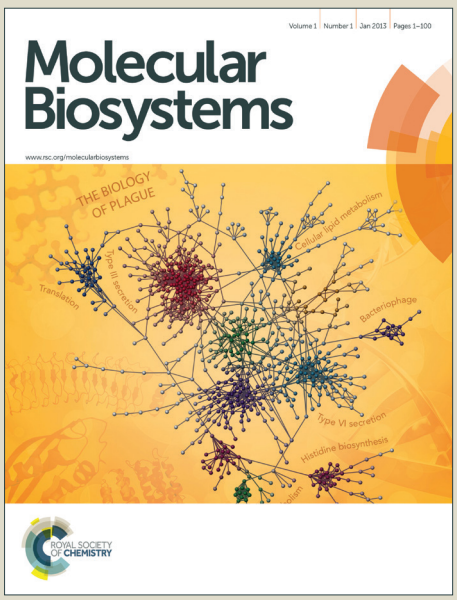
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Journal Name

RSCPublishing

ARTICLE

Role of a Remote Leucine Residue in the Catalytic Function of Polyol Dehydrogenase

Cite this: DOI: 10.1039/xoxxoooooox

Manish Kumar Tiwari,[a,‡] Vipin C. Kalia,[c] Yun Chan Kang,*[,b] Jung-Kul Lee*[,a]

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

Studies on the protein-metal binding sites have mainly focused on the residues immediately surrounding the reacting substrate, cofactors, and metal ions. The contribution of residues in remote layers to the highly optimized microenvironments of catalytic active sites is not well understood. To improve our understanding, the present study examined the role of remote residues on the structure and function of zinc-dependent polyol dehydrogenases. We used an integrated computational and biochemical approach to determine the role of L136 in the third shell of the L-arabinitol 4-dehydrogenase (LAD) from *Neurospora crassa*. Substitution of L136 with charged (Asp, Lys, or His) and bulky (Trp) side chain amino acids abolished enzyme activity. Whereas the L136A mutant exhibited a 95% decrease in catalytic efficiency (k_{cat}/K_m), the L136C mutant exhibited a 39% decrease in k_{cat}/K_m. Additionally, molecular docking and dynamic simulations on the mutant (L136A, L136C, L136H, and L136P) complexes showed the loss of crucial H-bonds between G77 and the corresponding mutated residue. It is evident from theoretical and biochemical studies that the L136 is part of the extensive hydrogen bonding network coupled to the reaction catalyzed at the active site. We propose that L136, critically positioned behind the active site residues H78 and E79 in the third shell of LAD, plays a crucial role in modulating catalysis or substrate binding by stabilizing the GHE motif in the LAD active site.

Introduction

Proteins set a gold standard for efficiency and selectivity that few other natural or artificial molecules can match. Consequently, the study of protein structure and function has been the focus of research for many years.[1-4] Over the past decades, visualization of the three dimensional structures of enzymes has revealed that active sites are located in crevices or pockets.[5, 6] Within these active sites are found coenzymes, cofactors, metal ions, and side chains that participate in a rich array of chemical reactions.[7] Previous studies established that active site pockets can be regarded as multiple layers of residues.[8, 9] Residues that interact directly with substrate, cofactors, and metal ions also interact with residues in a second shell. These, in turn, interact with residues in a third shell. Enzymes evolve new functions via mutations that affect the active site, its periphery, or even the second and third shell of residues around it, while leaving key catalytic residues intact.[10] Enzyme active sites provide a highly optimized microenvironment for the catalysis of biologically useful chemical transformations. The precisely positioned active site residues promote catalysis by providing reactive groups, such as nucleophiles or acids/bases, and by providing a microenvironment that stabilizes the transition state.[11] Do these highly optimized microenvironments arise solely from the first coordination sphere (the residues immediately surrounding the reacting substrate, cofactors and metal ions), or do remote residues in the second and third coordination spheres also contribute? Previous studies have focused predominantly on first-shell amino acid residues.[12, 13] They have shown that direct transition metal-ligand interactions have a profound effect on protein-metal affinity and metalloprotein function.[14] In recent years, focus has shifted to the role of the second-

shell residues in metalloprotein function.[9, 15-18] In our previous study, we proposed that the strictly conserved second-shell residue G77 in medium-chain dehydrogenase/reductase (MDR) enzymes provided the conformational flexibility required by the enzyme active site and maintained the optimum geometry of the active site during the MDR reaction.[9] These studies indicate that residues outside the active site can modulate protein function. Indirect ligands, i.e., second-shell residues that H-bond to direct metal ligands typically have more subtle effects on the chemical properties of a protein-metal complex.[14, 15] Although the studies described above identified some roles for second-shell residues in protein-metal recognition and catalysis, few remote residues have been identified, and the extent to which remote residues participate in catalysis has not been extensively investigated.

Zinc metalloenzymes, such as the MDR superfamily members L-arabinitol 4-dehydrogenase (LAD),[9] carbonic anhydrase II,[19] carboxypeptidase A,[20] and thermolysin,[21, 22] generally have a water molecule (solvent) bound to the zinc ion. H-bonding networks with zinc ligands can modulate the p*K*a and reactivity of the zinc-bound water molecule,[23, 24] which is implicated in the catalytic mechanism of the enzyme. These H-bond networks with metal ligands could be important for the structure and function of metalloenzymes. Identifying remote residues and understanding their interactions during enzymatic catalysis could be a step toward the systematic modulation of the properties of catalytic residues,[25] which would facilitate the structure-assisted design of protein-metal binding sites for better control of catalysis and substrate recognition in polyol dehydrogenases.

The objective of the present paper was to identify the role of remote residues that influence the catalytic activity of a polyol dehydrogenase, LAD. To elucidate the physical properties of the H-bonding interactions in the LAD active site, we constructed site-directed variants of LAD. Furthermore, we used molecular mechanical (MM) calculations to analyze changes in the H-bonding pattern in the variants. The role of the remotely located residue L136 was determined, using *in vitro* and *in silico* methods. A single mutation at position 136 in the third shell of LAD decreased enzyme activity, suggesting that both direct residues and indirect residues involved in the H-bond network must be included in the structure-assisted design of protein-metal binding sites.

Results and discussion

Selection of a mutant

The structure and chemical composition of the active site cavity in metalloenzymes, including the H-bond network, likely modulate the properties of the metal binding site.[15] Metalloenzyme active sites are regarded as layers of residues. The donor atoms (N, O or S) of the first layer of amino acid residues were located within 2.9 Å of the distance[13] from the catalytic zinc in the NcLAD active site. Followed by a 3.5 Å cutoff distance[26-28] from the heavy atoms of the first-layer residues were used to locate a second layer of amino acid residues in the substrate-binding pocket. Finally, a maximum distance of 10 Å from the catalytic zinc center was used to locate residue remote layers. Residues that interact directly with the substrate also interact with residues in a second shell. These, in turn, interact with residues in a third shell, creating a H-bond network.[8]

We previously observed that the introduction of a point mutation at position G77, a second-shell residue in the active site layer of NcLAD, abolished enzyme activity.[9] Further analysis of the NcLAD substrate binding pocket revealed that one particular residue, L136, lying in the third shell of the NcLAD substrate binding pocket, interacted with the crucial G77 via a H-bond (Fig. 1). G77 is strictly conserved in the MDR superfamily and is essential for stabilizing the critical Zn ion at the catalytic site. Multiple sequence alignment of the NcLAD (PDB entry 3M6I) protein sequence with the sequences of homologues enzymes from the NCBI

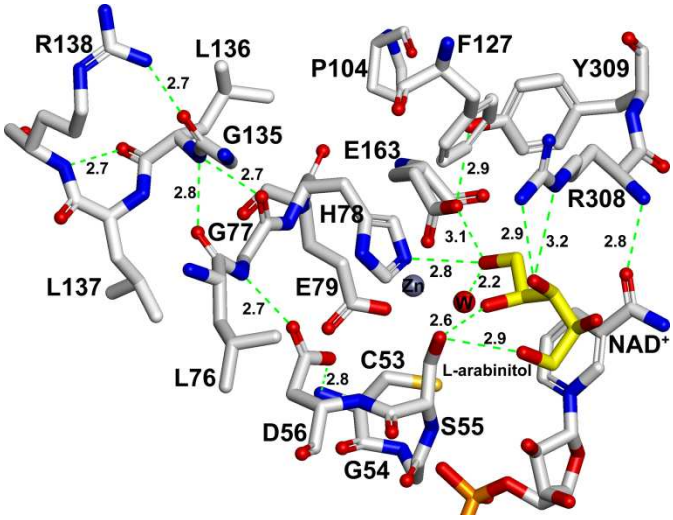


Fig. 1 Molecular docking of L-arabinitol to wild-type NcLAD. Binding interaction of L-arabinitol in the NcLAD active site. L-arabinitol (yellow carbon), NAD^+, and the amino acid residues in the NcLAD binding pocket are represented with a stick model. Active site zinc and water molecules are shown as spheres.

Table 1 Kinetic parameters of the L136 mutants

Enzyme	V_{max} (mM min^{-1})	K_m (mM)		k_{cat} (min^{-1})	k_{cat}/K_m, L-arabinitol ($mM^{-1}min^{-1}$)
		Cofactor	Substrate		
Wild-type	42	0.9	16	1640	102
L136A	2.8	0.6	22	109	5.0
L136C	24	0.8	15	936	62.4
L136I	8.69	1.3	18	339	18.8
L136M	11.6	1.1	28	451	16.1
L136V	6.8	0.79	22	265	12.0
L136W	N/D	N/D	N/D	N/D	N/D
L136K	N/D	N/D	N/D	N/D	N/D
L136E	N/D	N/D	N/D	N/D	N/D
L136H	N/D	N/D	N/D	N/D	N/D
L136P	N/D	N/D	N/D	N/D	N/D

NAD^+ and L-arabinitol were used as cofactor and substrate, respectively, with NcLAD or HjLAD. Results are the mean ± S.D. from three experiments. ND, not determined because of very low enzyme activity.

non-redundant database revealed that L136 was strictly conserved in 45 polyol dehydrogenase sequences (Fig. S1). Because of the H-bonding interaction, the strict conservation of L136 in polyol dehydrogenases, and the significance of remotely located residues in metalloproteins,[15, 16, 29] the role of L136 was further investigated by thorough site-directed mutagenesis. To probe the function of L136, we constructed a series of LAD variants with substitutions at position 136.

Site-directed mutagenesis and kinetic analysis of NcLAD wild-type and variants

A number of L136 mutants were constructed to delineate the functional role of the L136 residue. The mutants, including L136A, L136C, L136I, L136E, L136V, L136M, L136K, L136W, L136H, and L136P, were expressed in *E. coli* and purified to homogeneity, as assessed by SDS-PAGE (Fig. S2). Analysis of activity using the substrate L-arabinitol and comparison with wild-type NcLAD showed that mutation of L136 markedly decreased activity or abolished catalysis. Almost no measurable activity was observed when L136 was substituted with a charged amino acid (Asp, Lys, or His). Other neutral, non-polar, and aromatic substitutions (Ala, Ile, Val, Met, and Trp) at position L136 resulted in the loss of ~60% – 80% of the activity observed with the wild-type enzyme at saturating substrate concentrations. In contrast, only ~20% of the total wild-type enzyme activity was lost in the L136C mutant. Notably, similar results were observed when another enzyme in the polyol dehydrogenase family, HjLAD, was mutated at the position equivalent to L136 in NcLAD (L149A, L149C, L149I, L149E, L149V, L149M, L149K, L149W, L149H, and L149P) (Table S1). Presumably, this conserved leucine has a common role in polyol dehydrogenase enzymes. The role of L136 was further investigated through biochemical, biophysical, and computational analysis of two active mutants (L136C and L136A) and inactive mutants (L136H and L136P).

To understand the basis of the variations in the specific activity of the mutant enzymes, the kinetic constants k_{cat} and K_m were determined, as described in the Materials and Methods section. Of the 10 enzyme variants studied, five had detectable activity. The remaining five variants (L136E, L136K, L136H, L136W, and L136P) were almost inactive. Detailed kinetic parameters were impossible to determine because of the extremely low activity of the L136H and L136P variants (Table S1). L136C exhibited a 43% decrease in k_{cat} and a 39% decrease in k_{cat}/K_m. Another active mutant, L136A, exhibited a 93% decrease in k_{cat} and a 95% decrease in k_{cat}/K_m (Table 1).

Circular dichroism spectra of the wild-type and mutant enzymes

To examine how substitution at the remote residue L136 affected protein conformation globally, far UV-CD spectra were measured for four variants (L136A, L136C, L136H, and L136P) (Fig. 2A). Of the four variants, L136A and L136C exhibited CD spectra that were very similar, or identical, to that of wild-type protein, suggesting that these variants folded without significant alterations in secondary structure. However, L136H and L136P generated spectra were slightly different from those of the wild-type protein (Fig. 2A). The CD curves of L136H and L136P mutants exhibited changes in negative ellipticity at 209 and 222 nm,

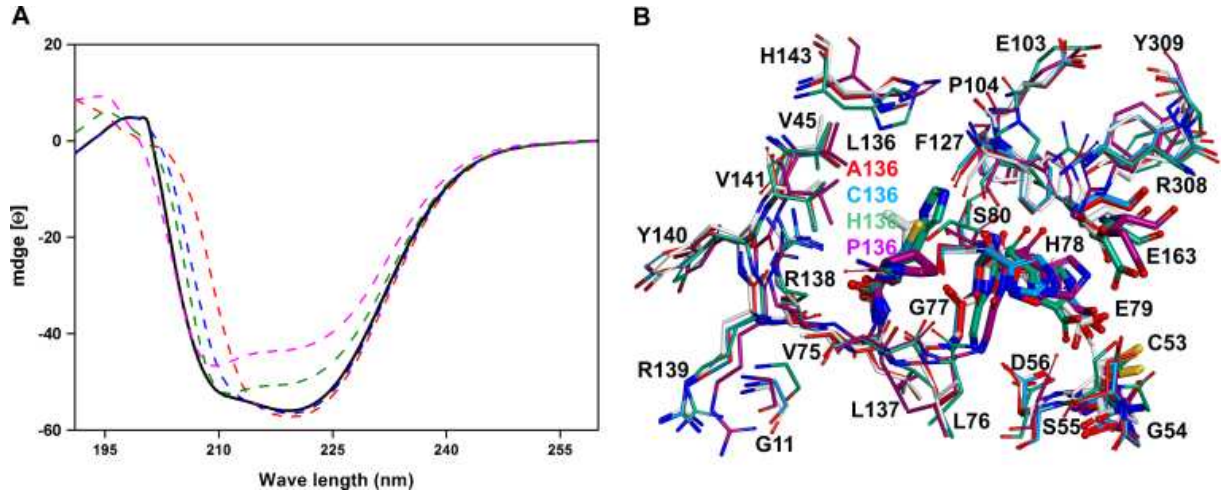


Fig. 2 Effect of point mutations on the secondary structure of NcLAD. (A) The anomalous circular dichroism spectra of the wild-type and mutant NcLADs. The CD spectra of wild-type (black line), L136A (red dash), L136C (blue dash), L136H (green dash), and L136P (pink dash) enzymes were recorded at 20 °C using a J-815 spectrophotometer. (B) Detailed view of the aligned active sites of wild-type NcLAD and the mutants (L136A, L136C, L136H, and L16P). Residues in the wild-type (L136) enzyme are shown in white, and the corresponding resides in the mutants are shown in red, blue, green, and pink, respectively.

suggesting a geometry shift of the polypeptide backbone.[30, 31] Superimposing the amino acid residues within 10Å of the substrate binding pocket of wild-type NcLAD and the L136A, L136C, L136H, and L136P mutants yielded a local conformational shift in the protein backbone with an RMSD value of 0.64 Å, 0.70 Å, 1.0 Å, and 1.3 Å, respectively (Fig. 2B). Furthermore, we investigated the effect of point mutation on the globular structure and secondary structure elements of wild-type and mutant proteins (Fig. S3). All the structures were superimposed by C-alpha pairs, where the wild-type was set as a reference structure, which suggests that the two inactive mutants, namely L136H and L136P, experience significant changes in the overall globular geometry (Fig. S3A) and altered secondary structure elements (Fig. S3B). In addition, we provide the RMSD values for the globular geometries of all the superimposed structures (Table S2). These RMSD values are in accordance with the UV-CD spectra of the wild-type and mutant enzymes (Fig. 2A).

It has been long known that intramolecular H-bonds are a ubiquitous and essential feature of the protein structure. As per Stickle et al., the backbone H-bonds are the most prevalent (68.1%) in the protein structure.[32] Upon the substitution of L136 with various residues (Table S1), the loss or complete abolishment of enzyme activity was observed. A further structural analysis (Table S2, Fig. S3) of the mutants of interest (L136C, L136A, L136H, and L136P) revealed that due to the induction of point mutations at position 136, there is a significant conformational shift in the backbone mutant enzymes. This observation has been further supported by a UV-vis CD analysis (Fig. 2A). Therefore, we propose that the remote residue L136 also plays a role in NcLAD structural stability to some extent by communicating through intervening residues.

Molecular dynamics simulations

Molecular dynamics simulations were carried out on five systems (wild-type, L136A, L136C, L136H, and L136P), using the "dynamic simulation cascade" module of DS3.5 with the CHARMm force field. The default protonation state of His in DS3.5 was adopted. To clarify the dynamic stability of the five complexes, the RMSD values of the protein backbone atoms were calculated from the starting snapshot and plotted (Fig. 3). The RMSD of the backbone atoms in the five complexes increased sharply within 500 ps and then remained stable to the end of the simulation. When the average RMSD values of the five systems were compared, the four mutant complexes (1.7 Å, 1.6 Å, 1.8 Å, and 2.2 Å for L136A, L136C, L136H, and L136P, respectively) were less stable than the wild-type complex (1.4 Å) during the MD simulations (Fig. 3A). Because of the instability, active site residues such as C53, H78, E79, and E163 in the NcLAD mutants lost their close contact with L-arabinitol and the orderly

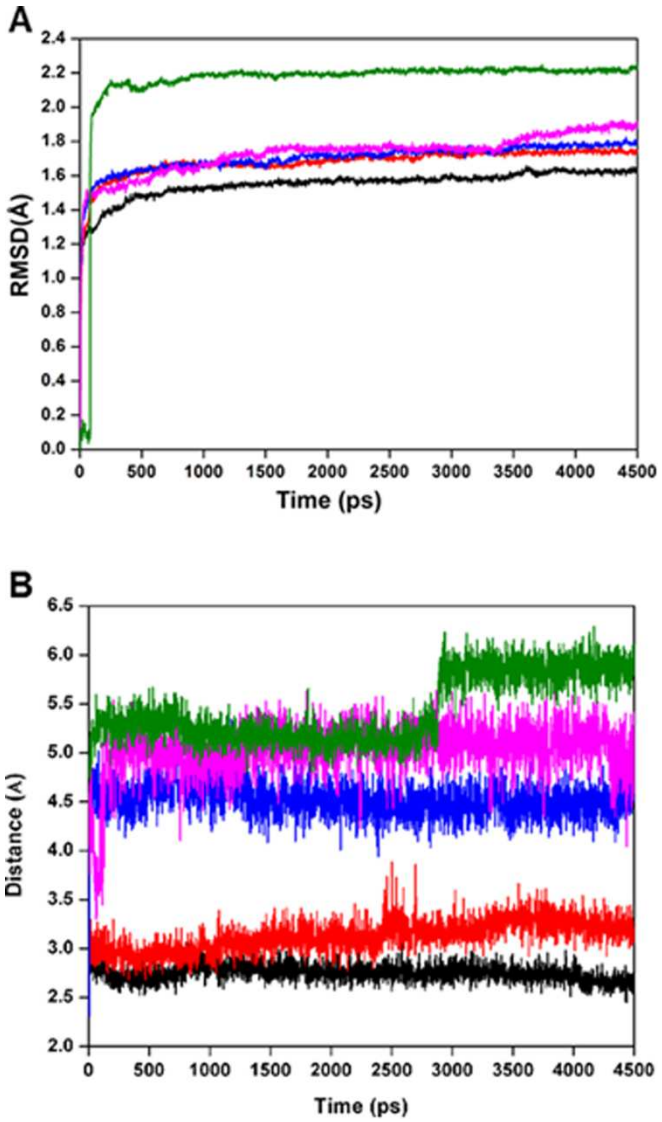


Fig. 3 Molecular dynamic simulations on NcLAD wild-type and mutant structures. (A) The root-mean-square displacements (RMSD) of the backbone atoms (Cα, N, C) of the complexes with respect to the first snapshot as a function of time. Wild type (black), L136C (red), L136A (blue), L136H (pink), and L136P (green) are shown. (B) Distances between atom N of L136 and the oxygen atom of G77 in the wild-type and mutant systems. Wild type (black), L136C (red), L136A (blue), L136H (pink), and L136P (green) are shown.

interactions observed originally dissociated. During the simulation, the distances between the nitrogen atom N of L136 and the oxygen atom of G77 in the wild-type system and the corresponding distance in the mutant systems were monitored (Fig. 3B). In the wild-type complex, the distance between the nitrogen atom of L136 and the oxygen atom of G77 fluctuated around 2.7 Å during the simulation, indicating the presence of a stable H-bond between L136 and G77. In comparison, the distance between the nitrogen atom of C136 and the oxygen atom of G77 increased slightly in the L136C mutant, to approximately 3.0 Å. In contrast, the distances in the L136A, L136H, and L136P complexes were approximately 4.5 to 6.5 Å, indicating the absence of a H-bond between the nitrogen atoms of the mutated residues and G77 (Fig. 3B).

Binding modes of the L-arabinitol-LAD complex

To understand the influence of the mutations, the binding modes of the five complexes (wild-type, L136A, L136C, L136H, and L136P) with L-arabinitol were investigated. The C1-OH and C2-OH atoms of the linear molecule of docked L-arabinitol were directly coordinated to the zinc atom. The orientation of L-arabinitol accounts for the polarization of C2 by the zinc atom, which is likely to facilitate hydride transfer from C2 to NAD^+, promoting the oxidation of L-arabinitol to L-xylulose.[33, 34] The MD calculation revealed that all of the hydroxyl groups of L-arabinitol were within H-bonding distance of the catalytically important residues C53, H78, E79, and E163 (Fig. 1). These interactions suggest that H-bonds[35] between the surrounding residues and L-arabinitol or a water molecule provide a path for the hydride transfer during the course of L-arabinitol oxidation.[34]

A hydrophobic cleft between the catalytic and coenzyme-binding domains is a common feature of MDR alcohol dehydrogenase enzymes.[36] A similar hydrophobic cleft is also observed in NcLAD during our structural analysis (Fig. S4). However, L-arabinitol binds to the NcLAD active site through numerous H-bonds, and it nestles tightly in the inner groove formed by residues C53, S55, D56, L76, G77, H78, E79, E163, R308, and Y309 in wild-type NcLAD (Fig. 1). In wild-type NcLAD, C1-OH and C2-OH of L-arabinitol are directly involved in the interaction with the active site residues H78 and E163 via H-bonds. C3-OH forms two H-bonds with the guanidine group of residue R308. S55 also forms two H-bonds with C2-OH and C5-OH. In addition to active site and substrate interactions, other residues lying in the second and third shell of the NcLAD binding pocket form several intra-molecular H-bonds, i.e., D56 H-bonds with active site residues C53 and G77, and the OH group of the Y309 side chain links to E613 via a single H-bond. In addition to these interactions, nitrogen atom N of L136 and the H atoms attached to it formed two stable H-bonds with oxygen atoms O1A and O1B of G77 and L76, respectively. This extensive H-bond network in wild-type NcLAD (Fig. 1) could stabilize the substrate binding motif of NcLAD. On the other hand, substitution of L136 with C136 or

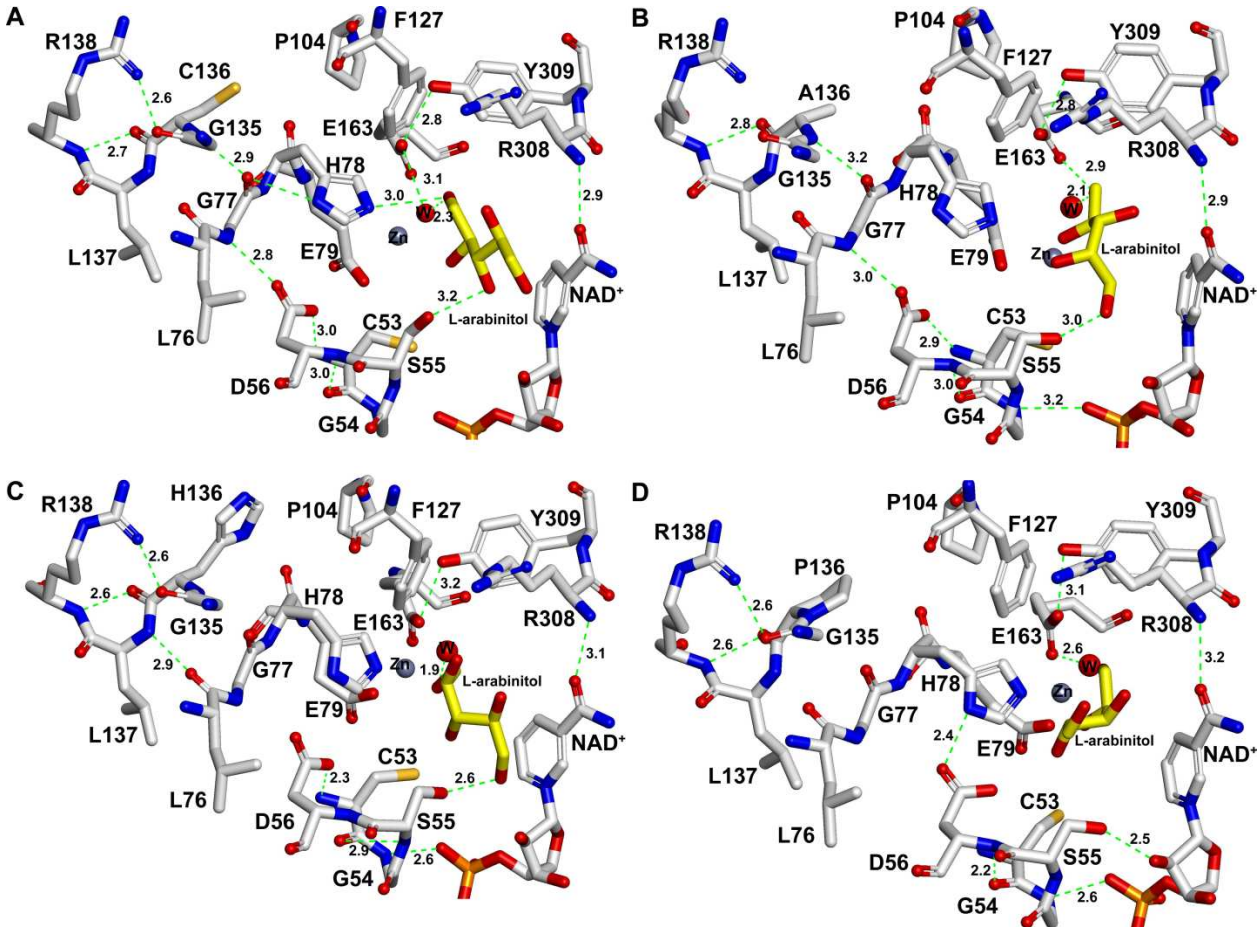


Fig. 4 Molecular docking of L-arabinitol with four NcLAD mutants. (A) L136C. (B) L136A. (C) L136H. (D) L136 P. L-arabinitol interaction in the NcLAD active site of NcLAD mutants. L-arabinitol (yellow carbons), NAD^+, and amino acid residues in the NcLAD mutants binding pocket are represented with a stick model. H-bonds are represe nted with green dash and distances are in Å.

A136 resulted in the loss of one of the two H-bonds between G77 and the corresponding mutated residue. The length of the remaining H-bond increased slightly to 2.9Å and 3.2Å for the C136 and A136 mutants, respectively (Fig. 4A and 4B). In contrast, both H-bonds were lost in the L136H and L136P mutants. The nitrogen atoms of H136 and P136 did not form H-bonds with G77 (Fig. 4C and 4D).

G77, a conserved residue in the catalytic GHE motif of the MDR superfamily, may maintain the metal binding affinity and the electronic state of the catalytic Zn ion during MDR enzyme catalysis.[9] The extensive H-bond network in the NcLAD wild-type active site (Fig. 1) stabilizes the substrate binding pocket. Mutation of residue L136 to C136 or A136 resulted in the loss of H-bonding between G77 and the corresponding mutated residue (Fig. 4A and 4B). Strikingly, both H-bonds were lost in the L136H and L136P mutants (Fig. 4C and 4D). The loss of these crucial H-bonds, followed by a change in the relative positions of the active site residues C53, H78, E79, and E163 (Fig. 2B), changed the overall geometry of the substrate binding pocket. Therefore, an unfavorable L-arabinitol binding conformation was observed in the mutant enzymes, along with the loss of enzymatic activity (Table S1), when compared to wild-type NcLAD (Fig. 1).

Previous studies focused on zinc-metalloenyzmes, such as human sorbitol dehydrogenase (hSDH),[37] Glucose dehydrogenase (GlcDH),[38] ketose reductase (KR),[39] and carbonic anhydrase II (CAII)[15] have suggested that the key residues at the active site form extensive hydrogen bonds in order to facilitate optimum catalysis. Specifically, theoretical and biochemical studies on CAII have revealed that the indirect residues in remote layers are capable of providing an optimal orientation to the residues directly interacting with metal ions at the active site. This, in turn, enhances the metal affinity and electrostatic

interactions. In fact, it has been suggested that such nested hydrogen bonds between direct (first layer) and indirect (second and third layer) residues at the catalytic interface of metalloenzymes are determinants of reactivity and catalytic efficiency.[15] In line with the significance of the H-bond network, a previous study on *E. coli* dihydrofolate reductase proposed a significant correlation between H-bonding ability and the rate of hydride transfer.[40] Perturbed, H-bonding ability may alter ligand (substrate) affinities, turnover, and the kinetically preferred pathway.[41] Therefore, loss of the H-bonding network in the substrate binding pocket of NcLAD resulting from mutation of L136 could impair the catalytic ability of NcLAD. The third-shell residue L136, with the greatest effect on the catalytic efficiency of NcLAD, is located behind the catalytic residues H78 and E79. This suggests that L136 is involved in the proper positioning of the catalytic residues H78 and E79, which in turn could impart the necessary chemical and electrostatic properties during catalysis (Fig. 1). Thus, L136 is part of a spatially extended H-bond network that facilitates catalytic events at the active site.

Binding free energies predicted by MM/GBSA

To estimate the effect of the L136 mutations on the binding free energies of L-arabinitol, the binding affinities were calculated by the MM/GBSA method. The calculated binding free energy of the wild-type complex (-27.18 kcal/mol) was more favorable than that of the L136C (-22.95 kcal/mol), L136A (-16.21 kcal/mol), L136H (-12.58 kcal/mol), and L136P (-10.98 kcal/mol) mutants. Table 2 lists the individual energy components of the MM/GBSA for the wild-type and four mutant enzymes. The van der Waals and electrostatic terms were the predominant favorable contributions to the binding of L-arabinitol to the LAD active site. The ΔE_{vdw} values of the four mutants were less favorable than that of the wild-type enzyme, by ~2–7 kcal/mol. The ΔE_{ele} term for L136H and L136P was less favorable for binding, which is in agreement with the loss of H-bonding between G77 and H136 or P136 in the L136H and L136P mutants, respectively (Fig. 4C and 4D). The calculated free energy of binding of the substrate was electrostatically driven (Table 2). The electrostatic interactions between LAD and L-arabinitol in the wild-type complex were almost three times stronger than those in the L136H and L136P complexes, while the intermolecular van der Waals interactions for the L136H and L136P complexes were almost the same (Table 2),which indicates that the L136 mutation predominantly weakens the electrostatic interactions between LAD and L-arabinitol (Table 2). This is consistent with the results of the MD simulation and docking analyses described above. The total binding energy ($\Delta G_{binding}$) between LAD and L-arabinitol in the wild-type complex was -27.18 kcal/mol, which is 16.2 kcal/mol lower than that in the L136P complex. This shows that the L136P mutant forms an energetically less favorable complex, resulting in the loss of NcLAD activity, which is consistent with experimental studies.

L136, an aliphatic and hydrophobic amino acid, which is critically positioned behind the GHE motif in the third shell of wild-type NcLAD, interacts via two main chain H-bonds with oxygen atoms of G77 and L76 in the wild-type complex. The substitution of hydrophobic L136 residue into less hydrophobic residues (L136C, L136M, L136A, L136W, L136P, L136H, L136E, and L136K; hydrophobicity of residues in decreasing order) in this study resulted in the loss or total abolishment of enzymatic activity (Table S1). This observation is further supported by the fact that the four mutants analyzed in this study (L136A, L136C, L136H, and L136P) exhibit ~2–7 kcal/mol ΔE_{vdw} unfavorable values than that of the wild-type complex, suggesting a hydrophobic packing defect. However, anomalies in the cases of L136I and L136V mutations were observed, where the L136I and L136V mutants resulted in 39% and 29% relative activity, respectively, compared to the wild-type complex. It suggests that any conformational change in the GHE motif due to relative hydrophobicity and polarity would probably lead to the unfavorable orientation of active site residues. Consequently, the substitution of L136 to charged (L136K, L136E, and L136H), aromatic, and bulky side chains (L136H and L136W), as well as with additional conformational constrains (L136P), might reinforce the local conformational changes at the active site, as seen in the cases of two analyzed mutants (L136H and L136P). It is evident from the L136H and L136P substitutions that the relative positioning of the active site motif (Fig. 2B) and the significant alternations to the secondary elements (Fig. S3) can directly be correlated with total attenuation or the elimination of enzyme activity. Similar observations have also been reported previously, where mutation occurs ~17Å away from the active site, as in the case of *Escherichia coli* dihydrofolate reductase (DHFR), which resulted in significant decreases in the rate of hydride transfer.[40, 42, 43]

To achieve a better understanding how the differences in the binding free energies between wild-type and mutant complexes for L-arabinitol ($\Delta\Delta G_{binding}$) are related, we correlated the relative activities (Table S1) of five enzyme-substrate complexes (wild-type,

Table 2 Binding free energies and individual energy terms of NcLAD with L-arabinitol (kcal/mol)

Enzymes	ΔE_{ele}	ΔE_{vdw}	ΔG_{np}	ΔG_{n}	ΔG_{gb}	$\Delta G_{binding}$	$\Delta\Delta G_{binding}$
Wild-type	-212.62	-17.56	-4.94	201.94	-10.68	-27.18	0
L136C	-205.64	-15.84	-4.74	199.27	-6.37	-22.95	4.23
L136A	-169.79	-15.28	-2.52	166.81	-2.98	-16.21	10.97
L136H	-87.96	-11.25	-1.98	86.95	-1.01	-12.58	14.60
L136P	-85.33	-10.18	-1.72	84.95	-0.38	-10.98	16.21

ΔE_{ele} are electrostatic energies calculated by the MM force field.
ΔE_{vdw} are van der Waals contributions from MM.
ΔG_{np} are nonpolar contributions to the solvation free energy.
ΔG_{gb} are electrostatic contributions to the solvation free energy calculated by GB.
$\Delta G_{binding}$ are binding energies.
$\Delta\Delta G_{binding}$ values are the differences in the binding free energies between wild-type and mutants for L-arabinitol.

L136C, L136A, L136H, and L136P) against their predicted free energies of substrate binding ($\Delta G_{binding;}$ Table 2). It is notable that these two parameters are related in a statistically significant linear relationship (R^2 = 0.97, p = 002; Fig. S5). The predicted relative binding free energies of five pairs of enzyme-substrate complexes of NcLAD showed an excellent correlation with the experimentally observed values (Fig. S5). This observation is in accordance with the loss of binding affinity for L-arabinitol in the case of mutants, as predicted by the free energy of binding calculation. Previous studies also provide support to such a paradigm that binding differences account for enzyme catalysis.[44]

Conclusion

In conclusion, we focused on remote residues in the third shell because these residues, which are not well understood, often have indirect, but significant effects on the structure and function of enzymes. Utilizing the PDB database coordinates for NcLAD, we located L136, which is critical for maintaining the H-bond network in the NcLAD binding pocket at the GHE triad[9] and is strictly conserved in 45 polyol dehydrogenases, as shown by BLAST and multiple sequence alignment using NcLAD as the sequence query.

The abolished kinetic efficiency observed with the L136H and L136P mutants suggests that L136 has a crucial role in the dynamics of the microenvironment in the substrate binding pocket and therefore in enzyme activity. The kinetics results reported here support the prediction that third-shell residues play a supporting role as anchors that help pre-orient and stabilize the GHE motif[9] at the NcLAD active site. To rationalize the role of L136, this study outlined how the L136 amino acid residue contributes to LAD catalysis. A proposed mechanism for the participation of L136 includes (i) shifts in the H-bonding networks and (ii) changes in dynamics that affect the active conformation of the active site. Along with a few examples in the literature, this study clearly demonstrates that long-range effects, including electrostatic effects, influence enzyme catalysis.[45-47] These results suggest that remotely located functional residues that contribute to the H-bond network should be included in structure-assisted protein design.

Experimental

Biochemical methods

Reagents

Unless otherwise stated, all chemicals were from Sigma-Aldrich (St. Louis, MO, USA). Oligonucleotide primers were obtained from Bioneer (Daejeon, South Korea). PCR reagents and Ex-Taq DNA polymerase were purchased from Takara (Takara Inc., Shiga, Japan). Restriction enzymes were obtained from New England BioLabs (Ipswich, MA, USA). The expression vector (pET-28a), plasmid isolation kit, and Ni-NTA Superflow column for protein purification were from Qiagen (Hilden, Germany) and electrophoresis reagents were from Bio-Rad.

Construction, expression, and site-directed mutagenesis of two polyol dehydrogenases

To investigate the general role of L136 in the polyol dehydrogenase family, plasmids containing the wild-type *Neurospora crassa* LAD (pET-28a-NcLAD) and *Hypocrea jecorina* LAD (pET-28a-HjLAD) genes were constructed and expressed as described in detail in our previous studies.[9, 48, 49] Wild-type recombinant plasmids were used as DNA template and site-directed

mutagenesis was performed using a QuikChange mutagenesis kit from Stratagene (La Jolla, CA, USA) and confirmed by DNA sequencing. Plasmids containing the correct mutant genes were then used to transform *Escherichia coli* BL21 (DE3), and colonies selected by kanamycin resistance were used for protein expression. Detailed IPTG (Isopropyl-β-D-thiogalactopyranoside) induced expression conditions are described in our previous studies.[9, 48, 49]

Protein purification and quantification

Wild-type and mutant enzymes were purified following a procedure described previously.[9, 49] The purity and identity of the purified enzymes were confirmed by 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), followed by staining with Coomassie blue R250. A prestained ladder marker (Invitrogen, Carlsbad, CA, USA) was used as reference. Protein concentrations were determined by the Bradford method using bovine serum albumin as a standard.[50]

Determination of kinetic parameters

Using a UV-visible spectrophotometer (Thermo Scientific, Waltham, MA, USA) at 25 °C in 50 mM Tris (pH 8.0), initial rates were determined at 340 nm. Kinetic measurements with substrate and the cofactor NAD^+ were performed as described earlier.[9, 49, 51] K_m and V_{max}, were determined from non-linear regression fitting of the Michaelis-Menten equation using Prism 5 (GraphPad Software, Inc., La Jolla, CA, USA).

Circular dichroism

Circular dichroism (CD) samples were prepared as described elsewhere[9] and experiments were carried out using a Jasco J-815 spectrophotometer (Jasco Corp., Tokyo, Japan) at 20 °C. At a rate of 100 nm/min, spectrum for each enzyme was recorded within the 190-300 nm wavelength range. Finally, baseline correction was performed for each collected CD spectra and ellipticity was expressed in millidegrees.

Computational methods

Structural setup

For structural analysis and computational calculations, the initial wild-type complex configuration was extracted from the Protein Data Bank (PDB entry code: 3M6I).[34] For further alignment, structure analysis and mutant modeling operations, as well as to obtain a reliable docking of L-arabinitol, the initial receptor structure (3M6I) was prepared using the "Prepare Protein tools" in Discovery Studio 3.5 (DS3.5; Accelrys Inc., San Diego, CA, USA). Briefly, the downloaded PDB structure was cleaned to remove any alternate conformations, and side chains were optimized. Hydrogen atoms were added to the receptor molecule using the CHARMm forcefield.[52] Finally, the protonation state and hydrogen atom positions were optimized using calculate protein ionization and residue pK protocol.

Mutant modeling

Using refine initial x-crystallography coordinates of wild-type NcLAD, mutant structures were achieved by replacing the side chain with the protein design extension of DS3.5. Using the "build mutant" protocol, mutant structures (L136A, L136C, L136H, and L136P) were modeled and then minimized with a molecular dynamics cascade, as described elsewhere.[9]

Molecular docking

Using CDOCKER, a CHARMm forcefield and grid-based docking protocol was implemented in DS3.5 where receptors (wild type and mutant NcLAD structures) were held rigid while the ligand (L-arabinitol) was allowed to be flexible in the protein-substrate complex. Following the setup, random orientations of the L-arabinitol conformations are produced by translating the center of the L-arabinitol within the NcLAD active site and performing a series of random rotations. Until the desired low-energy conformation of L-arabinitol is found, this process continues. For each protein-substrate complex, 100 poses of L-arabinitol were generated during every individual docking run and randomly distributed around the center of the NcLAD active site.[9, 33, 49] The substrate poses were then refined with a full-potential final minimization in the rigid receptor. The final L-arabinitol poses are retained based on the CHARMm energy and top CDOCKER scoring value. In this case it was mostly negative, thus favorable to binding, and was used for post-docking analysis.[53]

Molecular dynamic (MD) simulations

All simulations were performed according to the standard protocol for MD simulation with CHARMm force-field parameters as implemented in DS3.5. The MD simulation cascade in DS3.5 consists of energy minimization, followed by gradual heating of the system and isothermal isobaric ensemble (NPT) MD. Each system was initially minimized using the methods of steepest descent and conjugate gradient, with a

decreasing force constant applied to the heavy atoms, to remove unfavorable contacts. The resulting systems were then minimized without any constraints to adjust the water molecules and counter ions locally and eliminate any residual geometrical strain. The minimized solvated systems served as the initial structures in the subsequent molecular dynamics simulation. After relaxation, each system was gradually heated from 0 to 300 K in 15 ps and then maintained at 300 K for 35 ps.[54] Following the equilibration procedure, 2000-ps MD simulations were performed with a periodic boundary condition in the NPT ensemble, where the temperature (300 K) and constant pressure (1 atm) were maintained by Berendsen temperature coupling with a time constant of 1.0 ps and isotropic position scaling with a relaxation time of 2.0 ps, respectively. To observe the equilibration, root mean square deviations (RMSD) were recorded for all systems. MD simulations were analyzed using the "trajectory analysis" module of DS3.5. The RMSDs involving respective superposition were calculated by fitting each coordinate set from the production run to the starting structures. The criterion for a H-bond consisted of a distance between the acceptor (A) atom and the donor (D) atom smaller than 3.2 Å and an A-H-D angle larger than 120°. Only the last 500-ps sub-trajectory is appropriate for the main analysis because each system in equilibrium has stabilized enough for the sampling to be statistically valid. Snapshots without water molecules and counter ions for energy analysis were obtained during the last 500 ps at 10-ps intervals.

Binding free energy calculation

The binding free energies ($\Delta G_{binding}$) for each protein-ligand complex were calculated using Generalized Born Surface Area (MM/GBSA) methods[55] as implemented in DS3.5. Snapshot coordinates were extracted every 20 ps during the 4-ns trajectories. The binding free energy was computed from the difference between the protein, substrate, and the protein-substrate complex of a single MD simulation, where the protein and substrate were taken from the complex simulation.[56]

Acknowledgements

This research was supported by a grant from the Intelligent Synthetic Biology Center of Global Frontier Project (2011-0031955) funded by the Ministry of Science, ICT and Future Planning, Republic of Korea. This work was also supported by the Energy Efficiency & Resources Core Technology Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (201320200000420).

Notes and references

[a] Department of Chemical Engineering, 1 Hwayang–Dong, Gwangjin–Gu, Seoul 143–701, Republic of Korea. E-mail: jkrhee@konkuk.ac.kr; Fax: +82-2-458-3504.

[b] Department of Materials Science and Engineering, Korea University, Anam–Dong, Seongbuk–Gu, Seoul 136–713, Republic of Korea. E-mail: yckang@korea.ac.kr; Fax: +82-2-928-3584.

[c] Microbial Biotechnology and Genomics, CSIR-Institute of Genomics and Integrative Biology, Delhi University Campus, Mall Road, Delhi-110007, India.

† Electronic Supplementary Information (ESI) available: [Fig. S1 Multiple sequence alignment of polyol dehydrogenase enzymes, Fig. S2 Expression and purification of NcLAD wild-type and mutant enzymes, Table S1 Relative activities of NcLAD and HjLAD leucine mutant enzymes]. See DOI: 10.1039/b000000x/

‡ Current address: Department of Chemistry, Technical University of Denmark, Kongens Lyngby 2800, Denmark.

1. J. Kraut, *Science*, 1988, 242, 533-540.
2. A. Radzicka and R. Wolfenden, *Science*, 1995, 267, 90-93.
3. D. Ringe and G. A. Petsko, *Science*, 2008, 320, 1428-1429.
4. K. Bastard, A. A. T. Smith, C. Vergne-Vaxelaire, A. Perret, A. Zaparucha, R. De Melo-Minardi, A. Mariage, M. Boutard, A. Debard, C. Lechaplais, C. Pelle, V. Pellouin, N. Perchat, J. L. Petit, A. Kreimeyer, C. Medigue, J. Weissenbach, F. Artiguenave, V. De Berardinis, D. Vallenet and M. Salanoubat, *Nat Chem Biol*, 2014, 10, 42-U77.
5. H. Eklund, E. Horjales, H. Jornvall, C. I. Branden and J. Jeffery, *Biochemistry*, 1985, 24, 8005-8012.
6. S. Ramaswamy, M. ElAhmad, O. Danielsson, H. Jornvall and H. Eklund, *Protein Sci*, 1996, 5, 663-671.
7. J. P. Schwans, D. A. Kraut and D. Herschlag, *Proc Natl Acad Sci U S A*, 2009, 106, 14271-14275.
8. H. R. Brodkin, W. R. P. Novak, A. C. Milne, J. A. D'Aquino, N. M. Karabacak, I. G. Goldberg, J. N. Agar, M. S. Payne, G. A. Petsko, M. J. Ondrechen and D. Ringe, *Biochemistry*, 2011, 50, 4923-4935.
9. M. K. Tiwari, R. K. Singh, R. Singh, M. Jeya, H. Zhao and J. K. Lee, *J Biol Chem*, 2012, 287, 19429-19439.
10. N. Tokuriki, F. Stricher, L. Serrano and D. S. Tawfik, *PLoS Comput Biol*, 2008, 4, e1000002.
11. M. D. Toscano, K. J. Woycechowsky and D. Hilvert, *Angew Chem Int Ed Engl*, 2007, 46, 3212-3236.
12. W. N. Lipscomb and N. Strater, *Chem Rev*, 1996, 96, 2375-2434.
13. T. Dudev, Y. L. Lin, M. Dudev and C. Lim, *J Am Chem Soc*, 2003, 125, 3168-3180.
14. C. C. Huang, C. A. Lesburg, L. L. Kiefer, C. A. Fierke and D. W. Christianson, *Biochemistry*, 1996, 35, 3439-3446.
15. L. L. Kiefer, S. A. Paterno and C. A. Fierke, *J Am Chem Soc*, 1995, 117, 6831-6837.
16. C. A. Lesburg and D. W. Christianson, *J Am Chem Soc*, 1995, 117, 6838-6844.
17. P. Mertz, L. Yu, R. Sikkink and F. Rusnak, *J Biol Chem*, 1997, 272, 21296-21302.
18. Q.-Y. He, A. B. Mason, R. C. Woodworth, B. M. Tam, R. T. A. MacGillivray, J. K. Grady and N. D. Chasteen, *J Biol Chem*, 1998, 273, 17018-17024.

19. G. Behravan, B. H. Jonsson and S. Lindskog, *Eur J Biochem*, 1990, 190, 351-357.
20. D. W. Christianson and W. N. Lipscomb, *Proc Natl Acad Sci U S A*, 1986, 83, 7568-7572.
21. B. W. Matthews, J. N. Jansonius, P. M. Colman, B. P. Schoenborn and D. Dupourque, *Nat New Biol*, 1972, 238, 37-41.
22. C. L. Waller and G. R. Marshall, *J Med Chem*, 1993, 36, 2390-2403.
23. D. W. Christianson and J. D. Cox, *Annu Rev Biochem*, 1999, 68, 33-57.
24. D. Riccardi and Q. Cui, *J Phys Chem A*, 2007, 111, 5703-5711.
25. D. A. Kraut, K. S. Carroll and D. Herschlag, *Annu Rev Biochem*, 2003, 72, 517-571.
26. M. M. Harding, *Acta Crystallogr D Biol Crystallogr*, 2000, 56, 857-867.
27. M. M. Harding, *Acta Crystallogr D Biol Crystallogr*, 1999, 55, 1432-1443.
28. I. K. McDonald and J. M. Thornton, *J Mol Biol*, 1994, 238, 777-793.
29. S. Karlin, Z. Y. Zhu and K. D. Karlin, *Proc Natl Acad Sci U S A*, 1997, 94, 14225-14230.
30. N. J. Greenfield, *Methods Mol Biol*, 2004, 261, 55-78.
31. S. M. Kelly, T. J. Jess and N. C. Price, *Biochim Biophys Acta*, 2005, 1751, 119-139.
32. D. F. Stickle, L. G. Presta, K. A. Dill and G. D. Rose, *J Mol Biol*, 1992, 226, 1143-1159.
33. M. Tiwari and J. K. Lee, *J Mol Graph Model*, 2010, 28, 707-713.
34. B. Bae, R. P. Sullivan, H. Zhao and S. K. Nair, *J Mol Biol*, 2010, 402, 230-240.
35. I. Gitlin, J. D. Carbeck and G. M. Whitesides, *Angew Chem Int Ed Engl*, 2006, 45, 3022-3060.
36. H. Eklund and S. Ramaswamy, *Cell Mol Life Sci*, 2008, 65, 3907-3917.
37. T. A. Pauly, J. L. Ekstrom, D. A. Beebe, B. Chrunyk, D. Cunningham, M. Griffor, A. Kamath, S. E. Lee, R. Madura, D. McGuire, T. Subashi, D. Wasilko, P. Watts, B. L. Mylari, P. J. Oates, P. D. Adams and V. L. Rath, *Structure*, 2003, 11, 1071-1085.
38. P. J. Baker, K. L. Britton, M. Fisher, J. Esclapez, C. Pire, M. J. Bonete, J. Ferrer and D. W. Rice, *Proc Natl Acad Sci U S A*, 2009, 106, 779-784.
39. M. J. Banfield, M. E. Salvucci, E. N. Baker and C. A. Smith, *J Mol Biol*, 2001, 306, 239-250.
40. G. P. Miller and S. J. Benkovic, *Biochemistry*, 1998, 37, 6336-6342.
41. P. T. Rajagopalan, S. Lutz and S. J. Benkovic, *Biochemistry*, 2002, 41, 12618-12628.
42. C. E. Cameron and S. J. Benkovic, *Biochemistry*, 1997, 36, 15792-15800.
43. G. P. Miller and S. J. Benkovic, *Biochemistry*, 1998, 37, 6327-6335.
44. J. L. Radkiewicz and C. L. Brooks, *J Am Chem Soc*, 2000, 122, 225-231.
45. S. J. Benkovic and S. Hammes-Schiffer, *Science*, 2003, 301, 1196-1202.
46. R. D. Bach, C. Canepa and M. N. Glukhovtsev, *J Am Chem Soc*, 1999, 121, 6542-6555.
47. S. Hammes-Schiffer and S. J. Benkovic, *Annu Rev Biochem*, 2006, 75, 519-541.
48. R. K. Singh, M. K. Tiwari, R. Singh, J. R. Haw and J. K. Lee, *Appl Microbiol Biotechnol*, 2013, 98, 1095-1104.
49. M. K. Tiwari, R. K. Singh, H. Gao, T. Kim, S. Chang, H. S. Kim and J. K. Lee, *Bioorg Med Chem Lett*, 2014, 24, 173-176.
50. M. M. Bradford, *Anal Biochem*, 1976, 72, 248-254.
51. R. Sullivan and H. Zhao, *Appl Microbiol Biotechnol*, 2007, 77, 845-852.
52. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J Comp Chem*, 1983, 4, 187-217.
53. M. K. Tiwari, R. K. Singh, J. K. Lee and H. Zhao, *Bioorg Med Chem Lett*, 2012, 22, 1344-1347.
54. L. Desheng, G. Jian, C. Yuanhua, C. Wei, Z. Huai and J. Mingjuan, *Bioorg Med Chem Lett*, 2011, 21, 6630-6635.
55. H. Sun, Y. J. Jiang, Q. S. Yu, C. C. Luo and J. W. Zou, *Biochem Biophys Res Commun*, 2008, 377, 962-965.
56. N. Zhang, Y. Jiang, J. Zou, S. Zhuang, H. Jin and Q. Yu, *Proteins*, 2007, 67, 941-949.