant and the three others (Ile-Tyr-Val, Val-Tyr-Leu and Val-Tyr-Val) had native structure, were considerably more stable, and bound to DNA, although the cooperativity of binding was reduced from 890 to 0.3-7.8 ω (expressed as the ratio of full-operator binding affinity to half-operator binding affinity of the protein). The study by Waldburger *et al.* (1995) showed convincingly that it is possible to remove ionic networks from the protein interior without disrupting structure, stability or even function. Also, in *Leishmana mexicana* TIM there is a glutamate (Glu65) buried in the dimer interface which is Gln in most other TIMs. Mutating this Glu to Gln results in a vast increase in T_m by 26° C at pH 7 (Williams *et al.* 1999).

In contrast, in RNase T1, the buried and non-ion-paired, but hydrogen-bonded Asp contributes favorably to stability (Giletto and Pace, 1999). Here, Asp 76 was replaced with Asn, Ser, or Ala, and the stability of the protein was observed to decrease by 3.1, 3.2, or 3.7 kcal/mol, respectively. The environment of this 99% buried Asn carboxylate is quite polar as it makes three hydrogen bonds to Tyr11, Asn9 and Thr91 and a fourth bond to a conserved structural water molecule. The stabilizing effect of the buried charge could be explained by its efficient solvation by the protein; the pK_a of Asp76 in the folded state is 0.5. In addition, Asp76 locks the RNase N-terminal β-hairpin (residues 1-12) to the body of the protein through hydrogen bonding to Asn9 and Tyr11. Thus it

seems the hydrogen bonds from Asp76 are important perhaps because they link the N-terminus to the protein and it does not seem possible to accommodate other residues that can compensate for its hydrogen bonding, while deleting it destabilizes the structure (Giletto and Pace 1999). Estimation of the energies associated with the dipole rearrangements by computer calculations, as suggested by Schutz and Warshel (2001), would probably be very useful. Asp76 also contributes to the structure of the unfolded state of RNase T1 and the mutants have less structured unfolded states (Giletto and Pace 1999).

Earlier studies have also found buried charged residues to both increase and decrease stability: in staphylococcal nuclease a single Val to Lys mutation is very destabilizing at neutral pH, while at high pH, where the Lys is neutral, it was only slightly less stable than wild type (Stites *et al.* 1991). The destabilizing effect comes mostly from the desolvation of the charge. The X-ray structure of this Val66Lys-mutant shows that the lysine is buried without hydrogen bonding or charge-interactions, and the pK_a in the folded state was approximated to be less than 6.4. Therefore, the lysine must be neutral when buried. The desolvation energy for the charged Lys66 was estimated to be 5.1 kcal/mol or more. Dao-pin *et al.* (1991a) introduced a buried Lys in place of Met102 and a buried Asp in place of Leu133 in T4 lysozyme, and found both to be significantly destabilizing over the whole pH-range, although the mutants did fold at neutral pH. The pK_a of the buried Lys was about 6.5 (comparable to 6.4 found by Stites *et al.* (1991) in staphylococcal nuclease). However, these results are hardly surprising, as the incorporation of a buried charge would require a preorganised environment.

The role of two buried salt bridges in barnase (Arg69-Asp93 and Arg83-Asp75) on stability and folding were investigated by Tissot *et al.* (1996) by mutating each Asp to Asn, while various replacements were made for the arginines. All single variants were clearly destabilizing compared to the wild type (within the range of 1.3-5.4 kcal/mol), and coupling energies for the salt bridges from double mutant cycles were found to be 3.2-3.3 kcal/mol. These salt bridges thus were clearly stabilizing compared to the mutants. The pK_a for Asp93 in the folded state was 0.7; it is thus charged while buried, and therefore is clearly very well solvated by the protein! The results above would seem to imply that to know whether buried charges are destabilizing or not we also have to know what to compare the buried charge to.

A consensus still needs to be reached on which kinds of experiments should be carried out. Buried salt-bridges may often be less stable than their hydrophobic counter-parts, but it is not clear what the actual polarity of the protein interior is, nor is it clear to what extent hydrogen bonds contribute to protein stability. In addition, the buried polar groups will almost always be networked to other

polar atoms which will be left partly unsatisfied if a certain group is removed. This will complicate calculations and measurements. It is also true that conservative replacements, such as Asp to Asn, do not give an estimate of what the contribution of the polar groups might be relative to hydrophobic isosteres (or other more stabilizing alternatives).

As another alternative, the use of double-mutant cycles (mutating each ion-pair partner to Ala individually and together; such as Glu-Arg to Ala-Arg, Glu-Arg and Ala-Ala; Carter *et al.* 1984, Horovitz and Fersht 1990) should give the interaction energy between chosen residues as the difference in $\Delta\Delta G_{folding}$ of (1) wild-type relative to Ala-Charge and (2) the double mutant relative to one of the point mutants (Figure 14) (Charge-Ala to Ala-Ala). Thus, we get:

$$\Delta G \text{ of interaction} = (1) - (2)$$

However, this will include all changes introduced (including packing rearrangements), but would be equal to the interaction energy, if the other structural changes in the double mutant are an exact sum of those in each of the single mutants. Observations fairly close to this have been reported in one case (Vaughan *et al.* 2002). Ideally to get a full description of the contributions of various effects by one ion pair, the stabilities for a neutral hydrogen bond, a charged hydrogen bond, an ion pair and a hydrophobic alternative should be measured and compared, the replacements should be analyzed in terms of structural rearrangements, and relaxation of the electrostatic interactions should also be estimated from computer calculations.

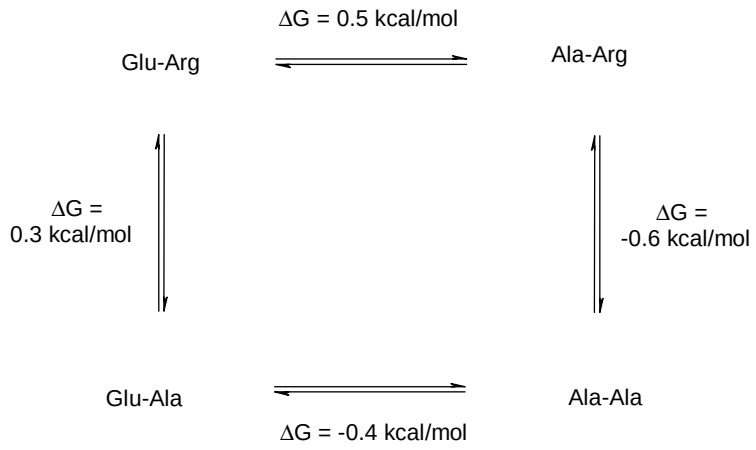

Figure 14. A Glu-Arg double mutant cycle. The energy of the Glu-Arg interaction here is 0.9 kcal/mol.

2.6. Packing of the protein core, stability and function

In this thesis work interesting differences were found in core packing in bacterial MLEs. Previous studies indicate that the protein interior is more like crystals of amino acids than the liquid organic phase in terms of its density and compressibility against pressure (Creighton 1993). However, the packing of proteins is not perfect and cavities of variable sizes are observed frequently. Based on the cavity, or free volume, distribution, protein molecules are more like liquids or glasses (amorphous materials) and appear, according to Liang and Dill (2001), more like "randomly packed spheres near their percolation threshold rather than like jigsaw-puzzle". Thus, although the average packing of proteins and compressibility appear to be similar to organic crystals, the actual packing is not as ordered or compact as in a crystalline solid (Liang and Dill, 2001). This is not surprising considering the heterogeneity of the protein matrix, and the more complex structure of the folded polypeptide than that of a small molecule crystal.

The distribution of cavities in proteins has been analyzed statistically (Rashin *et al.* 1986, Hubbard *et al.* 1994, Williams *et al.* 1994, Hubbard and Argos 1994) and in relation to protein folding and dynamics (Kocher *et al.* 1997, Rashin *et al.* 1997). Cavities have been suggested to be related to function, in particular at domain interfaces (Hubbard and Argos 1996). The analyses have been based on the accessibility concepts of Lee and Richards (1971) and have often used the Voronoi-method to divide space occupied by non-uniform distribution of positions (of *e.g.* spheres or atoms) into geometric cells. Another method is that of Connolly (1983), which is based on the calculation of the so-called molecular surface, further based on the principles of Lee and Richards (1971). Several algorithms have been developed for detection of cavities; *e.g.* "flood-filling" (Ho and Marshall 1990), "fattening" (Kleywegt and Jones 1994), and filling void regions with spheres (Smart *et al.* 1993).

Cavities were observed as new methods were developed to calculate the surface area and volume (Lee and Richards 1971). Later studies revealed the generality of their existence (Rashin *et al.* 1986). Hubbard *et al.* (1994) and Williams *et al.* (1994) extended the analysis on intramolecular cavities with larger data sets describing also the hydration, shape considerations and composition of flanking residues. Most of the small cavities in proteins (82 %) are non-hydrated (Williams *et al.* 1994). As would be expected, the number and total volume of cavities increases slowly with the size of the protein. The number of cavities does not seem to be associated with the type of structure or resolution; multi-domain proteins do not have more cavities than single domain proteins. The cavity volume associated with proteins of same size can, however, vary substantially (Hubbard *et al.* 1994, Williams *et al.* 1994).

Hubbard and Argos (1994) also examined cavities at protein interfaces. Cavities within domains, at domain-domain interfaces and at protein subunit interfaces differ with respect to hydration and mean size. Hydrated cavities possess more polar surface and are typically larger than non-hydrated cavities. Cavities at domain and subunit interfaces are larger than cavities within domains and are more often water-filled. On average, solvated intra-domain, inter-domain and inter-subunit cavities were 39.8, 43.8 and 86.5 $Å^3$ in volume, respectively. Few non-solvated intra-domain cavities were more than 50 $Å^3$ in size. Cavities occur more often at domain interfaces associated with shear-type motions than at interfaces associated with hinge-type motions. Hubbard and Argos (1996) examined the relationship of inter-domain cavities and domain movements in detail and found cavities associated with functional domain movements both in hinge and shear-type motions.

The effects of packing on protein stability and structure have also been studied by site-directed mutagenesis. In general, correct hydrophobic core packing is not necessary for the formation of a native-like fold, but several alternative arrangements of hydrophobic residues can be tolerated. More dramatic mutations result in increasingly non-native properties, but overall secondary and tertiary structure is retained even then. This has been shown by randomization of hydrophobic core positions in arc repressor (Lim and Sauer 1989, 1991; Lim *et al.* 1992). Similar results have been obtained on barnase: all 13 hydrophobic core residues of barnase can be randomly replaced by other hydrophobic residues and a significant fraction of the variants will retain enzymatic function (~20 %) (Axe *et al.* 1996). Crude activity can thus be achieved even with non-optimal core packing, and this clearly supports the idea that packing is not the most important determinant of protein folding. Results from the comparison of thermophilic proteins and mesophilic proteins suggest that they have similar packing densities (Karshikoff and Ladenstein 1998).

The cost of creating cavities has been reported to range between 0.5 and 5.0 kcal/mol. Experimental measurements on barnase, T4 lysozyme and lambda-repressor indicate that creation of a cavity the size of a single CH_2-group costs around 1.0 kcal/mol. The cost of a Leu to Ala mutation is ca. 2.0 kcal/mol plus a term resulting from the size of the cavity created (33 cal/mol/ $Å^3$) (Kellis *et al.* 1988, Eriksson *et al.* 1992, Lim *et al.* 1992). The destabilizing effect results from two components, firstly the loss of buried hydrophobic surface, and secondly from the altered packing in the protein core and its energetic consequences.

3. Aims of the study

The aims of the work described were to:

1) Study further the role of buried charges in MLE and their number and distribution in proteins in general.

2) Solve the structure of *Neurospora crassa* CMLE, the first structure of an eucaryotic MLE.

3) Understand the catalytic mechanism of and dehalogenation by MLEs from a structural point of view.

To achieve the last goal, several attempts were initially made to solve complex structures of MLEs with substrates and inhibitors, both of PpMLE and ReCl-MLE and P51 Cl-MLE. In all cases only crystal forms without the substrate were obtained or the crystals always cracked upon addition of substrate, even in the presence of the inhibitory metal cofactor substituent Sm^{+3}, which was known to bind to the MLE active site (Goldman 1985). However, the results on the Cl-MLE high resolution structure will help in revealing the dehalogenation mechanism. The second important goal (above) of solving the CMLE structure provided a more integrated view of the MLE catalytic activity, and gave insight into what is indispensable for MLE catalysis.

4. Materials and Methods

4.1. Protein expression and purification

PpMLE, its variants, and the P51 Cl-MLE were expressed in *E. coli* BL21 (DE3) pLysS (Novagen) at 30 °C as published (Vollmer *et al.* 1998) and purified on a Pharmacia FPLC Q-sepharose anion exchange column. This single step yielded protein of >80% purity which was suitable for MLE crystallization. For crystallization of P51 Cl-MLE an additional hydrophobic interaction chromatography step with a phenyl-sepharose column was included to obtain pure protein.

Recombinant *Neurospora crassa* CMLE was expressed under the P_L-promoter in *E. coli* N4830-1 (Mazur *et al.* 1994) and induced with a temperature increase from 30 to 42 °C at around OD600 = 0.8, as the P_L-repressor in *E. coli* N4830-1 is temperature sensitive. Protein purification was done as described by Mazur *et al.* (1994) with some modifications and simplifications (Paper IV). Selenomethionine (SeMet) protein was produced from new minimal medium (NMM; Hädener *et al.* 1993) using the *E. coli* BL21 (DE3) B834 methionine auxotrophic strain (Novagen). With the original plasmid, the promoter leaked, as there is not enough P_L-repressor produced in this *E. coli* strain, and so production was constitutive. Nevertheless, this approach worked well with addition of 50 mg/ml SeMet at OD600 = 0.8 and growth at 37 °C. Simultaneously, the metabolic inhibition method (Van Duyne *et al.* 1993) was used with the original *E. coli* N4830-1 strain, which gave somewhat larger quantities of protein and apparently worked as well. Any differences in SeMet incorporation were not quantified, but all crystals tested gave good Se-absorption at the ESRF synchrotron (A. Thompson, personal communication). Mass spectroscopy was utilized to verify full incorporation of SeMet into CMLE for one purified protein batch from minimal medium.

4.2. Site-directed mutagenesis

Mutagenesis was performed with the Stratagene QuickChange™ mutagenesis kit, or with equivalent materials according to manufacturers instructions, sometimes with some modification towards the conditions used elsewhere for MLE (Kaulmann *et al.* 2001). Essentially no effect was observed for the different PCR cycling temperature conditions tried. Sequencing was performed in one direction with the *Pseudomonas putida* MLE gene *catB* oligonucleotides (see Papers II and V). In the case of *Neurospora* CMLE mutants His148Ala and Glu212Ala mutagenesis was done as in the QuickChange-instructions (Paper IV). For materials, Pfu DNA polymerase and DpnI restriction enzyme were used as recommended, while oligonucleotide mixes were from various sources.

4.3. Crystallization and data collection

The MLE wild type and variants were crystallized as published (Goldman *et al.* 1985), sometimes

with modifications (Paper I and Paper II). Several screens were made to find alternative crystal forms using the screening conditions of Jancarik and Kim (1991) and other similar sparse matrix screens for protein crystallization. In addition to the conditions reported by Hasson *et al.* (1998), one new crystal form was found in I4 (Table 2) without twinning, and without metal ligands present, although 1-2 mM $MnCl_2$ was always added as well as one of the substrates or inhibitors of MLE (*cis,cis*-muconate, 2,4-diCl-muconate, 2-Cl-muconate, 2-F-muconate and 3-methyl-muconate). The new I4 crystal form diffracted to *ca.* 2.3 Å, and for the variant Lys169Ala it also showed residual density for the loop 19-30 for one monomer (of two) in the asymmetric unit, not visible in the original twinned I4 crystals.

Table 2. Data for the new non-twinned I4 MLE (wild type and variant Lys169Ala) crystal form ["S9 and "S41" stand for conditions 9 and 41 in the Jancarik and Kim (1991) screen].

Unit cell dimensions	a=126, b=126, c=92.0 (space group I4)
"S9" Conditions for I4 crystals	100 mM Citrate pH 5.6, 200 mM NH_4Acetate, 8% PEG 600, 1mM $MnCl_2$. For flash freezing/mounting 35 % ethylene glycol and 3 mM 3-methyl-muconate were added. 1 h soak with substrate. Crystals were grown at +21 °C
"S41"Conditions for I4 crystals	50 mM K_2PO_4 pH 7.5, 10 % PEG 6000, 1mM $MnCl_2$, 0,5 mM 2,4-diCl-muconate. For flash freezing/mounting buffer exchanged to 50 mM Tris-HCl pH 7.5, 50 mM Na_2SO_4, 35% PEG 6000, 2 mM $MnCl_2$, 0,2 mM 2,4-diCl-muconate (in a large vol.); buffer exchange for mounting was done with a gradient of six five minute steps to the mounting conditions and a further soak with substrate for 15 min. Crystals were grown at +4 °C

The *Pseudomonas* P51 Cl-MLE was crystallized as reported in Paper V. Several crystal forms appeared initially but none of them diffracted particularly well, usually to 3.0-3.5 Å at best with the in-house Rigaku/RAXIS II generator and detector system at the Turku Center for Biotechnology.

The bacterial *Pseudomonas* MLE and Cl-MLE wild type and variant data were mostly collected on the in-house R-AXIS II detector in the Turku Center for Biotechnology, except for the final data on Cl-MLE, which were collected at the Lund MaxLab II beam line 711, in Sweden, and the data on MLE variant Asp178Asn, which were collected on the EMBL Hamburg beam line X31, Germany. All data were collected from flash-frozen crystals. The MLE wild type and Asp178Asn crystals were dipped into crystallization solution supplemented with 30% MPD prior flash-freezing for data collection. Initial data on wild type PpMLE were also collected on the ESRF ID14 beam line, which indicated that the crystals diffracted to 1.5 Å resolution.

The *Neurospora* CMLE was crystallized as published (Glumoff *et al.* 1995) and the SeMet form also crystallized from the same conditions (1.5-1.9 M $(NH_4)_2SO_4$, 100 mM PIPES pH 5.7-6.0). The crystallization process was generally slow (taking from weeks to months), so streak seeding from crushed crystals was attempted to speed up the crystallization (nucleation). This worked well with serial dilutions (Stura and Wilson 1992), but the SeMet crystals also grew a bit faster than the native ones (Paper III). The crystals typically had high mosaicity (around 1.0°) and did not last in the X-ray beam unless they were flash frozen. PEG 400 (40%) was the best cryo-protectant as also found by Glumoff *et al.* (1995); in essence no other organic cryo-solvents gave processable diffraction data. For flash freezing it was further found that the PIPES buffer itself crystallized in 40% PEG 400, and thus it was exchanged to 100mM citrate pH 6.0 in the cryo-protectant solution, into which the crystals were always quickly dipped before cryo-cooling. The crystals did not tolerate prolonged incubation in 40% PEG 400 very well. Crystal annealing by repeated freezing-thawing cycles (of seconds to minutes) was tried, but this did not produce any increase in diffraction quality, contrary to observations on some other proteins (Harp *et al.* 1998). Later it was found that small crystals could also be flash-frozen directly from the crystallization solution (A. Thompson, personal communication).

The crystals also tolerated exchange from 1.5-1.7 M $(NH_4)_2SO_4$ to 1.6 M citrate pH 6.0 in stepwise exchange with 5 minute soaks at 5-10 intermediate solution conditions (*e.g.* with 0.2-0.3 M changes of $(NH_4)_2SO_4$ and citrate concentrations simultaneously). After this they sometimes still diffracted to high resolution like non-soaked crystals. This was attempted for purposes of soaking in the citrate, as it ought to be a good inhibitor of CMLE with K_i ~20 µM (P. Mazur, personal communication). However, no citrate was visible in (F_o-F_c) difference Fourier maps in the case when $(NH_4)_2SO_4$ was present (at 3.0 Å) and with buffer totally exchanged we were not able to collect a full data set due to failure of the cooling system with the best crystal that diffracted to at least 2.5 Å; similarly a power-failure at the Lund synchrotron eliminated a 2.5 Å data set on 100 mM citrate-buffer soaked crystal. Most often the crystals simply did not either tolerate the cooling procedure or the soaking. However, "cryo-salts" (Rubinson *et al.* 2000) were not tested as cryo-protectants, except for the obvious case of $(NH_4)_2SO_4$, which indeed gave encouraging results.

DENZO and SCALEPACK (Otwinowski and Minor 1997) were used for all data indexing, merging and scaling in this work. With CMLE, scaling always resulted in some overlaps at higher resolution and thus loss of data because of the large unit cell, so XDS (Kabsch 1993) was tried as its scaling algorithm allows the use of thinner slices of diffraction space per frame, but it failed to process the CMLE data when tried. For data collection from SeMet-CMLE crystals for MAD phasing, the

crystal was aligned so that each Friedel pair was always collected on the same image to avoid errors in the measurement of the anomalous differences, as generally has been recommended (Paper III). The three wavelengths were processed with the same crystal parameters, but separately.

4.4. Structure solution, refinement and analysis

4.4.1. CMLE phasing

The CMLE structure was solved as reported in Paper III. Shortly, the anomalous diffraction data in the form of structure factor amplitudes from the peak data set at 0.9786 Å were pre-processed with the DREAR suite (Blessing and Smith 1999), incorporated in the Shake-n-Bake vs. 2.0 (SnB) software used (Weeks and Miller 1999), to generate normalized structure factors (|E|s). As input to SnB 80 random atoms were given, because this was the number of Se-atoms expected inside the asymmetric unit: 10 SeMets per monomer in eight monomers. Altogether 5428 |E|-values (12.2 % of reflections) were generated from the peak wavelength data set and 2400 were chosen for phasing (80 x 3 phases; standard parameterization; Howell *et al.* 2000), these were further used by the program to generate 24000 triplet invariants. For structure factor amplitudes $|\mathbf{F}|_{\mathbf{h}}$ and $|\mathbf{F}|_{\mathbf{k}}$ with indices **h** and **k** the triplet invariant is given by

$$\Phi_{hk} = \phi_h + \phi_{-k} + \phi_{-h+k}$$

where ϕ is the phase of the reflection. Each structure factor is involved in a number of triplets and they are therefore linked. The probability distributions of the invariants $P(\Phi_{hk})$ are used in direct methods to calculate the most probable phases. In the SnB algorithm, this is done using a minimal function (R_{min}) to identify a convergence to a solution during repeated cycles of dual-space refinement in one trial (De Titta *et al.* 1994, Hauptmann 1997); each trial begins with a new set of random atoms. The dual space refinement proceeds with phase refinement in reciprocal space and peak picking (density modification) in real space. The input number of atoms is used as a real space constraint in peak picking (it is the number of peaks picked).

Two solutions were found for the CMLE data with a very low success rate (about 1:2000 trials). The solutions were identified on the basis of a bimodal distribution of trials in the histogram and a constrained drop in R_{min} (Figure 15). The solutions were cross-checked against each other and tested for convergence with a few sites and identical sites from the two solutions, altogether 60, were refined in MLPHARE (CCP4 1994). The use of direct methods was necessary because of the large number of sites: 80 Se-sites in the asymmetric unit would create 4 x 80 = 320 sites in the unit cell, which yields 320^2-320 = 102080 Patterson cross-peaks between the sites in the unit cell and so

12760 in one asymmetric unit. Further, in $P2_12_12_1$ each unique section of the Harker plane in the anomalous difference Patterson maps would be expected to have 80 peaks from the 80 sites. The large total number of peaks in the asymmetric will produce a very high origin peak, and the high total density of peaks further will add non-Harker peaks close to or on the Harker planes. Further, in the CMLE $P2_12_12_1$ crystal several NCS-twofolds lie close to the x-y plane (Paper III), generating more unwanted signals perpendicular to this Harker plane. Not surprisingly anomalous difference Pattersons did not give meaningful signal in the form of discrete peaks.

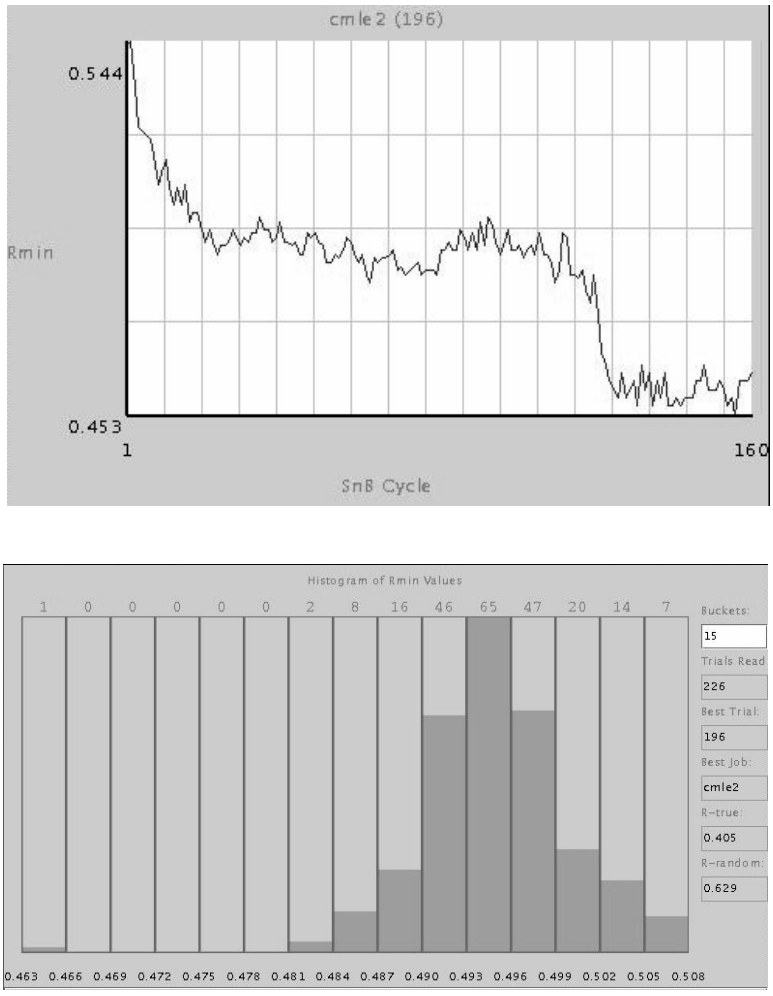

Figure 15. The 80 Se-site CMLE direct methods solution by SnB. A constrained drop in R_{min} (upper figure) and the bimodal distribution of trials, indicative of solution (lower figure).

The handedness of the solution was determined by the result of the solvent flattening applied during phase refinement. Subsequently, the sites that were stable during the phase refinement with MLPHARE (CCP4, 1994) were picked and this yielded a final set of 57 sites, which were used for generation of the initial maps and envelopes and to find the first 2 NCS-operators computationally (with program FindNCS; Lu 1999). These were utilized to generate initial dimer envelopes and to

identify the first NCS related pairs of molecules in the 4 Å solvent-flattened maps. The use of "dummy"-model molecules (here, TIM-barrels) and manual translation to fill the NCS and solvent envelopes, allowed us to find the rest of the NCS operators (Figure 16). The operators were refined with the IMP and MAMA programs of the RAVE-package (Kleywegt and Jones 1999). These operators together with the $P2_12_12_1$ space group symmetry were used to expand the 57 sites into a unit cell to create clusters of sites. The resulting 80 clusters were taken as the correct full set; 56 sites generated the full set of 80 and these 80 positions gave the full set of NCS-operators. These were refined with MLPHARE and the resulting phases were solvent flattened, averaged and phase extended to give a final 3 Å experimental MAD-map into which the model was built (Figure 16; Papers III and IV).

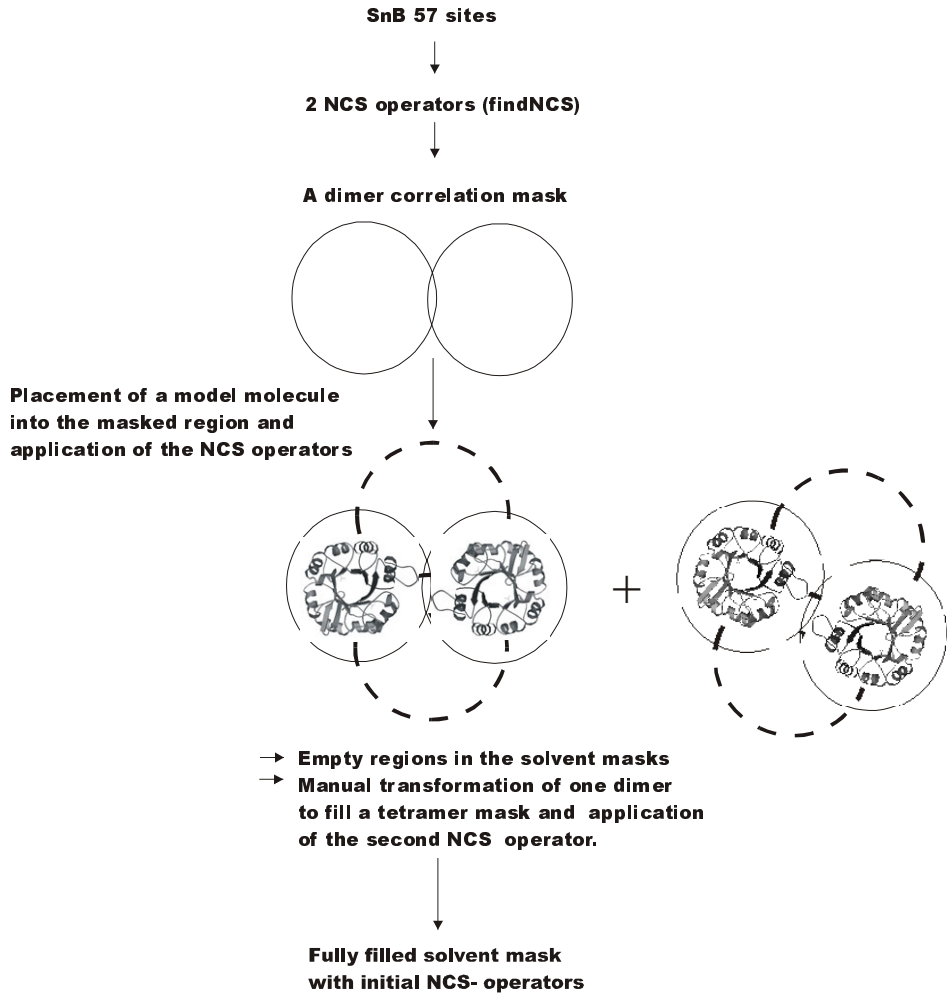

Figure 16. The solution of CMLE's NCS-operators. The spheres represent the solvent and NCS correlation masks. Dashes spheres represent the empty regions of the solvent mask.

4.4.2. Molecular replacement

The molecular replacements for the MLE crystal forms and variants were done with X-PLOR (Brünger *et al.* 1992) and later with CNS (Brünger *et al.* 1998). However, the peak rotation search was very slow for CMLE, and it did not produce any solution between the two CMLE crystal forms. Therefore AMoRe was used instead (Navaza 1994), which was very fast and gave the correct solutions within a few hours. However, an octamer model did not find a solution, but it was necessary to use a tetramer and to place them individually into the new unit cell. With CNS (and its predecessor X-PLOR), even tetramer models failed. After molecular replacement using a tetramer search model and rigid body refinement with AMoRe the R-factor was 37.2 %.

4.4.3. Refinement of MLE structures

Refinement of the CMLE structure was done with CNS. After rigid body refinement refinement I used torsion angle dynamics refinement with simulated annealing starting from 5000 K, and applied anisotropic B-factor correction with bulk solvent modification, as implemented in CNS (Brünger *et al.* 1998). As a result of the B-factor correction both conventional and free R-factors dropped by 3% to 26.9 % and 27.7 %. The 80 Se-sites located in the other crystal form were also rotated and translated to the new unit cell, and these phases were refined for the second 3 Å MAD-data set, for which an independent solution was not found, and phase extension was done to 2.5 Å resolution with the high resolution SeMet-data used in refinement (Papers III and IV). This yielded consistent maps, although the phasing power was significantly lower (1.1) than for the original set (2.0). This was used as an "experimental NCS-averaged reference" in model building and refinement and checks were made to look for consistencies in the various Fourier maps. The other structures in this work were refined either by X-PLOR or its successor CNS (Brünger *et al.* 1998), except for the P51 Cl-MLE 1.95 Å structure, where initial building and refinement after molecular replacement with X-PLOR were done with ARP/wARP vs. 5.1 (Perrakis *et al.* 1999). This successfully improved the R-factors, while refinement and manual rebuilding did not produce a drop in the R_{free} from 32% (see Paper V). ARP/wARP was also used for some rounds of addition of water molecules to the CMLE octamer structure.

4.5. Spectroscopy measurements

The change in fluorescence on unfolding was studied using a Specs Fluorolog Model F1T11 instrument with excitation at wavelength of 278 nm. The tryptophan emission maximum of 337 nm was followed as a function of temperature to construct the MLE mutant melting curves (Paper II). The CD spectra for CMLE and variants were measured on a Jasco-720 spectropolarimeter (paper IV). The activity of CMLE and MLE was measured as the disappearance of substrate absorbance in

the UV-wavelengths, at 270 nm for 3-carboxy-*cis,cis*-muconate and 260 nm for *cis,cis*-muconate, respectively. Extinction coefficients of 6 mM^{-1} cm^{-1} for 3-carboxy-*cis,cis*-muconate and 18 mM^{-1} cm^{-1} for *cis,cis*-muconate (Mazur *et al.* 1994) were used.

4.6. Sequence comparisons and figures

Sequence homology searches for the CMLE work were performed with PSI-BLAST as implemented at NCBI (http://www.ncbi.nlm.nih.gov/BLAST) (Altschul *et al.* 1997) and multiple sequence alignment was done with ClustalW and manually hand edited (Paper IV). Figure 5 in Paper IV was prepared with Alscript (Barton 1993). Structure figures were prepared with Molscript (Kraulis 1991), Raster3D (Merritt and Bacon 1997) and Bobscript (Esnouf 1997).

5. Results and Discussion

5.1. The structure of Neurospora crassa CMLE

5.1.1. Structure solution

The initial experimental Fourier-maps (F_o) were obtained as described in Materials and Methods and Paper III. To our knowledge the CMLE structure solution represent one of the largest Se-substructures solved by direct methods. The structure was then solved in the dominant unit cell at higher resolution by molecular replacement. AMoRe (Navaza 1994) was the only program successful in finding a solution. Nevertheless, the major difference between the two unit cells is a slight difference in the intertetramer angles (Paper III). Further, as predicted by Glumoff *et al.* (1995) from the first native data processing in space group $P2_12_12_1$, the asymmetric unit turned out to contain two tetramers related by an intertetramer two-fold (Figure 17). This is consistent with size exclusion chromatography studies (Mazur *et al.* 1994) showing that the biological, active oligomer state was a tetramer. Thus, altogether seven NCS-operators defined the CMLE crystal structure. The intertetramer twofold operator was not perfect and did not refine with IMP (Kleywegt and Jones 1999), which suggests that the two tetramers were not identical. Indeed, the r.m.s.d. for all Cα-atoms was 0.532 Å between tetramers, while the intratetramer r.m.s.ds were on the order of 0.05-0.09 Å.

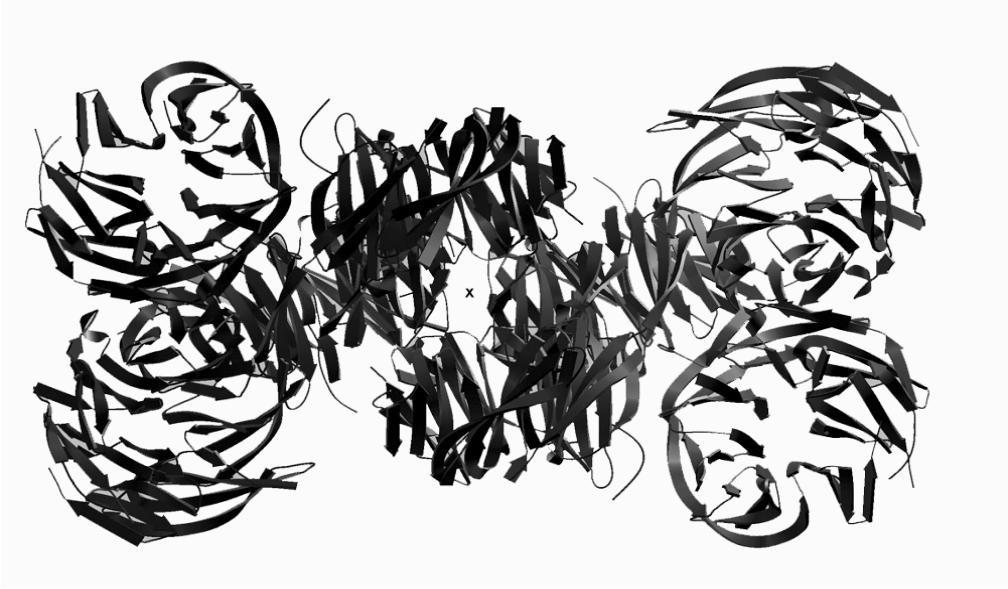

Figure 17. The initial 3 Å model for the CMLE crystal asymmetric unit contents with two tetramers and 328 kDa. The inter-tetramer twofold axis is indicated by an "X" (view along the axis). One tetramer is on the left and the other on the right.

5.1.2. Description of the crystal structure

From the crystal structure CMLE was found to have a quite regular seven bladed β-propeller fold (Figure 18), with the biological unit being a D_2-homotetramer with two distinct subunit interfaces (the third intra-tetramer twofold was along the channel that is formed at the centre of the tetramer (Paper IV)). The monomer consists of the common repeating four-stranded antiparallel β-sheet present in all β-propeller proteins. The side connecting the chains is, following established conventions, termed the "top" (Varghese *et al.* 1983). The N- and C-termini are located on the same blade (numbered as the last one, VII), such that the C-terminus forms the innermost strand, while the N-terminus is the second innermost. The chain runs from inside to out and is connected through large loops from the outermost strand to the innermost of the next blade (strands being labeled A to D in-to-out). The active site was deduced to reside on the internal pseudo-sevenfold symmetry axis, at the N-terminal "top-side" of the β-propeller innermost β-strands – as is usual for β-propeller enzymes (Jawad and Paoli 2002). In this particular case, it contained a cluster of potentially active catalytic residues (Figure 18), which were also absolutely conserved in weakly homologous sequences. Furthermore, site-directed mutagenesis of two of the five conserved residues (Glu212 to Ala, His148 to Ala) confirmed that they had a functional role. No changes in structure were observed in the mutants, as shown by native PAGE and CD spectroscopy (Figure 19). Finally, it was also possible to model the substrate into the active site so-defined in a catalytically rational and consistent way (*e.g.* in terms of correct product stereochemistry) (Figure 20). Interestingly, on one twofold subunit interface of the tetramer an unexpected ligand, a PIPES buffer molecule, was found (Figure 21). The binding site is reminiscent of a real binding site with symmetrically arranged non-bonded arginines and glutamates close to the 2-fold symmetry axis. PIPES does not however fit here very well, but a similar symmetrical ligand (such as a sugar-bis-phosphate) might.

In conclusion, the structure of *Neurospora crassa* CMLE, representative of the fungal, eucaryotic MLE-class (Mazur *et al.* 1994), was shown to be structurally completely unrelated to the other two known structural classes of MLEs, the bacterial MLEs and CMLEs. I have therefore identified the structure of enzymes of the third branch of convergent functional evolution in MLEs. This is particularly interesting as the fungal and bacterial MLEs both catalyze a stereochemically identical reaction, with and without a metal-cofactor and in a completely different structural framework. Nonetheless the active sites have some conserved features.

5.1.3. Implications for the catalytic mechanism of MLEs

The *Neurospora* CMLE catalyzes the lactonization reaction with an analogous mechanism to that of bacterial MLEs. In both cases acidic and basic residues are important in catalysis, and positively

charged residues are important in binding the carboxylic acid substrate. While the bacterial MLEs need a Mn^{2+} cofactor for catalytic activity, no metal cofactor is required for *Neurospora* CMLE. This difference raises the question as to whether the metal cofactor might actually have a direct role in catalysis in bacterial MLEs, and how its function is replaced in catalysis by CMLE.

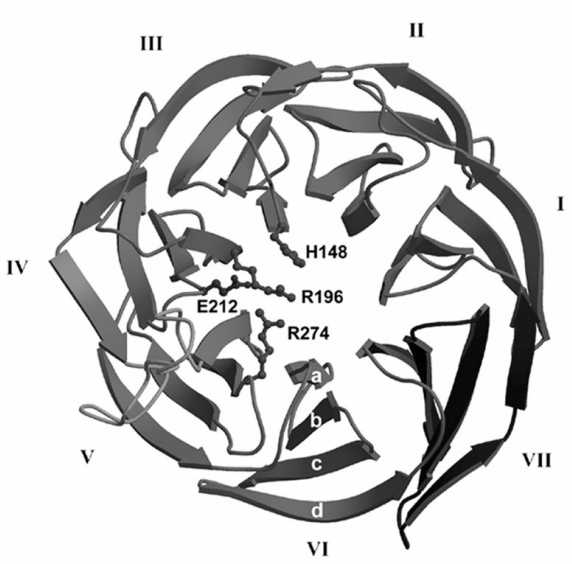

Figure 18. The fold of CMLE. The blades are numbered I-VII, and strand a-d, with N- terminus at blade VII strand b and C-terminus at blade VII strand a. The cluster of charged side chains on top-side of the propeller is shown.

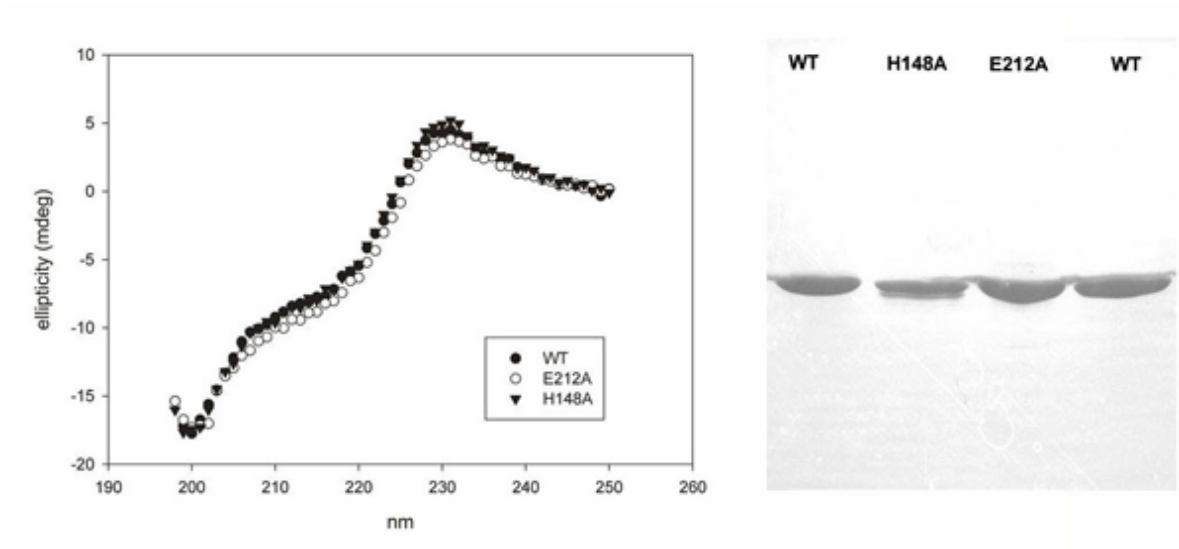

Figure 19. CD-spectra and native-PAGE on the His148Ala and Glu212Ala mutants of CMLE.

In *Neurospora* CMLE it seems probable that Glu212 is used as a general acid catalyst and His148 as the general base, as the arginines (Arg 196 and Arg 274) probably mainly have a binding function and are not involved in transition state stabilization (see Paper IV). In any case the structure of *Neurospora* CMLE seems to validate the catalytic mechanism proposed for PpMLE by Gerlt and

Gassman (1992), where a general acid is used to protonate the enolate intermediate. It shows that a metal ion can be effectively replaced by protein groups. This presumably requires precise hydrogen bonding to the catalytic acid, Glu212 (Figure 19; Paper IV).

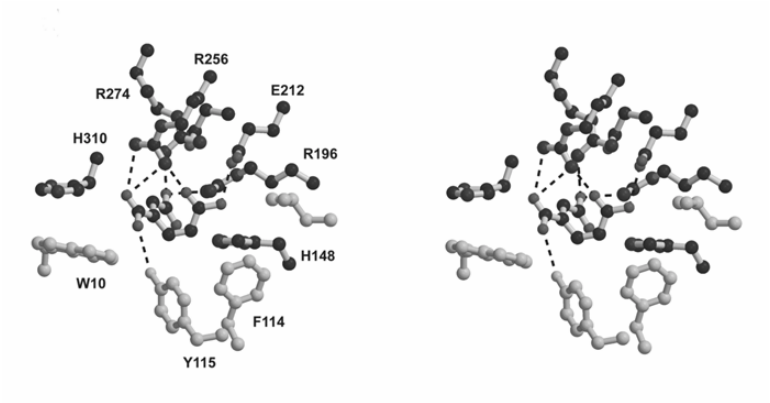

Figure 20. Modelling of the 3-carboxy-*cis,cis*-muconate substrate into the active site (stereo). Proposed hydrogen bonds are shown as dashed lines.

Whether the metal cofactor in bacterial MLEs has a binding role analogous to the active site arginines in *Neurospora* CMLE or a catalytic role remains to be studied. In the related GlucD enzyme one of the metal-ion ligands is neutral (Gulick *et al.* 2000), which was speculated to possibly have a catalytic effect. In MLE making such neutralizing metal-ligand mutations (Asp to Asn, Glu to Gln) might be useful in estimating the effect of the metal ligand on catalysis and binding. The mutant Glu327 to Gln should also be made to verify the role of this residue in catalysis.

Figure 21. Binding of a PIPES-molecule (arrow) at a twofold subunit interface of the CMLE tetramer.

5.2. Substrate specificity and dehalogenating activity of Cl-MLEs

5.2.1. Substrate specificity and binding

The capping domains (the N-terminal α+β domain and the C-terminal subdomain) confer substrate specificity for the bacterial MLEs, as is evident from mutagenesis studies, as well as from the structural studies of the Ile54Val and Phe329Ile variants of MLE (Paper I, Vollmer *et al.* 1998). The same is true for the other enzymes in the family (Gerlt and Babbitt 2001). Mutagenesis and modelling indicate Phe329 and Ile54 as positions for binding pockets (Paper I). They define substrate specificity in terms of substituents, and therefore also the direction of cycloisomerisation (1,4- or 3,6-cycloisomerisation) (Figure 12). This information was also the basis for positioning of the modelled substrates into the active site (Paper I). Substrate modelling suggests that Lys169 is the only residue available to act as the general base: it is on the correct face of the substrate to produce *syn*-addition and the correct product stereoisomer (4*S*)-muconolactone (Figure 12). The inactivity of the Lys169Ala variant of PpMLE has confirmed this to be very likely (Kaulmann *et al.* 2001). Lys273 of PpMLE is homologous to the enolase general base, but is apparently not functional as a general base in MLE. For stereochemical reasons it cannot act as a catalytic base, and it is also ion-paired, while the Lys169 side chain is free in the active site (Helin *et al.* 1995, Hasson *et al.* 1998).

A hypothesis was presented for lactonization and dehalogenation by a single base, Lys169 of MLE, as based on the modeling it seems clear that Lys169 is the only base accessible for the substrate. This requires rotation of the lactone ring 180° before the second step to yield dienelactones (Paper I). This is necessary to produce the correct *trans*-dienelactone product of dehalogenation. Theoretically, either the lactone ring or the reactive carboxylate could rotate to produce the correct product, but the latter is fixed by its interaction with the active site ionic groups (Figure 12).

5.2.2. Co-evolution of the hydrophobic core in bacterial MLEs with dehalogenation

During the comparison of MLE and ReCl-MLE structures it was noted that ReCl-MLE contained a large cavity in its interior. It was realized that these “packing defects” were also present, but smaller, in MLE: 50 + 33 $Å^3$ in MLE *versus* 210 $Å^3$ in ReCl-MLE (Figure 22). Site-directed mutagenesis was done to study the possible effect on activity, as we hypothesized that these might have a role in dehalogenation. It was also noticed that the residues bordering the cavities were different between MLE and Cl-MLE sequences (Paper V). Further, the *Pseudomonas* Cl-MLE was crystallized and the structure solved at 1.95 Å. This high resolution structure confirmed the conserved structural features, and allowed visualization of structural features not visible in the 3 Å

ReCl-MLE structure *e.g.* B-factor distribution by residue and atom (Paper V). Residues Phe103, Thr52 and Glu304 and Ser312 in PpMLE were mutated to the Cl-MLE equivalents (Ser99, Gly48, Asp300 and Ala308, respectively) and activities measured. None of the mutants had dehalogenating activity, but other interesting effects were observed for Thr52Gly and the double mutant Thr52Gly-Phe103Ser: Thr52Gly had a k_{cat} of 2% wild type, while the double mutant had a k_{cat} of about 10% of wild type, similar to the k_{cat} of P51 Cl-MLE for *cis,cis*-muconate, while substrate binding was only slightly affected (Paper V).

The effects of the above mutations can be explained by observed structural differences between MLEs and Cl-MLEs: the Cl-MLEs exist in open and closed conformations, where a major switch is observed at Gly48 (homologous to Thr52 of PpMLE). This results in the large cavity observed in the ReCl-MLE structure with the closed active site structure, where Ser99 (homologous to Phe103) is hydrogen bonded to Gly48 (Figure 22). I suggest these changes from MLE to Cl-MLE lead to changes in binding dynamics which in turn can affect dehalogenation. Firstly, the mutants showed that k_{cat}, not K_m, is mainly affected. Secondly, with Thr52 in place of Gly48 such conformational changes are not likely, as the polypeptide cannot twist as with Gly and because Thr52 is hydrogen bonded to Glu50 (Paper V).

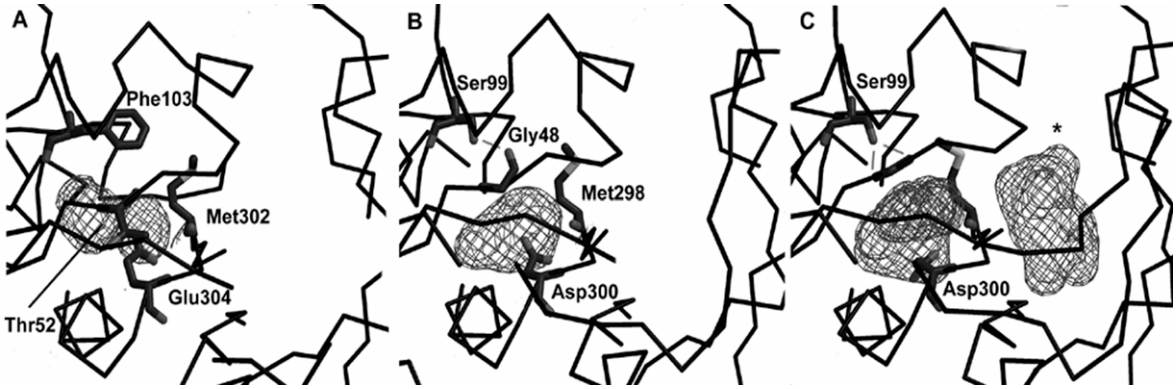

Figure 22. The hydrophobic core packing in MLEs and Cl-MLEs near the active site. A) MLE B) P51 Cl-MLE (open conformation) C) ReCl-MLE (closed conformation). Light grey lines represent hydrogen bonds; the second mesh (marked with *) in the ReCl-MLE structure represents the volume of the closed active site. Note the movement of Asp300 and Gly48 from B to C as a result of the active site loop closure.

Finally, a major difference between the MLE and Cl-MLE active sites resides in the mobile 21-30 loop. It is clearly less polar in Cl-MLEs; in particular Ile19 and Met21 point towards the active site in the ReCl-MLE structure, while modelling the MLE loop shows that His22 of MLE is in the same position as Ile19 of Cl-MLE (Paper V). This may also account for functional differences with Cl-substituted muconates (Paper V). The differences in the loop and in the hydrophobic core structure near the active site at Thr52/Gly48 (MLE/Cl-MLE) may be connected. They may thus determine

differences in binding dynamics and active site properties, resulting in the functional differences observed.

5.3. Charged residues in MLE and globular proteins

5.3.1. Data base analysis

When the PpMLE structure was analyzed (Helin *et al.* 1995), it was noted that a surprisingly large percentage of the total charged surface area was buried (64%). This was extended to an analysis of the distribution of charged residues in a larger set of globular proteins, as earlier studies had included only relatively small sets of proteins and did not include larger globular proteins (Barlow and Thornton 1983, Rashin and Honig 1984, Miller *et al.* 1987). It was found that large proteins bury an increasing amount of charged surface in general, and this happens at a rate comparable, or faster, than for aliphatic surface, whereas polar uncharged and aromatic surfaces are largely buried even for small proteins (Paper II, Figure 1).

To measure the actual distribution of individual groups within proteins a smaller set of 11 proteins was analyzed; for five large proteins of more than 500 amino acids 17-30 fully buried charged residues were found, while small proteins seem to bury 2 charges on average (Paper II, Rashin and Honig 1984, Barlow and Thornton 1983). These results and the analysis of charge distribution within MLE and two other examples of larger proteins in the set, hemocyanin and cyclodextrin glycosyltransferase, indicated that protein cores are necessarily quite polar (to solvate the large number of charges adequately). They also may not be fully optimized in terms of charge stabilization, as hydrogen bonding potential is not always fully satisfied, or charges are not necessarily neutralized, or both (Paper II).

5.3.2. Stability of MLE variants

Candidate residues among the buried charges found in MLE were selected for further mutagenesis to neutral or hydrophobic variants to study the effect of the electrostatics on structural stability. To test the hypothesis that some buried charges might not be optimized for stability three MLE mutants, Glu50Gln, Asp178Asn, and Asp150Asn, were constructed. The variants were studied by fluorescence unfolding to estimate the melting temperature compared to wild type, and the structure of the Asp178Asn variant was solved. The Glu50 is fully buried and acts as a helix cap; as a result the Glu50Gln-mutant turned out to be unstable and inactive with a $T_m < 40$ °C and $k_{cat} \sim 5\ s^{-1}$, $K_m \sim 50 \mu M$ (D. Cohen, personal communication). The protein also went into inclusion bodies and

gave a low yield (ca 1 mg/l; U. Schell, personal communication). The Asp178Asn had a T_m of 55.8 °C (4.2 °C higher than wild type) and was also close to wild type in activity (>50% activity). The structure of the Asp178Asn variant is essentially identical with wild type, with one extra hydrogen bond to Asn178 (Figure 23). The Asp150Asn mutant on the other hand had no change in T_m, while the activity was compromised (≤10% wild type activity), but purification gave good yields (Paper II). Mutants Arg196Met and Glu304Gln were also constructed. Both gave very low yields (~1 mg/l) and activities were low, but no thermostability measurements have been carried on these mutants.

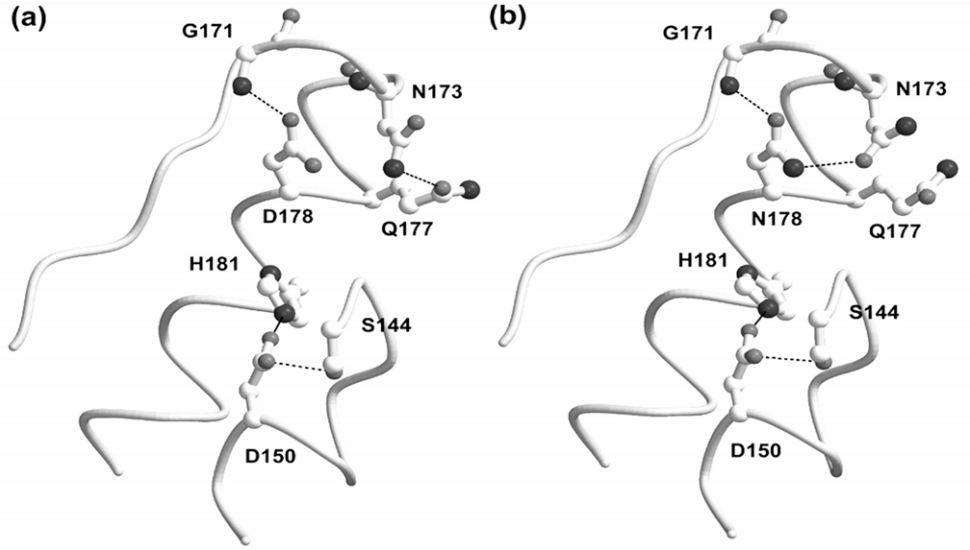

Figure 23. The structure of the PpMLE Asp178Asn-variant. Hydrogen bonded interactions of (a) the wild type enzyme and (b) the Asn178Asn-variant around the mutated residue.

Concluding remarks

The MLEs provide a collection of interesting enzymes, because of their convergent functional evolution from three completely different protein structures to catalyze the same reaction. The opportunity to look at the same catalytic reaction in different structural frameworks (proteins) can be helpful for figuring out what are the essential parts. This is now possible in MLEs with our new *Neurospora* CMLE structure (Paper IV). Crystal structures in complex with substrates and further site directed mutagenesis would be needed to reveal exactly what is conserved and what is not. The crystallization of substrate complexes remains the major challenge for this system. Finally, the crystal structure of a bacterial CMLE, the third class of MLEs, would complete the picture for MLE structures. The enzyme has been crystallized (Hasson *et al.* 1997) and the structure will hopefully become available in near future (D. Ringe, personal communication).

The MLEs on the other hand also give an example of evolution of a novel catalytic function on top of an existing one in the dehalogenating Cl-MLEs. This function may also have practical applications in bioremediation by bacteria. The mechanism of how this function has evolved seems intriguing, possibly by affecting both the dynamics of binding and the polarity of the active site. Nevertheless, the function seems hidden in subtle changes not necessarily directly connected to substrate binding or the catalytic residues themselves. Although substrate specificity determining changes have been identified, the catalytic residues used seem to be the same for both lactonization and dehalogenation. Future mutagenesis work hopefully will bring conclusive evidence on what is essential for dehalogenation activity, while our structural analysis thus far have revealed the essential, while not obvious, conserved differences between the MLE and Cl-MLE structures (Paper I, Paper V).

Finally, the PpMLE structure reveals in a fascinating way how proteins are made of finely balanced networks of interaction. The MLE barrel is hydrophilic, with a buried metal cofactor and intricate networks of hydrogen bonds, to stabilize all the ionic residues involved in catalysis. However, it is now clear from an increasing number of studies, including ours (Paper II), that proteins in their interior are not like oil droplets even outside the active sites. Polar interactions are important and buried charges are common. Buried charges are often solvated very well by the protein. This will also be a very interesting field to follow, and understanding the electrostatics of proteins will no doubt be central to understanding folding and thermostability.

Acknowledgements

This work has been carried out for the first two years in the Turku Centre for Biotechnology and then at the Institute for Biotechnology, University of Helsinki and partly “on the side” during my “siviilipalvelus” at the Haartman institute and afterwards.

I wish to acknowledge the directors of the institutes, both at BTK and BI, Prof. Markku Jalkanen and Prof. Riitta Lahesmaa in Turku and Prof. Mart Saarma in Helsinki, for the excellent facilities and environment for the work.

I also wish to give my thanks to:

Prof. Carl Gahmberg for the opportunity to complete my studies at the Department of Biosciences.

My supervisor Prof. Adrian Goldman for the excellent introduction to protein chemistry and structural biology, encouraging attitude, and secure funding for the project.

All the people in the X-ray lab during the years: especially the "senior" Ph. D. students Pirkko Heikinheimo, Veli-Matti Leppänen, and Sari Helin as well as Drs. Tuomo Glumoff and Michael Merckel, for advice in the lab and with the computers.

Veli-Matti, Vesa, Pirkko, and Michael also for help with data collections among other things, and Michael in particular for his work on the CMLE NCS operators.

Sari Helin, Ursula Schell and Lari Lehtiö and Tuomo Glumoff for help and discussions on the project. Especially to Lari for doing a lot of good work on MLEs for his Master’s thesis.

Sirpa Valjakka and Kaija Söderlund for secretarial help and various other very practical advice outside their exact duties.

The computer staff during the years: Dan-Johann "Dante" Still, Petri Vähäkoski, Markus Tujula and the BI-helpers.

Raija Andersen, Merja Nissinen, Elaine Connor and Maria Rehn for skillful technical help.

All the co-authors of the publications, in particular Andy Thompson for sacrificing a lot of beam time to the CMLE crystals at the ESRF.

Dr. Sarah Butcher and Prof. Juha Rouvinen for kindly reviewing my thesis and their constructive critisism.

I wish also to warmly thank Prof. Peter Kahn and Jennifer Kahn for being such excellent hosts during my stay in New Jersey in 1998, and Peter Kahn also for partly guiding my work.

Thanks also to Prof. Timo Hyypiä and the Picorna-lab girls (Maaria, Vilja, Camilla, Annu, Åse, Päivi, Olya, Hanna-Mari, Pia and Anne) for very nice working environment and friendly atmosphere in the lab at the Virology Department.

Thanks to My Mom for sponsoring me during my studies (and for the rest of the upbringing, in particular before all this).

Most of all, thanks to my lovely wife, Merja, and my son, Tuomas, for their love and support and keeping me on the track. Indeed, a special thank you to my dear little Tuomas for making the priorities in life a bit clearer.

Helsinki, in January 2003.

Tommi Kajander

References

Abraham W-R, Nogales B, Golyshin PN, Pieper DH and Timmis KN (2002). Polychlorinated biphenyl-degrading microbial communities in soils and sediments. Curr Opin Mirobiol 5:246-253.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-402.

Amyes TL and Kirby AJ (1988). Intramolecular nucleophilic addition of phenolate oxygen to double bonds activated by carboxyl and carboxylate groups. Relative reactivity, stereochemistry, and mechanism. J Am Chem Soc 110:6505-6514.

Anandarajah K, Kiefer PM Jr, Donohoe BS and Copley SD (2000). Recruitment of a double bond isomerase to serve as a reductive dehalogenase during biodegradation of pentachlorophenol. Biochemistry 39:5303-5311.

Anderson SR, Anderson VE, Knowles JR (1994). Primary and secondary kinetic isotope effects as probes of the mechanism of yeast enolase. Biochemistry 33:10545-55.

Argos P, Rossmann MG, Grau UM, Zuber H, Frank G and Tratschin JD (1979). Thermal stability and protein structure. Biochemistry 25:5698–5703.

Axe DD, Foster NW and Fersht AR (1996). Active barnase variants with completely random hydrophobic cores. Proc Natl Acad Sci USA 93:5590-5594.

Babbitt PC, Hasson MS, Wedekind JE, Palmer DRJ, Barrett WC, Reed GH, Rayment I, Ringe D, Kenyon GL and Gerlt JA (1996). The enolase superfamily: a general strategy for enzyme-catalysed abstraction of the α-protons of carboxylic acids. Biochemistry 35:16489-16501.

Babbitt PC, Mrachko GT, Hasson MS, Huisman GW, Kolter R, Ringe D, Petsko GA, Kenyon GL and Gerlt JA (1995). A functionally diverse enzyme superfamily that abstracts the α protons of carboxylic acids. Science 267:1159-61.

Barlow DJ and Thornton JM (1983). Ion pairs in proteins. J Mol Biol 168:867-885.

Barton GJ (1993). ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng 6:37-40.

Bearne SL and Wolfenden R (2000). Mandelate racemase in pieces: concentrations of enzyme functional groups in the transition state. Biochemistry 36:1646-1656.

Betz SF, Bryson JW and DeGrado WF (1995). Native-like and structurally characterized designed α-helical bundles. Curr Opin Struct Biol 5:457-463.

Betz SF, Raleigh DP and DeGrado WF (1993). *De novo* protein design: from molten globules to native-like states. Curr Opin Struct Biol 3:601-610.

Blasco R, Wittich RM, Mallavarapu M, Timmis KN and Pieper DH (1995). From xenobiotic to antibiotic, formation of protoanemonin from 4-chlorocatechol by enzymes of the 3-oxoadipate pathway. J Biol Chem 270:29229-35.

Blessing RH and Smith GD (1999). Difference structure factor normalization for heavy-atom or anomalous-scattering substructure determinations. J Appl Cryst 32:664-670.

Boersma MG, Dinarieva TY, Middelhoven WJ, van Berkel WJ, Doran J, Vervoort J and Rietjens IM (1998). 19F nuclear magnetic resonance as a tool to investigate microbial degradation of fluorophenols to fluorocatechols and fluoromuconates. Appl Environ Microbiol.64:1256-1263.

Bolon DN and Mayo SL (2001). Polar residues in the protein core of *Eschericia coli* thioredoxin are important for fold specificity. Biochemistry 40:10047-10053.

Brünger AT (1992). X-PLOR Version 3.1. A system for X-ray crystallography and NMR, Yale University Press, New Haven.

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T and Warren GL (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D54:905-921.

Byrne MP, Manuel RL, Lowe LG and Stites WE (1995). Energetic contribution of side chain hydrogen bonding to the stability of staphylococcal nuclease. Biochemistry 34:13949-13960.

Cain RB (1980). The uptake and catabolism of lignin related aromatic compounds and their regulation in micro-organisms. pp. 21-60 *in* Lignin biodegradation: microbiology, chemistry, and potential applications. vol I. Kirk TK, Higuchi T and Chang H (eds.) CRC press, Boca Raton.

Cambillau C and Claverie J-M (2000). Structural and genomic correlates of hyperthermostability. J Biol Chem 275:32383-32386.

Carter PJ, Winter G, Wilkinson AJ and Fersht AR (1984). The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). Cell 38:835-840.

Cavagnero S, Debe DA, Zhou ZH, Adams MW and Chan SI (1998). Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins. Biochemistry 37:3369-3376.

Chakravarty and Varadarajan (1999). Residue depth: a novel parameter for the analysis of protein structure and stability. Structure Fold Des 7:723-732.

Chothia C (1974). Hydrophobic bonding and accessible surface area in proteins. Nature 248:338-339.

Chothia C (1975). Structural invariants in protein folding. Nature 254:304-308.

Chothia C (1976). The nature of accessible and buried surfaces in proteins. J Mol Biol 105:1-14.

Cohen BE, McAnaney TB, Park ES, Jan YN, Boxer SG and Jan LY (2002). Probing protein electrostatics with a synthetic fluorescent amino acid. Science 296:1700-1703.

Collaborative Computational Project, Number 4 (1994). The CCP4 Suite: Programs for protein crystallography. Acta Crystallogr D50:760-763.

Conolly M (1983). Analytical molecular surface calculation. J Appl Cryst 16:548-558.

Copley RR and Bork P (2000). Homology among (βα)(8) barrels: implications for the evolution of metabolic pathways. J Mol Biol 303:627-641.

Cordes MHJ, Burton RE, Walsh NP, McKnight J and Sauer RT (2000). An evolutionary bridge to a new protein fold. Nat Struct Biol 7:1129-1132.

Creighton TE (1993). Proteins, structure and molecular properties. WH Freeman and Company, New York.

Creighton DJ and Hamilton DS (2001). Brief history of glyoxalase I and what we have learned about metal ion-dependent, enzyme-catalyzed isomerizations. Arch Biochem Biophys. 387:1-10.

Dao-pin S, Sauer U, Nicholson H and Matthews BW (1991a). Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. Biochemistry 30: 7142-7153.

Dao-pin S, Anderson DE, Baase WA, Dahlquist FW and Matthews BW (1991b). Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. Biochemistry 30:11521-9.

DeGrado WF, Wasserman ZR and Lear JD (1989). Protein design, a minimalist approach. Science. 243:622-8.

DeTitta GT, Weeks CM, Thuman P, Miller R and Hauptman HA (1994). Structure solution by minimal-function phase refinement and Fourier filtering. I. Theoretical basis. Acta Crystallogr A50:203-210.

Dill KA (1990). Dominant forces in protein folding. Biochemistry 29:7133-7155.

Elcock AH (1998). The stability of salt bridges at high temperatures — implications for hyperthermophilic proteins. J Mol Biol 284:289-503.

Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP and Matthews BW (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science 255:178-183.

Esnouf RM (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. J Mol Graph Model 15:132-134.

Eulberg D, Kourbatova EM, Golovleva LA and Schlomann M (1998). Evolutionary relationship between chlorocatechol catabolic enzymes from Rhodococcus opacus 1CP and their counterparts in proteobacteria: sequence divergence and functional convergence. J Bacteriol 180:1082-1094.

Faller LD, Baroudy BM, Johnson AM and Ewall RX (1977). Magnesium ion requirements for yeast enolase activity. Biochemistry 16:3864-3869.

Fetzner S (1998). Bacterial dehalogenation. Appl Microbiol Biotechnol 50:633-57.

Galperin MY, Bairoch A and Koonin EV (1998). A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases. Protein Sci 7:1829-35.

Gerlt JA and Babbitt PC (2001). Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. Annu Rev Biochem 70:209-246.

Gerlt JA and Gassman PG (1992). Understanding enzyme-catalyzed proton abstraction from carbon acids: details of stepwise mechanisms for β-elimination reactions. J Am Chem Soc 114:5928-5934.

Gerlt JA and Gassman PG (1993). An explanation for rapid enzyme-catalyzed proton abstraction from carbon acids: importance of late transition states in concerted mechanisms. J Am Chem Soc 115:11552-11568.

Gerlt JA, Kozarich JW, Kenyon Gl and Gassman PG (1991). Electrophilic catalysis can explain the unexpected acidity of carbon acids in enzyme-catalyzed reactions. J Am Chem Soc 113: 9667-9669.

Giletto A and Pace NC (1999). Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T1. Biochemistry 38:13379-13384.

Glumoff T, Helin S, Mazur P, Kozarich JW and Goldman A (1996). Crystallization and preliminary crystallographic analysis of 3-carboxy-*cis,cis*-muconate lactonizing enzyme from *Neurospora crassa*. Acta Crystallogr D52:221-223.

Goldman A (1985). Crystal structure of muconate lactonizing enzyme from *Pseudomonas putida* PRS2000 at 3 Å resolution. Ph.D. Thesis. Yale University. New Haven.

Goldman, A. (1995). How to make my blood boil. Structure 3, 1277-1279.

Goldman A, Ollis D, Ngai KL and Steitz TA (1985). Crystal structure of muconate lactonizing enzyme at 6.5 Å resolution. J Mol Biol 182:353-5.

Goldman A, Ollis DL and Steitz TA (1987). Crystal Structure of Muconate Lactonizing Enzyme at 3 Å Resolution. J Mol Biol 194:143-153.

Gonzales L, Woolfson DN and Alber T (1996). Buried polar residues and structural specificity in the GCN4 leucine zipper. Nat Struct Biol 3:1011-1018.

Gribble GW (1998). Naturally occurring organohalogen compounds. Acc Chem Res 31:141-152.

Grimsley GR, Shaw KL, Fee LR, Alston RW, Huyghues-Despointes BM, Thurlkill RL, Scholtz JM and Pace CN (1999). Increasing protein stability by altering long-range coulombic interactions. Protein Sci 8:1843-1849.

Gulick AM, Hubbard BK, DM, Gerlt JA and Rayment I (2000). Evolution of enzymatic activities in the enolase superfamily: crystallographic and mutagenesis studies on the reaction catalysed by D-glucarate dehydratase from *Escherichia coli.* Biochemistry 39:4590-602.

Gulick AM, Schmidt DM, Gerlt JA and Rayment I (2001). Evolution of enzymatic activities in the enolase superfamily: crystal structures of the L-Ala-D/L-Glu epimerases from *Escherichia coli* and *Bacillus subtilis.* Biochemistry 40:15716-15724.

Guthrie PJ and Kluger R (1993). Electrostatic stabilization can explain the unexpected acidity of carbon acids in enzyme-catalysed reactions. J Am Chem Soc 115:11569-11572.

Harp JM, Timm DE and Bunick GJ (1998). Macromolecular crystal annealing: overcoming increased mosaicity associated with cryocrystallography. Acta Crystallogr D54:622-628

Harris JM, Gonzalez-Bello C, Kleanthous C, Hawkins AR, Coggins JR and Abell C (1996). Evidence from kinetic isotope studies for an enolate intermediate in the mechanism of type II dehydroquinases. Biochem J 319:333-336.

Harwood CS and Parales RE (1996). The β-ketoadipate pathway and the biology of self-identity. Annu Rev Microbiol 50:553-90.

Hasson MS, Schlichting I, McGowen MM, Woolridge EM, Kozarich JW, Petsko GA and Ringe D (1997). Characterization of two crystal forms of 3-carboxy-*cis,cis*-muconate lactonizing enzyme from *Pseudomonas putida*. Acta Crystallogr D53:352-353.

Hasson MS, Schlichting I, Moulai J, Taylor K, Barrett W, Kenyon GL, Babbitt PC, Gerlt JA, Petsko GA and Ringe D (1998). Evolution of an enzyme active site: the structure of a new crystal form of muconate lactonizing enzyme compared with mandelate racemase and enolase. Proc Natl Acad Sci USA 95:10396-10401.

Hauptman HA (1997). Shake-and-bake: an algorithm for automatic solution *ab initio* of crystal structures. Methods Enzymol 277:3-13.

Hebert EJ, Giletto A, Sevcik J, Urbanikova L, Wilson KS, Dauter Z and Pace CN (1998). Contribution of a conserved asparagine to the conformational stability of ribonucleases Sa, Ba, and T1. Biochemistry 37:16192-16200.

Hegeman GD, Rosenberg EY and Kenyon GL (1970). Mandelic acid racemase from *Pseudomonas putida*. Purification and properties of the enzyme. Biochemistry 9:4029-4036.

Helin S, Kahn PC, Guha BL, Mallows DG and Goldman A (1995). The refined X-ray structure of muconate lactonizing enzyme from *Pseudomonas putida* PRS2000 at 1.85 angstrom resolution. J Mol Biol 254:918-941.

Hendsch ZS and Tidor B (1994). Do salt bridges stabilise proteins? A continuum electrostatic analysis. Protein Sci 3:211-226.

Hendsch ZS, Jonsson T, Sauer RT and Tidor B (1996). Protein stabilisation by removal of unsatisfied polar groups: computational approaches and experimental tests. Biochemistry 35:7621-7625.

Hilal SH, Brewer JM, Lebioda L, Carreira LA (1995). Calculated effects of the chemical environment of 2-phospho-D-glycerate on the pKa of its carbon-2 and correlations with the proposed mechanism of action of enolase. Biochem Biophys Res Commun 211:607-613.

Hill RB and DeGrado WF (2000). A polar, solvent exposed residue can be essential for native protein structure. Structure 8:471-479.

Ho SP and DeGrado WF (1987). Design of a 4-helix bundle protein: synthesis of peptides which selfassociate into a helical protein. J Amer Chem Soc 109:6751–6758.

Ho CM and Marshall GR (1990). Cavity search: an algorithm for the isolation and display of cavity-like binding regions. J Comput Aided Mol Des 4:337-454.

Hoier H, Schlömann M, Hammer A, Glusker JP, Carrell HL, Goldman A, Stezowski JJ and Heinemann U (1994). Crystal structure of chloromuconate cycloisomerase from *Alcaligenes eutrophus* JMP134 (pJP4) at 3 Å resolution. Acta Crystallogr D50:75-84.

Honig B and Nicholls A (1995). Classical electrostatics in biology and chemistry. Science 268:1144-1149.

Horowitz A and Fersht AR (1990). Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. J Mol Biol 214:613-617.

Horovitz A, Serrano L, Avron B, Bycroft M and Fersht AR (1990). Strength and co-operativity of contributions of surface salt bridge to protein stability. J Mol Biol 216:1031-1044.

Howell PL, Blessing RH, Smith GD and Weeks CM (2000). Optimizing DREAR and SnB parameters for determining Se-atom substructures. Acta Crystallogr D56:604-617.

Hubbard SJ and Argos P (1994). Cavities and packing at protein interfaces Protein Sci 3:2194-2206.

Hubbard SJ and Argos P (1996). A functional role for protein cavities in domain:domain motions. J Mol Biol 261:289-300.

Hubbard SJ, Gross KH and Argos P (1994). Intramolecular cavities in globular proteins. Protein Eng 7:613-626.

Hwang J-K and Warshel A (1988). Why ion pair reversal by protein engineering is unlikely to succeed. Nature 334:270-272.

Hädener A, Matzinger PK, Malashkevich VN, Louie GV, Wood SP, Oliver P, Alefounder PR, Pitt AR, Abell C and Battersby AR (1993). Purification, characterization, crystallisation and X-ray analysis of selenomethionine-labelled hydroxymethylbilane synthase from *Escherichia coli*. Eur J Biochem 211:615-624.

Jancarik J and Kim S-H (1991). Sparse matrix sampling: a screening method for crystallization of proteins. J Appl Cryst 24:409-411.

Janin J (1997). Ångströms and calories. Structure 5:473-479.

Janssen DB, Oppentocht JE and Poelarends GJ (2001). Microbial dehalogenation. Curr Opin Biotechnol 12:254-258.

Jawad Z and Paoli M (2002). Novel sequences propel familiar folds. Structure 10:447-454.

Jencks WP (1980). When is an intermediate not an intermediate? Enforced mechanisms of general acid-base catalyzed, carbocation, carbanion, and ligand exchange reactions. Chem Soc Rev 10:345-375.

de Jong E, Field JA, Spinnler H-E, Wijnberg JBPA and de Bont JAM (1994). Significant biogenesis of chlorinated aromatics by fungi in natural environments. Appl Environ Microbiol 60:264-270.

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

Kallarakal AT, Mitra B, Kozarich JW, Gerlt JA, Clifton JG, Petsko GA and Kenyon GL (1995). Mechanism of the reaction catalyzed by mandelate racemase: structure and mechanistic properties of the K166R mutant. Biochemistry. 34:2788-2797.

Karplus PA (1997). Hydrophobicity regained. Protein Sci 6:1302-1307.

Karshikoff A and Ladenstein R (1998). Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. Protein Eng 11:867-872.

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

Kato Y and Asano Y (1995). 3-methylaspartate ammonia-lyase from a facultative anaerobe, strain YG-1002. Appl Microbiol Biotechnol 43:901–907.

Kaulmann U, Kaschabek S and Schlomann (2001). Mechanism of chloride elimination from 3-chloro- and 2,4-dichloro-*cis,cis*-muconate: new insight obtained from analysis of muconate cycloisomerase variant CatB-K169A. J Bacteriol 183:4551-4561.

Kauzmann W (1959). Some factors in the interpretation of protein denaturation. Adv Prot Chem 14:1-57.

Kellis JT Jr, Nyberg K, Sali D, Fersht AR (1988). Contribution of hydrophobic interactions to protein stability. Nature 333:784-786.

Key BD, Howell RD and Criddle CS (1997). Fluorinated organics in the biosphere. Envinron Sci Technol 31:2445-2454.

Khechinashvili NN, Janin J and Rodier F (1995). Thermodynamics of the temperature-induced unfolding of globular proteins. Protein Sci 4:1315-1324.

Kleywegt GJ Hoier H and Jones TA (1996). A re-evaluation of the crystal structure of chloromuconate cycloisomerase. Acta Crystallogr D52:858-863.

Kleywegt GJ and Jones TA (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. Acta Crystallogr D50:178-185.

Kleywegt GJ and Jones TA (1999). Software for handling macromolecular envelopes. Acta Crystallogr D55:941-944.

Kocher J-P, Prevost M, Wodak SJ and Lee B (1997). Properties of the protein matrix revealed by the free energy of cavity formation. Structure 4:1517-1529.

Koh JT, Cornish VW, Schultz PG (1997). An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. Biochemistry 36:11314-11322.

Kraulis P (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J Appl Cryst 24:946-950.

Landro JA, Gerlt JA, Kozarich JW, Koo CW, Shah VJ, Kenyon GL, Neidhart DJ, Fujita S and Petsko GA (1994). The role of lysine 166 in the mechanism of mandelate racemase from *Pseudomonas putida*: mechanistic and crystallographic evidence for stereospecific alkylation by (*R*)-α-phenylglycidate. Biochemistry 33:635-643.

Landro JA, Kallarakal AT, Ransom SC, Gerlt JA, Kozarich JW, Neidhart DJ and Kenyon GL (1991). Mechanism of the reaction catalyzed by mandelate racemase. 3. Asymmetry in reactions catalyzed by the H297N mutant. Biochemistry 30:9274-9281.

Larsen TM, Wedekind JE, Rayment I and Reed GH (1996). A carboxylate oxygen of the substrate bridges the magnesium ions at the active site of enolase: structure of the yeast enzyme complexed with the equilibrium mixture of 2-phosphoglycerate and phosphoenolpyruvate at 1.8 Å resolution. Biochemistry 35:4349-4358.

Lebioda L and Stec B (1988). Crystal structure of enolase indicates that enolase and pyruvate kinase evolved from a common ancestor. Nature 333:683-686.

Lebioda L, Stec B and Brewer JM (1989).The structure of yeast enolase at 2.25-A resolution. An 8-fold beta + alpha-barrel with a novel ββαα(βα)6 topology. J Biol Chem 264:3685-3693.

Lee B and Richards FM (1971). The interpretation of protein structures: estimation of static accessibility. J Mol Biol 55:379-400.

Lee SH, Hidaka T, Nakashita H, and Seto H (1995). The carboxyphosphonoenolpyruvate synthase-encoding gene from the bialaphos-producing organism *Streptomyces hygroscopicus*. Gene 153:143-144.

Levy CW, Buckley PA, Sedelnikova S, Kato Y, Asano Y, Rice DW and Baker PJ (2002). Insights into enzyme evolution revealed by the structure of methylaspartate ammonia lyase. Structure 10:105-113.

Liang J and Dill KA (2001). Are proteins well packed? Biophys J 81:751-766.

Lim WA and Sauer RT (1989). Alternative packing arrangements in the hydrophobic core of lambda repressor. Nature 339:31-36.

Lim WA and Sauer RT (1991). The role of internal packing interactions in determining the structure and stability of a protein. J Mol Biol 219:359-376.

Lim WA, Farruggio DC and Sauer RT (1992). Structural and energetic consequences of disruptive mutations in a protein core. Biochemistry 31:4324-4333.

Lu GJ (1999). FINDNCS: a program to detect non-crystallographic symmetries in protein crystals from heavy-atom sites. J Appl Cryst 32:365-368

Maes D, Zeelen JP, Thanki N, Beaucamp N, Alvarez M, Thi MHD, Backmann J, Martial JA, Wyns L, Jaenicke R and Wierenga RK (1999). The crystal structure of triosephosphate isomerase (TIM) from *Thermotoga maritima*: a comparative thermostability structural analysis of ten different TIM structures. Proteins 37:441-453.

Mazur P, Henzel, WJ, Mattoo S and Kozarich JW (1994). 3-Carboxy-*cis,cis*-muconate lactonizing enzyme from *Neurospora crassa*: an alternate cycloisomerase motif. J Bacteriol 176:1718-1728.

Merritt EA and Bacon DJ (1997). Raster3D: photorealistic molecular graphics. Methods Enzymol 277:505-524.

Miller S, Janin J, Lesk AM and Chothia C (1987). Interior and surface of monomeric proteins. J Mol Biol 196:641-656.

Mitra B, Kallarakal AT, Kozarich JW, Gerlt JA, Clifton JG, Petsko GA and Kenyon GL (1995). Mechanism of the reaction catalyzed by mandelate racemase: importance of electrophilic catalysis by glutamic acid 317. Biochemistry 34:2777-2787.

Miyauchi K, Suh SK, Nagata Y and Takagi M (1998). Cloning and sequencing of a 2,5-dichlorohydroquinone reductive dehalogenase gene whose product is involved in degradation of gamma-hexachlorocyclohexane by *Sphingomonas paucimobilis*. J Bacteriol 180:1354-1359.

Myers JK and Pace NC (1996). Hydrogen bonding stabilises globular proteins. Biophys J 71:2033-2039.

Navaza J (1994). AMoRe: an automated package for molecular replacement. Acta Crystallogr A50:157-163.

Neidhart DJ, Kenyon GL, Gerlt JA and Petsko GA (1990). Mandelate racemase and muconate lactonizing enzyme are mechanistically distinct and structurally homologous. Nature 347:692-694.

Neidhart DJ, Howell PL, Petsko GA, Powers VM, Li R, Kenyon GL and Gerlt JA (1991). Mechanism of the reaction catalyzed by mandelate racemase. 2. crystal structure of mandelate racemase at 2.5-Å resolution: identification of the active site and possible catalytic residues. Biochemistry 30:9264-9273.

Newman J, Peat TS, Richard R, Kan L, Swanson PE, Affholter JA, Holmes IH, Schindler JF, Unkefer CJ and Terwilliger TC (1999). Haloalkane dehalogenase: Structure of a *Rhodococcus* enzyme. Biochemistry 38:16105-16114.

Ngai K-L, Ornston LN and Kallen RG (1983). Enzymes of the β-ketoadipate pathway in *Pseudomonas putida*: kinetic and magnetic resonance studies of the *cis,cis*-muconate cycloisomerase catalyzed reaction. Biochemistry 22:5223-5230.

Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J and Goldman A (1992). The α/β-hydrolase fold. Protein Eng 5:197-211.

O'Shea EK, Klemm JD, Kim PS and Alber T (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science 254:539-544.

Otwinowski Z and Minor W (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276:307-326.

Pace NC (2001). Polar group burial contributes more to protein stability than nonpolar group burial. Biochemistry 40:310-313.

Pace NC, Horn G, Hebert EJ, Bechert J, Shaw K, Urbanikova L, Scholtz JM and Sevcik J (2001). Tyrosine hydrogen bonds make a large contribution to protein stability. J Mol Biol 312:393-404.

Packter NM and Ward AC (1972). Biosynthesis of fatty acids and phenols by stationary-phase cultures of *Aspergillus fumigatus*. Biochem. J 127:14-15.

Palmer DR, Hubbard BK and Gerlt JA (1998). Evolution of enzymatic activities in the enolase superfamily: partitioning of reactive intermediates by (D)-glucarate dehydratase from *Pseudomonas putida*. Biochemistry 37:14350-14357.

Palmer DR, Garrett JB, Sharma V, Meganathan R, Babbitt PC and Gerlt JA (1999). Unexpected direvenge of enzyme function and sequence: "N-acylamino acid racemase" is *o*-succinylbenzoate synthase.

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

Perrakis A, Morris R and Lamzin VS (1999). Automated protein model building combined with iterative structure refinement. Nat Struct Biol 6:458-463.

Perutz MF and Raidt H (1975). Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. Nature 255:256-259.

Pham TN, Koide A and Koide S (1998). A stable single-layer β-sheet without a hydrophobic core. Nat Struct Biol 5:115-119.

Powers VM, Koo CW, Kenyon GL, Gerlt JA and Kozarich JW (1991). Mechanism of the reaction catalyzed by mandelate racemase. 1. Chemical and kinetic evidence for a two-base mechanism. Biochemistry 30:9255-9263.

Poyner RR, Laughlin LT, Sowa GA and Reed GH. (1996). Toward identification of acid/base catalysts in the active site of enolase: comparison of the properties of K345A, E168Q, and E211Q variants. Biochemistry 35:1692-1699.

Poyner RR, Cleland WW and Reed GH (2001). Role of metal ions in catalysis by enolase: an ordered kinetic mechanism for a single substrate enzyme. Biochemistry 40:8009-8017.

Quiocho FA, Sack JS and Vyas NK (1987). Stabilization of charges on isolated ionic groups sequestered in proteins by polarized peptide units. Nature 329:562-564.

Raleigh DP, Bezt SF and DeGrado WF (1995). *De novo* designed protein mimics the native state of natural proteins. J Am Chem Soc 117:7558-7669.

Rashin AA, Iofin M and Honig B (1986). Internal cavities and buried waters in globular proteins. Biochemistry 25:3619-3625.

Rashin AA, Rashin BH, Rashin A and Abagyan R (1997). Evaluating the energetics of empty cavities and internal mutations in proteins. Protein Sci 6:2143-2158.

Rashin AA and Honig B (1984). On the enviroment of ionizable groups in globular proteins. J Mol Biol 173:515-521.

Richard JP and Amyes TL (2002). Proton transfer at carbon. Curr Opin Chem Biol 5:626-633.

Regan L and DeGrado W (1988). Characterization of a helical protein designed from first principles. Science 241:976-978.

Roseman MA (1988). Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. J Mol Biol 200:513-522.

Rubinson KA, Ladner JE, Tordova M and Gilliland GL (2000). Cryosalts: suppression of ice formation in macromolecular crystallography. Acta Crystallogr D56:996-1001.

Schmidt E and Knackmuss H-J (1980). Chemical structure and biodegradability of halogenated aromatic compounds: Conversion of chlorinated muconic acids into maleoylacetic acid. Biochem. J 192:339-347.

Schlömann M, Fischer P, Schmidt E and Knackmuss HJ (1990). Enzymatic formation, stability, and spontaneous reactions of 4-fluoromuconolactone, a metabolite of the bacterial degradation of 4-fluorobenzoate. J Bacteriol 172:5119-5129.

Schlömann M (1994). Evolution of chlorocatechol catabolic pathways. Conclusions to be drawn from comparisons of lactone hydrolases. Biodegradation 5:301-321.

Schutz CN and Warshel A (2001). What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins 44:400-417.

Serrano L, Horovitz A, Avron B, Bycroft M and Fersht AR (1990). Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double mutant cycles. Biochemistry 29:9343-9352.

Shoichet BK, Baase WA, Kuroki R and Matthews BW (1995). A relationship between protein stability and protein function. Proc Natl Acad Sci USA 92:452-456.

Simpson A, Bateman O, Driessen H, Lindley P, Moss D, Mylvaganam S, Narebor E and Slingsby C (1994). The structure of avian eye lens delta-crystallin reveals a new fold for a superfamily of oligomeric enzymes. Nat Struct Biol 1:724-34.

Smart OS, Goodfellow JM and Wallace BA (1993). The pore dimensions of gramicidin A. Biophys J 65:2455-60.

Solyanikova IP, Maltseva OV, Vollmer MD, Golovleva LA, Schlomann M (1995). Characterization of muconate and chloromuconate cycloisomerase from *Rhodococcus erythropolis* 1CP: indications for functionally convergent evolution among bacterial cycloisomerases. J Bacteriol 177:2821-2826.

Spassov VZ, Karshikoff AD and Ladenstein R (1995). The optimization of protein-solvent interactions: thermostability and the role of hydrophobic and electrostatic interactions. Protein Sci 4:1516-1527.

Spector S, Wang M, Carp SA, Robblee J, Hendsch ZS, Fairman R, Tidor B and Raleigh DP (2000). Rational modification of protein stability by the mutation of charged surface residues. Biochemistry 39:872-879.

Stites EW, Gittis GA, Lattman EE and Shortle D (1991). In a staphylococcal nuclease mutant the side-chain of a lysine replacing valine 66 is fully buried in the hydrophobic core. J Mol Biol 221:7-14.

Stubbe J and Abeles RH (1980). Mechanism of action of enolase: effect of the β-hydroxyl group on the rate of dissociation of the α-carbon hydrogen bond. Biochemistry 19:5505-5512.

Stura E and Wilson IA (1992). Seeding techniques. pp. 99-126 *in* Crystallisation of nucleic acids and proteins. A practical approach. Ducruix and Giege (eds.). IRL PRESS, Oxford university press, Oxford, Great Britain.

Tanford C (1970). Protein denaturation. Adv Prot Chem 24:1-98.

Thompson TB, Garrett JB, Taylor EA, Meganathan R, Gerlt JA and Rayment I (2000). Evolution of enzymatic activity in the enolase superfamily: structure of *o*-succinylbenzoate synthase from *Escherichia coli* in complex with Mg2+ and *o*-succinylbenzoate. Biochemistry 39:10662-10676.

Tiedje JM, Quensen JF 3rd, Chee-Sanford J, Schimel JP and Boyd SA (1993). Microbial reductive dechlorination of PCBs. Biodegradation 4:231-240.

Tissot AC Vuilleumier S and Fersht AR (1996). Importance of two buried salt bridges in the stability and folding pathway of barnase. Biochemistry 35:6786-6794.

Toth EA and Yeates TO (2000). The structure of adenylosuccinate lyase, an enzyme with dual activity in the *de novo* purine biosynthetic pathway. Structure Fold Des 8:163-74.

Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL and Clardy J (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J Mol Biol 229:105-124.

Varghese JN, Laver WG and Colman PM (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. Nature 303:35-40.

Vaughan CK, Harryson P, Buckle AM and Fersht AR (2002). A structural double-mutant cycle: estimating the strength of a buried salt bridge in barnase. Acta Crystallogr D58:591-600.

Verschueren KH, Seljee F, Rozeboom HJ, Kalk KH and Dijkstra BW (1993). Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. Nature 363:693-698.

Vetriani C, Maeder DL, Tolliday N, Yip KS, Stillman TJ, Britton KL, Rice DW, Klump HH and Robb FT (1998). Protein thermostability above 100 °C: a key role for ionic interactions. Proc Natl Acad Sci USA 95:12300-12305.

Vollmer MD, Fischer P, Knackmuss HJ and Schlomann M (1994). Inability of muconate cycloisomerases to cause dehalogenation during conversion of 2-chloro-*cis,cis*-muconate. J Bacteriol 176:4366-4375.

Vollmer MD, Hoier H, Hecht H-J, Schell U, Gröning J, Goldman A and Schlömann M (1998). Substrate specificity and product formation of muconate cycloisomerases: an analysis of wild type enzymes and engineered variants. Appl Env Microbiol 64:3290-3299.

Vollmer MD, Schell U, Seibert V, Lakner S and Schlömann M (1999). Substrate specificities of the chloromuconate cycloisomerases from *Pseudomonas* sp. B13, *Ralstonia eutropha* JMP134 and *Pseudomonas* sp. P51. Appl Microbiol Biotechnol 51:598-605.

Vollmer MD and Schlömann M (1995). Conversion of 2-chloro-*cis,cis*-muconate and its metabolites 2-chloro- and 5-chloromuconolactone by chloromuconate cycloisomerases of pJP4 and pAC27. J Bacteriol 177:2938-2941.

Waldburger CD, Schildbach JF and Sauer RT (1995). Are buried salt bridges important for protein stability and conformational specificity? Nature Struct Biol 2:122-128.

Warshel A (1978). Energetics of enzyme catalysis. Proc Natl Acad Sci USA 75:5250-5254.

Warshel A (1981). Calculations of enzymatic reactions: calculations of pKa, proton transfer reactions, and general acid catalysis reactions in enzymes. Biochemistry 20:3167-3177.

Warshel A and Papazyan A (1996). Energy considerations show that low-barrier hydrogen bonds do not offer a catalytic advantage over ordinary hydrogen bonds. Proc Natl Acad Sci USA 93:13665-13670.

Warshel A and Papazyan A (1998). Electrostatic effects in macromolecules: fundamental concepts and practical modeling. Curr Opin Struct Biol 8:211-217.

Warshel A, Russell ST and Churg AK (1984). Macroscopic models for studies of electrostatic interactions in proteins: limitations and applicability. Proc Natl Acad Sci USA 81:4785-4789.

Warshel A, Åqvist J and Creighton S (1989). Enzymes work by solvation substitution rather than by desolvation. Proc Natl Acad Sci U S A 86:5820-5824.

Weeks CM and Miller R (1999). Design and implementation of SnB vs. 2.0. J Appl Cryst 32:120-124.

Wedekind JE, Poyner RR, Reed GH and Rayment I (1994). Chelation of serine 39 to Mg2+ latches a gate at the active site of enolase: structure of the bis(Mg2+) complex of yeast enolase and the intermediate analog phosphonoacetohydroxamate at 2.1 Å resolution. Biochemistry 33:9333-9342.

Weaver TM, Levitt DG, Donnelly MI, Stevens PP and Banaszak LJ (1995). The multisubunit active site of fumarase C from *Escherichia coli*. Nat Struct Biol 2:654-62.

Wimley WC, Creamer TP and White SH (1996a). Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. Biochemistry 35:5109-5124.

Wimley WC, Gawrisch K, Creamer TP and White SH (1996b). Direct measurement of salt-bridge solvation energies using a peptide model system: Implications for protein stability. Proc Natl Acad Sci USA 93:2985-2990.

Williams MA, Goodfellow JM, and Thornton JM (1994). Buried waters and internal cavities in monomeric proteins. Protein Sci 3:1224-1235.

Williams SE, Woolridge EM, Ransom SC, Landro JA, Babbitt PC and Kozarich JW (1992). 3-carboxy-cis,cis-muconate lactonizing enzyme from *Pseudomonas putida* is homologous to the class II fumarase family: a new reaction in the evolution of a mechanistic motif. Biochemistry 31:9768-9776.

Williams JC, Zeelen JP, Neubauer G, Vriend G, Backmann J, Michels PAM, Lambeir A-M and Wierenga RK (1999). Structural and mutagenesis studies of *Leishmania* triose phosphate isomerase: a point mutation can covert a mesophilic enzyme into a superstable enzyme without losing catalytic power. Protein Eng 12:243-250.

Wittich RM (1998). Degradation of dioxin-like compounds by microorganisms. Appl Microbiol Biotechnol 49:489-499.

Wohlfarth G and Diekert G. (1997). Anaerobic dehalogenases. Curr Opin Biotechnol 8:290-295.

Xiao L and Honig B (1999). Electrostatic contributions to the stability of hyperthermophilic proteins. J Mol Biol 289:1435-1444.

Yip KS, Stillman TJ, Britton KL, Artymiuk PJ, Baker PJ, Sedelnikova SE, Engel PC, Pasquo A, Chiaraluce R and Consalvi V (1995). The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure 3:1147-1158.

Zhang E, Brewer JM, Minor W, Carreira LA and Lebioda L (1997). Mechanism of enolase: the crystal structure of asymmetric dimer enolase-2-phospho-D-glycerate/enolase-phosphoenolpyruvate at 2.0 Å resolution. Biochemistry 36:12526-12534.

Zhang X, Meining W, Fischer M, Bacher A and Ladenstein R (2001). X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile *Aquifex aeolicus* at 1.6 Å resolution: determinants of thermostability revealed from structural comparisons. J Mol Biol 306:1099-114.