Bioinformatics of Protein Bound Water

A Dissertation

Presented to

the Faculty of the Graduate School

University of Missouri-Columbia

In Partial Fufillment

Of the Requirements for the Degree

Doctor of Philosophy

by

CHRISTOPHER A. BOTTOMS

Dr. John J. Tanner, Dissertation Supervisor

July 2005

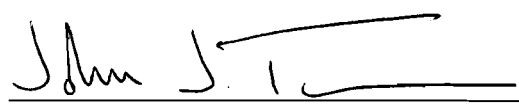
The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

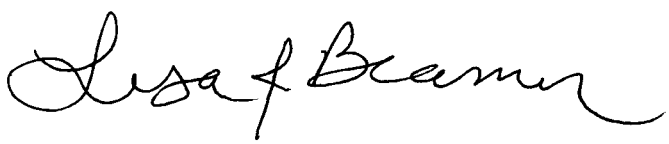
BIOINFORMATICS OF PROTEIN-BOUND WATER

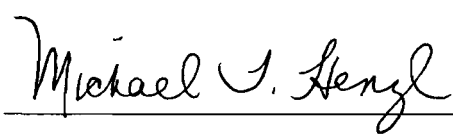
Presented by Christopher A. Bottoms

A candidate for the degree of Doctor of Philosophy

And hereby certify that in their opinion it is worthy of acceptance.







ACKNOWLEDGEMENTS

I really appreciate Dr. Jack Tanner's tutelage. If I had known about the field of Structural Bioinformatics before coming to MU, I would probably have gone to another university. Fortunately, I did not. I could not have received better training and become so involved in a project so dear to my heart than I have here. I really appreciate Jack's patience with me. I am somewhat of a "Lone Ranger" by nature and he probably often wondered what I was up to. He has really given me the room to spread my wings and I thank him for it.

In the summer of 2002, we had just completed our first paper, "A Structurally Conserved Water Molecule in Rossmann-fold Dinucleotide Binding Proteins." Finding this conserved water was a fortuitous accident, and I wished for a way to find conserved water molecules in other proteins. So, Jack and I began discussing the idea. If I remember correctly, it took several weeks of brainstorming before we came up with the current algorithm described herein. He taught me FORTRAN, helping me develop new code as well as providing some of his own. Though the process was still rough, within a few weeks we had recognizable results.

I have really enjoyed writing papers with Jack. This is probably the first time in my life that I have actually written collaboratively. I cannot count the number of times we sat down together to work on text, either writing notes on a manuscript or actually typing in the computer as we composed.

I want to thank Jack for occasionally staying late. Sometimes ideas just started rolling and everything else went on hold for a while. I apologize to Jack's wife, Evelyn. I

believe that spending time with one's family is very important and I've kept Jack late too often. So, I apologize and thank her for her patience. I wish the best of luck to Jack, Evelyn and family.

Many thanks to the Interdisciplinary Plant Group for providing the bulk of my funding during my graduate training. Their funding gave me the freedom to pursue a very novel line of research, one I never would have anticipated. I am also grateful to the Genetics Area Program for also providing financial assistance.

I doubt Matthew L. Klein realizes how much he has assisted me in my research. Never in my wildest dreams did I expect that one of my college roommates would have such an influence on my research. Matthew, a young undergraduate computer science student, could not believe that I was using such an archaic language as FORTRAN. He kept contrasting it with some new language called C# (pronounced C sharp). After hearing about C# advantages like being "object-oriented," having "garbage collection," and other terms and features that were completely novel to me, I finally gave in and tried it. My research has not been the same since. He is currently serving as a Latter-day Saint missionary in Mexico and will return to school in Fall 2005 or Winter 2006. I wish him success wherever he goes.

Our first sucessful test set, after Rossmann-fold proteins, was parvalbumins. My first rotation had been with Dr. Michael Henzl, so I had developed an affection for parvalbumins. In fact, he introduced me to my first molecular graphics program (Rasmol) and parvalbumin was the first protein model I ever viewed on a computer. It was really gratifying to go to him and show him something novel about his protein. We later collaborated in solving a 1.05Å structure of rat α-parvalbumin.

With respect to an education that led up to my current research, I would like to thank all of my former teachers, from kindergarten through graduate school. I still remember you. I would especially like to thank Mr. William Burns, my computer science teacher at Taft Middle School, in Oklahoma City, Oklahoma, for making computer programming fun. Because of his classes, I aspired to become a computer programmer. Experiences in high school diverted my path from computer science, but now I am back and enjoying every minute of it. I would also like to thank Dr. Charles Graham of Piedmont High School in Piedmont, Oklahoma. Learning physiology in his Biology II class led me to study biochemistry in college. I would also like to specifically thank Drs. Robert Fink, Terry Conley, Susan Barber, and Terry Phelps from Oklahoma City University for their instruction and continued friendship. There are many, many more who deserve my thanks and I regret not being able to name everyone.

Most of all, I would like to thank my dear mother and father. They have always encouraged me in my studies. They helped me feel that I could acheive whatever I desired. In particular, I would like to mention the memory of my father bringing home our first computer, an Atari 400. We have had a computer in the home ever since. He taught us children some basic programming. He planted a seed that is now flourishing. Thanks Dad.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

Exploring Structurally Conserved Solvent Sites in Protein Families

(Adapted from manuscript by Bottoms, White, and Tanner submitted to Proteins: Structure, Function and Bioinformatics.)

Chapter 1 Abstract

Protein-bound water molecules are important components of protein structure, and therefore, protein function and energetics. Although structural conservation of solvent has been studied in a few protein families, a lack of suitable computational tools has hindered more comprehensive analyses. Here we present a semi-automated computational approach for identifying solvent sites that are conserved among proteins sharing a common three-dimensional structure. This method is tested on six protein families: (1) monodomain cytochrome *c*, (2) fatty-acid binding protein, (3) lactate/malate dehydrogenase, (4) parvalbumin, (5) phospholipase A_2, and (6) serine protease. For each family, the method successfully identified previously known conserved solvent sites. Moreover, the method discovered several novel conserved solvent sites, some of which have higher degrees of conservation than the previously known sites. Analyses of these novel sites led to new structure/function hypotheses for the protein families studied. Our results suggest that every protein family will have highly conserved solvent sites, and that these sites should be considered as one of the defining three-dimensional structural characteristics of protein families and folds.

Key Words

Bioinformatics; cytochrome *c*; fatty-acid binding protein; lactate dehydrogenase; malate dehydrogenase; parvalbumin; EF-hand; phospholipase A_2; Rossmann fold; serine protease; water molecules

Abbreviations used

DOC, degree of conservation; SSM, Secondary Structure Matching

1.1 Introduction

Water is the solvent of biological chemistry and so it is not surprising that water molecules underlie many fundamental biochemical processes, including protein folding, enzymatic catalysis and biomolecular recognition. The major driving force of protein folding, the hydrophobic effect, is an entropic gain associated with the solvent. Catalytic water molecules poised in enzyme active sites directly participate in the bond-breaking and bond-making steps of many enzymatic reactions. Other water molecules in active sites may be involved in binding and stabilizing the substrate during catalysis. Water is critical for biomolecular recognition and association since molecular surfaces must be desolvated during complexation, and this shedding of solvent contributes significantly to the free energy of association. Lastly, and most relevant to our work, ordered water molecules that are tightly bound to proteins mediate intramolecular interactions within proteins and intermolecular interactions within biomolecular complexes.

The role of water as a mediator of noncovalent interactions is supported by high-resolution X-ray crystallographic data. Analysis of protein crystal structures has shown that water-mediated hydrogen bonds are abundant in protein complexes with DNA, protein, and small-molecule ligands. It is estimated that 40% of all protein-DNA hydrogen bonds are water mediated,[1] and that protein-protein interfaces contain an average of 22 water molecules and 11 water-mediated hydrogen bonds.[2] There are many examples of protein-ligand complexes in which water molecules bridge the protein and a bound small molecule ligand. For example, our study of Rossmann dinucleotide-binding domains revealed that 30% of the hydrogen bonds between the protein and adenine dinucleotide cofactors ($NAD(P)^+$, FAD) were water-mediated.[3] Similarly, Babor *et al*.

showed that water-mediated hydrogen bonds were important for recognition of the ribose moieties of ATP, ADP, and FAD.[4]

Importantly, several examples from the structure-based drug design and medicinal chemistry literature have demonstrated the essential nature of water-mediated hydrogen bonds in protein-inhibitor recognition, including the class C β-lactamase AmpC,[5] HIV-1 integrase,[5] HIV protease,[6-8] cyclin-dependent kinase-2,[8] Factor Xa,[9] thrombin,[10-12] Herpes simplex virus type I thymidine kinase,[13] thymidylate synthase,[14,15] neuraminidase,[16] heat-labile enterotoxin,[16] adenosine deaminase[5] and FKBP12.[17] In these cases, water molecules play a crucial role in ligand recognition by mediating hydrogen bonds between the ligand and the receptor protein. The overarching conclusion from this body of work was that successful drug design often requires explicit consideration of ordered water molecules in the binding pocket.

Given that protein-bound water molecules are important for maintaining proteins in their native conformations and for mediating biomolecular associations, there is considerable interest in elucidating exactly which water molecules might be particularly critical for protein structure and function. One hypothesis is that structurally and functionally important water molecules occupy solvent sites that are highly conserved among proteins sharing a common three-dimensional fold or active site structure. Conserved solvent sites have been examined only in a few protein families, such as fatty-acid binding proteins,[18] cytochrome *c*,[19] lectins,[20] phospholipase A_2,[21] ribonucleases,[22] Rossmann dinucleotide-binding proteins,[3] serine proteases[12,23,24] and parvalbumins.[25]

As part of our ongoing interest in probing the roles of solvent in protein structure and function, we have developed a structural bioinformatics methodology for identifying

and analyzing conserved solvent sites in protein structures. The method is tested on six protein families for which conserved solvent sites have been previously identified: monodomain cytochrome *c*, fatty-acid binding protein, lactate/malate dehydrogenase, parvalbumin, phospholipase A_2 and serine proteases. Our method not only identified known conserved solvent sites, but also revealed many novel ones. Analysis of the novel sites generated new hypotheses about protein structure, stability and function. Since our method has the potential to become fully automated, it could eventually be used to perform comprehensive analyses of conserved solvent sites in all protein families represented in the Protein Data Bank (PDB).[26]

1.2 Results

1.2.1 Introduction to the Methodology (see also Materials and Methods)

Our method of analyzing conserved solvent is based on identifying equivalent water molecules in a superimposed set of structurally-related proteins. For example, in the cytochrome *c* data set, the superimposed structures are shown in Figure 1.1a. It is readily apparent that identification of conserved water molecules by manual inspection is an extremely challenging task and that robust computational methods are preferred, particularity when considering protein families with many members. Note that our method is structure-based, allowing for the comparison of proteins that have similar structures despite limited sequence similarities.

Equivalence of water molecules is determined by considering two criteria: (1) the spatial locations of water molecules relative to the protein, and (2) the noncovalent interactions that water molecules form with the protein. Spatial equivalence is assessed by calculating the distribution of water molecules within a non-redundant set

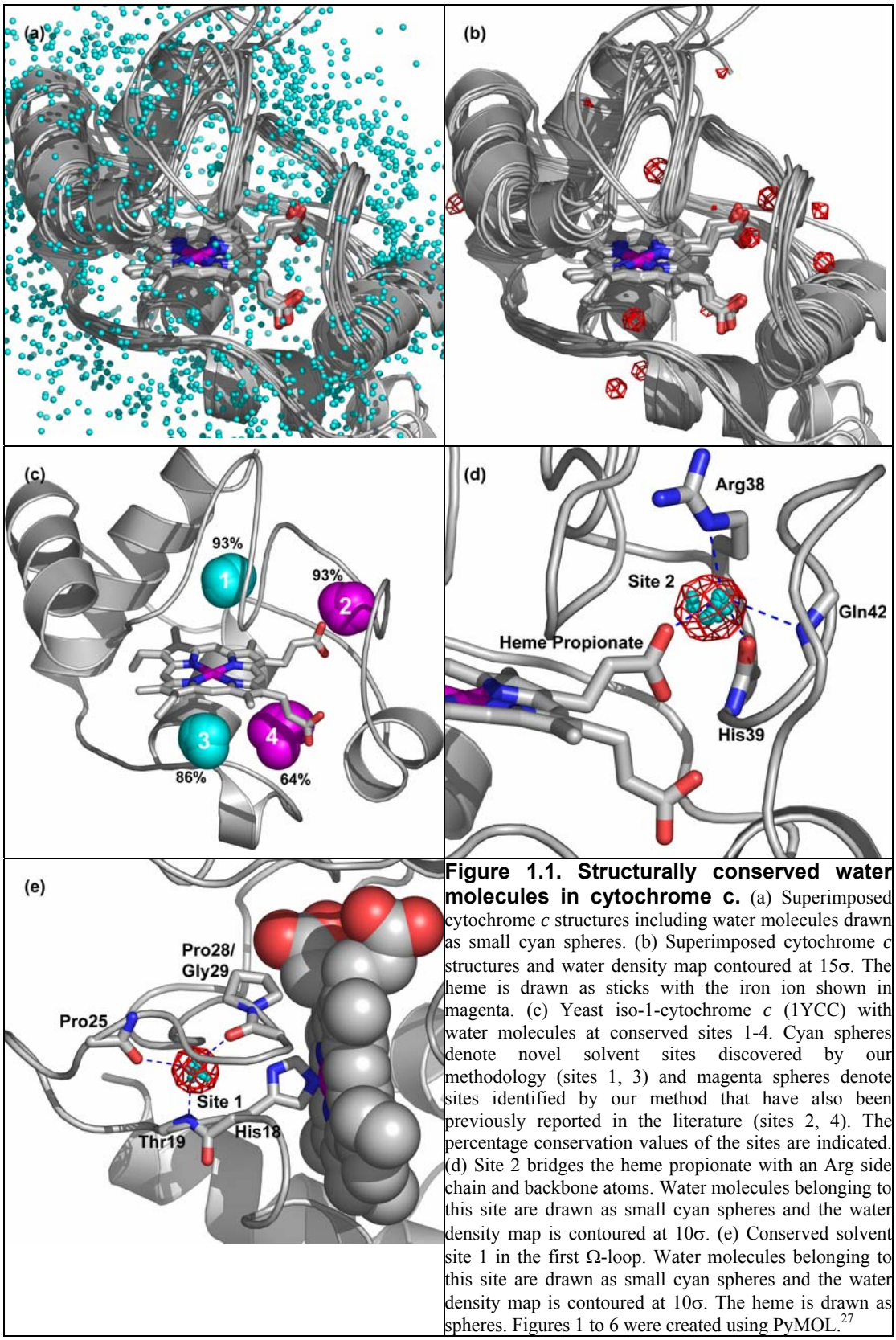


Figure 1.1. Structurally conserved water molecules in cytochrome c. (a) Superimposed cytochrome *c* structures including water molecules drawn as small cyan spheres. (b) Superimposed cytochrome *c* structures and water density map contoured at 15σ. The heme is drawn as sticks with the iron ion shown in magenta. (c) Yeast iso-1-cytochrome *c* (1YCC) with water molecules at conserved sites 1-4. Cyan spheres denote novel solvent sites discovered by our methodology (sites 1, 3) and magenta spheres denote sites identified by our method that have also been previously reported in the literature (sites 2, 4). The percentage conservation values of the sites are indicated. (d) Site 2 bridges the heme propionate with an Arg side chain and backbone atoms. Water molecules belonging to this site are drawn as small cyan spheres and the water density map is contoured at 10σ. (e) Conserved solvent site 1 in the first Ω-loop. Water molecules belonging to this site are drawn as small cyan spheres and the water density map is contoured at 10σ. The heme is drawn as spheres. Figures 1 to 6 were created using PyMOL.[27]

of superimposed protein structures. This distribution is represented as a pseudo-electron density map, which is calculated from the water molecules of the superimposed family of structures. For example, the water distribution map for cytochrome *c* is shown in Figure 1b. Note that there are several prominent features in the map. These regions of high electron density represent potential conserved solvent sites, and the map provides an intuitive visual representation of the structural context of these sites. One can see, for example, that there are several strong features near the heme group in this case (Figure 1b). Water molecules near these peaks are considered spatially equivalent.

Interaction equivalence is assessed by comparing the noncovalent interactions that each water molecule in the family makes with its respective protein. This criterion is quite general, and could include, for example, hydrogen bonding based on angle and distance cutoffs, and van der Waals interactions. For this initial study, a simple hydrogen-bonding criterion based on a 3.2 Å distance cutoff was used.

Water molecules in different structures of the family are considered structurally equivalent if (1) they are close to same peak in the water distribution map, and (2) they have at least one interaction with the protein in common. This analysis leads to a quantitative measure that expresses the degree of conservation (DOC) of a solvent site, as explained in the Materials and Methods section. The DOC-value is the percentage of structures in the protein family possessing an equivalent water molecule in a given solvent site. For example, the strongest peak in the cytochrome *c* map had a density value of 33σ (σ = standard deviations above the mean density, see A3.3.2) and the corresponding site had a DOC-value of 93% (Table 1.1, site 1). Thus, 93% (13/14) of the

Table 1.1. Selected conserved solvent sites of cytochrome *c*.

Site	Novel Site?	DOC (%)	Density Peak (σ)	Representative Interactions[a]		Water ID[a]
				Backbone	Side Chain	
1	✓	93	33	N-Thr19 O-Pro25 O-Gly29		110
2		93	32	O-His39 N-Gln42	NE-Arg38 O1A-Hem104	121
3	✓	86	30	O-Lys79 N-Ala81		122
4		64	20		ND2-Asn52 OH-Tyr67 OG1-Thr78	166

[a]Representative interactions and water ID for structure with PDB code 1YCC.

cytochrome *c* structures surveyed possessed an equivalent water molecule in this solvent site. Note that the DOC-value is analogous to the percent conservation of an individual amino acid residue within a sequence alignment.

1.2.2 Cytochrome c

Using a yeast mitochondrial cytochrome *c* (1YCC) as a query structure, we obtained a non-redundant cytochrome *c* data set consisting of 14 structures with resolution of at least 2.2 Å (see Materials and Methods). These matching structures were all monodomain cytochrome *c* proteins, comprising five eukaryotic mitochondrial cytochrome *c* proteins, seven bacterial cytochrome c_2 proteins, one bacterial cytochrome c_H and one bacterial cytochrome c_{552}.

The top three conserved solvent sites had DOC-values of 86-93% and peak heights in the water distribution map of 30-33σ (Table 1.1, sites 1-3). The high DOC-values suggest that these sites are highly conserved within this family. Site 4 has lower, but nonetheless significant, conservation (DOC = 64%). All four listed sites are near the

heme (Figure 1.1c), which suggests possible roles for these water molecules in stabilizing protein structural elements that bind the heme.

The conservation of sites 2 and 4 in the monodomain cytochrome *c* family has been discussed previously.[19,28,29] Thus, our method successfully identified known conserved solvent sites. Site 2 has been described by Benning *et al.* as conserved in eukaryotic mitochondrial cytochrome *c* and in bacterial cytochrome c_2.[30] This site is interesting because it bridges the heme propionate with the surrounding protein (Figure 1.1d). This interaction likely helps to stabilize the heme and orient it correctly in the active site.

In addition, our method identified two novel sites with DOC-values of 93 % and 86 % (Table 1.1, sites 1 and 3). A water molecule located in site 1 has been discussed for cytochrome c_2 separately for *Rhodopila globiformis*,[30] *Rhodopseudomonas palustris*[31] and *Paracoccus denitrificans*.[32] To our knowledge, the wider conservation of this site among monodomain cytochrome *c* structures has not been previously appreciated. This site is located in the middle of the first Ω-loop in cytochrome *c* (Figure 1.1e), and the water molecule is buried by the protein backbone. Within members of our data set, site 1 typically forms a bridge between a backbone amide nitrogen and two backbone carbonyls within the middle of this loop (Figure 1.1e).

It is interesting to note that this conserved site exists despite variations of sequence positions, or of spatial positions, of the solvent-coordinating residues. For example, the length of this Ω-loop is variable in our data set (Figure 1.1b). More specifically, given that the amino interaction is with residue number *i* of the sequence, the

first carbonyl interaction is located at a position within *i+3* to *i+6*, and the second carbonyl interaction is located at a position within *i+7* to *i+13*.

Despite the structural variation of the Ω-loop, the loop does contain a conserved Gly-Pro motif that interacts with the conserved water molecule. The second interacting carbonyl (i.e. at a position within *i+7* and *i+13*) belongs to the Gly residue of the conserved Gly-Pro motif. Therefore, this carbonyl lies in the same plane as the N, C_α and C_β atoms of the proline. The conserved Gly-Pro structural feature of the Ω-loop probably contributes to the observed high level of conservation of solvent site 1.

1.2.3 Fatty-acid binding proteins

Prendergast's group has extensively studied a conserved, internal solvent site in fatty-acid binding proteins using NMR, molecular dynamics simulations, and analysis of crystal structure data.[18,33] In our study, this site corresponded to the top peak in the water distribution map (62σ), and the site's DOC-value was 100% (Table 1.2, Figure 1.2a). Thus, as in the monodomain cytochrome *c* case, our method successfully identified a known conserved solvent site.

In addition, our method identified two novel sites with DOC-values greater than 80%. Site 3 is near the Prendergast site (Figure 1.2a), and site 2 is near the bound fatty acid.

Table 1.2. Selected conserved solvent sites of fatty-acid binding proteins.

Site	Novel Site?	DOC (%)	Density Peak (σ)	Representative Interactions[a] Backbone	Side Chain	Water ID[a]
1		100	62	O-Lys65 O-Val68 N-Val84		143
2	✓	91	29	O-Ala33 O-Thr36		135
3	✓	82	35	N-Lys65 O-Val68		212

[a]Representative interactions and water ID for structure with PDB code 1HMT.

Site 2 is particularly interesting because it appears to play a role in fatty-acid binding (Figure 1.2b). It typically donates hydrogen bonds to backbone carbonyl oxygen atoms that are *i* and *i+3* with respect to each other. These two residues lie at the end of the second helix of a pair of helices that have been described as a "lid" for the fatty-acid binding site.[34] In three structures, this structurally conserved water accepts a hydrogen bond from a conserved arginine residue (Arg126 in 1HMT) that, in turn, interacts with a negatively charged group of the bound fatty acid (Figure 1.2b).

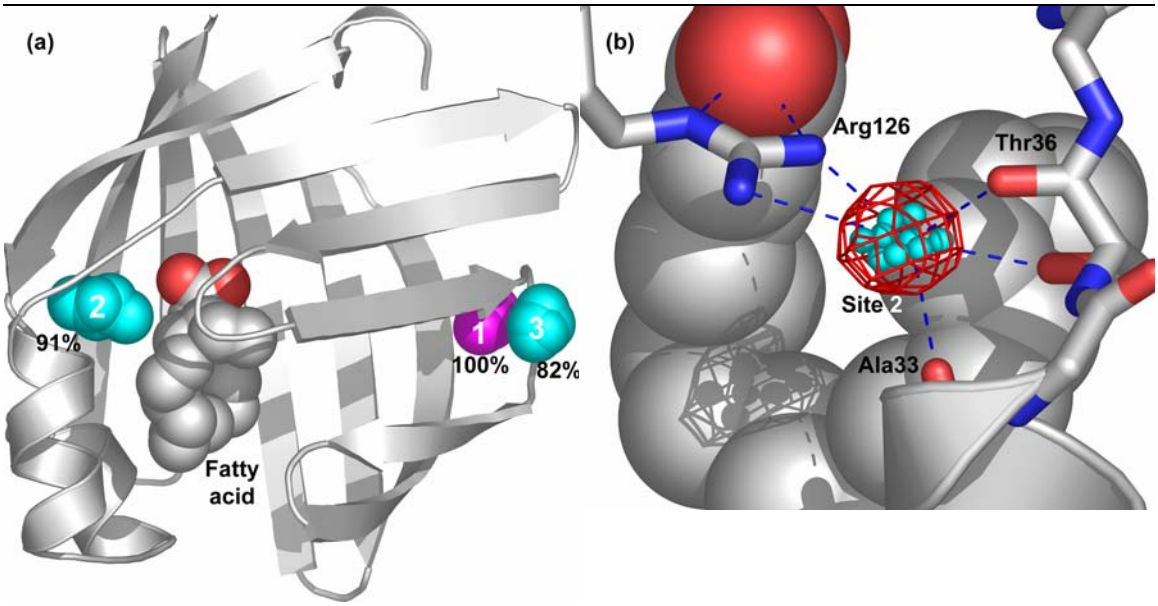


Figure 1.2. Structurally conserved water molecules in fatty-acid binding proteins. (a) Human muscle fatty-acid binding protein (1HMT) with water molecules from the superposition at conserved sites 1-3 (Table 1.2). Cyan spheres denote novel solvent sites discovered by our methodology (sites 2, 3) and magenta spheres denote a site identified by our method that has also been previously reported in the literature (site 1). The percentage conservation values of the sites are indicated. (b) Close-up view of conserved solvent site 2 (Table 1.2). Water molecules belonging to this site are drawn as small cyan spheres and the water density map is contoured at 10σ.

1.2.4 Lactate/malate dehydrogenase

The lactate/malate dehydrogenase family features a structurally conserved water molecule in the dinucleotide-binding site. This site was previously described in the lactate dehydrogenase family[35,36] and, more generally, in proteins containing the dinucleotide-binding Rossmann fold.[3] In our study, this site corresponded to the top peak

in the density map, and the DOC-value was 100% (Figure 1.3a, Table 1.3 site 1). Thus, a water molecule was found in this site in all 14 of the structures in our data set.

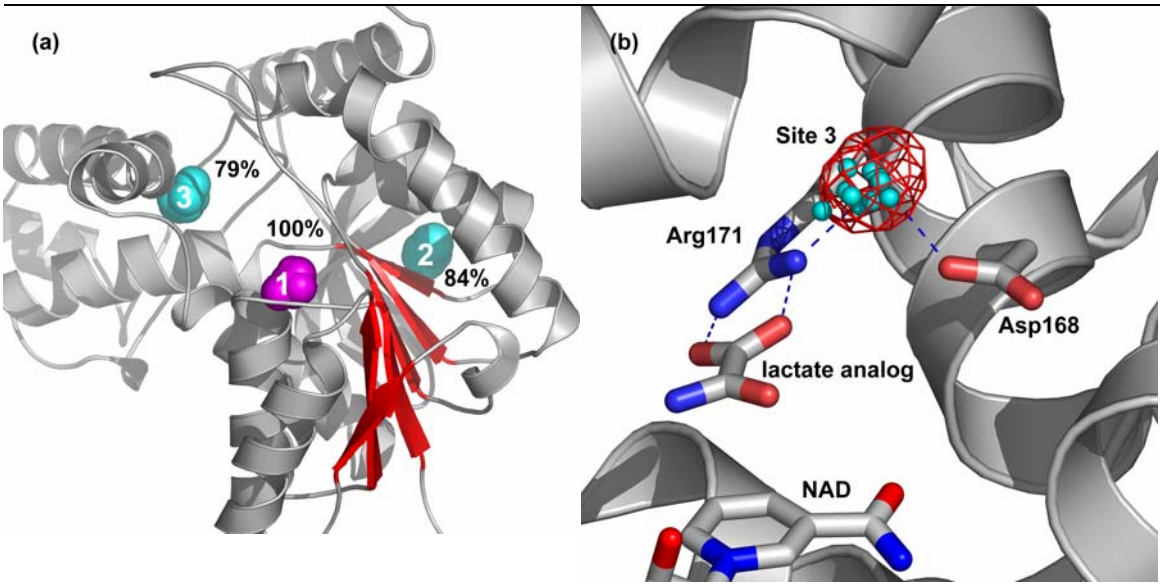


Figure 1.3. Structurally conserved water molecules in lactate/malate dehydrogenases. (a) Lactate dehydrogenase of the malaria-causing *Plasmodium falciparum* (1LDG) with highly conserved water molecules from the superposition. Cyan spheres denote novel solvent sites discovered by our methodology (sites 2, 3) and magenta spheres denote a site identified by our method that has also been previously reported in the literature (site 1). The percentage conservation values of the sites are indicated. (b) Close-up view of conserved solvent site 3, which is in the active site (Table 1.3, site 3). Water molecules belonging to this site are drawn as small cyan spheres and the water density map is contoured at 10σ.

Site 3, one of the interesting novel sites, had a DOC-value of 79%. Water at this site often serves as a salt bridge link, forming hydrogen bonds to both an arginine and an aspartate residue (Figure 1.3b). These residues are located at positions i and i+3 with respect to each other and are 100% conserved by sequence. It is well known that, within this family, the positively charged guanidinium group of this arginine stabilizes the negatively charged carboxylate of the substrate.[35] Thus, we see a correlation between amino acid sequence conservation and solvent structural conservation. The water-mediated salt-bridge occurs in 8 of the 14 proteins studied. In three of the remaining proteins, a water molecule near the peak interacts with either the arginine or aspartate, but not both. Using the Electron Density Server,[37] we observed that in another structure

Table 1.3. Selected conserved solvent sites of lactate/malate dehydrogenases.

Site	Novel Site?	DOC (%)	Density Peak (σ)	Representative Interactions[a]		Water ID[a]
				Backbone	Side Chain	
1		100	62	N-Gly29 N-Gly32	OG1-Thr97 NO2-NAD401[b]	13
2	✓	86	39	O-Lys157 O-Ile160 N-Leu274	OG1-Thr273	11
3	✓	79	35	O-Thr232	OD1-Asp168 NH1-Arg171	59

[a]Representative interactions and water ID for structure with PDB code 1LDG.
[b]NO2 is a pyrophosphate oxygen of the nicotinamide half of NAD. It is equivalent to atom name O2P in Schultze and Feigon.[38]

(1HYH) there was experimental density in both subunits that could be attributed to this conserved water. There was a sulfate ion in 6LDH and a crystal contact in 1LLD that could be attributed to disrupting the normal interactions.

An analysis on the fold level for NAD(P)-binding Rossmann-fold proteins (n=126 structures) was also performed (see Materials and Methods) in order to test our method with a set of structures with low overall sequence identity. The 126-structure data set had mean pairwise sequence identity of only 14 %. Nonetheless, our method clearly showed that the most conserved site on the fold level corresponded to site 1 of the LDH/MDH family (Table 1.3, site 1). This result agrees with our previous survey of water molecules in Rossmann dinucleotide-binding domains, which was performed using laborious manual inspection of superimposed structures.[3] This highly conserved solvent site is critical for cofactor recognition because it bridges the dinucleotide pyrophosphate with the Gly-rich loop of the Rossmann fold.[3]

1.2.5 Parvalbumins

Parvalbumins are soluble calcium-binding proteins containing two Ca^{2+}-binding sites: the CD-loop and the EF-loop. Conservation of solvent sites in the parvalbumin family has

been extensively studied,[25,39-41] and thus represents a good test case for our method. Our calculation revealed many sites with DOC-values > 70% that corresponded to previously identified sites (Figure 1.4a and Table 1.4, sites 1-3, 7, 9, 12, 16-19). Nine novel solvent

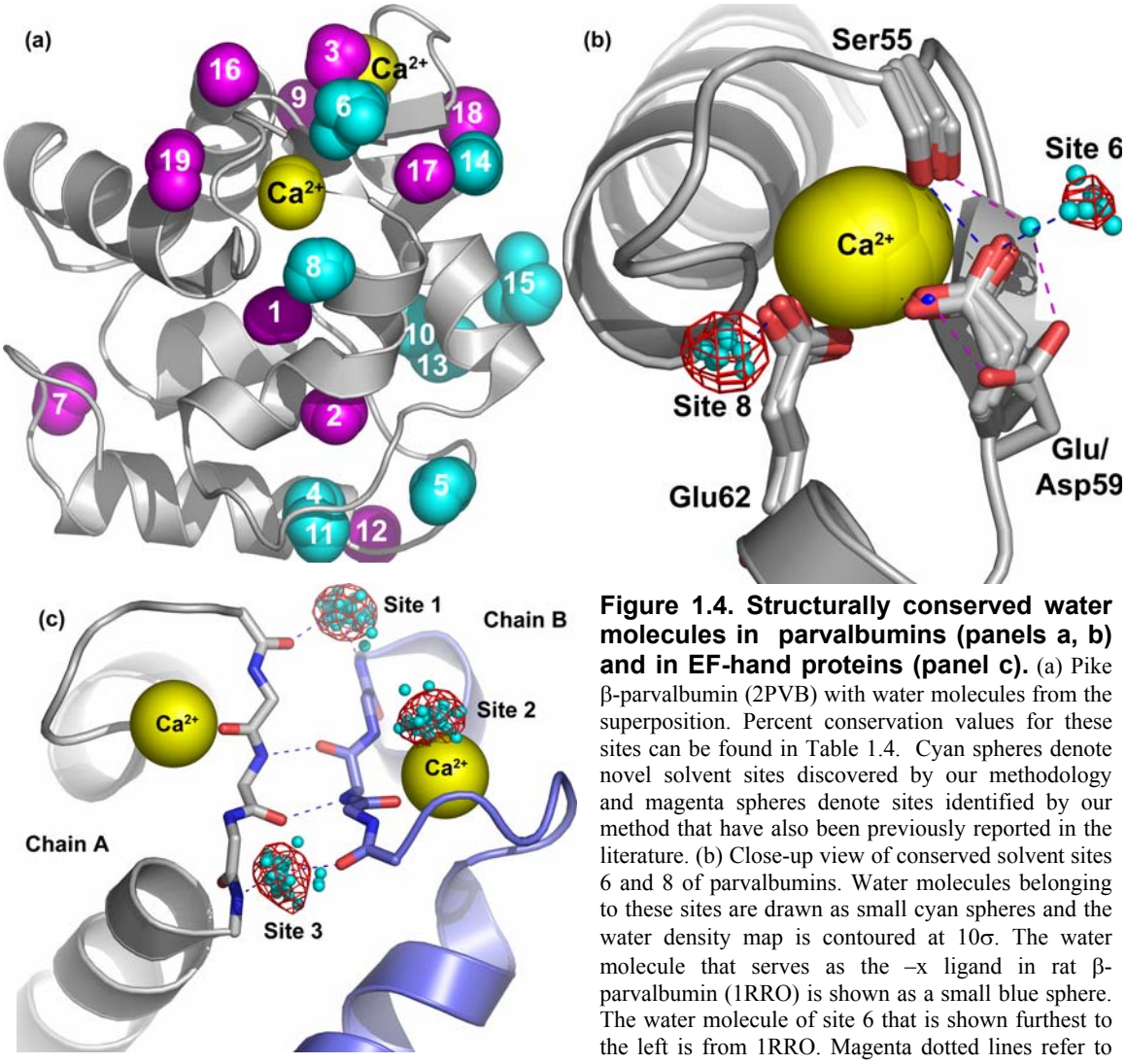


Figure 1.4. Structurally conserved water molecules in parvalbumins (panels a, b) and in EF-hand proteins (panel c). (a) Pike β-parvalbumin (2PVB) with water molecules from the superposition. Percent conservation values for these sites can be found in Table 1.4. Cyan spheres denote novel solvent sites discovered by our methodology and magenta spheres denote sites identified by our method that have also been previously reported in the literature. (b) Close-up view of conserved solvent sites 6 and 8 of parvalbumins. Water molecules belonging to these sites are drawn as small cyan spheres and the water density map is contoured at 10σ. The water molecule that serves as the –x ligand in rat β-parvalbumin (1RRO) is shown as a small blue sphere. The water molecule of site 6 that is shown furthest to the left is from 1RRO. Magenta dotted lines refer to interactions found in 1RRO. (c) Close-up view of the top three conserved solvent sites of EF-hand proteins. EF-hand sites 1, 2 and 3 correspond to sites 16, 3 and 17 of the parvalbumin family, respectively (Table 1.4). Water molecules belonging to these sites are drawn as small cyan spheres and the water density map is contoured at 10σ. The protein shown is timothy grass (*Phleum pratense*) pollen allergen Phl p 7 (1K9U). Unlike other EF-hand proteins, the structurally adjacent calcium-binding sites are from different polypeptide chains.

sites were also found (Table 1.4). Sites 4-6 are particularly notable because they have DOC-values of 100%, yet they have not been mentioned, to our knowledge, as being conserved in this family.

Several conserved solvent sites are close to the Ca^{2+}-binding sites, and some of

Table 1.4. Selected conserved solvent sites of parvalbumins.

Site	Novel Site?	DOC (%)	Density Peak (σ)	Representative Interactions[a]		Water ID[a]
				Backbone	Side Chain	
1		100	30	O-Ile50 O-Glu62		204
2		100	30	O-Lys64 N-Leu67 O-Arg75		202
3		100	28		OD1-Asp94 OD2-Asp94 OE1-Glu101	201
4	✓	100	27	O-Leu15		211
5	✓	100	24	O-Ala76		284
6	✓	100	15		OE2-Glu59	295
7		86	26	N-Lys7 O-Val33	OD2-Asp10	216
8	✓	86	26		OE2-Glu62	212
9		86	25	O-Gly89 O-Glu101		203
10	✓	86	25	O-Glu81		246
11	✓	86	24	N-Arg75		230
12		86	23	O-Ala17 N-Ala20		218
13	✓	86	20	O-Phe24		261
14	✓	86	19	O-Gly95		257
15	✓	86	17	O-Ala80		205
16		71	21	O-Gly56 N-Ile99		208[b]
17		71	21	N-Glu60 O-Gly95		224
18		71	17		OD2asp90	210
19		71	16		OD2asp51	221

[a]Representative interactions and water ID for structure with PDB code 2PVB.
[b]Symmetry-mate of water 208.

these water molecules play direct or indirect roles in binding Ca^{2+}. In this family of proteins, the ligands to the calcium ion are labelled with a letter corresponding to each of the three Cartesian coordinate axes and a positive or negative sign to indicate whether the ligand is "above" or "below" a specific plane. Within this coordination sphere, therefore, the –x ligand lies on the x axis and below the yz plane. Hence, there are +x, –x, +y, –y, +z, and -z ligands.[42] In the present study, site 3 is the –x ligand for the EF Ca^{2+}-binding

site. Two other known conserved water molecules, sites 18 and 19,[41] interact with the +x ligand of the EF- and CD-loops, respectively. Site 8, a novel site with DOC = 86%, hydrogen bonds to the CD –z ligand.

Novel site 6 is particularly interesting because it plays an indirect role in binding Ca^{2+} at the CD-loop (Figure 1.4b). This water molecule hydrogen bonds to the side chain of residue 59, which is Asp in 1RRO or Glu in the remaining structures. One of the carboxylate oxygen atoms of Glu59 is the –x ligand of the Ca^{2+}-binding site and the other hydrogen bonds to the hydroxyl of Ser55, which is the +z ligand.[41] In 1RRO, however, the carboxylate oxygen atoms of Asp59 each use a water molecule extension to allow them to fill the same roles as in Glu59. One water serves as the –x ligand to the Ca^{2+} ion,[41] and the site 6 water forms a hydrogen bond with Ser55. These two water molecules are also within hydrogen bonding distance of each other. Thus, we see that despite some variation in the Ca^{2+}-binding site, the site-6 water molecule still interacts with the –x ligand.

We also performed an analysis of a nonredundant set of proteins (n=38) representing the EF-hand superfamily, as defined by SCOP.[43] The three sites with the highest DOC values (71-76%) were equivalent to sites 3, 16 and 17 of the parvalbumin family. In a family mostly comprised of Ca^{2+} binding proteins, it is not surprising that site 3, the –x ligand of the EF-loop calcium ion, should be conserved. However, sites 16 and 17 are not obvious candidates for very high conservation. As noted by Strynadka and James over fifteen years ago, these solvent sites effectively extend the β-sheet between the calcium-binding loops.[41] They also mentioned that loops lacking calcium tended to form direct hydrogen bonds instead of water-mediated hydrogen bonds at these positions

within the sheet. Their observations were made in a study that included five proteins: carp parvalbumin, turkey troponin C, chicken troponin C, bovine calmodulin and bovine intestinal calcium-binding protein. In our study, all three sites are also seen in frequenin (1G8I),[44] osteonectin (1SRA),[45] the calcium-binding pollen allergen Phl p 7 (1K9U),[46] S100 proteins (1E8A,[47] 1MHO,[48] 1IRJ[49]), and sarcoplasmic calcium-binding protein (2SCP).[50] One or more of the three top sites are also seen in many of the other proteins within the EF-hand superfamily.

Conservation of the β-sheet-extending solvent sites (sites 16 and 17 of parvalbumin) is particularly interesting in Phl p 7 (1K9U). In contrast to the other calcium-binding proteins, the two calcium-binding sites that form the β-sheet are from two different polypeptide chains (Figure 1.4c). Thus, these highly conserved water molecules extend an *inter*molecular β-sheet in Phl p 7, in contrast to an intramolecular β-sheet in the other proteins. Figure 1.4c depicts Phl p 7 with the three most highly conserved water molecules of the EF-hand superfamily. Despite this unique structural difference, this protein contains all three of the most conserved solvent sites of this family. This is an excellent example of structural conservation that would not likely be predicted by amino acid sequence alone.

1.2.6 Phospholipase A_2

In the phospholipase A_2 family, two known conserved solvent sites are part of an important hydrogen-bonding network that typically involves the N-terminal residue (e.g. Ser1), His48, Tyr52, Ser68, and Asp99.[21,51,52] One is considered to have more of a structural role (Table 1.5, site 3), while the other has an important catalytic role (Table 1.5, site 9). The structural water molecule typically hydrogen bonds to all of these

Table 1.5. Selected conserved solvent sites of phospholipase A_2.

Site	Novel Site?	DOC (%)	Density Peak (σ)	Representative Interactions[a]		Water ID[a]
				Backbone	Side Chain	
1	✓	100	45	O-Cys1044		2005
2	✓	100	43	O-Cys1098		2002
3		100	39	N-Ser1001 O-Ser1068	OD2-Asp1099	2006
4	✓	92	39	N-Thr1041	OD2-Asp1039 OG1-Thr1041 OG1-Thr1112	2003
5	✓	92	33	O-Ala1101		2173
6	✓	83	38	N-Cys1084	OE2-Glu1097	2016
7	✓	83	36	O-Leu1106	OH-Tyr1022 OG1-Thr1041	2001
8	✓	83	27	O-Lys1049		2083
9		67	26		ND1-His48[b] OD1-Asp49[b]	69[b]

[a]Except where noted, representative interactions and water ID for structure with PDB code 1MC2.
[b]Interactions and solvent ID from 1G4I, since it more accurately represents the family in this instance.

residues except for His48. His48 interacts with Asp99 and with the catalytic water molecule.

The structural water is conserved in all of the proteins in our data set (DOC-value = 100%) and the catalytic water is conserved in most of the proteins in our data set (DOC

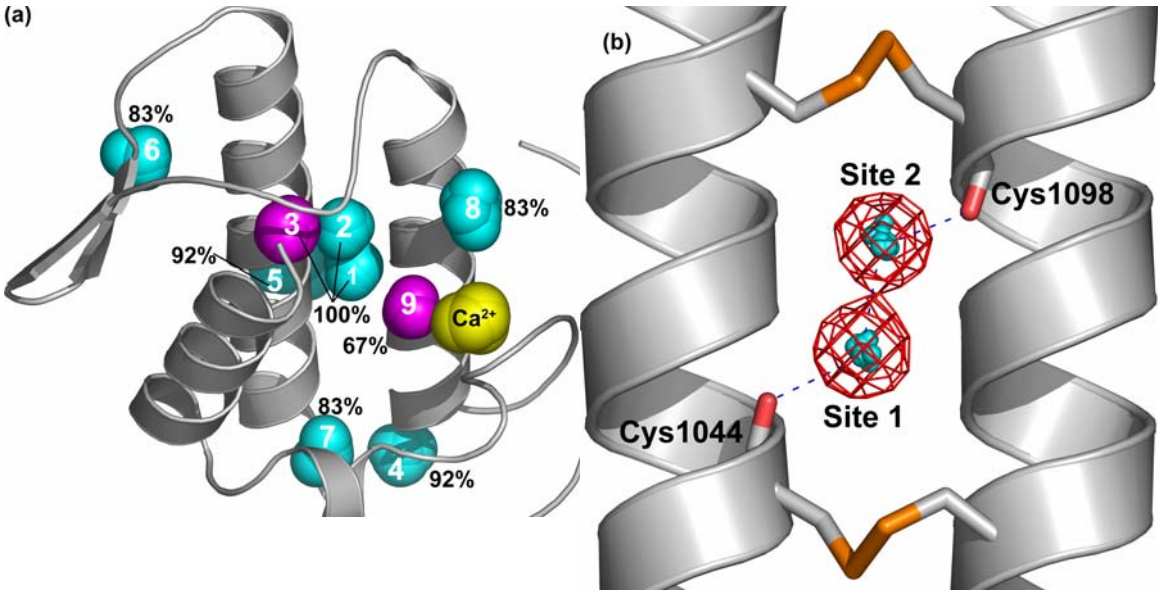


Figure 1.5. Structurally conserved water molecules in the phospholipase A2 family. (a) Phospholipase A_2 from *Agkistrodon acutus* venom (1MC2) with water molecules from the superposition. Cyan spheres denote novel solvent sites discovered by our methodology and magenta spheres denote sites identified by our method that have also been previously reported in the literature. The percentage conservation values of the sites are indicated. (b) Close-up view of sites 1 and 2 (Table V), which bridge backbone carbonyls of disulfide-linked cysteines. Water molecules belonging to these sites are drawn as small cyan spheres and the water density map is contoured at 10σ.

value = 67%), which attests to their importance (Figure 1.5 and Table 1.5, sites 3 and 9). In two cases (1HN4[53] and 1KVO[54]), the catalytic water was displaced by an inhibitor.

Interestingly, our method uncovered two highly conserved novel solvent sites (Table 1.5, sites 1 and 2) nestled between two α-helices (Figures 1.5a, 1.5b). These sites had DOC-values of 100%, thus they are at least as well conserved as the previously described structural and catalytic water molecules.

The two helices are linked by two conserved disulfide bonds, which are always separated by two helical turns (Figure 1.5b). The pair of conserved water molecules bridges backbone carbonyls of the two disulfide links. These water molecules most likely fulfill a structural role, and their ubiquity among members of this family invites further study.

1.2.7 Serine Proteases

Several studies have been published on conserved solvent sites in serine proteases.[12,23,24,55,56] Our study discovered many conserved solvent sites in this protein, ten of which had DOC values greater than 80% (Table 1.6). All of these had already been described in the literature. Perhaps this should not be surprising, since serine proteases are among the most studied with respect to conserved solvent. Two of these conserved sites (5 and 10) were found near the catalytic triad (Figure 1.6). These two sites represent water molecules that are typically in hydrogen bonding contact with each other. Sometimes they are replaced by three water molecules that perform similar roles. Site 5 bridges the amino nitrogen of either residue 101 or 102 to the N-terminus of a β-sheet (Figure 1.6b). Site 10 bridges the carbonyl oxygen of residue 99 to the hydroxyl of

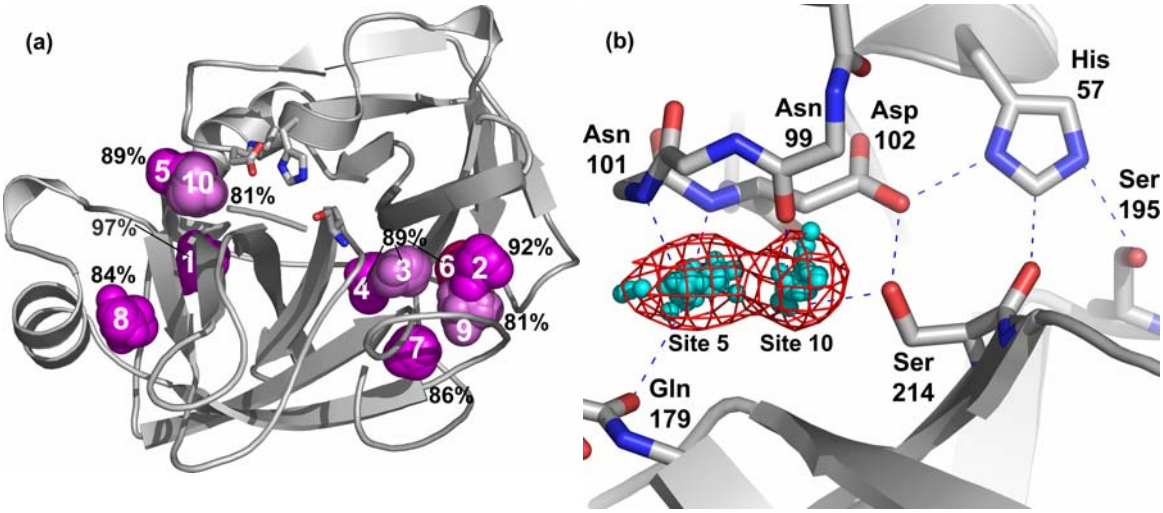


Figure 1.6. Structurally conserved water molecules in serine proteases. (a) *Fusarium oxysporum* trypsin (1FY4) with water molecules from the superposition. Residues of the catalytic triad are depicted as sticks. Magenta spheres denote solvent sites identified by our method that have also been previously reported in the literature. The percentage conservation values of the sites are indicated. (b) Close-up view of sites 5 and 10 (Table 1.6), which are near the catalytic triad (Asp 102, His 57, Ser 195). Water molecules belonging to these sites are drawn as small cyan spheres and the water density map is contoured at 10σ.

Table 1.6. Selected conserved solvent sites of serine proteases.

Site	Novel Site?	DOC (%)	Density Peak (σ)	Representative Interactions[a]		Water ID[a]
				Backbone	Side Chain	
1		97	40	O-Ile210 N-Gly232		1020
2		92	37	O-Ile30 O-Arg-66[b] N-Ser70		1011
3		89	43	N-Trp141 O-Gly193		1003
4		89	42	O-Ala139 O-Asp194		1002
5		89	36	N-Asn101 N-Asp102 O-Gln179[c]	OG1-Thr229[c,d]	1005[e]
6		89	35	O-Pro28 O-Ile30 N-Gly69		1024
7		86	42	N-Ile16 O-Gly140 O-Gly142		1014
8		84	26	O-Val163 N-Cys182		1018
9		81	34	O-Glu70[b]		67[b]
10		81	33	O-Asn99	OGser214	1008

[a]Except where noted, representative interactions and water ID for structure with PDB code 1FY4. [b]From 1HJ8, since it more accurately represents the family in this instance., [c]Found in less than half of the structures, [d]From 1GVK, since it more accurately represents the family in this instance. [e]This site has two waters in some structures.

Ser214. Note that Ser214 interacts with the aspartate and histidine of the catalytic triad.

1.3 Discussion

1.3.1 Advantages and limitations of the method

Structural conservation of solvent sites in six protein families was studied using a method that represents the distribution of water in superimposed structures as a pseudo-electron density map. The method described here has advantages over other methods that have been used to identify conserved solvent sites. For example, manual inspection using a graphics program has been the most commonly used approach. The main advantages of our computational approach are speed and objectivity. The only *a priori* knowledge required is a set of superimposed structures. No previous knowledge of conserved structural elements is required, such as the location of the active site or binding site. Thus, our method provides a powerful discovery tool. Manual analysis tends to focus on water molecules near the active sites or binding sites of proteins. Our method, however, is unbiased. It will discover conserved solvent sites in any part of the structure.

Sanschagrin and Kuhn have described a semi-automated computational method for identifying conserved solvent sites.[12] Using complete cluster linkage analysis, they found conserved water molecules among a set of thrombin structures (minimum pairwise sequence identity > 96%) and among a set of trypsin structures (100% sequence identity). The results of each of these two analyses were compared to find solvent sites conserved between thrombin and trypsin. However, complete linkage cluster analysis was not directly used to compare thrombin and trypsin. Linkage cluster analysis has also been used to study conserved water in structures of *Bacillus stearothermophilus* alanine racemase that were crystalized with different ligands bound.[57]

In contrast to these two studies using cluster linkage analysis, our method has been applied to directly compare much more diverse protein structures (Table 1.7). Since linkage cluster analysis studies have focused on proteins with very high sequence identity, a direct comparison between our method and linkage cluster using diverse data sets is not possible at this time. We did, however, apply our method to the set of serine protease structures used in the Sanschagrin and Kuhn work, and we were able to reproduce their results with excellent quantitative agreement (results not shown).

The current implementation of our method does have a few limitations. First, it is dependent on the superposition method. We used a method that superimposes proteins by their global fit to other proteins. Since global fitting is based on similar folds, it is an inherently low resolution method. To improve results, fitting of domains and smaller substructures could be implemented. Iterative superpositioning could focus on protein regions near water density peaks.

Another limitation of this method is that only water molecules seen in crystal structures are analyzed. Of these, only the most ordered solvent molecules will be consistently seen in multiple structures. In principle, there could potentially be dynamic water molecules that are also conserved but that are not observed in crystal structures. To identify such mobile, yet conserved, solvent molecules will require additional techniques such as molecular dynamics simulation, which has been used previously to identify preferred hydration sites around proteins.[58,59]

1.3.2 Other applications

This paper focused on conserved solvent in protein families, but the methods employed could be applied to other problems involving protein-bound water. For example,

conserved solvent sites could be analyzed within a single structure that has a repeating structural unit. In this context, our method has been used recently to show that a particular β-propeller protein[60] has multiple solvent sites that are structurally conserved among all six individual blades (L.J. Beamer, X. Li, C.A. Bottoms, and M. Hannink unpublished results).

Knowledge of conserved solvent could also be used to selectively model highly conserved water molecules in low/moderate resolution crystal structures. Difference electron density maps for low/moderate resolution structures will likely exhibit features corresponding to highly conserved solvent sites, although most crystallographers would be reluctant to model any solvent. For example, using the Electron Density Server[37] we observed an electron density peak in a 3.0 Å structure of human class IV alcohol dehydrogenase (1AGN)[61] that corresponds to the highly conserved water molecule of the Rossmann-fold dinucleotide-binding domain (peak 1, Table 1.3). This feature is especially apparent in the D chain of 1AGN. Thus, a conserved water molecule could probably be built in this low-resolution structure, even if no other water molecules were included. Doing so in this case would have been substantiated by a 2.5Å resolution structure (1D1S[62]) of the same protein deposited later in the PDB, which contained the structurally conserved water molecule modeled in all four chains in the asymmetric unit.

Protein structure modeling is another area of potential application. Due to advances in whole genome sequencing, there are currently many more protein sequences known than three-dimensional protein structures, making protein structure modelling ever more important.[63,64] Homology modeling programs often treat solvent as a continuum fluid because of the computational and theoretical challenges associated with modeling

explicit water molecules. Yet, bound water molecules are integral components of protein folds, and thus there are clear benefits to including water in protein models. For example, addition of conserved water to a homology model of rat submaxillary kallikrein greatly improved the Ramachandran statistics and led to a more accurate model.[56] We suggest that knowledge of highly conserved water molecules and the interactions that they form with the protein could provide a basis for selective incorporation of water into homology models. This application will be explored in the future.

1.4 Conclusion

For each protein family studied, our method successfully identified previously known conserved solvent sites, which validated our approach. Moreover, our method discovered several novel conserved solvent sites in protein families that have been intensely studied for decades. These discoveries suggested new structure/function relationships. For example, site 1 of cytochrome *c* probably serves a structural role within an Ω-loop that coordinates the iron of the heme group. Site 3 of fatty-acid binding proteins apparently assists the establishment of protein-fatty-acid interactions. Site 3 of the LDH/MDH family helps stabilize a conserved side chain that binds the substrate. Sites 6 and 8 of parvalbumins provide supplementary interactions that may help stabilize the calcium ion of the CD site. Lastly, sites 1 and 2 of phospholipase A_2 provide a compelling picture of structural solvent integrated into a protein structure.

All of the families studied had conserved solvent sites with degree of conservation values greater than 90 %. This result is significant because it suggests the hypothesis that all protein families have highly conserved solvent sites. Moreover, we assert that these conserved water molecules are integral to the protein structure because they provide

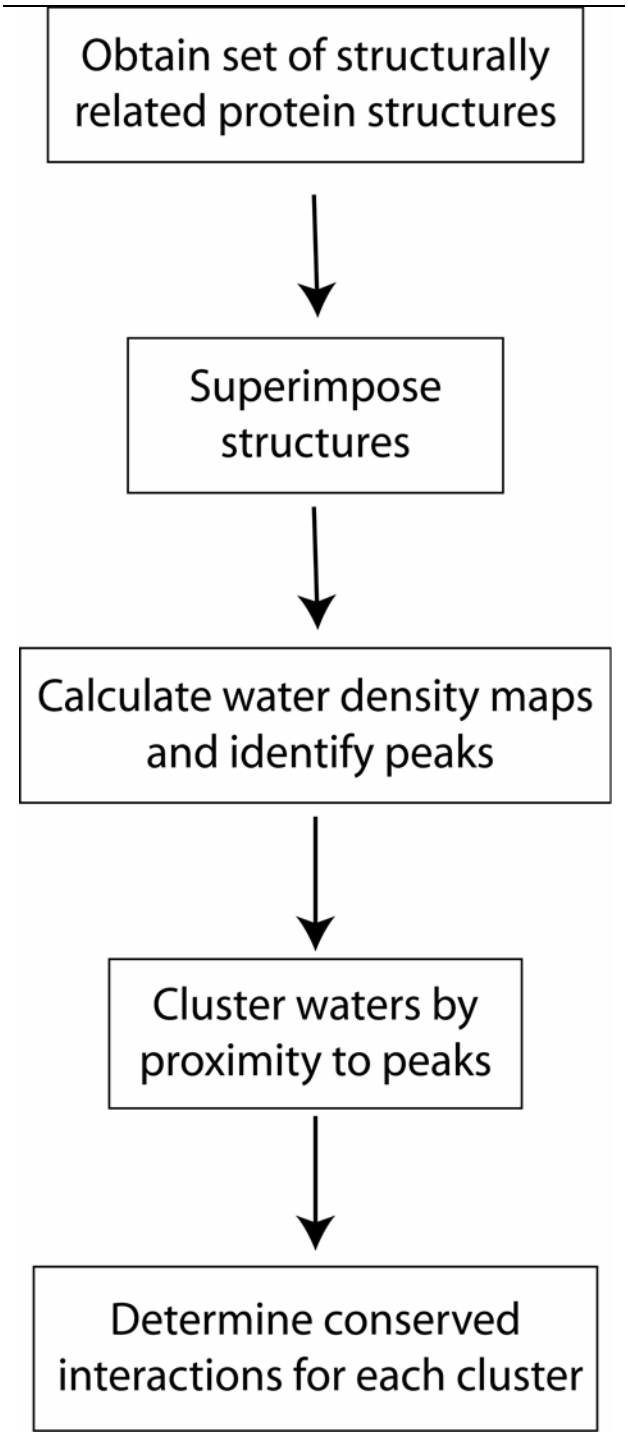


Figure 1.7. Schematic of the computational algorithm used in this study for discovery of structurally conserved water molecules.

noncovalent interactions that help stabilize specific regions of the protein fold. Thus, we suggest that highly conserved solvent sites should be considered as defining features of protein families, in addition to the better known secondary and tertiary structural features of the polypeptide chain. We plan to test and elaborate these ideas in the future by performing a PDB-wide analysis of protein-bound water using the approach described here.

1.5 Materials and Methods

1.5.1 Description of Method

Our strategy involves a pipeline of five calculations, as shown in Figure 1.7. First, a query protein structure is input to the Secondary Structure Matching (SSM)[65] service in order to retrieve a set of protein structures related by a common three-dimensional architecture. The default parameters of SSM were used for this study. A non-redundant subset of the structures returned by SSM is identified by comparison to the PDB_SELECT list,[66] and structures with a high

resolution diffraction limit of 2.2 Å or better were retained. Coordinate files are retrieved from the Protein Quaternary Structures (PQS) database. These files contain the full quaternary structure and all crystallographically related water molecules near the protein. The SSM service (also called MSDfold) and the PQS database can be found at the European Bioinformatics Institute's Macromolecular Structure Database web-site (http://www.ebi.ac.uk/msd/).

The non-redundant protein structures, including water molecules, are next superimposed using the transformation matrices obtained from SSM. An electron density map representing the distribution of water molecules in the data set is calculated using the fast Fourier transform algorithms in CNS[67] or the CCP4 suite.[68] For this calculation, all atoms except water molecules are removed, and the B-factors of all water molecules are set to 20 $Å^2$. Setting all of the B-factors to the same value avoids weighting effects due to variations in B-factors from structure to structure. Water molecules extremely distant from the surface of the reference structure were also excluded. The number of water molecules used to calculate each density map ranged from 1208 for parvalbumins to 38845 for the Rossmann fold data set. Peaks for the resulting electron density map are found using the peak picking modules of CNS or the CCP4 suite.[68] Each peak represents a potential conserved solvent site of the protein family.

The water molecules in the data set are next organized into clusters based on their proximity to the peaks of the map using a 2.0 Å cutoff. Finally, the noncovalent interactions formed by each water molecule in its respective structure are tabulated. For this study, two atoms were defined as interacting with each other if they were separated by no more than 3.2 Å. This criterion focuses on hydrogen bonding, but it could easily be

modified to include van der Waals interactions, or it could be made more restrictive by adding a hydrogen bond angle cutoff.

Water molecules in different structures are considered to be structurally conserved if (1) they belong to the same peak cluster, and (2) they have at least one equivalent interaction with the protein in common. This approach allows one to quantitatively express the degree of conservation (DOC) of a solvent site. We define DOC as the percentage of structures within a data set that contain structurally equivalent water molecules within a given cluster. This value is analogous to "percent conservation" of an amino acid residue at a specific position within a sequence alignment.

Though many conserved solvent sites exist, we have limited ourselves to listing only those that have the highest conservation (e.g. cytochrome *c* sites 1-3, Table 1.1) or, in several cases, those that have interesting features or that have had considerable attention in the literature (e.g. cytochrome *c* site 4, Table 1.1).

Programs used for molecular visualization during the course of this project include O,[69] Protein Explorer,[70] PyMOL,[27] and Coot.[71]

1.5.2 Data sets used in this study

The method was tested on six protein families for which conserved solvent sites have been previously identified: cytochrome *c*, fatty-acid binding protein, lactate/malate dehydrogenase, parvalbumin, phospholipase A_2 and serine proteases. For each data set containing *n* structures, $n(n-1)/2$ pairwise sequence identities were obtained using SSM.[65] In the case of serine proteases, ten structures were removed from the data set because they contained greater than 90% sequence identity. The resulting sequence comparison statistics for each of these data sets is given in Table 1.7.

The representative structure of the cytochrome *c* family that was input to SSM was 1YCC[72], and the resulting non-redundant data set consisted of 14 structures: 1CCR,[73] 1CO6,[29] 1COT,[32] a revised 1CXC structure,[74] 1HRO,[30] 1I8O,[31] 1JDL,[75] 1QL3,[76] 1QN2,[77] 1WEJ,[78] 1YCC,[72] 1YTC,[79] 3C2C,[80] 5CYT.[81] For the fatty-acid binding protein family, the representative structure was 1HMT[82] and the non-redundant data set included 11 structures: 1B56,[83] 1CBS,[84] 1CRB,[85] 1FDQ,[86] 1FTP,[87] 1HMT,[82] 1KQW,[88] 1LID,[89] 1LPJ,[90] 1MDC,[91] 1OPB.[92] For lactate/malate dehydrogenases, the representative structure was 1LDG[93] and the resulting data set consisted of 14 structures: 1A5Z,[81] 1BDM,[94] 1GUY,[95] 1GUZ,[95] 1HYE,[96] 1HYH,[36] 1I0Z,[97] 1LDG,[93] 1LLD,[98] 1MLD,[99] 1O6Z,[100] 2CMD,[101] 6LDH,[102] 9LDT.[103] The representative of the parvalbumin family was 2PVB[39] and the data set consisted of 7 structures: 1A75,[104] 1BU3,[105] 1PVA,[106] 1RRO,[107] 2PVB,[39] 4CPV,[108] 5PAL.[40] The phospholipase A_2 data set included 12 structures, resulting from using 1MC2[109] as the representative structure: 1FV0,[110] 1G4I,[111] 1HN4,[53] 1JIA,[112] 1JLT,[113] 1KVO,[54] 1LE6,[114] 1M8R,[115] 1MC2,[109] 1QLL,[116] 1VAP,[117] 1VIP.[118] The serine protease data set included 37 structures, with 1FY4[119] as the reference: 1A0J,[120] 1A7S,[121] 1BIO,[122] 1CGH,[123] 1CQQ,[124] 1DDJ,[125] 1EAX,[126] 1ELT,[127] 1EQ9,[128] 1F7Z,[129] 1FI8,[130]

Table 1.7. Pairwise amino acid sequence identity statistics of data sets studied[a]

Family	*n*	Mean	Median	Min	Max	Stdev	$\binom{n}{2}$
Cytochrome *c*	14	44	42	31	84	9	91
Fatty-acid binding protein	11	39	38	15	74	13	55
Lactate/malate dehydrogenase	14	30	28	12	77	11	91
Rossmann-fold dinucleotide-binding	126	14	13	1	88	6	7875
Parvalbumins	7	59	55	44	89	14	21
EF-hand Superfamily	38	20	18	1	86	11	703
Phospholipase A_2	12	49	48	33	82	11	66
Serine Proteases	37	35	35	9	87	11	666

[a]*n* = number of nonredundant structures in the data set. The last column is the number of one-on-one comparisons.

1FIW,[131] 1FUJ,[132] 1FXY,[133] 1FY4,[119] 1G2L,[134] 1GJ7,[135] 1GL1,[136] 1GVK,[137] 1H4W,[138] 1H8D,[139] 1HJ8,[140] 1HJ9,[140] 1IAU,[141] 1KLI,[142] 1LO6,[143] 1MBM,[144] 1MCT,[145] 1NN6,[146] 1NPM,[147] 1PPF,[148] 1SGT,[148] 1TON,[149] 1TRN,[150] 1UCY,[151] 2HLC,[152] 3RP2.[153]

Analysis of conserved solvent was also performed for Rossmann-fold proteins and for the EF-hand superfamily. In this case, the PDB Beta web site (http://pdbbeta.rcsb.org/pdb)[26] was used to identify structurally similar proteins. Using selection features of this new site, PDB files were chosen such that they (1) were within the same SCOP classification, (2) were crystal structures with 2.2 Å or better resolution, and (3) were not more than 90% identical in sequence (using the CD-HIT algorithm).[154]

The EF-hand data set was based on the SCOP superfamily classification "EF-hand" (SCOP code: a.39.1). The EF-hand superfamily included 38 structures: 1ALV,[155] 1AUI,[156] 1BU3,[105] 1CDP,[157] 1DTL,[158] 1E8A,[47] 1EG3,[159] 1EXR,[160] 1G4Y,[161] 1G8I,[44] 1GGZ,[162] 1IG5,[163] 1IRJ,[49] 1J55,[164] 1JF0,[165] 1K8U,[166] 1K94,[167] 1K9U,[46] 1KSO,[168] 1M45,[169] 1MHO,[48] 1MR8,[170] 1NCX,[171] 1OE9,[172] 1OMR,[173] 1PSR,[174] 1PVA,[106] 1QV1,[175] 1RRO,[107] 1RWY,[25] 1S6C,[176] 1SRA,[45] 1UHN,[177] 1WDC,[178] 2CBL,[179] 2PVB,[39] 2SCP,[50] 5PAL.[40]

The Rossmann-fold protein data set was based on the SCOP fold classification "NAD(P)-binding Rossmann-fold domains" (SCOP code: c.2). The nonredundant data set (n=126) of structures within the SCOP fold classification "NAD(P)-binding Rossmann-fold domains" included the following: 1A4I,[180] 1A5Z,[181] 1B16,[182] 1B7G,[183] 1B8P,[184] 1BDB,[185] 1BDM,[94] 1BG6,[186] 1BGV,[187] 1BXK,[188] 1CDO,[189] 1CF2,[190] 1CYD,[191] 1D7O,[192] 1DLJ,[193] 1DPG,[194] 1DSS,[195] 1DXY,[35] 1E3I,[196] 1E5Q,[197] 1E6U,[198] 1E6W,[199] 1E7W,[200] 1EK6,[201] 1EQ2,[202] 1EUD,[203] 1EVY,[204] 1F06,[205] 1F0Y,[206] 1FJH,[207] 1FMC,[208]

1G0O,[209] 1GAD,[210] 1GD1,[211] 1GEE,[212] 1GEG,[213] 1GPJ,[214] 1GR0,[215] 1GU7,[216] 1GUZ,[95] 1GY8,[217] 1H2B,[218] 1H5Q,[219] 1H6D,[220] 1HDO,[221] 1HEU,[222] 1HT0,[223] 1HXH,[224] 1HYE,[96] 1I0Z,[97] 1I24,[225] 1I36,[226] 1IUK,[227] 1IY8,[228] 1J3V,[229] 1J4A,[230] 1J5P,[231] 1JA9,[232] 1JAY,[233] 1JQB,[234] 1JTV,[235] 1JVB,[236] 1K3T,[237] 1K6X,[238] 1KEW,[239] 1KOL,[240] 1KS9,[241] 1L7D,[242] 1LC0,[243] 1LDM,[102] 1LI4,[244] 1LJ8,[245] 1LLD,[98] 1LUA,[246] 1M6H,[247] 1MB4,[248] 1MG5,[249] 1MLD,[99] 1MV8,[250] 1MX3,[251] 1N2S,[252] 1N7H,[253] 1NP3,[254] 1NPY,[255] 1NVM,[256] 1NWH,[257] 1NXQ,[258] 1NYT,[259] 1O0S,[260] 1O6Z,[100] 1OAA,[261] 1OBB,[262] 1OBF,[263] 1OC2,[264] 1OI7,[265] 1ORR,[266] 1P0F,[267] 1P1J,[268] 1P77,[269] 1PJ3,[270] 1PJC,[271] 1PL8,[272] 1PR9,[273] 1PX0,[274] 1PZG,[275] 1Q7B,[276] 1QMG,[277] 1QSG,[278] 1R6D,[279] 1RKX,[280] 1RPN,[281] 1T2A,[282] 1T2D,[283] 1UAY,[284] 1UDC,[285] 1UR5,[286] 1UUF,[287] 1VI2,[288] 1VJ0,[289] 1VJ1,[290] 1VJP,[291] 2AE2,[292] 2CMD,[101] 2NAC,[293] 2PGD,[294] 9LDT.[103]

Chapter 1 Acknowledgments

We thank Dr. Herbert Axelrod for providing revised coordinates of 1CXC, and Drs. Eugene Krissinel and Kim Henrick for help with SSM and MSD.

Chapter 1 References

1. Luscombe NM, Laskowski RA, Thornton JM. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res 2001;29(13):2860-2874.

2. Janin J. Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. Structure Fold Des 1999;7(12):R277-279.

3. Bottoms CA, Smith PE, Tanner JJ. A structurally conserved water molecule in Rossmann dinucleotide-binding domains. Protein Sci 2002;11(9):2125-2137.

4. Babor M, Sobolev V, Edelman M. Conserved positions for ribose recognition: importance of water bridging interactions among ATP, ADP and FAD-protein complexes. J Mol Biol 2002;323(3):523-532.

5. Powers RA, Shoichet BK. Structure-based approach for binding site identification on AmpC beta-lactamase. J Med Chem 2002;45(15):3222-3234.

6. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins 2002;46(1):34-40.

7. Rarey M, Kramer B, Lengauer T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. Proteins 1999;34(1):17-28.

8. Johnson LN, De Moliner E, Brown NR, Song H, Barford D, Endicott JA, Noble ME. Structural studies with inhibitors of the cell cycle regulatory kinase cyclin-dependent protein kinase 2. Pharmacol Ther 2002;93(2-3):113-124.

9. Rao MS, Olson AJ. Modelling of factor Xa-inhibitor complexes: a computational flexible docking approach. Proteins 1999;34(2):173-183.

10. Engh RA, Brandstetter H, Sucher G, Eichinger A, Baumann U, Bode W, Huber R, Poll T, Rudolph R, von der Saal W. Enzyme flexibility, solvent and 'weak' interactions characterize thrombin-ligand interactions: implications for drug design. Structure 1996;4(11):1353-1362.

11. Katz BA, Elrod K, Luong C, Rice MJ, Mackman RL, Sprengeler PA, Spencer J, Hataye J, Janc J, Link J, Litvak J, Rai R, Rice K, Sideris S, Verner E, Young W. A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site. J Mol Biol 2001;307(5):1451-1486.

12. Sanschagrin PC, Kuhn LA. Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteases and contributions to ligand specificity. Protein Sci 1998;7(10):2054-2064.

13. Vogt J, Perozzo R, Pautsch A, Prota A, Schelling P, Pilger B, Folkers G, Scapozza L, Schulz GE. Nucleoside binding site of herpes simplex type 1 thymidine kinase analyzed by X-ray crystallography. Proteins 2000;41(4):545-553.

14. Rutenber EE, Stroud RM. Binding of the anticancer drug ZD1694 to *E. coli* thymidylate synthase: assessing specificity and affinity. Structure 1996;4(11):1317-1324.

15. Sage CR, Rutenber EE, Stout TJ, Stroud RM. An essential role for water in an enzyme reaction mechanism: the crystal structure of the thymidylate synthase mutant E58Q. Biochemistry 1996;35(50):16270-16281.

16. Minke WE, Diller DJ, Hol WG, Verlinde CL. The role of waters in docking strategies with incremental flexibility for carbohydrate derivatives: heat-labile enterotoxin, a multivalent test case. J Med Chem 1999;42(10):1778-1788.

17. Faerman CH, Karplus PA. Consensus preferred hydration sites in six FKBP12-drug complexes. Proteins 1995;23(1):1-11.

18. Likic VA, Juranic N, Macura S, Prendergast FG. A "structural" water molecule in the family of fatty acid binding proteins. Protein science 2000;9(3):497-504.

19. Berghuis AM, Guillemette JG, McLendon G, Sherman F, Smith M, Brayer GD. The role of a conserved internal water molecule and its associated hydrogen bond network in cytochrome *c*. J Mol Biol 1994;236(3):786-799.

20. Loris R, Stas PP, Wyns L. Conserved waters in legume lectin crystal structures. The importance of bound water for the sequence-structure relationship within the legume lectin family. J Biol Chem 1994;269(43):26722-26733.

21. Kumar A, Sekharudu C, Ramakrishnan B, Dupureur CM, Zhu H, Tsai MD, Sundaralingam M. Structure and function of the catalytic site mutant Asp 99 Asn of phospholipase A_2: absence of the conserved structural water. Protein Science 1994;3(11):2082-2088.

22. Loris R, Langhorst U, De Vos S, Decanniere K, Bouckaert J, Maes D, Transue TR, Steyaert J. Conserved water molecules in a large family of microbial ribonucleases. Proteins 1999;36(1):117-134.

23. Krem MM, Di Cera E. Conserved water molecules in the specificity pocket of serine proteases and the molecular mechanism of Na^+ binding. Proteins 1998;30(1):34-42.

24. Sreenivasan U, Axelsen PH. Buried water in homologous serine proteases. Biochemistry 1992;31(51):12785-12791.

25. Bottoms CA, Schuermann JP, Agah S, Henzl MT, Tanner JJ. Crystal structure of rat α-parvalbumin at 1.05 Å resolution. Protein Sci 2004;13:1724-1734.

26. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank (http://www.rcsb.org/). Nucleic Acids Res 2000;28(1):235-242.

27. DeLano WL. "The PyMOL Molecular Graphics System." San Carlos, CA, USA: DeLano Scientific; 2002.

28. Lett CM, Berghuis AM, Frey HE, Lepock JR, Guillemette JG. The role of a conserved water molecule in the redox-dependent thermal stability of iso-1-cytochrome *c*. J Biol Chem 1996;271(46):29088-29093.

29. Sogabe S, Miki K. Refined crystal structure of ferrocytochrome c_2 from *Rhodopseudomonas viridis* at 1.6 Å resolution. J Mol Biol 1995;252(2):235-247.

30. Benning MM, Meyer TE, Holden HM. Molecular Structure of a High Potential Cytochrome c_2 Isolated from *Rhodopila globiformis*. Arch Biochem Biophys 1996;333(2):338-348.

31. Geremia S, Garau G, Vaccari L, Sgarra R, Viezzoli MS, Calligaris M, Randaccio L. Cleavage of the iron-methionine bond in c-type cytochromes: Crystal structure of oxidized and reduced cytochrome c_2 from *Rhodopseudomonas palustris* and its ammonia complex. Protein Sci 2002;11(1):6-17.

32. Benning MM, Meyer TE, Holden HM. X-Ray structure of the cytochrome c_2 isolated from *Paracoccus denitrificans* refined to 1.7-Å resolution. Arch Biochem Biophys 1994;310(2):460-466.

33. Likic VA, Prendergast FG. Dynamics of internal water in fatty acid binding protein: computer simulations and comparison with experiments. Proteins 2001;43(1):65-72.

34. LaLonde JM, Bernlohr DA, Banaszak LJ. The up-and-down β-barrel proteins. Faseb J 1994;8(15):1240-1247.

35. Dengler U, Niefind K, Kiess M, Schomburg D. Crystal structure of a ternary complex of D-2-hydroxyisocaproate dehydrogenase from *Lactobacillus casei*, NAD^+ and 2-oxoisocaproate at 1.9 Å resolution. J Mol Biol 1997;267(3):640-660.

36. Niefind K, Hecht HJ, Schomburg D. Crystal structure of L-2-hydroxyisocaproate dehydrogenase from *Lactobacillus confusus* at 2.2 Å resolution. An example of strong asymmetry between subunits. J Mol Biol 1995;251(2):256-281.

37. Kleywegt G, Harris M, Zou J, Taylor T, Wählby A, Jones T. The Uppsala Electron Density Server. accepted (Acta Cryst D/CCP4 Proceedings) 2004.

38. Schultze P, Feigon J. Chirality errors in nucleic acid structures. Nature 1997;387(6634):668.

39. Declercq JP, Evrard C, Lamzin V, Parello J. Crystal structure of the EF-hand parvalbumin at atomic resolution (0.91 Å) and at low temperature (100 K). Evidence for conformational multistates within the hydrophobic core. Protein Sci 1999;8(10):2194-2204.

40. Roquet F, Declercq JP, Tinant B, Rambaud J, Parello J. Crystal structure of the unique parvalbumin component from muscle of the leopard shark (*Triakis semifasciata*). The first X-ray study of an α-parvalbumin. J Mol Biol 1992;223(3):705-720.

41. Strynadka NC, James MN. Crystal structures of the helix-loop-helix calcium-binding proteins. Annu Rev Biochem 1989;58:951-998.

42. Kretsinger RH, Nockolds CE. Carp Muscle Calcium-binding Protein. II. Structure determination and general description. J Biol Chem 1973;248(9):3313-3326.

43. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247(4):536-540.

44. Bourne Y, Dannenberg J, Pollmann V, Marchot P, Pongs O. Immunocytochemical localization and crystal structure of human frequenin (neuronal calcium sensor 1). J Biol Chem 2001;276(15):11949-11955.

45. Hohenester E, Maurer P, Hohenadl C, Timpl R, Jansonius JN, Engel J. Structure of a novel extracellular Ca^{2+}-binding module in BM-40. Nat Struct Biol 1996;3(1):67-73.

46. Verdino P, Westritschnig K, Valenta R, Keller W. The cross-reactive calcium-binding pollen allergen, Phl p 7, reveals a novel dimer assembly. EMBO J 2002;21(19):5007-5016.

47. Moroz OV, Antson AA, Murshudov GN, Maitland NJ, Dodson GG, Wilson KS, Skibshoj I, Lukanidin EM, Bronstein IB. The three-dimensional structure of human S100A12. Acta Crystallog sect D 2001;57(Pt 1):20-29.

48. Kilby PM, Van Eldik LJ, Roberts GC. The solution structure of the bovine S100B protein dimer in the calcium-free state. Structure 1996;4(9):1041-1052.

49. Itou H, Yao M, Fujita I, Watanabe N, Suzuki M, Nishihira J, Tanaka I. The crystal structure of human MRP14 (S100A9), a Ca^{2+}-dependent regulator protein in inflammatory process. J Mol Biol 2002;316(2):265-276.

50. Vijay-Kumar S, Cook WJ. Structure of a sarcoplasmic calcium-binding protein from *Nereis diversicolor* refined at 2.0 Å resolution. J Mol Biol 1992;224(2):413-426.

51. Brunie S, Bolin J, Gewirth D, Sigler PB. The refined crystal structure of dimeric phospholipase A_2 at 2.5 Å. Access to a shielded catalytic center. J Biol Chem 1985;260(17):9742-9749.

52. Dijkstra BW, Drenth J, Kalk KH. Active site and catalytic mechanism of phospholipase A_2. Nature 1981;289(5798):604-606.

53. Epstein T, Yu B, Pan Y, Tutton S, Maliwal B, Jain M, Bahnson B. The Basis for k^*_{cat} Impairment in Prophospholipase A_2 from the Anion-Assisted Dimer Structure. Biochemistry 2001;40(38):11411-11422.

54. Cha SS, Lee D, Adams J, Kurdyla JT, Jones CS, Marshall LA, Bolognese B, Abdel-Meguid SS, Oh BH. High-resolution X-ray crystallography reveals precise binding interactions between human nonpancreatic secreted phospholipase A_2 and a highly potent inhibitor (FPL67047XX). J Med Chem 1996;39(20):3878-3881.

55. Di Cera E, Guinto ER, Vindigni A, Dang QD, Ayala YM, Wuyi M, Tulinsky A. The Na^+ binding site of thrombin. J Biol Chem 1995;270(38):22089-22092.

56. Henriques EF, Ramos MJ, Reynolds CA. Inclusion of conserved buried water molecules in the model structure of rat submaxillary kallikrein. J Comput Aided Mol Des 1997;11(6):547-556.

57. Mustata G, Briggs JM. Cluster analysis of water molecules in alanine racemase and their putative structural role. Protein Eng Des Sel 2004;17(3):223-234.

58. Henchman RH, McCammon JA. Extracting hydration sites around proteins from explicit water simulations. J Comput Chem 2002;23(9):861-869.

59. Lounnas V, Pettitt BM, Phillips GN, Jr. A global model of the protein-solvent interface. Biophys J 1994;66(3 Pt 1):601-614.

60. Li X, Zhang D, Hannink M, Beamer LJ. Crystal structure of the Kelch domain of human Keap1. J Biol Chem 2004;279(52):54750-54758.

61. Xie P, Parsons SH, Speckhard DC, Bosron WF, Hurley TD. X-ray structure of human class IV sigmasigma alcohol dehydrogenase. Structural basis for substrate specificity. J Biol Chem 1997;272(30):18558-18563.

62. Xie PT, Hurley TD. Methionine-141 directly influences the binding of 4-methylpyrazole in human sigma sigma alcohol dehydrogenase. Protein Sci 1999;8(12):2639-2644.

63. Guo J-t, Ellrott K, Chung WJ, Xu D, Passovets S, Xu Y. PROSPECT-PSPP: an automatic computational pipeline for protein structure prediction. Nucl Acids Res 2004;32(suppl_2):W522-525.

64. Zhang Y, Skolnick J. Automated structure prediction of weakly homologous proteins on a genomic scale. Proc Natl Acad Sci 2004;101(20):7594-7599.

65. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallog sect D 2004;60(Pt 12 Pt 1):2256-2268.

66. Hobohm U, Sander C. Enlarged representative set of protein structures. Protein Sci 1994;3(3):522-524.

67. 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, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallog sect D 1998;54 (Pt 5):905-921.

68. Collaborative Computational Project N. The CCP4 suite: programs for protein crystallography. Acta Crystallog sect D 1994;50(Pt 5):760-763.

69. Jones TA, Zou JY, Cowan SW, Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallog sect A 1991;47 (Pt 2):110-119.

70. Martz E. Protein Explorer: easy yet powerful macromolecular visualization. Trends Biochem Sci 2002;27(2):107-109.

71. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallog sect D 2004;60(Pt 12 Pt 1):2126-2132.

72. Louie GV, Brayer GD. High-resolution refinement of yeast iso-1-cytochrome *c* and comparisons with other eukaryotic cytochromes *c*. J Mol Biol 1990;214:527.

73. Ochi H, Hata Y, Tanaka N, Kakudo M, Sakurai T, Aihara S, Morita Y. Structure of rice ferricytochrome *c* at 2.0 Å resolution. J Mol Biol 1983;166(3):407-418.

74. Axelrod HL, Feher G, Allen JP, Chirino AJ, Day MW, Hsu BT, Rees DC. Crystallization and X-ray structure determination of cytochrome c_2 from *Rhodobacter sphaeroides* in three crystal forms. Acta Crystallog sect D 1994;50(4):596-602.

75. Camara-Artigas A, Williams JC, Allen JP. Structure of cytochrome c_2 from *Rhodospirillum centenum*. Acta Crystallog sect D 2001;57(11):1498-1505.

76. Harrenga A, Reincke B, Ruterjans H, Ludwig B, Michel H. Structure of the soluble domain of cytochrome c_{552} from *Paracoccus denitrificans* in the oxidized and reduced states. J Mol Biol 2000;295(3):667-678.

77. Read J, Gill R, Dales SL, Cooper JB, Wood SP, Anthony C. The molecular structure of an unusual cytochrome c_2 determined at 2.0 Å; the cytochrome c_H from *Methylobacterium extorquens*. Protein Sci 1999;8(6):1232-1240.

78. Mylvaganam SE, Paterson Y, Getzoff ED. Structural basis for the binding of an anti-cytochrome *c* antibody to its antigen: crystal structures of FabE8-cytochrome *c* complex to 1.8 Å resolution and FabE8 to 2.26 Å resolution. J Mol Biol 1998;281(2):301-322.

79. McGee WA, Rosell FI, Liggins JR, Rodriguez-Ghidarpour S, Luo Y, Chen J, Brayer GDM, A.G., Nall BT. Thermodynamic cycles as probes of structure in unfolded proteins. Biochemistry 1996;13(35):1995-2007.

80. Bhatia GE. Refinement of the Crystal Structure of Oxidized *Rhodospirillum Rubrum* Cytochrome c_2. San Diego: University of California, San Diego; 1981.

81. Takano T. Refinement of myoglobin and cytochrome *c*. In: Hall S, Ashida wT, editors. Methods and Applications in Crystallographic Computing. Oxford, England: Oxford University Press; 1984. p 527-.

82. Young AC, Scapin G, Kromminga A, Patel SB, Veerkamp JH, Sacchettini JC. Structural studies on human muscle fatty acid binding protein at 1.4 Å resolution: binding interactions with three C18 fatty acids. Structure 1994;2(6):523-534.

83. Hohoff C, Borchers T, Rustow B, Spener F, van Tilbeurgh H. Expression, purification, and crystal structure determination of recombinant human epidermal-type fatty acid binding protein. Biochemistry 1999;38(38):12229-12239.

84. Kleywegt GJ, Bergfors T, Senn H, Le Motte P, Gsell B, Shudo K, Jones TA. Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid. Structure 1994;2(12):1241-1258.

85. Cowan SW, Newcomer ME, Jones TA. Crystallographic studies on a family of cellular lipophilic transport proteins. Refinement of P2 myelin protein and the structure determination and refinement of cellular retinol-binding protein in complex with all-trans-retinol. J Mol Biol 1993;230(4):1225-1246.

86. Balendiran GK, Schnutgen F, Scapin G, Borchers T, Xhong N, Lim K, Godbout R, Spener F, Sacchettini JC. Crystal structure and thermodynamic analysis of human brain fatty acid-binding protein. J Biol Chem 2000;275(35):27045-27054.

87. Haunerland NH, Jacobson BL, Wesenberg G, Rayment I, Holden HM. Three-dimensional structure of the muscle fatty-acid-binding protein isolated from the desert locust *Schistocerca gregaria*. Biochemistry 1994;33(41):12378-12385.

88. Calderone V, Folli C, Marchesani A, Berni R, Zanotti G. Identification and Structural Analysis of a Zebrafish Apo and Holo Cellular Retinol-binding Protein. J Mol Biol 2002;321(3):527-535.

89. Xu Z, Bernlohr DA, Banaszak LJ. The adipocyte lipid-binding protein at 1.6-Å resolution. Crystal structures of the apoprotein and with bound saturated and unsaturated fatty acids. J Biol Chem 1993;268(11):7874-7884.

90. Folli C, Calderone V, Ramazzina I, Zanotti G, Berni R. Ligand binding and structural analysis of a human putative cellular retinol-binding protein. J Biol Chem 2002;277(44):41970-41977.

91. Benning MM, Smith AF, Wells MA, Holden HM. Crystallization, structure determination and least-squares refinement to 1.75 Å resolution of the fatty-acid-binding protein isolated from *Manduca sexta L*. J Mol Biol 1992;228(1):208-219.

92. Winter NS, Bratt JM, Banaszak LJ. Crystal structures of holo and apo-cellular retinol-binding protein II. J Mol Biol 1993;230(4):1247-1259.

93. Dunn CR, Banfield MJ, Barker JJ, Higham CW, Moreton KM, Turgut-Balik D, Brady RL, Holbrook JJ. The structure of lactate dehydrogenase from *Plasmodium falciparum* reveals a new target for anti-malarial design. Nature structural biology 1996;3(11):912-915.

94. Kelly CA, Nishiyama M, Beppu T, Birktoft JJ. The Structure at 1.8 Angstroms Resolution of a Single Site Mutant (T189I) of Malate Dehydrogenase from *Thermus flavus* with Increased Enzymatic Activity. Protein Data Bank deposition (PDB code 1BDM) 1993.

95. Dalhus B, Saarinen M, Sauer UH, Eklund P, Johansson K, Karlsson A, Ramaswamy S, Bjork A, Synstad B, Naterstad K. Structural Basis for Thermophilic Protein Stability: Structures of Thermophilic and Mesophilic Malate Dehydrogenases. J Mol Biol 2002;318(3):707-721.

96. Lee BI, Chang C, Cho S-J, Eom SH, Kim KK, Yu YG, Suh SW. Crystal structure of the MJ0490 gene product of the hyperthermophilic archaebacterium *Methanococcus jannaschii*, a novel member of the Lactate/Malate family of dehydrogenases. J Mol Biol 2001;307(5):1351-1362.

97. Read JA, Winter VJ, Eszes CM, Sessions RB, Brady RL. Structural basis for altered activity of M- and H-isozyme forms of human lactate dehydrogenase. Proteins 2001;43(2):175-185.

98. Iwata S, Ohta T. Molecular basis of allosteric activation of bacterial L-lactate dehydrogenase. J Mol Biol 1993;230(1):21-27.

99. Gleason WB, Fu Z, Birktoft J, Banaszak L. Refined crystal structure of mitochondrial malate dehydrogenase from porcine heart and the consensus structure for dicarboxylic acid oxidoreductases. Biochemistry 1994;33(8):2078-2088.

100. Irimia A, Ebel C, Madern D, Richard SB, Cosenza LW, Zaccai G, Vellieux FM. The Oligomeric states of *Haloarcula marismortui* malate dehydrogenase are modulated by solvent components as shown by crystallographic and biochemical studies. J Mol Biol 2003;326(3):859-873.

101. Hall MD, Levitt DG, Banaszak LJ. Crystal structure of *Escherichia coli* malate dehydrogenase. A complex of the apoenzyme and citrate at 1.87 Å resolution. J Mol Biol 1992;226(3):867-882.

102. Abad-Zapatero C, Griffith JP, Sussman JL, Rossmann MG. Refined crystal structure of dogfish M4 apo-lactate dehydrogenase. J Mol Biol 1987;198(3):445-467.

103. Dunn CR, Wilks HM, Halsall DJ, Atkinson T, Clarke AR, Muirhead H, Holbrook JJ. Design and synthesis of new enzymes based on the lactate dehydrogenase framework. Philos Trans R Soc Lond B Biol Sci 1991;332(1263):177-184.

104. Declercq JP, Baneres JL, Rambaud J, Parello J. Tertiary Structure of a Trp-Containing Parvalbumin from Whiting (*Merlangius merlangus*). Protein Data Bank deposition (PDB code 1A75) 1998.

105. Revett SP, King G, Shabanowitz J, Hunt DF, Hartman KL, Laue TM, Nelson DJ. Characterization of a helix-loop-helix (EF hand) motif of silver hake parvalbumin isoform B. Protein Sci 1997;6(11):2397-2408.

106. Roquet F, Rambaud J, Declercq JP, Tinant B, Baldellon C, Padilla A, Cave A, Parello J. Parvalbumin (Pike, 5.0, Component) Complexed With Two Calcium Ions (Synchrotron X-Ray Diffraction). Protein Data Bank deposition (PDB code 1PVA) 1998.

107. Ahmed FR, Rose DR, Evans SV, Pippy ME, To R. Refinement of recombinant oncomodulin at 1.30 Å resolution. J Mol Biol 1993;230(4):1216-1224.

108. Kumar VD, Lee L, Edwards BF. Refined crystal structure of calcium-liganded carp parvalbumin 4.25 at 1.5-Å resolution. Biochemistry 1990;29(6):1404-1412.

109. Liu Q, Huang Q, Teng M, Weeks CM, Jelsch C, Zhang R, Niu L. The crystal structure of a novel, inactive, lysine 49 PLA2 from *Agkistrodon acutus* venom: an ultrahigh resolution, *ab initio* structure determination. J Biol Chem 2003;278(42):41400-41408.

110. Chandra V, Jasti J, Kaur P, Srinivasan A, Betzel C, Singh TP. Structural Basis of Phospholipase A_2 Inhibition for the Synthesis of Prostaglandins by the Plant Alkaloid Aristolochic Acid from a 1.7 Å Crystal Structure. Biochemistry 2002;41(36):10914 -10919.

111. Steiner RA, Rozeboom HJ, de Vries A, Kalk KH, Murshudov GN, Wilson KS, Dijkstra BW. X-ray structure of bovine pancreatic phospholipase A_2 at atomic resolution. Acta Crystallographica Section D, Biological crystallography 2001;57(4):516-526.

112. Zhao K, Song S, Lin Z, Zhou Y. Structure of a Basic Phospholipase A_2 from *Agkistrodon halys pallas* at 2.13 Å Resolution. Acta Crystallographica Section D, Biological crystallography 1998;54(4):510-521.

113. Banumathi S, Rajashankar KR, Notzel C, Aleksiev B, Singh TP, Genov N, Betzel C. Structure of the neurotoxic complex vipoxin at 1.4 Å resolution. Acta Crystallog sect D 2001;57(11):1552-1559.

114. Pan YH, Yu B-Z, Singer AG, Ghomashchi F, Lambeau G, Gelb MH, Jain MK, Bahnson BJ. Crystal Structure of Human Group X Secreted Phospholipase A_2. Electrostatically Neutral Interfacial Binding Surface Targets Zwitterionic Membranes. J Biol Chem 2002;277(32):29086-29093.

115. Xu S, Gu L, Jiang T, Zhou Y, Lin Z. Structures of cadmium-binding acidic phospholipase A_2 from the venom of *Agkistrodon halys Pallas* at 1.9Å resolution. Biochem Biophys Res Commun 2003;300(2):271-277.

116. Lee W-H, da Silva Giotto MT, Marangoni S, Toyama MH, Polikarpov I, Garratt RC. Structural Basis for Low Catalytic Activity in Lys49 Phospholipases A_2-A Hypothesis: The Crystal Structure of Piratoxin II Complexed to Fatty Acid. Biochemistry 2001;40(1):28-36.

117. Han SK, Yoon ET, Scott DL, Sigler PB, Cho W. Structural aspects of interfacial adsorption. A crystallographic and site-directed mutagenesis study of the phospholipase A_2 from the venom of *Agkistrodon piscivorus piscivorus*. J Biol Chem 1997;272(6):3573-3582.

118. Carredano E, Westerlund B, Persson B, Saarinen M, Ramaswamy S, Eaker D, Eklund H. The three-dimensional structures of two toxins from snake venom throw light on the anticoagulant and neurotoxic sites of phospholipase A_2. Toxicon 1998;36(1):75-92.

119. Rypniewski WR, Ostergaard PR, Norregaard-Madsen M, Dauter M, Wilson KS. *Fusarium oxysporum* trypsin at atomic resolution at 100 and 283 K: a study of ligand binding. Acta Crystallog sect D 2001;57(Pt 1):8-19.

120. Schroder HK, Willassen NP, Smalas AO. Structure of a non-psychrophilic trypsin from a cold-adapted fish species. Acta Crystallog sect D 1998;54 (Pt 5):780-798.

121. Karlsen S, Iversen LF, Larsen IK, Flodgaard HJ, Kastrup JS. Atomic resolution structure of human HBP/CAP37/azurocidin. Acta Crystallog sect D 1998;54 (Pt 4):598-609.

122. Jing H, Babu YS, Moore D, Kilpatrick JM, Liu XY, Volanakis JE, Narayana SV. Structures of native and complexed complement factor D: implications of the atypical His57 conformation and self-inhibitory loop in the regulation of specific serine protease activity. J Mol Biol 1998;282(5):1061-1081.

123. Hof P, Mayr I, Huber R, Korzus E, Potempa J, Travis J, Powers JC, Bode W. The 1.8 Å crystal structure of human cathepsin G in complex with Suc-Val-Pro-PheP-(OPh)2: a Janus-faced proteinase with two opposite specificities. EMBO J 1996;15(20):5481-5491.

124. Matthews DA, Dragovich PS, Webber SE, Fuhrman SA, Patick AK, Zalman LS, Hendrickson TF, Love RA, Prins TJ, Marakovits JT, Zhou R, Tikhe J, Ford CE, Meador JW, Ferre RA, Brown EL, Binford SL, Brothers MA, DeLisle DM, Worland ST. Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. Proc Natl Acad Sci U S A 1999;96(20):11000-11007.

125. Wang X, Terzyan S, Tang J, Loy JA, Lin X, Zhang XC. Human plasminogen catalytic domain undergoes an unusual conformational change upon activation. J Mol Biol 2000;295(4):903-914.

126. Friedrich R, Bode W. Crystal Structure of Mtsp1 (Matriptase). Protein Data Bank deposition (PDB code 1EAX) 2001.

127. Berglund GI, Willassen NP, Hordvik A, Smalas AO. Structure of native pancreatic elastase from North Atlantic salmon at 1.61 Å resolution. Acta Crystallog sect D 1995;51:925-937.

128. Botos I, Meyer E, Nguyen M, Swanson SM, Koomen JM, Russell DH, Meyer EF. The structure of an insect chymotrypsin. J Mol Biol 2000;298(5):895-901.

129. Pasternak A, White A, Jeffery CJ, Medina N, Cahoon M, Ringe D, Hedstrom L. The energetic cost of induced fit catalysis: Crystal structures of trypsinogen mutants with enhanced activity and inhibitor affinity. Protein Sci 2001;10(7):1331-1342.

130. Waugh SM, Harris JL, Fletterick R, Craik CS. The structure of the pro-apoptotic protease granzyme B reveals the molecular determinants of its specificity. Nat Struct Biol 2000;7(9):762-765.

131. Tranter R, Read JA, Jones R, Brady RL. Effector sites in the three-dimensional structure of mammalian sperm beta-acrosin. Structure Fold Des 2000;8(11):1179-1188.

132. Fujinaga M, Chernaia MM, Halenbeck R, Koths K, James MN. The crystal structure of PR3, a neutrophil serine proteinase antigen of Wegener's granulomatosis antibodies. J Mol Biol 1996;261(2):267-278.

133. Hopfner KP, Kopetzki E, Kresse GB, Bode W, Huber R, Engh RA. New enzyme lineages by subdomain shuffling. Proc Natl Acad Sci U S A 1998;95(17):9813-9818.

134. Nar H, Bauer M, Schmid A, Stassen JM, Wienen W, Priepke HW, Kauffmann IK, Ries UJ, Hauel NH. Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors. Structure (Camb) 2001;9(1):29-37.

135. Katz BA, Sprengeler PA, Luong C, Verner E, Elrod K, Kirtley M, Janc J, Spencer JR, Breitenbucher JG, Hui H, McGee D, Allen D, Martelli A, Mackman RL. Engineering inhibitors

highly selective for the S1 sites of Ser190 trypsin-like serine protease drug targets. Chem Biol 2001;8(11):1107-1121.

136. Roussel A, Mathieu M, Dobbs A, Luu B, Cambillau C, Kellenberger C. Complexation of two proteic insect inhibitors to the active site of chymotrypsin suggests decoupled roles for binding and selectivity. J Biol Chem 2001;276(42):38893-38898.

137. Katona G, Wilmouth RC, Wright PA, Berglund GI, Hajdu J, Neutze R, Schofield CJ. X-ray structure of a serine protease acyl-enzyme complex at 0.95-Å resolution. J Biol Chem 2002;277(24):21962-21970.

138. Katona G, Berglund GI, Hajdu J, Graf L, Szilagyi L. Crystal structure reveals basis for the inhibitor resistance of human brain trypsin. J Mol Biol 2002;315(5):1209-1218.

139. Skordalakes E, Dodson GG, Green DS, Goodwin CA, Scully MF, Hudson HR, Kakkar VV, Deadman JJ. Inhibition of human alpha-thrombin by a phosphonate tripeptide proceeds via a metastable pentacoordinated phosphorus intermediate. J Mol Biol 2001;311(3):549-555.

140. Leiros HK, McSweeney SM, Smalas AO. Atomic resolution structures of trypsin provide insight into structural radiation damage. Acta Crystallog sect D 2001;57(Pt 4):488-497.

141. Rotonda J, Garcia-Calvo M, Bull HG, Geissler WM, McKeever BM, Willoughby CA, Thornberry NA, Becker JW. The three-dimensional structure of human granzyme B compared to caspase-3, key mediators of cell death with cleavage specificity for aspartic acid in P1. Chem Biol 2001;8(4):357-368.

142. Sichler K, Banner DW, D'Arcy A, Hopfner KP, Huber R, Bode W, Kresse GB, Kopetzki E, Brandstetter H. Crystal structures of uninhibited factor VIIa link its cofactor and substrate-assisted activation to specific interactions. J Mol Biol 2002;322(3):591-603.

143. Bernett MJ, Blaber SI, Scarisbrick IA, Dhanarajan P, Thompson SM, Blaber M. Crystal structure and biochemical characterization of human kallikrein 6 reveals that a trypsin-like kallikrein is expressed in the central nervous system. J Biol Chem 2002;277(27):24562-24570.

144. Barrette-Ng IH, Ng KK, Mark BL, Van Aken D, Cherney MM, Garen C, Kolodenko Y, Gorbalenya AE, Snijder EJ, James MN. Structure of arterivirus nsp4. The smallest chymotrypsin-like proteinase with an alpha/beta C-terminal extension and alternate conformations of the oxyanion hole. J Biol Chem 2002;277(42):39960-39966.

145. Huang Q, Liu S, Tang Y. Refined 1.6 Å resolution crystal structure of the complex formed between porcine beta-trypsin and MCTI-A, a trypsin inhibitor of the squash family. Detailed comparison with bovine beta-trypsin and its complex. J Mol Biol 1993;229(4):1022-1036.

146. Reiling KK, Krucinski J, Miercke LJ, Raymond WW, Caughey GH, Stroud RM. Structure of human pro-chymase: a model for the activating transition of granule-associated proteases. Biochemistry 2003;42(9):2616-2624.

147. Kishi T, Kato M, Shimizu T, Kato K, Matsumoto K, Yoshida S, Shiosaka S, Hakoshima T. Crystal structure of neuropsin, a hippocampal protease involved in kindling epileptogenesis. J Biol Chem 1999;274(7):4220-4224.

148. Bode W, Wei AZ, Huber R, Meyer E, Travis J, Neumann S. X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. EMBO J 1986;5(10):2453-2458.

149. Fujinaga M, James MN. Rat submaxillary gland serine protease, tonin. Structure solution and refinement at 1.8 Å resolution. J Mol Biol 1987;195(2):373-396.

150. Gaboriaud C, Serre L, Guy-Crotte O, Forest E, Fontecilla-Camps J-C. Crystal Structure of Human Trypsin 1: Unexpected Phosphorylation of Tyr151. J Mol Biol 1996;259(5):995-1010.

151. Martin PD, Malkowski MG, DiMaio J, Konishi Y, Ni F, Edwards BF. Bovine thrombin complexed with an uncleavable analog of residues 7-19 of fibrinogen A alpha: geometry of the catalytic triad and interactions of the P1', P2', and P3' substrate residues. Biochemistry 1996;35(40):13030-13039.

152. Broutin I, Arnoux B, Riche C, Lecroisey A, Keil B, Pascard C, Ducruix A. 1.8 Å Structure of *Hypoderma lineatum* Collagenase: a Member of the Serine Proteinase Family. Acta Crystallog sect D 1996;52:380-392.

153. Remington SJ, Woodbury RG, Reynolds RA, Matthews BW, Neurath H. The structure of rat mast cell protease II at 1.9-Å resolution. Biochemistry 1988;27(21):8097-8105.

154. Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics 2001;17(3):282-283.

155. Lin GD, Chattopadhyay D, Maki M, Wang KK, Carson M, Jin L, Yuen PW, Takano E, Hatanaka M, DeLucas LJ, Narayana SV. Crystal structure of calcium bound domain VI of calpain at 1.9 Å resolution and its role in enzyme assembly, regulation, and inhibitor binding. Nat Struct Biol 1997;4(7):539-547.

156. Kissinger CR, Parge HE, Knighton DR, Lewis CT, Pelletier LA, Tempczyk A, Kalish VJ, Tucker KD, Showalter RE, Moomaw EW, et al. Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex. Nature 1995;378(6557):641-644.

157. Swain AL, Kretsinger RH, Amma EL. Restrained least squares refinement of native (calcium) and cadmium-substituted carp parvalbumin using X-ray crystallographic data at 1.6-Å resolution. J Biol Chem 1989;264(28):16620-16628.

158. Li Y, Love ML, Putkey JA, Cohen C. Bepridil opens the regulatory N-terminal lobe of cardiac troponin C. Proc Natl Acad Sci U S A 2000;97(10):5140-5145.

159. Huang X, Poy F, Zhang R, Joachimiak A, Sudol M, Eck MJ. Structure of a WW domain containing fragment of dystrophin in complex with β-dystroglycan. Nat Struct Biol 2000;7(8):634-638.

160. Wilson MA, Brunger AT. The 1.0 Å crystal structure of Ca^{2+}-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. J Mol Biol 2000;301(5):1237-1256.

161. Schumacher MA, Rivard AF, Bachinger HP, Adelman JP. Structure of the gating domain of a Ca^{2+}-activated K^{+} channel complexed with Ca^{2+}/calmodulin. Nature 2001;410(6832):1120-1124.

162. Han BG, Han M, Sui H, Yaswen P, Walian PJ, Jap BK. Crystal structure of human calmodulin-like protein: insights into its functional role. FEBS Lett 2002;521(1-3):24-30.

163. Andersson M, Malmendal A, Linse S, Ivarsson I, Forsen S, Svensson LA. Structural basis for the negative allostery between Ca^{2+} - and Mg^{2+}-binding in the intracellular Ca^{2+} -receptor calbindin D_{9k}. Protein Sci 1997;6(6):1139-1147.

164. Zhang H, Wang G, Ding Y, Wang Z, Barraclough R, Rudland PS, Fernig DG, Rao Z. The crystal structure at 2Å resolution of the Ca^{2+} -binding protein S100P. J Mol Biol 2003;325(4):785-794.

165. Deng L, Vysotski E, Liu Z-J, Markova S, Lee J, Rose J, Wang B-C. The Crystal Structure Of Obelin From *Obelia geniculata* At 1.82 Å Resolution. Protein Data Bank deposition (PDB code 1JF0) 2001.

166. Otterbein LR, Kordowska J, Witte-Hoffmann C, Wang CL, Dominguez R. Crystal structures of S100A6 in the Ca^{2+}-free and Ca^{2+}-bound states: the calcium sensor mechanism of S100 proteins revealed at atomic resolution. Structure (Camb) 2002;10(4):557-567.

167. Jia J, Borregaard N, Lollike K, Cygler M. Structure of Ca^{2+}-loaded human grancalcin. Acta Crystallog sect D 2001;57(Pt 12):1843-1849.

168. Mittl PR, Fritz G, Sargent DF, Richmond TJ, Heizmann CW, Grutter MG. Metal-free MIRAS phasing: structure of apo-S100A3. Acta Crystallog sect D 2002;58(Pt 8):1255-1261.

169. Terrak M, Wu G, Stafford WF, Lu RC, Dominguez R. Two distinct myosin light chain structures are induced by specific variations within the bound IQ motifs-functional implications. EMBO J 2003;22(3):362-371.

170. Ishikawa K, Nakagawa A, Tanaka I, Suzuki M, Nishihira J. The structure of human MRP8, a member of the S100 calcium-binding protein family, by MAD phasing at 1.9 Å resolution. Acta Crystallog sect D 2000;56 (Pt 5):559-566.

171. Rao ST, Satyshur KA, Greaser ML, Sundaralingam M. X-ray Structures of Mn, Cd and Tb Metal Complexes of Troponin C. Acta Crystallog sect D 1996;52:916-922.

172. Coureux PD, Wells AL, Menetrey J, Yengo CM, Morris CA, Sweeney HL, Houdusse A. A structural state of the myosin V motor without bound nucleotide. Nature 2003;425(6956):419-423.

173. Weiergraber OH, Senin, II, Philippov PP, Granzin J, Koch KW. Impact of N-terminal myristoylation on the Ca2+-dependent conformational transition in recoverin. J Biol Chem 2003;278(25):22972-22979.

174. Brodersen DE, Etzerodt M, Madsen P, Celis JE, Thogersen HC, Nyborg J, Kjeldgaard M. EF-hands at atomic resolution: the structure of human psoriasin (S100A7) solved by MAD phasing. Structure 1998;6(4):477-489.

175. Liu ZJ, Vysotski ES, Deng L, Lee J, Rose J, Wang BC. Atomic resolution structure of obelin: soaking with calcium enhances electron density of the second oxygen atom substituted at the C2-position of coelenterazine. Biochem Biophys Res Commun 2003;311(2):433-439.

176. Zhou W, Qian Y, Kunjilwar K, Pfaffinger PJ, Choe S. Structural insights into the functional interaction of KChIP1 with Shal-type K(+) channels. Neuron 2004;41(4):573-586.

177. Nagae M, Nozawa A, Koizumi N, Sano H, Hashimoto H, Sato M, Shimizu T. The crystal structure of the novel calcium-binding protein AtCBL2 from Arabidopsis thaliana. J Biol Chem 2003;278(43):42240-42246.

178. Houdusse A, Cohen C. Structure of the regulatory domain of scallop myosin at 2 Å resolution: implications for regulation. Structure 1996;4(1):21-32.

179. Meng W, Sawasdikosol S, Burakoff SJ, Eck MJ. Structure of the amino-terminal domain of Cbl complexed to its binding site on ZAP-70 kinase. Nature 1999;398(6722):84-90.

180. Allaire M, Li Y, MacKenzie RE, Cygler M. The 3-D structure of a folate-dependent dehydrogenase/cyclohydrolase bifunctional enzyme at 1.5 Å resolution. Structure 1998;6(2):173-182.

181. Auerbach G, Ostendorp R, Prade L, Korndorfer I, Dams T, Huber R, Jaenicke R. Lactate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima*: the crystal structure at 2.1 Å resolution reveals strategies for intrinsic protein stabilization. Structure 1998;6(6):769-781.

182. Benach J, Atrian S, Gonzalez-Duarte R, Ladenstein R. The catalytic reaction and inhibition mechanism of *Drosophila* alcohol dehydrogenase: observation of an enzyme-bound NAD-ketone adduct at 1.4 Å resolution by X-ray crystallography. J Mol Biol 1999;289(2):335-355.

183. Isupov MN, Fleming TM, Dalby AR, Crowhurst GS, Bourne PC, Littlechild JA. Crystal structure of the glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. J Mol Biol 1999;291(3):651-660.

184. Kim SY, Hwang KY, Kim SH, Sung HC, Han YS, Cho Y. Structural basis for cold adaptation. Sequence, biochemical properties, and crystal structure of malate dehydrogenase from a psychrophile *Aquaspirillium arcticum*. J Biol Chem 1999;274(17):11761-11767.

185. Hulsmeyer M, Hecht HJ, Niefind K, Hofer B, Eltis LD, Timmis KN, Schomburg D. Crystal structure of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase from a PCB degrader at 2.0 Å resolution. Protein Sci 1998;7(6):1286-1293.

186. Britton KL, Asano Y, Rice DW. Crystal structure and active site location of N-(1-D-carboxylethyl)-L-norvaline dehydrogenase. Nat Struct Biol 1998;5(7):593-601.

187. Stillman TJ, Baker PJ, Britton KL, Rice DW. Conformational flexibility in glutamate dehydrogenase. Role of water in substrate recognition and catalysis. J Mol Biol 1993;234(4):1131-1139.

188. Hegeman AD, Gross JW, Frey PA. Probing catalysis by Escherichia coli dTDP-glucose-4,6-dehydratase: identification and preliminary characterization of functional amino acid residues at the active site. Biochemistry 2001;40(22):6598-6610.

189. Ramaswamy S, el Ahmad M, Danielsson O, Jornvall H, Eklund H. Crystal structure of cod liver class I alcohol dehydrogenase: substrate pocket and structurally variable segments. Protein Sci 1996;5(4):663-671.

190. Charron C, Talfournier F, Isupov MN, Branlant G, Littlechild JA, Vitoux B, Aubry A. Crystallization and preliminary X-ray diffraction studies of D-glyceraldehyde-3-phosphate

dehydrogenase from the hyperthermophilic archaeon *Methanothermus fervidus*. Acta Crystallog sect D 1999;55 (Pt 7):1353-1355.

191. Tanaka N, Nonaka T, Nakanishi M, Deyashiki Y, Hara A, Mitsui Y. Crystal structure of the ternary complex of mouse lung carbonyl reductase at 1.8 Å resolution: the structural origin of coenzyme specificity in the short-chain dehydrogenase/reductase family. Structure 1996;4(1):33-45.

192. Roujeinikova A, Levy CW, Rowsell S, Sedelnikova S, Baker PJ, Minshull CA, Mistry A, Colls JG, Camble R, Stuitje AR, Slabas AR, Rafferty JB, Pauptit RA, Viner R, Rice DW. Crystallographic analysis of triclosan bound to enoyl reductase. J Mol Biol 1999;294(2):527-535.

193. Campbell RE, Mosimann SC, van De Rijn I, Tanner ME, Strynadka NC. The first structure of UDP-glucose dehydrogenase reveals the catalytic residues necessary for the two-fold oxidation. Biochemistry 2000;39(23):7012-7023.

194. Rowland P, Basak AK, Gover S, Levy HR, Adams MJ. The three-dimensional structure of glucose 6-phosphate dehydrogenase from *Leuconostoc mesenteroides* refined at 2.0 Å resolution. Structure 1994;2(11):1073-1087.

195. Song SY, Xu YB, Lin ZJ, Tsou CL. Structure of active site carboxymethylated D-glyceraldehyde-3-phosphate dehydrogenase from *Palinurus versicolor*. J Mol Biol 1999;287(4):719-725.

196. Svensson S, Hoog JO, Schneider G, Sandalova T. Crystal structures of mouse class II alcohol dehydrogenase reveal determinants of substrate specificity and catalytic efficiency. J Mol Biol 2000;302(2):441-453.

197. Johansson E, Steffens JJ, Lindqvist Y, Schneider G. Crystal structure of saccharopine reductase from *Magnaporthe grisea*, an enzyme of the α-aminoadipate pathway of lysine biosynthesis. Structure Fold Des 2000;8(10):1037-1047.

198. Rosano C, Bisso A, Izzo G, Tonetti M, Sturla L, De Flora A, Bolognesi M. Probing the catalytic mechanism of GDP-4-keto-6-deoxy-d-mannose Epimerase/Reductase by kinetic and crystallographic characterization of site-specific mutants. J Mol Biol 2000;303(1):77-91.

199. Powell AJ, Read JA, Banfield MJ, Gunn-Moore F, Yan SD, Lustbader J, Stern AR, Stern DM, Brady RL. Recognition of structurally diverse substrates by type II 3-hydroxyacyl-CoA dehydrogenase (HADH II)/amyloid-β binding alcohol dehydrogenase (ABAD). J Mol Biol 2000;303(2):311-327.

200. Gourley DG, Schuttelkopf AW, Leonard GA, Luba J, Hardy LW, Beverley SM, Hunter WN. Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites. Nat Struct Biol 2001;8(6):521-525.

201. Thoden JB, Wohlers TM, Fridovich-Keil JL, Holden HM. Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase. Biochemistry 2000;39(19):5691-5701.

202. Deacon AM, Ni YS, Coleman WG, Jr., Ealick SE. The crystal structure of ADP-L-glycero-D-mannoheptose 6-epimerase: catalysis with a twist. Structure Fold Des 2000;8(5):453-462.

203. Fraser ME, James MN, Bridger WA, Wolodko WT. Phosphorylated and dephosphorylated structures of pig heart, GTP-specific succinyl-CoA synthetase. J Mol Biol 2000;299(5):1325-1339.

204. Suresh S, Turley S, Opperdoes FR, Michels PA, Hol WG. A potential target enzyme for trypanocidal drugs revealed by the crystal structure of NAD-dependent glycerol-3-phosphate dehydrogenase from *Leishmania mexicana*. Structure Fold Des 2000;8(5):541-552.

205. Cirilli M, Scapin G, Sutherland A, Vederas JC, Blanchard JS. The three-dimensional structure of the ternary complex of *Corynebacterium glutamicum* diaminopimelate dehydrogenase-NADPH-L-2-amino-6-methylene-pimelate. Protein Sci 2000;9(10):2034-2037.

206. Barycki JJ, O'Brien LK, Strauss AW, Banaszak LJ. Sequestration of the active site by interdomain shifting. Crystallographic and spectroscopic evidence for distinct conformations of L-3-hydroxyacyl-CoA dehydrogenase. J Biol Chem 2000;275(35):27186-27196.

207. Grimm C, Maser E, Mobus E, Klebe G, Reuter K, Ficner R. The crystal structure of 3α -hydroxysteroid dehydrogenase/carbonyl reductase from *Comamonas testosteroni* shows a novel oligomerization pattern within the short chain dehydrogenase/reductase family. J Biol Chem 2000;275(52):41333-41339.

208. Tanaka N, Nonaka T, Tanabe T, Yoshimoto T, Tsuru D, Mitsui Y. Crystal structures of the binary and ternary complexes of 7α-hydroxysteroid dehydrogenase from *Escherichia coli*. Biochemistry 1996;35(24):7715-7730.

209. Liao D, Basarab GS, Gatenby AA, Valent B, Jordan DB. Structures of trihydroxynaphthalene reductase-fungicide complexes: implications for structure-based design and catalysis. Structure (Camb) 2001;9(1):19-27.

210. Duee E, Olivier-Deyris L, Fanchon E, Corbier C, Branlant G, Dideberg O. Comparison of the structures of wild-type and a N313T mutant of *Escherichia coli* glyceraldehyde 3-phosphate dehydrogenases: implication for NAD binding and cooperativity. J Mol Biol 1996;257(4):814-838.

211. Skarzynski T, Moody PC, Wonacott AJ. Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. J Mol Biol 1987;193(1):171-187.

212. Yamamoto K, Kurisu G, Kusunoki M, Tabata S, Urabe I, Osaki S. Crystal structure of glucose dehydrogenase mutant Q252L complexed with NAD+. Protein Data Bank deposition (PDB code 1GEE) 2000.

213. Otagiri M, Kurisu G, Ui S, Takusagawa Y, Ohkuma M, Kudo T, Kusunoki M. Crystal structure of meso-2,3-butanediol dehydrogenase in a complex with NAD^+ and inhibitor mercaptoethanol at 1.7 Å resolution for understanding of chiral substrate recognition mechanisms. J Biochem (Tokyo) 2001;129(2):205-208.

214. Moser J, Schubert WD, Beier V, Bringemeier I, Jahn D, Heinz DW. V-shaped structure of glutamyl-tRNA reductase, the first enzyme of tRNA-dependent tetrapyrrole biosynthesis. EMBO J 2001;20(23):6583-6590.

215. Norman RA, McAlister MS, Murray-Rust J, Movahedzadeh F, Stoker NG, McDonald NQ. Crystal structure of inositol 1-phosphate synthase from *Mycobacterium tuberculosis*, a key enzyme in phosphatidylinositol synthesis. Structure (Camb) 2002;10(3):393-402.

216. Airenne TT, Torkko JM, Van den plas S, Sormunen RT, Kastaniotis AJ, Wierenga RK, Hiltunen JK. Structure-function analysis of enoyl thioester reductase involved in mitochondrial maintenance. J Mol Biol 2003;327(1):47-59.

217. Shaw MP, Bond CS, Roper JR, Gourley DG, Ferguson MA, Hunter WN. High-resolution crystal structure of *Trypanosoma brucei* UDP-galactose 4'-epimerase: a potential target for structure-based development of novel trypanocides. Mol Biochem Parasitol 2003;126(2):173-180.

218. Guy JE, Isupov MN, Littlechild JA. The structure of an alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix*. J Mol Biol 2003;331(5):1041-1051.

219. Horer S, Stoop J, Mooibroek H, Baumann U, Sassoon J. The crystallographic structure of the mannitol 2-dehydrogenase $NADP^+$ binary complex from *Agaricus bisporus*. J Biol Chem 2001;276(29):27555-27561.

220. Nurizzo D, Halbig D, Sprenger GA, Baker EN. Crystal structures of the precursor form of glucose-fructose oxidoreductase from *Zymomonas mobilis* and its complexes with bound ligands. Biochemistry 2001;40(46):13857-13867.

221. Pereira PJ, Macedo-Ribeiro S, Parraga A, Perez-Luque R, Cunningham O, Darcy K, Mantle TJ, Coll M. Structure of human biliverdin IXβ reductase, an early fetal bilirubin IXβ producing enzyme. Nature structural biology 2001;8(3):215-220.

222. Meijers R, Morris RJ, Adolph HW, Merli A, Lamzin VS, Cedergren-Zeppezauer ES. On the enzymatic activation of NADH. J Biol Chem 2001;276(12):9316-9321.

223. Niederhut MS, Gibbons BJ, Perez-Miller S, Hurley TD. Three-dimensional structures of the three human class I alcohol dehydrogenases. Protein Sci 2001;10(4):697-706.

224. Benach J, Filling C, Oppermann UC, Roversi P, Bricogne G, Berndt KD, Jornvall H, Ladenstein R. Structure of bacterial 3beta/17beta-hydroxysteroid dehydrogenase at 1.2 Å resolution: a model for multiple steroid recognition. Biochemistry 2002;41(50):14659-14668.

225. Theisen MJ, Sanda SL, Ginell SL, Benning C, Garavito RM. High Resolution Crystal Structure Of The Wild-Type Protein SQD1, With NAD And UDP-Glucose. Protein Data Bank deposition (PDB code 1I24) 2001.

226. Korolev SV, Dementieva IS, Christendat D, Edwards A, Joachimiak A. Structural similarities of Mth1747 hypothetical protein from *Methanobacterium thermoautotrophicum* with 3-hydroxyacid dehydrogenases. Protein Data Bank deposition (PDB code 1I36) 2001.

227. Wada T, Shirouzu M, Terada T, Ishizuka Y, Matsuda T, Kigawa T, Kuramitsu S, Park SY, Tame JR, Yokoyama S. Structure of a conserved CoA-binding protein synthesized by a cell-free system. Acta Crystallog sect D 2003;59(Pt 7):1213-1218.

228. Sogabe S, Yoshizumi A, Fukami TA, Shiratori Y, Shimizu S, Takagi H, Nakamori S, Wada M. The crystal structure and stereospecificity of levodione reductase from Corynebacterium aquaticum M-13. J Biol Chem 2003;278(21):19387-19395.

229. Lokanath NK, Shiromizu I, Nodake Y, Sugahara M, Yokoyama S, Kuramitsu S, Miyano M, Kunishima N. Crystallization and preliminary X-ray crystallographic studies of NADP-dependent 3-hydroxyisobutyrate dehydrogenase from Thermus thermophilus HB8. Acta Crystallog sect D 2003;59(Pt 12):2294-2296.

230. Razeto A, Kochhar S, Hottinger H, Dauter M, Wilson KS, Lamzin VS. Domain closure, substrate specificity and catalysis of D-lactate dehydrogenase from *Lactobacillus bulgaricus*. J Mol Biol 2002;318(1):109-119.

231. Genomics JCfS. Crystal structure of aspartate dehydrogenase (TM1643) from *Thermotoga maritima* at 1.9 Å resolution. Protein Data Bank deposition (PDB code 1J5P) 2002.

232. Liao DI, Thompson JE, Fahnestock S, Valent B, Jordan DB. A structural account of substrate and inhibitor specificity differences between two naphthol reductases. Biochemistry 2001;40(30):8696-8704.

233. Warkentin E, Mamat B, Sordel-Klippert M, Wicke M, Thauer RK, Iwata M, Iwata S, Ermler U, Shima S. Structures of F420H2:$NADP^+$ oxidoreductase with and without its substrates bound. EMBO J 2001;20(23):6561-6569.

234. Bogin O, Levin I, Hacham Y, Tel-Or S, Peretz M, Frolow F, Burstein Y. Structural basis for the enhanced thermal stability of alcohol dehydrogenase mutants from the mesophilic bacterium *Clostridium beijerinckii*: contribution of salt bridging. Protein Sci 2002;11(11):2561-2574.

235. Gangloff A, Shi R, Nahoum V, Lin SX. Pseudo-symmetry of C19 steroids, alternative binding orientations, and multispecificity in human estrogenic 17beta-hydroxysteroid dehydrogenase. Faseb J 2003;17(2):274-276.

236. Esposito L, Sica F, Raia CA, Giordano A, Rossi M, Mazzarella L, Zagari A. Crystal structure of the alcohol dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus* at 1.85 Å resolution. J Mol Biol 2002;318(2):463-477.

237. Pavao F, Castilho MS, Pupo MT, Dias RL, Correa AG, Fernandes JB, da Silva MF, Mafezoli J, Vieira PC, Oliva G. Structure of Trypanosoma cruzi glycosomal glyceraldehyde-3-phosphate dehydrogenase complexed with chalepin, a natural product inhibitor, at 1.95 A resolution. FEBS Lett 2002;520(1-3):13-17.

238. Stammers DK, Ren J, Leslie K, Nichols CE, Lamb HK, Cocklin S, Dodds A, Hawkins AR. The structure of the negative transcriptional regulator NmrA reveals a structural superfamily which includes the short-chain dehydrogenase/reductases. EMBO J 2001;20(23):6619-6626.

239. Allard ST, Beis K, Giraud MF, Hegeman AD, Gross JW, Wilmouth RC, Whitfield C, Graninger M, Messner P, Allen AG, Maskell DJ, Naismith JH. Toward a structural understanding of the dehydratase mechanism. Structure (Camb) 2002;10(1):81-92.

240. Tanaka N, Kusakabe Y, Ito K, Yoshimoto T, Nakamura KT. Crystal structure of formaldehyde dehydrogenase from *Pseudomonas putida*: the structural origin of the tightly bound cofactor in nicotinoprotein dehydrogenases. J Mol Biol 2002;324(3):519-533.

241. Matak-Vinkovic D, Vinkovic M, Saldanha SA, Ashurst JL, von Delft F, Inoue T, Miguel RN, Smith AG, Blundell TL, Abell C. Crystal structure of *Escherichia coli* ketopantoate reductase at 1.7 Å resolution and insight into the enzyme mechanism. Biochemistry 2001;40(48):14493-14500.

242. Prasad GS, Wahlberg M, Sridhar V, Sundaresan V, Yamaguchi M, Hatefi Y, Stout CD. Crystal structures of transhydrogenase domain I with and without bound NADH. Biochemistry 2002;41(42):12745-12754.

243. Whitby FG, Phillips JD, Hill CP, McCoubrey W, Maines MD. Crystal structure of a biliverdin IXalpha reductase enzyme-cofactor complex. J Mol Biol 2002;319(5):1199-1210.

244. Yang X, Hu Y, Yin DH, Turner MA, Wang M, Borchardt RT, Howell PL, Kuczera K, Schowen RL. Catalytic strategy of S-adenosyl-L-homocysteine hydrolase: transition-state stabilization and the avoidance of abortive reactions. Biochemistry 2003;42(7):1900-1909.

245. Kavanagh KL, Klimacek M, Nidetzky B, Wilson DK. Crystal structure of *Pseudomonas fluorescens* mannitol 2-dehydrogenase binary and ternary complexes. Specificity and catalytic mechanism. J Biol Chem 2002;277(45):43433-43442.

246. Ermler U, Hagemeier CH, Roth A, Demmer U, Grabarse W, Warkentin E, Vorholt JA. Structure of methylene-tetrahydromethanopterin dehydrogenase from *Methylobacterium extorquens* AM1. Structure (Camb) 2002;10(8):1127-1137.

247. Sanghani PC, Robinson H, Bosron WF, Hurley TD. Human glutathione-dependent formaldehyde dehydrogenase. Structures of apo, binary, and inhibitory ternary complexes. Biochemistry 2002;41(35):10778-10786.

248. Blanco J, Moore RA, Kabaleeswaran V, Viola RE. A structural basis for the mechanism of aspartate-β-semialdehyde dehydrogenase from *Vibrio cholerae*. Protein Sci 2003;12(1):27-33.

249. Benach J, Winberg JO, Svendsen JS, Atrian S, Gonzalez-Duarte R, Ladenstein R. *Drosophila* alcohol dehydrogenase: acetate-enzyme interactions and novel insights into the effects of electrostatics on catalysis. J Mol Biol 2005;345(3):579-598.

250. Snook CF, Tipton PA, Beamer LJ. Crystal structure of GDP-mannose dehydrogenase: a key enzyme of alginate biosynthesis in *P. aeruginosa*. Biochemistry 2003;42(16):4658-4668.

251. Kumar V, Carlson JE, Ohgi KA, Edwards TA, Rose DW, Escalante CR, Rosenfeld MG, Aggarwal AK. Transcription corepressor CtBP is an NAD^+-regulated dehydrogenase. Mol Cell 2002;10(4):857-869.

252. Blankenfeldt W, Kerr ID, Giraud MF, McMiken HJ, Leonard G, Whitfield C, Messner P, Graninger M, Naismith JH. Variation on a theme of SDR. dTDP-6-deoxy-L- lyxo-4-hexulose reductase (RmlD) shows a new Mg^{2+}-dependent dimerization mode. Structure (Camb) 2002;10(6):773-786.

253. Mulichak AM, Bonin CP, Reiter WD, Garavito RM. Structure of the MUR1 GDP-mannose 4,6-dehydratase from *Arabidopsis thaliana*: implications for ligand binding and specificity. Biochemistry 2002;41(52):15578-15589.

254. Ahn HJ, Eom SJ, Yoon HJ, Lee BI, Cho H, Suh SW. Crystal structure of class I acetohydroxy acid isomeroreductase from *Pseudomonas aeruginosa*. J Mol Biol 2003;328(2):505-515.

255. Korolev S, Koroleva O, Zarembinski T, Collart F, Joachimiak A. Structure of shikimate 5-dehydrogenase-like protein HI0607. Protein Data Bank deposition (PDB code 1NPY) 2003.

256. Manjasetty BA, Powlowski J, Vrielink A. Crystal structure of a bifunctional aldolase-dehydrogenase: sequestering a reactive and volatile intermediate. Proc Natl Acad Sci U S A 2003;100(12):6992-6997.

257. Blanco J, Moore RA, Viola RE. Capture of an intermediate in the catalytic cycle of L-aspartate-beta-semialdehyde dehydrogenase. Proc Natl Acad Sci U S A 2003;100(22):12613-12617.

258. Niefind K, Muller J, Riebel B, Hummel W, Schomburg D. The crystal structure of R-specific alcohol dehydrogenase from *Lactobacillus brevis* suggests the structural basis of its metal dependency. J Mol Biol 2003;327(2):317-328.

259. Michel G, Roszak AW, Sauve V, Maclean J, Matte A, Coggins JR, Cygler M, Lapthorn AJ. Structures of shikimate dehydrogenase AroE and its Paralog YdiB. A common structural framework for different activities. J Biol Chem 2003;278(21):19463-19472.

260. Rao GS, Coleman DE, Karsten WE, Cook PF, Harris BG. Crystallographic studies on *Ascaris suum* NAD-malic enzyme bound to reduced cofactor and identification of an effector site. J Biol Chem 2003;278(39):38051-38058.

261. Auerbach G, Herrmann A, Gutlich M, Fischer M, Jacob U, Bacher A, Huber R. The 1.25 Å crystal structure of sepiapterin reductase reveals its binding mode to pterins and brain neurotransmitters. EMBO J 1997;16(24):7219-7230.

262. Lodge JA, Maier T, Liebl W, Hoffmann V, Strater N. Crystal structure of *Thermotoga maritima* α-glucosidase AglA defines a new clan of NAD^+-dependent glycosidases. J Biol Chem 2003;278(21):19151-19158.

263. Antonyuk SV, Eady RR, Strange RW, Hasnain SS. The structure of glyceraldehyde 3-phosphate dehydrogenase from *Alcaligenes xylosoxidans* at 1.7 Å resolution. Acta Crystallog sect D 2003;59(Pt 5):835-842.

264. Beis K, Allard ST, Hegeman AD, Murshudov G, Philp D, Naismith JH. The structure of NADH in the enzyme dTDP-d-glucose dehydratase (RmlB). J Am Chem Soc 2003;125(39):11872-11878.

265. Takahashi H, Tokunaga, Y., Kuroishi, C.,, Babayeba N, Kuramitsu, S., Yokoyama, S.,, Miyano M, Tahirov, T.H. The Crystal Structure of Succinyl-Coa Synthetase from *Thermus thermophilus*. Protein Data Bank deposition (PDB code 1OI7) 2003.

266. Koropatkin NM, Liu HW, Holden HM. High resolution x-ray structure of tyvelose epimerase from *Salmonella typhi*. J Biol Chem 2003;278(23):20874-20881.

267. Rosell A, Valencia E, Pares X, Fita I, Farres J, Ochoa WF. Crystal structure of the vertebrate NADP(H)-dependent alcohol dehydrogenase (ADH8). J Mol Biol 2003;330(1):75-85.

268. Jin X, Geiger JH. Structures of NAD^+- and NADH-bound 1-l-myo-inositol 1-phosphate synthase. Acta Crystallog sect D 2003;59(Pt 7):1154-1164.

269. Ye S, Von Delft F, Brooun A, Knuth MW, Swanson RV, McRee DE. The crystal structure of shikimate dehydrogenase (AroE) reveals a unique NADPH binding mode. J Bacteriol 2003;185(14):4144-4151.

270. Tao X, Yang Z, Tong L. Crystal structures of substrate complexes of malic enzyme and insights into the catalytic mechanism. Structure (Camb) 2003;11(9):1141-1150.

271. Baker PJ, Sawa Y, Shibata H, Sedelnikova SE, Rice DW. Analysis of the structure and substrate binding of *Phormidium lapideum* alanine dehydrogenase. Nat Struct Biol 1998;5(7):561-567.

272. Pauly TA, Ekstrom JL, Beebe DA, Chrunyk B, Cunningham D, Griffor M, Kamath A, Lee SE, Madura R, McGuire D, Subashi T, Wasilko D, Watts P, Mylari BL, Oates PJ, Adams PD, Rath VL. X-ray crystallographic and kinetic studies of human sorbitol dehydrogenase. Structure (Camb) 2003;11(9):1071-1085.

273. El-Kabbani O, Ishikura S, Darmanin C, Carbone V, Chung RP, Usami N, Hara A. Crystal structure of human L-xylulose reductase holoenzyme: probing the role of Asn107 with site-directed mutagenesis. Proteins 2004;55(3):724-732.

274. de Jong RM, Tiesinga JJ, Rozeboom HJ, Kalk KH, Tang L, Janssen DB, Dijkstra BW. Structure and mechanism of a bacterial haloalcohol dehalogenase: a new variation of the short-chain dehydrogenase/reductase fold without an NAD(P)H binding site. EMBO J 2003;22(19):4933-4944.

275. Kavanagh KL, Elling RA, Wilson DK. Structure of Toxoplasma gondii LDH1: active-site differences from human lactate dehydrogenases and the structural basis for efficient APAD+ use. Biochemistry 2004;43(4):879-889.

276. Price AC, Zhang YM, Rock CO, White SW. Cofactor-induced conformational rearrangements establish a catalytically competent active site and a proton relay conduit in FabG. Structure (Camb) 2004;12(3):417-428.

277. Thomazeau K, Dumas R, Halgand F, Forest E, Douce R, Biou V. Structure of spinach acetohydroxyacid isomeroreductase complexed with its reaction product dihydroxymethylvalerate, manganese and (phospho)-ADP-ribose. Acta Crystallographica Section D, Biological crystallography 2000;56 (Pt 4):389-397.

278. Stewart MJ, Parikh S, Xiao G, Tonge PJ, Kisker C. Structural basis and mechanism of enoyl reductase inhibition by triclosan. J Mol Biol 1999;290(4):859-865.

279. Allard ST, Cleland WW, Holden HM. High resolution X-ray structure of dTDP-glucose 4,6-dehydratase from Streptomyces venezuelae. J Biol Chem 2004;279(3):2211-2220.

280. Powers JP, Rozek A, Hancock RE. Structure-activity relationships for the beta-hairpin cationic antimicrobial peptide polyphemusin I. Biochim Biophys Acta 2004;1698(2):239-250.

281. Webb NA, Mulichak AM, Lam JS, Rocchetta HL, Garavito RM. Crystal structure of a tetrameric GDP-D-mannose 4,6-dehydratase from a bacterial GDP-D-rhamnose biosynthetic pathway. Protein Sci 2004;13(2):529-539.

282. Vedadi M, Walker JR, Sharma S, Houston S, Wasney G, Loppnau P, Oppermann U. Crystal Structure and Biophysical Characterization of Human GDP-D-mannose 4,6-dehydratase. Protein Data Bank deposition (PDB code 1T2A) 2004.

283. Cameron A, Read J, Tranter R, Winter VJ, Sessions RB, Brady RL, Vivas L, Easton A, Kendrick H, Croft SL, Barros D, Lavandera JL, Martin JJ, Risco F, Garcia-Ochoa S, Gamo FJ, Sanz L, Leon L, Ruiz JR, Gabarro R, Mallo A, Gomez de las Heras F. Identification and activity of a series of azole-based compounds with lactate dehydrogenase-directed anti-malarial activity. J Biol Chem 2004;279(30):31429-31439.

284. Kunishima NA, Y. Yokoyama, S. Kuramitsu, S. Miyano, M. Crystal structure of type II 3-hydroxyacyl-CoA dehydrogenase from *Thermus thermophilus* HB8. Protein Data Bank deposition (PDB code 1UAY) 2003.

285. Thoden JB, Hegeman AD, Wesenberg G, Chapeau MC, Frey PA, Holden HM. Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from *Escherichia coli*. Biochemistry 1997;36(21):6294-6304.

286. Bjork A, Dalhus B, Mantzilas D, Eijsink VG, Sirevag R. Stabilization of a tetrameric malate dehydrogenase by introduction of a disulfide bridge at the dimer-dimer interface. J Mol Biol 2003;334(4):811-821.

287. Kobayashi N, Kigawa, T., Koshiba, S., Tochio,, N. I, M., Yokoyama, S. Solution structure of the first SH3 domain of human intersectin2 (KIAA1256). Protein Data Bank deposition (PDB code 1UFF) 2003.

288. GenomiX S. Crystal structure of shikimate-5-dehydrogenase with NAD. Protein Data Bank deposition (PDB code 1VI2) 2003.

289. Genomics JCfS. Crystal structure of alcohol dehydrogenase (TM0436) from *Thermotoga maritima* at 2.00 Å resolution. Protein Data Bank deposition (PDB code 1VJ0) 2003.

290. Levin I, Schwarzenbacher R, McMullan D, Abdubek P, Ambing E, Biorac T, Cambell J, Canaves JM, Chiu HJ, Dai X, Deacon AM, DiDonato M, Elsliger MA, Godzik A, Grittini C, Grzechnik SK, Hampton E, Jaroszewski L, Karlak C, Klock HE, Koesema E, Kreusch A, Kuhn P, Lesley SA, McPhillips TM, Miller MD, Morse A, Moy K, Ouyang J, Page R, Quijano K, Reyes R, Robb A, Sims E, Spraggon G, Stevens RC, van den Bedem H, Velasquez J, Vincent J, von Delft F, Wang X, West B, Wolf G, Xu Q, Hodgson KO, Wooley J, Wilson IA. Crystal structure of a putative NADPH-dependent oxidoreductase (GI: 18204011) from mouse at 2.10 A resolution. Proteins 2004;56(3):629-633.

291. Genomics JCfS. Crystal structure of MYO-inositol-1-phosphate synthase-related protein (TM1419) from Thermotoga maritima at 1.70 A resolution. Protein Data Bank deposition (PDB code 1VJP) 2004.

292. Yamashita A, Kato H, Wakatsuki S, Tomizaki T, Nakatsu T, Nakajima K, Hashimoto T, Yamada Y, Oda J. Structure of tropinone reductase-II complexed with $NADP^+$ and pseudotropine at 1.9 Å resolution: implication for stereospecific substrate binding and catalysis. Biochemistry 1999;38(24):7630-7637.

293. Lamzin VS, Dauter Z, Popov VO, Harutyunyan EH, Wilson KS. High resolution structures of holo and apo formate dehydrogenase. J Mol Biol 1994;236(3):759-785.

294. Adams MJ, Gover S, Leaback R, Phillips C, Somers DO. The structure of 6-phosphogluconate dehydrogenase refined at 2.5 Å resolution. Acta Crystallogr B 1991;47 (Pt 5):817-820.

Chapter 2

Analysis of Structure and Solvent of Rat α-Parvalbumin at 1.05 Å Resolution

(adapted from Bottoms CA, Schuermann JP, Agah S, Henzl MT, Tanner JJ. Protein Sci 2004;13:1724-1734.)

Chapter 2 Abstract

The crystal structure of rat alpha-parvalbumin has been determined at 1.05 Å resolution, using synchrotron data collected at Advanced Photon Source beamline 19-ID. After refinement with SHELX, employing anisotropic displacement parameters and riding hydrogen atoms, R = 0.132 and Rfree = 0.162. The average coordinate estimated standard deviations are 0.021 Å and 0.038 Å for backbone atoms and side-chain atoms, respectively. Besides providing a more precise view of the alpha isoform than previously available, these data permit comparison with the 0.91 Å structure determined for pike beta-parvalbumin. Visualization of the anisotropic displacement parameters as thermal ellipsoids yields insight into the atomic motion within the Ca^{2+}-binding sites. The asymmetric unit includes three parvalbumin molecules. Interestingly, the EF site in one displays uncharacteristic flexibility. The ellipsoids for Asp-92 are particularly large and nonspherical, and the shape of the Ca^{2+} ellipsoid implies significant vibrational motion perpendicular to the plane defined by the four *y* and *z* ligands. The relative dearth of crystal-packing interactions in this site suggests that the heightened flexibility may be the result of diminished intermolecular contacts. The implication is that, by impeding conformational mobility, crystal-packing forces may cause serious overestimation of EF-hand rigidity. The high quality of the data permitted eleven residues to be modeled in alternative side-chain conformations, including the two core residues, Ile-97 and Leu-105. The discrete disorder observed for Ile-97 may have functional ramifications, providing a mechanism for communicating binding status between the CD and EF binding loops and between the parvalbumin metal ion-binding domain and the N-terminal AB region.

Keywords: parvalbumin, EF-hand, Ca^{2+}-binding proteins, atomic resolution crystallography

Abbreviations: PV, parvalbumin; RMSD, root mean square difference; e.s.d., estimated standard deviation; PDB, Protein Data Bank; ADP, anisotropic displacement parameter.

2.1 Introduction

EF-hand proteins are essential components of the eukaryotic Ca^{2+} signaling machinery.[1] Certain members of this large protein family function as Ca^{2+}-dependent signal transducers, calmodulin and troponin C being the archetypal examples. Others function as soluble Ca^{2+} buffers, shaping the size and duration of the Ca^{2+} signal. The small (M_r 12,000), vertebrate-specific proteins called parvalbumins (PVs) exemplify the latter. These proteins occupy a prominent place in EF-hand protein lore – the first family members detected, purified to homogeneity, sequenced, and crystallized.[2] Importantly, the X-ray structure of carp parvalbumin, reported by Kretsinger *et al.*[3] established the EF-hand structural paradigm.

Kretsinger observed that the parvalbumin molecule consists of three homologous 30-residue segments – each consisting of a central loop flanked by short, amphipathic helices. These were termed the AB, CD, and EF domains, in reference to the flanking helical elements. Only the latter two display Ca^{2+}-binding activity; it is thought that a two-residue deletion abolished the metal ion-binding capacity of the AB domain. Within the CD and EF binding loops, the six ligands to the bound Ca^{2+} are located at the approximate vertices of an octahedron and are indexed by Cartesian axes. Side-chain oxygen atoms furnish the $+x$, $+y$, $+z$, and $-z$ ligands. A main-chain carbonyl provides the invariant $-y$ ligand. The $-x$ position is occupied by a glutamyl side-chain in the CD site and by a water molecule in the EF site. Ligation by the nearly invariant glutamyl carboxylate at $-z$ is bidentate, so that the Ca^{2+} coordination geometry is actually pentagonal bipyramidal. Kretsinger also made the paradigmatic observation that the CD and EF sites are related by a two-fold symmetry axis and are physically joined by a short

segment of anti-parallel β-sheet structure. With very few exceptions, the 30-residue EF-hand sub-structures occur in pairs, forming a so-called EF-domain. The parvalbumin tertiary structure, then, includes two structural domains – the metal ion-binding domain, formed by the paired CD and EF sites, and the N-terminal AB domain. The AB domain packs tightly against the hydrophobic aspect of the CD-EF domain, preventing calmodulin-like interactions with target proteins.

In the three decades since Kretsinger and Nockolds published the structure of carp parvalbumin, over 20 additional parvalbumin structures have been deposited in the PDB. Despite our familiarity with the parvalbumin molecule, however, our understanding of parvalbumin structure-function relationships remains embryonic. For example, the rat α- and β-parvalbumin isoforms display 49% sequence identity,[4,5] and the overall RMSD for their superimposed CA backbones is just 0.8-0.9 Å (based on 88 equivalent residues). Nevertheless, these two proteins exhibit markedly different divalent ion-binding properties [6,7]. It is also well known that typical parvalbumin CD and EF sites, including those in rat α, behave identically in titrations with Ca^{2+} or Mg^{2+}. However, just one of the sites in α is capable of binding Na^{+},[8] evidence that this apparent functional equality is the result of cooperativity and not a reflection of any intrinsic equivalence.

Insight into these, and other, issues of parvalbumin behavior will require additional high-precision physical and structural measurements. Toward that goal, Declercq *et al.* recently reported a 0.91 Å structure for the pike 4.10 β isoform.[9] The superior resolution of that structure enabled the investigators to discern multiple side-chain conformations for several internal residues, indicative of slow structural rearrangement of the hydrophobic core. They suggested that the internal mobility of the protein core is linked

to the dynamics of the entire protein molecule and that the millisecond conformational dynamics of the core could influence the kinetics of Ca^{2+} and Mg^{2+} association/dissociation.

We herein describe a 1.05 Å structure of rat α-parvalbumin. The substantially higher resolution permits significant refinement of the 2.0 Å structure presently in the PDB.[10] Thus, the uncertainties in Ca^{2+}-ligand distances and ligand-Ca^{2+}-ligand bond angles are considerably smaller. Analysis of solvent at atomic resolution reveals eleven water molecules common to α-PV and β-PV structures. Moreover, the quality of the data allows refinement of the anisotropic displacement parameters (ADP), and visualization of the ADPs as thermal ellipsoids facilitates an evaluation of the atomic motions in the Ca^{2+}-binding sites. Finally, as observed for the pike 4.10 β isoform, alternative side-chain conformations are also evident for several internal residues in the high-resolution structure of the rat α isoform.

2.2 Results and Discussion

2.2.1 Quality of the structure and overall fold

The crystallographic asymmetric unit contains three α-PV molecules, labeled A, B, and C. With few exceptions, the protein electron density was unambiguous throughout. The entire backbone could be modeled with full occupancies in all three molecules, except for the terminal residue Ser-C109. Several surface side chains could not be modeled completely due to weak density (Lys-A44, Lys-A52, Lys-A80, Lys-B13, Lys-C36, Lys-C37, Glu-C108, and Ser-C109) and were therefore truncated. We note that

Table 2.1. Data collection and refinement statistics

Wavelength (Å)	0.946
Space group	$P2_12_12_1$
Unit cell dimensions (Å)	a = 33.8, b = 54.7, c = 153.6
No. of protein molecules per asymmetric unit	3
Diffraction resolution (Å)	50-1.05 (1.07-1.05) [a]
No. of observations	937, 912
No. of unique reflections	132, 869
Redundancy	7.1 (3.9) [a]
Completeness (%)	99.2 (97.5) [a]
Mean I/σ_I	36.2 (4.6) [a]
R_{merge}	0.06 (0.319) [a]
No. of protein atoms	2515
No. of water molecules	373
No. of calcium ions	6
No. of sulfate ions	4
No. of ammonium ions	1
No. of PEG molecules	1
R_{cryst}	0.1325
R_{free} [b]	0.1620
RMSD from restraint target values	
Bond lengths (Å)	0.015
Angle distances (Å)	0.029
Distances from restraint planes (Å)	0.028
Zero chiral volumes (Å^3)	0.079
Non-zero chiral volumes (Å^3)	0.089
Antibumping (Å)	0.076
Rigid-bond ADP components (Å^2)	0.005
Similar ADP components (Å^2)	0.034
Approximate isotropic ADPs (Å^2)	0.064
Ramachandran plot[c]	
Favored (%)	95.6%
Allowed (%)	4.1%
Generous (%)	0.3%
Disallowed (%)	0.0%
Average isotropic B-factors (Å^2)	
Protein	11
Solvent	21

[a] Values for the outer resolution shell of data are given in parenthesis.

[b] 5% random test set (6703 reflections).

[c] Calculated using PROCHECK [13]).

Lys-13, Lys-44, and Lys-52 are truncated in the 1.54 Å shark α-PV structure (PDB code 5PAL[11]).

As expected, the protein molecules display the characteristic parvalbumin fold, which consists of six α-helices (labeled A-F) connected by loops (Figure 2.1).[1,3] The helices associate in pairs to form three domains labeled AB, CD,

and EF. Calcium ions are bound in the helix-loop-helix, or EF-hand, substructures of the CD-EF domain. The two Ca^{2+}-binding sites are joined by a short segment of anti-parallel β-sheet structure formed by Phe-57 and Ile-58 from the CD loop and Lys-96 and Ile-97 from the EF loop.

Several statistical indicators attest to the quality of the structure. For example, 96% percent of the non-glycine residues occupy the favored region of the Ramachandran plot; and there are no residues in the disallowed regions (Table 2.1). Most notably, the estimated standard

deviations (e.s.d.s) of atomic coordinates determined from blocked matrix refinement in SHELX are indicative of a well-refined high resolution structure (Figure 2.2). The average coordinate e.s.d. is 0.021 Å for backbone atoms, and 0.038 Å for side-chain atoms.

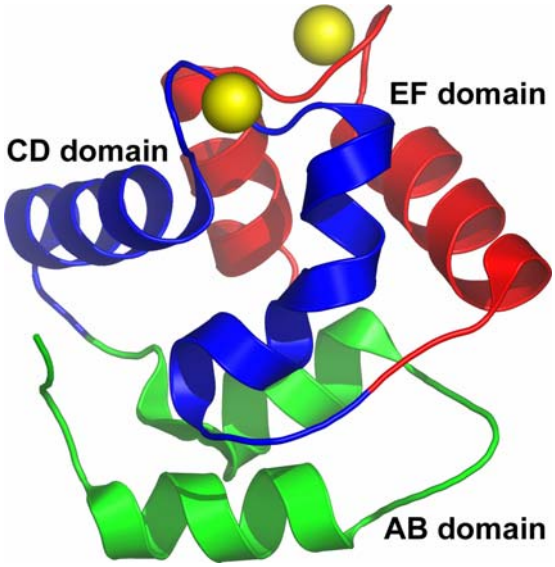


Figure 2.1. Ribbon drawing of one of the three protein molecules in the asymmetric unit. The bound Ca ions are indicated by spheres. Figures 2.1, 2.3, and 2.5 created using PyMol[12]).

The e.s.d. values are strongly correlated with the density of crystal contacts. Whereas regions of low e.s.d.

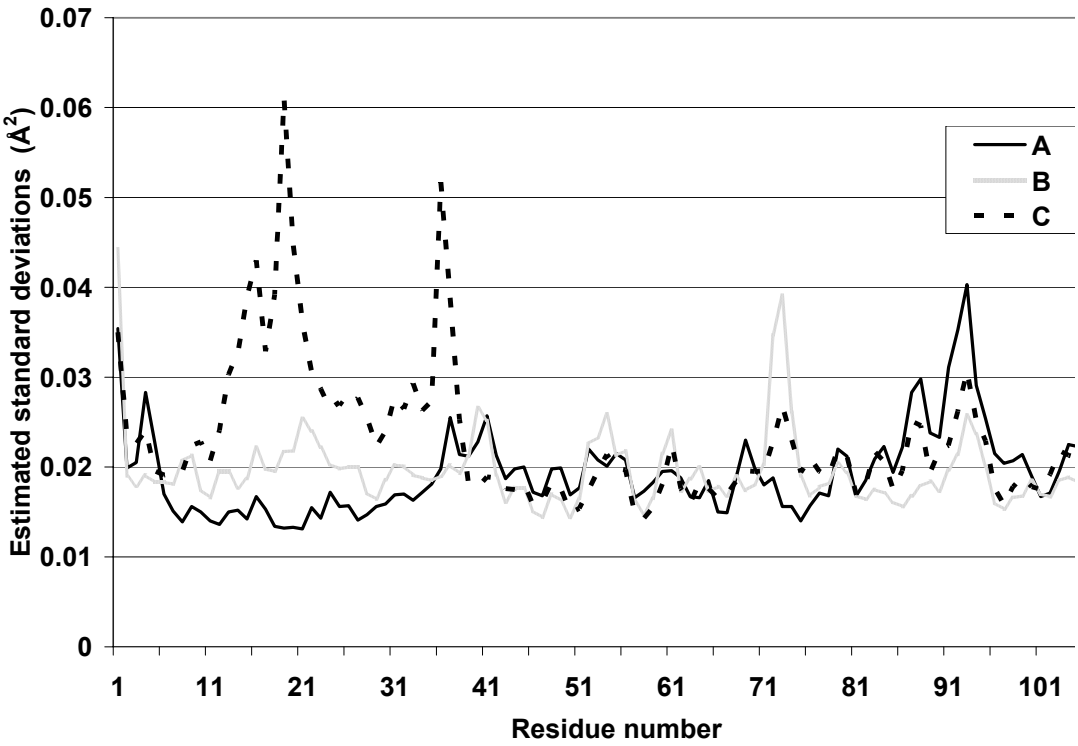


Figure 2.2. Estimated standard deviations (e.s.d.s) of coordinates for backbone atoms. The e.s.d.s were obtained from blocked matrix refinement performed in SHELX.

show numerous contacts, regions of high e.s.d. display a relative dearth of contacts. For example, consider the e.s.d.s for residues 13-37, which are much higher in the C chain than in the A and B chains (Figure 2.2). Residues C13-C37 form only 12 crystal contacts, while A13-A37 and B13-B37 make 40 and 34 crystal contacts, respectively. Likewise, residues B73-B75, which have relatively high e.s.d.s, form only a single crystal contact, while A73-A75 and C73-C75 form 9 and 4 crystal contacts, respectively. Finally, residues 87-95, which include a portion of the EF site, display higher e.s.d.s in the A chain (only 3 crystal contacts) than in the B (11 crystal contacts) and C chains (9 crystal contacts).

The protein backbone conformation is very similar to that of the 2.0 Å room temperature structure (PDB code 1RTP[10]). For example, the root mean square difference (RMSD) between the chain A backbones of the 1.05 Å and 2.0 Å structures is only 0.34 Å (following superposition of the A chain backbone atoms). The corresponding values for the B and C chains are 0.29 Å and 0.41 Å, respectively. The backbone differences between the 1.05 Å and 2.0 Å structures within the protein core are even smaller, with RMSD values of 0.27 Å, 0.22 Å, and 0.24 Å for chains A, B, and C. The RMSDs (C_α) between our structures and the 1.54 Å shark α-PV structure[11] are 0.7-0.8 Å; the RMSDs with the 0.91 Å pike β-PV structure (PDB code 2PVB[9]) are 0.8-0.9 Å.

The RMSDs between pairs of protein molecules in the asymmetric unit of the 1.05 Å structure are 0.6-0.7 Å, approximately two times greater than those between corresponding molecules in the 2.0 Å and 1.05 Å structures (0.3-0.4 Å). The relatively large RMSDs within the asymmetric unit are believed to reflect conformational differences resulting from disparate crystal packing environments. Indeed, this crystal

form – characterized by a relatively low solvent content – contains an unusually large number of intermolecular contacts. Chains A, B, and C form 116, 119, and 118 protein-protein crystal contacts, respectively. Moreover, 45% of the non-core residues participate in at least one crystal contact. In other words, roughly half of the solvent-exposed residues engage in protein-protein crystal packing interactions.

Cryogenic treatment has evidently provoked a contraction of the unit cell (see Materials and Methods). This alteration produces an interesting difference between the 1.05 Å and 2.0 Å structures. If the asymmetric units in the two are superimposed so as to minimize the deviations in the B and C chains, the resulting RMSDs for the A, B, and C chains are 1.0 Å, 0.33 Å, and 0.43 Å. The uniquely high RMSD value for the A chain indicates a relative rigid-body movement of the A chain. This movement is significant because it breaks a crystal contact that links the EF Ca^{2+}-binding site of chain A and the backbone of chain B in the room temperature structure (discussed below).

2.2.2 Solvent structure: H_2O, NH_4^+, SO_4^{2-}, and PEG

The refined 1.05 Å structure includes 373 water molecules (362 at full occupancy and 11 at half occupancy). The isotropic B-factors for these water molecules are in the range 6-54 $Å^2$, with an average of 21 $Å^2$. For comparison, the 2.0 Å structure has 110 water molecules. Thus, as expected, the increased resolution affords a more detailed and complete view of protein solvation. Nine of the peaks modeled as water in the 2.0Å structure correspond to non-aqueous solvent molecules or side chains in the 1.05Å structure. The 1.05 Å and 2.0 Å structures have 92 water molecules in common. The average B-factor of these common water molecules in the 1.05 Å structure is 14 $Å^2$, which is lower than average B-factor of 23 $Å^2$ for the remaining water molecules in the

1.05 Å structure. The water molecules common to the two structures, then, are generally the more ordered ones, so that the increased resolution appears to provide a better picture of the less well-ordered protein-associated solvent.

Of the 373 water molecules in the asymmetric unit, 292 form at least one hydrogen bond to protein. Forty-seven water molecules form bridging hydrogen bonds between two proteins, thus directly mediating crystal packing. The low solvent content of this crystal form (38%) prompted interest in the relative influence of crystal packing on solvent structure. The crystalline environments of the three polypeptides in the unit cell differ substantially. Thus, if crystal packing considerations strongly impact solvent structure, one would anticipate a relatively small number of structurally conserved water molecules among the three chains. In fact, superposition of the A, B, and C chains with their hydration shells reveals a common set of just 29 water molecules (1.5 Å cutoff, at least one interaction in common). This finding implicates crystal packing as a major determinant of solvent structure.

A comparison of the solvent in this structure with that of the atomic resolution β-PV structure was undertaken to identify solvent sites conserved in both lineages. For this analysis, all three α-PV chains and the β-PV structure were superimposed onto a common reference structure. Conserved water molecules were strictly defined as those that (1) occupied the same location in at least one of our proteins and in the β-PV structure, and (2) had at least two common hydrogen bonding partners. The analysis revealed 11 structurally conserved water molecules and 25 conserved protein-water hydrogen bonds (Table 2.2). For reference, Table 2.2 also lists the equivalent water

molecules in two other high resolution PV structures, 1.3 Å rat β-PV (PDB code 1RRO[14]) and 1.54 Å shark α-PV[11].

Table 2.2. Water molecules common to four high resolution parvalbumin crystal structures

	Water-ID						Hydrogen Bond Partners in 1RWY		
	1RWY[a]			2PVB[b]	1RRO[c]	5PAL[d]			
site	A	B	C						
1	252	246	228	201	111	246	OD1-Asp-94	OD2-Asp-94	OE1-Glu-101
2	34	105	230	204	120	203	O-Leu-50	O-Glu-62	
3	57		9	203	194	236	O-Gly-89	O-Glu-101	
4	47	43	10	224		264	N-Glu-60	O-Gly-95	
5		32	33	208[e]	131	272	O-Gly-56	N-Val-99	
6	22	14	70	216	117	202	N-Ser-7	O-Val-33	OD2-Asp-10
7	35	41	18	202	115	204	O-Gly-64	N-Leu-67	O-Arg-75
8	221		196	319			O-Ser-72	O-Ala-74	
9		72		218	123	215	O-Ala-17	N-Ala-20	
10	238	51		286			O-Ile-49	NZ-Lys-54	
11	169	64	92	235			N-Ser-39	OD2-Asp-42	

[a]1.05 Å rat α-PV (PDB code 1RWY, this work)
[b]0.91Å pike β-PV (PDB code 2PVB,[9]))
[c]1.30 Å rat β-PV (PDB code 1RRO,[14]))
[d]1.54 Å shark α-PV (PDB code 5PAL,[11]))
[e]Symmetry mate of water 208 of the PDB file for 2PVB.

Most of the conserved hydrogen bonds involve main chain atoms and the carbonyl oxygen, in particular. Seven of the eleven conserved water molecules are likewise present in both the shark α-PV and rat β-PV structures. Although conserved solvent sites 1-7 in Table 2.2 have been reported earlier[9,11], sites 8-11 represent previously unidentified structurally conserved water molecules. Site 1 has a clear functional role as the -*x* ligand of the EF site (see Table 2.3). Waters 2-7 bridge residues distant in sequence, while waters 8-11 link residues relatively close in sequence.

Site 2 hydrogen bonds to the backbone of Glu-62 and Leu-50 (Table 2.2), which are both located in the CD Ca^{2+}-binding loop. Glu-62 provides the conserved bidentate coordination with Ca^{2+}, and Leu 50 precedes the first coordinating residue (Asp-51).

Thus, site 2 joins the beginning and end of the CD loop. Site 3 forms analogous interactions in the EF domain. It hydrogen bonds to the backbone of Gly-89 and Glu-101. Thus, sites 2 and 3 are related by the pseudo-two-fold symmetry of the CD-EF domain.

Sites 4 and 5 are also related by the two-fold symmetry of the CD-EF domain; however, instead of being part of a single EF-hand, they bridge the CD and EF domains. These water molecules hydrogen bond to residues preceding and following the short β-sheet that joins the CD and EF loops, effectively extending the sheet in both directions and concomitantly doubling the number of hydrogen bonds between backbone atoms of the sheet. By virtue of their location, these water molecules could be involved in communication between the CD and EF sites.

Conserved solvent site 6 links residues quite distant in sequence, Ser-7, Asp-10, and Val-33. This water molecule bridges the amino terminus of helix-A to the carboxy terminus of helix-B. Note that one of its hydrogen bonding partners is the carboxylate of highly conserved Asp-10.

Sites 7 and 8 hydrogen bond to residues in the loop connecting the D and E helices. Site 9 bridges residues within the loop connecting helices A and B. Site 10 bridges the side chain of Lys-54 (a residue within the CD Ca^{2+}-binding site) and the backbone of Ile-49. Conserved water 11 helps stabilize the N-terminus of helix C by bridging the first residue of the helix, Ser-39, with the *i+3* residue, Asp-42. Interestingly, the same bridge is formed in rat β-PV and shark α-PV, except that in those structures Gln-42 replaces *both* Asp-42 and the water molecule.

A solvent molecule in a crystal contact near the CD site of chain A was modeled as an ammonium ion (B-value = 10 $Å^2$), based on hydrogen bonding environment. It interacts with five protein atoms that probably serve as hydrogen bond acceptors, namely two backbone carbonyl atoms of chain C (Thr-3 and Leu-6) and three carboxylate oxygen atoms of chain A (Asp-53, Glu-59, and Asp-61). The sixth hydrogen bonding partner, a water molecule, could theoretically be either an acceptor or a donor. This site is modeled as a water molecule in the B and C chains. The structures of rat α-PV at 2.0 Å resolution[10] and pike 4.10 β-PV also have ammonium ions modeled at this location.[9,15]

The solvent structure includes four sulfate ions (E1-E4). All four are located in intermolecular contact regions, and each interacts with at least one basic residue and two water molecules. Sulfate E1 interacts with His-A26, Thr-A84, Lys-A12, and three water molecules. Sulfate E2 is stabilized by His-B48, Lys-B12, Lys-B37, and four water molecules. Though not very clear, the density near LYS37 suggests a possible alternate binding mode in which NZ forms hydrogen bonds with two sulfate oxygen atoms. Sulfate E3 interacts with Lys-B54, Lys-A36, and two water molecules. Within 3.0Å, is additional 2Fo-Fc density that appears to be a water molecule. However, it looks quite messy at 0.8σ ($2F_o$-F_c) and was not modelled. Sulfate E4 is bound by Lys-B28, Ala-C79, and two water molecules.

A polyethylene glycol (PEG) fragment with molecular weight 194 Da ($HO(CH_2CH_2O)_4H$) was modeled, on the basis of a very strong tubular electron density feature in a crystal-contact region near residue Lys-C52 (Figure 2.3). The PEG molecule adopts a horseshoe-shaped conformation, allowing it to wrap around the Lys-52 ammonium group. There are 3 hydrogen bonds between this ammonium group and the

PEG molecule. HOH-75 and HOH-271 provide other hydrogen bonds to the PEG molecule (Figure 2.3). The clarity and strength of the electron density suggested that this PEG molecule is very tightly bound. The high degree of order, along with the location between protein molecules, suggested that the PEG molecule participates in noncovalent interactions that are essential for maintenance of the crystal lattice. Indeed, this PEG molecule may be, in part, responsible for the ability of the crystal to diffract x-rays to such high resolution.

Although the electron density maps suggested the presence of a second PEG fragment near residue Lys-B27, this feature was less convincing than that in Figure 2.3. When attempts to model PEG into this feature were unsuccessful, an overlapping set of water molecules (i.e. 366-371) were modeled at half occupancy within this density instead.

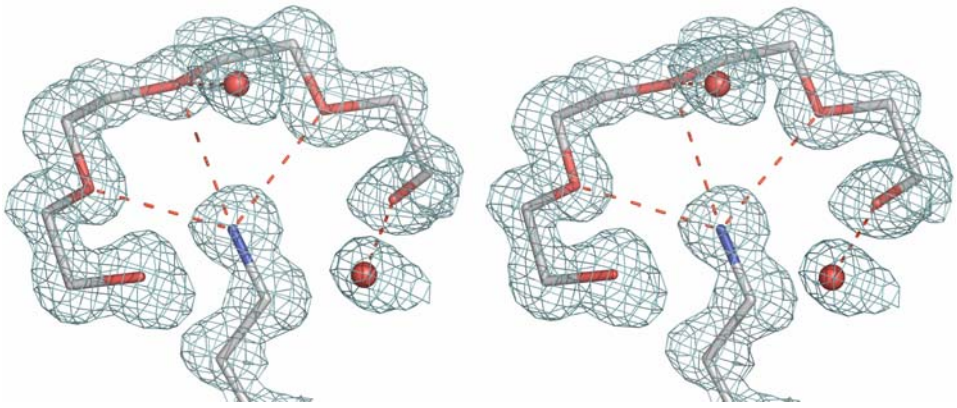


Figure 2.3. Stereoscopic view depicting the PEG fragment bound in a crystal contact, with 2Fo-Fc density contoured at 1σ. Lys—C52 can be clearly seen interacting with three PEG oxygen atoms.

Table 2.3. Ca^{+2}–binding geometry of rat α-parvalbumin determined at 1.05 Å resolution[a]

Ca^{2+} - ligand distances (Å)

CD site position	Coordinating atom	A	B	C	Mean
+x	Asp-51-OD1	2.26	2.28	2.27	2.27
+y	Asp-53-OD1	2.32	2.31	2.32	2.32
+z	Ser-55-OG	2.56	2.51	2.51	2.53
–y	Phe-57-O	2.33	2.33	2.34	2.33
–x	Glu-59-OE1	2.37	2.35	2.35	2.36
–z	Glu-62-OE1	2.41	2.43	2.42	2.42
–z	Glu-62-OE2	2.54	2.48	2.53	2.52
	Average	2.40	2.38	2.39	

EF site position	Coordinating atom	A	B	C	Mean
+x	Asp-90-OD1	2.33	2.31	2.33	2.32
+y	Asp-92-OD1	2.34	2.38	2.34	2.35
+z	Asp-94-OD1	2.34	2.36	2.33	2.34
–y	Lys-96-O	2.37	2.35	2.36	2.36
–x	HOH-O[b]	2.46	2.37	2.38	2.40
–z	Glu-101-OE1	2.42	2.47	2.43	2.44
–z	Glu-101-OE2	2.53	2.51	2.54	2.53
	Average	2.40	2.39	2.39	

Ligand – Ca^{2+} – ligand angles (°)

CD site			A	B	C	Mean
Asp-51-OD1	Ca^{2+}	Glu-59-OE1	166.0	167.0	165.4	166.1
Asp-53-OD1	Ca^{2+}	Glu-62-OE2	74.4	73.9	73.7	74.0
Ser-55-OG	Ca^{2+}	Asp-53-OD1	78.6	78.7	80.0	79.1
Phe-57-O	Ca^{2+}	Ser-55-OG	75.8	75.8	75.4	75.6
Glu-62-OE1	Ca^{2+}	Phe-57-O	79.2	81.0	80.5	80.2
Average, equatorial angles			77.0	77.3	77.4	

EF site			A	B	C	Average
Asp-90-OD1	Ca^{2+}	HOH-O[b]	158.7	163.9	161.4	161.3
Asp-92-OD1	Ca^{2+}	Glu-101-OE2	75.6	73.6	76.1	75.1
Asp-94-OD1	Ca^{2+}	Asp-92-OD1	80.7	78.7	83.4	80.9
Lys-96-O	Ca^{2+}	Asp-94-OD1	81.0	78.6	78.8	79.5
Glu-101-OE1	Ca^{2+}	Lys-96-O	80.3	80.6	84.9	81.9
Average, equatorial angles			79.4	77.9	80.8	

[a] Axial ligands occupy ±x positions (i.e. Asp-51 and Glu-59 in the CD site, Asp-90 and HOH in the EF site). Equatorial ligands occupy ±y and ±z positions.
[b] Water residue numbers are 252, 246, and 228 for A, B, and C, respectively.

2.2.3 Ca^{2+}-binding sites

The structural features of the EF-hand Ca^{2+}-binding motif have been discussed extensively [16,17]. The Ca^{2+} ions in both helix-loop-helix sites are heptacoordinate with pentagonal bipyramidal geometry. The Ca^{2+} ligands are listed in Table 2.3. The carboxyl groups of Asp and Glu account for five of the ligands in each site. Backbone carbonyl and Ser hydroxyl groups complete the coordination sphere of the CD site; the EF site features a backbone carbonyl and a water molecule.

The improved resolution of the present structure provides a more precise determination of the coordination geometry within the Ca^{2+}-binding sites than was previously available at 2.0

Å resolution. The bond distances and angles describing the Ca^{2+}-binding geometry are listed in Table 2.3. Although the average bond lengths and angles are similar to those of the 2.0 Å structure, there is significantly less molecule-to-molecule variation in the 1.05 Å structure. For example, in the 2.0 Å structure, the Asp-50 – Ca^{2+} bond length varies from 2.08 - 2.35 Å.[10] The variation in this bond length at 1.05 Å resolution is just 2.26 – 2.28 Å. Likewise, the variation in bond angles among the three molecules in the asymmetric unit is lower in the 1.05 Å structure than in the 2.0 Å structure. Whereas the axial bond angle in the EF site ranges from 152° - 168° in the 2.0 Å structure, the variation is just 159°-164° in the present structure.

The bonds and angles of the Ca^{2+}-binding sites are virtually identical to those of the 0.91 Å β-PV structure.[9] The average bond length difference between the two structures is 0.01 Å, and the average angle difference is 1.1°. The largest bond length difference is 0.03 Å (-*x* bonds), and the largest angle difference is 1.6° (Asp94-Ca^{2+}-Asp92 angle).

Very high-resolution crystallographic data permit refinement of anisotropic displacement parameters (ADPs). The ADPs can be visualized as thermal ellipsoids, thereby providing insight into the amplitude and direction of atomic motions.[18] For five of the six Ca^{2+}-binding sites in the asymmetric unit, the thermal ellipsoids for the Ca^{2+} ions and their oxygen ligands are generally small and nearly spherical. By contrast, the thermal ellipsoids for the atoms in the EF site of chain A are much larger than those in the other sites and decidedly nonspherical, indicating a substantially less rigid Ca^{2+}-binding site (Figure 2.4). Note, in particular, that the ellipsoids for Asp-A92 are much larger and more anisotropic that those of the other Ca^{2+}-binding ligands. Also, the Ca^{2+}ion ellipsoid for the chain A EF site suggests vibration perpendicular to the

equatorial plane (defined by the ligands at the ±*y* and ±*z* positions). The isotropic B-values also suggest increased flexibility in the EF site of chain A. The average isotropic B-value of the Ca^{2+} ion and its oxygen atom ligands is 13 $Å^2$ for the EF site of chain A, but only 7-8 $Å^2$ for the other five sites.

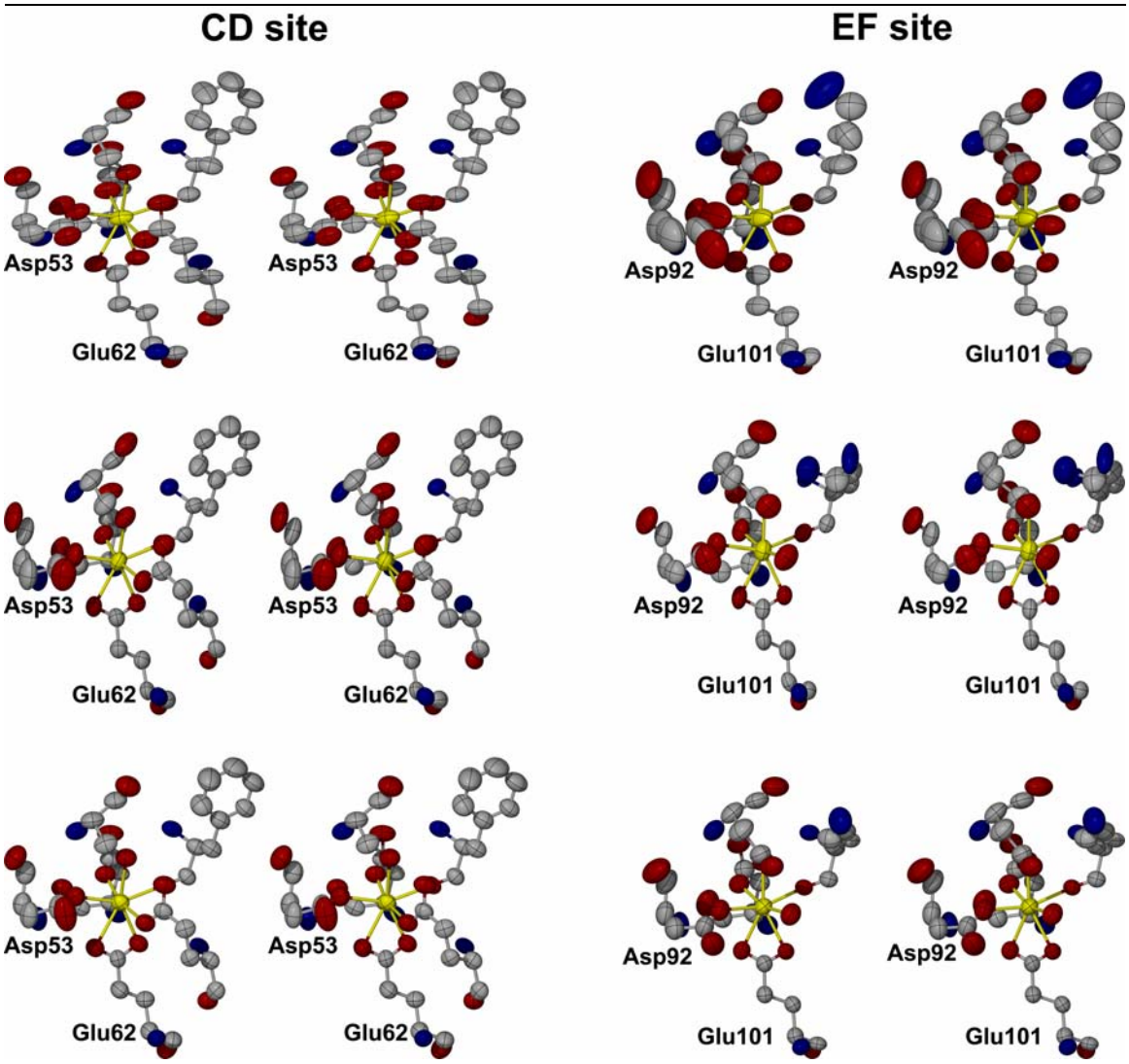


Figure 2.4. The six Ca^{2+}-binding sites in the asymmetric unit. These are shown in stereographic pairs, with the atoms represented as thermal ellipsoids displayed at the 50% probability level. The CD sites appear on the left side of the figure, and the EF sites appear on the right side. The top, middle, and bottom rows correspond to the sites of chains A, B, and C, respectively. This figure was made with X-seed.[19]

In all likelihood, the larger thermal parameters associated with this site reflect a paucity of noncovalent interactions with surrounding residues relative to the other Ca^{2+}-

binding sites. Whereas the EF site of chain A (residues A90-A101) forms 9 protein-protein crystal contacts, the EF sites of chains B and C form 12 and 16 crystal contacts, respectively. The CD sites of chains A, B, and C are even more highly constrained by the crystalline environment, forming 34, 13, and 26 protein-protein crystal contacts, respectively. Careful inspection of the Ca^{2+}-binding sites in the context of the crystalline environment revealed interactions that appear to be particularly important in stabilizing the EF sites of chains B and C. For example, Asp-B92 hydrogen bonds to a symmetry mate of Thr-B19, while Asp-B94 ion pairs with Lys-B96. These interactions help stabilize the EF site of the B chain, resulting in lowered thermal parameters and a more rigid binding site. This situation is even more evident in the EF site of chain C, where a PEG molecule is wedged between the Ca^{2+}-binding site and a neighboring protein molecule. The PEG molecule is within van der Waals contact (3.7-3.9 Å) of side-chain atoms of Asp-C92 and Glu-C101. Moreover, Asp-C94 hydrogen bonds to a crystallographic symmetry mate of Ser-C72, while Lys-C96 ion pairs to a symmetry mate of Asp-C73. Again, these noncovalent interactions stabilize the EF Ca^{2+}-binding site in the C chain.

These contacts are not observed in the EF site of chain A, and thus this Ca^{2+}-binding site displays higher thermal parameters. We therefore suggest 1) that the crystal structures of the EF sites in the B and C chains overestimate the rigidity of the EF site and 2) that the intrinsic flexibility of this site is more accurately depicted by thermal ellipsoids in the A chain EF site. Furthermore, the mobility of the A chain is probably more representative of the solution behavior.

Interestingly, the A chain EF Ca^{2+}-binding site in the room temperature 2.0 Å structure does not display reduced rigidity. The average B-value for that site in the 2.0 Å structure is 8 $Å^2$, as compared to 5-13 $Å^2$ for the other sites. This discrepancy between the 1.05 Å and 2.0 Å structures arises from a difference in the crystal-packing environment around the EF site of chain A. In the 2.0 Å structure, a hydrogen bond links OD2 of Asp-A94 and the backbone N-H of Ala-B79 (OD2-N distance = 2.9 Å). This interaction is absent in the 1.05 Å structure (OD2-N distance = 4.4 Å), due to the rigid-body movement of chain A caused by freezing the crystal in liquid nitrogen. Evidently, this intermolecular hydrogen bond contributes to the EF-site rigidity in the 2.0 Å structure, and its absence in the present structure allows greater flexibility.

2.2.4 Residues with alternative side-chain conformations

The quality of the 1.05 Å electron density map permitted the identification of eleven residues having alternative side-chain conformations. The side-chain dihedral angles for these residues – all of which were modeled by single side-chain conformations in the 2.0 Å structure – are listed in Table 2.4. Although most of these occur on the surface of the protein, Ile-97 and Leu-105 are notable because 1) they contribute to the hydrophobic core of the protein and 2) because the corresponding residues display two side-chain conformations in other high resolution EF-hand protein structures. Ile-97 has been modeled with dual side-chain conformations in the 1.44 Å structure of the rat α CD-EF fragment (PVratΔ37, PDB code 1G33[20]), the 0.91 Å β-PV structure[9], and the 1.0 Å calmodulin structure (PDB code 1EXR, residue Ile-136[21]). Leu-105 displays two side-chain conformations in 2PVB (Met-105) and in 1EXR (Met-144), but only a single conformation in PVratΔ37.

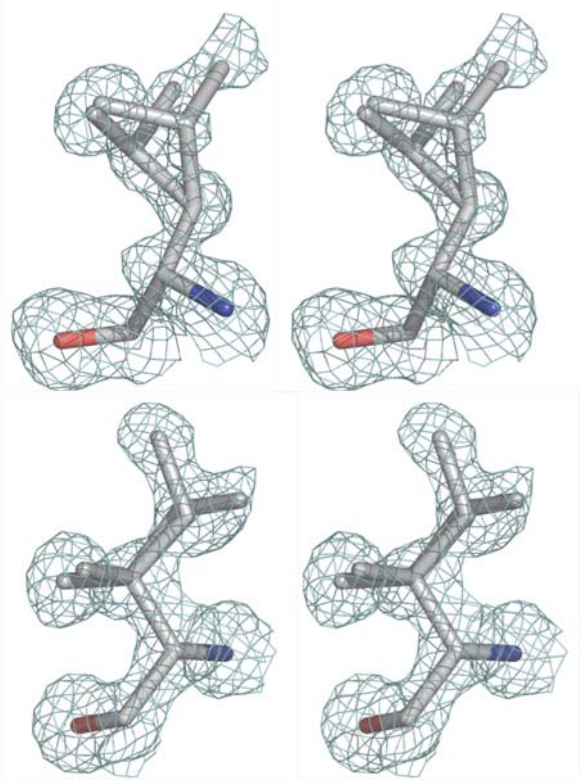


Figure 2.5. Stereoscopic view of core residues having dual side-chain conformations. (a) Ile-A97. (b) Leu-A105. The 2Fo-Fc map is contoured at 1σ in each figure.

It should be emphasized that alternative conformations have been modeled for Ile-97 and Leu-105 only in chain A of our structure (Figure 2.5). Although the electron density maps suggested minor secondary conformations for B97 and C97, the density was not sufficiently convincing to warrant the inclusion of two conformations. The unmistakable presence of two side-chain conformations for A97 may reflect the relatively greater flexibility of the A-chain EF site (see previous section).

The presence of buried residues with alternative side-chain conformations in β-PV

Table 2.4. Dihedral angles (°) for side-chains with alternative conformations.

Residue	Conformation A					Conformation B				
	Occ.	χ_1	χ_2	χ_3	χ_4	Occ.	χ_1	χ_2	χ_3	χ_4
Lys A12	0.69	172	-171	173	164	0.31	-168	-98	-174	-179
Gln A31	0.52	-69	158	27		0.48	172	69	54	
His A48	0.61	-72	-66			0.39	171	-110		
Ser A72	0.50	59				0.50	-49			
Thr A82	0.48	63				0.52	-52			
Ile A97	0.56	-71	167			0.44	-47	-60		
Leu A105	0.46	-175	-178			0.54	-175	63		
Ile B15	0.58	-63	-76			0.42	-63	-164		
Lys B38	0.51	-64	-82	106	-172	0.49	-64	-40	-64	-168
Lys B96	0.74	60	-166	-177	-66	0.26	60	-166	-177	-170
Gln C31	0.50	-65	157	-154		0.50	-159	-91	-23	

has been interpreted as evidence for conformational multistates within the hydrophobic core, and it has been proposed that conversion between these multistates might represent a slow dynamical process directly related to Ca^{2+} binding and release.[9] Our results would seem to support this idea and extend it to include α-PV as well. The calmodulin structure has a relatively large number of residues with multiple conformations (36 total). These discretely disordered residues may provide the structural plasticity that enables calmodulin to interact with a broad spectrum of effector molecules.[21]

There is presently no crystal structure available for Ca^{2+}-free rat α-PV. However, it is apparent that divalent ion binding must provoke significant conformational alterations in the molecule. Although residues 21 and 80 are well separated in the Ca^{2+}-bound protein (CA-CA distance = 9 Å), they clash in the apo-protein[22] – indicating that substantial inter-domain rearrangement must accompany divalent ion binding. Furthermore, the highly cooperative divalent ion-binding behavior[23] displayed by rat α-PV requires significant inter-site rearrangement.

That Ile-97 should display alternative conformations in three distinct PV structures suggests that the dynamics of this residue might be vital to parvalbumin function. Ile-97 is ideally positioned to transmit/receive binding information – simultaneously a component of the EF Ca^{2+}-binding loop, the short β-sheet segment connecting the CD and EF sites, and the hydrophobic core. We therefore suggest that the discrete disorder of Ile-97 may provide a mechanism for transmitting binding information between the two Ca-binding sites and between the EF site and the core. Targeted site-specific mutagenesis studies could provide insight into this issue. For example, replacement of Ile-97 by Val or Ala would reduce the side-chain volume and the number of possible

side-chain conformations, perhaps reducing the cooperative interaction between the two binding sites.

2.3 Materials and Methods

The purification of recombinant rat α-PV has been described previously.[24] The homogeneity of the isolated material was assessed by SDS-PAGE and UV absorbance spectroscopy. Because rat α-PV lacks tyrosine and tryptophan, the UV spectrum is a valuable indicator of purity. Assuming an average extinction coefficient at 280 nm of 1.0 $(mg/mL)^{-1}$ cm^{-1} for contaminants, the purity of the α preparation used for crystallization exceeded 99%.

Rat α-PV was dialyzed to equilibrium against 50 mM sodium acetate, 100 μM $CaCl_2$, pH 4.90, prior to use. Crystallization conditions were adapted from McPhalen *et al.*[10] The crystals were grown at 22°C by vapor diffusion. Reservoir solutions contained $(NH_4)_2SO_4$ (80% saturation), 50 mM sodium acetate (pH 4.55 – 4.66), 12 mM $CaCl_2$, and PEG 600 (1 – 3%). Hanging drops were formed by combining 2 μL of the protein solution (12 – 20 mg/mL) with an equal volume of reservoir solution. The largest crystals were produced in 3% PEG at pH 4.55. A crystal was prepared for cryogenic data collection by soaking it in a solution of 80% saturated ammonium sulfate, 11% glycerol, and 50 mM sodium acetate for a few minutes. The crystal was then picked up with a cryoloop and plunged into liquid nitrogen.

The crystals belong to the space group $P2_12_12_1$ with three molecules per asymmetric unit, as reported previously[10] The unit cell dimensions are a=33.8 Å, b=54.7 Å, c =153.6 Å. The Matthews coefficient[25] is 2.0 with a solvent content of only 38%. Note that the c

axis reported here is 2.5 Å shorter than that of the room temperature cell reported previously. Freezing of the crystal presumably caused a compression of the c-axis.

A 1.55 Å data set was collected from a single crystal using an R-axis IV detector coupled to a Rigaku RU-H3R copper rotating-anode generator equipped with Osmic MaxFlux confocal optics and an X-stream cryogenic system. The data set consisted of 377 frames with an oscillation angle of 0.5^{o} per frame, detector distance of 100 mm, detector theta of zero, and an exposure time of 5 minutes per frame. The data were processed using d*trek, as implemented in the program CrystalClear.[26] The 1.55 Å data set contained 41,649 unique reflections.

A higher resolution data set was subsequently collected from the same crystal at beamline 19-ID of the Structural Biology Center at the Advanced Photon Source (APS). The data collection consisted of three scans corresponding to different values of the detector theta and goniostat kappa angles. The detector distance was 120 mm, and the oscillation range was 0.5° per frame for all three scans. Scan 1 consisted of 240 frames with theta and kappa set to zero, and an exposure time of 1 second per frame. Scan 2 consisted of 240 frames with theta=20°, kappa=0, and an exposure time of 3 seconds per frame. The final scan consisted of 280 frames with theta=20°, kappa=-50°, and an exposure time of 3 seconds per frame. The data were processed using HKL2000 [27]. See Table 2.1 for data processing statistics. Note that the 1.05 Å data set provided a three-fold increase in the number of unique reflections compared to the 1.55 Å data set.

Initial model building and refinement efforts were done with the 1.55 Å R-axis data set because the APS data set was not available at the time. The coordinates of the 2.0 Å structure (PDB code 1RTP, solvent removed) were input to rigid body refinement,

positional refinement, and B-factor refinement performed using CNS,[28] which resulted in an R-factor of 0.289 for all data to 1.55 Å. No geometrical bonding restraints were imposed on the Ca^{2+} ions during refinement. The resulting phases were input to the automated building program ARP/wARP,[29] which correctly built the backbone of 317 of the expected 327 residues in the asymmetric unit, with a Connectivity Index of 0.98. Side chains were added to the ARP/wARP model with guiSIDE via CCP4i.[30] The auto-built model was improved with several rounds of manual building in O[31] followed by refinement in CNS.

When the 1.05 Å data set became available, the best 1.55 Å model was input to refinement in SHELX-97,[32] and the model was completed over several rounds of iterative model building and refinement in SHELX. See Table 2.1 for refinement statistics. The SHELX calculations consisted of conjugate gradient least-squares (CGLS) minimization, using the default effective standard deviations for geometrical restraints. The riding hydrogen and anisotropic displacement parameter (ADP) features of SHELX were used, with the ADPs of solvent molecules restrained to be approximately isotropic. No geometrical bonding restraints were imposed on the Ca^{2+} ions during refinement. After the refinement had converged, a blocked-matrix least-squares refinement calculation was performed to obtain estimated standard deviations (e.s.d.s) for atomic coordinates. For this calculation, each protein molecule was divided into overlapping blocks of 9-10 residues, with an overlap of two residues between each block. Atomic coordinate parameters and ADPs were refined for each block.

Structure analysis was performed with X-PLOR,[33] CNS, O, Pymol,[12] and Protein Explorer.[34] For purposes of core RMSD calculations, core residues for rat α-PV were

defined previously[10] as those having solvent accessibility of less than 5%: 2, 6, 11, 14, 29, 30, 33-35, 46, 47, 50, 58, 63, 66, 67, 70, 74, 85, 89, 97, 102, 105, and 106. RMSD calculations were performed with CNS and the CE website.[35] Structural alignments utilized transformation matrices obtained from the secondary structure matching service available at the European Bioinformatics Institute website (http://www.ebi.ac.uk/msd-srv/ssm).[36] Analysis of protein-protein crystal contacts was done with CNS using a 3.9-Å atom-based cutoff.

2.4 PDB accession code

The atomic coordinates and structure factors have been deposited in the PDB as entry 1RWY.

2.5 Acknowledgements

We thank the personnel of APS beamline 19-ID for assistance with data collection, especially Younchang Kim. Use of the Argonne National Laboratory Structural Biology Center beamlines at the APS was supported by the U. S. Department of Energy, Office of Energy Research, under Contract No. W-31-109-ENG-38.

Chapter 2 References

1. Celio MR, Pauls T, Schwaller B. Guidebook to the Calcium-binding Proteins. Oxford University Press; 1996.

2. Wnuk W, Cox JA, Stein EA. Parvalbumins and other soluble high-affinity calcium-binding proteins from muscle. Calcium Cell Func. Volume 2; 1982. p 243-278.

3. Kretsinger RH, Nockolds CE. Carp muscle calcium-binding protein. II. Structure determination and general description. J Biol Chem. Volume 248; 1973. p 3313-3326.

4. Epstein P, Means AR, Berchtold MW. Isolation of a rat parvalbumin gene and full length cDNA. J Biol Chem. Volume 261; 1986. p 5886-5891.

5. Gillen MF, Banville D, Rutledge RG, Narang S, Seligy VL, Whitfield JF, MacManus JP. A complete complementary DNA for the oncodevelopmental calcium-binding protein, oncomodulin. J Biol Chem. Volume 262; 1987. p 5308-5312.

6. Eberhard M, Erne P. Calcium and magnesium binding to rat parvalbumin. Eur J Biochem. Volume 222; 1994. p 21-26.

7. Hapak RC, Lammers PJ, Palmisano WA, Birnbaum ER, Henzl MT. Site-specific substitution of glutamate for aspartate at position 59 of rat oncomodulin. *J Biol Chem*. Volume **264**; 1989. p 18751-18760.

8. Henzl MT, Larson JD, Agah S. Influence of monovalent cations on rat alpha- and beta-parvalbumin stabilities. Biochemistry. Volume 39; 2000. p 5859-5867.

9. Declercq JP, Evrard C, Lamzin V, Parello J. Crystal structure of the EF-hand parvalbumin at atomic resolution (0.91 A) and at low temperature (100 K). Evidence for conformational multistates within the hydrophobic core. Protein Sci. Volume 8; 1999. p 2194-2204.

10. McPhalen CA, Sielecki AR, Santarsiero BD, James MN. Refined crystal structure of rat parvalbumin, a mammalian alpha-lineage parvalbumin, at 2.0 A resolution. J Mol Biol. Volume 235; 1994. p 718-732.

11. Roquet F, Declercq JP, Tinant B, Rambaud J, Parello J. Crystal structure of the unique parvalbumin component from muscle of the leopard shark (Triakis semifasciata). The first X-ray study of an alpha-parvalbumin. J Mol Biol. Volume 223; 1992. p 705-720.

12. DeLano WL. The PyMOL Molecular Graphics System (http://www.pymol.org). 2002.

13. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr. Volume 26; 1993. p 283-291.

14. Ahmed FR, Rose DR, Evans SV, Pippy ME, To R. Refinement of recombinant oncomodulin at 1.30 A resolution. J Mol Biol. Volume 230; 1993. p 1216-1224.

15. Declercq JP, Tinant B, Parello J, Rambaud J. Ionic interactions with parvalbumins. Crystal structure determination of pike 4.10 parvalbumin in four different ionic environments. J Mol Biol. Volume 220; 1991. p 1017-1039.

16. McPhalen CA, Strynadka NC, James MN. Calcium-binding sites in proteins: a structural perspective. Adv Protein Chem. Volume 42; 1991. p 77-144.

17. Strynadka NC, James MN. Crystal structures of the helix-loop-helix calcium-binding proteins. Annu Rev Biochem. Volume 58; 1989. p 951-998.

18. Stout GH, Jensen LH. X-ray Structure Determination: A Practical Guide. 2nd ed. New York: John Wiley & Sons; 1989.

19. Barbour LJ. X-Seed — A Software Tool for Supramolecular Crystallography. J Supramol Chem. Volume 1; 2001. p 189–191.

20. Thepaut M, Strub MP, Cave A, Baneres JL, Berchtold MW, Dumas C, Padilla A. Structure of rat parvalbumin with deleted AB domain: implications for the evolution of EF hand calcium-binding proteins and possible physiological relevance. Proteins. Volume 45; 2001. p 117-128.

21. Wilson MA, Brünger AT. The 1.0 A crystal structure of Ca(2+)-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. J Mol Biol. Volume 301; 2000. p 1237-1256.

22. Agah S, Larson JD, Henzl MT. Impact of proline residues on parvalbumin stability. Biochemistry. Volume 42; 2003. p 10886-10895.

23. Henzl MT, Larson JD, Agah S. Influence of monovalent cation identity on parvalbumin divalent ion-binding properties. Biochemistry. Volume 43; 2004. p 2747-2763.

24. Henzl MT, Graham JS. Conformational stabilities of the rat alpha- and beta-parvalbumins. FEBS Lett. Volume 442; 1999. p 241-245.

25. Matthews BW. Solvent content of protein crystals. J Mol Biol. Volume 33; 1968. p 491-497.

26. Pflugrath JW. The finer things in X-ray diffraction data collection. Acta Crystallogr. Volume D55; 1999. p 1718-1725.

27. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. Volume 276; 1997. p 307-326.

28. 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, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. Volume D54; 1998. p 905-921.

29. Morris RJ, Perrakis A, Lamzin VS. ARP/wARP's model-building algorithms. I. The main chain. Acta Crystallogr. Volume D58; 2002. p 968-975.

30. CCP4. The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallogr. Volume D50; 1994. p 760-763.

31. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. Volume A47; 1991. p 110-119.

32. Sheldrick GM, Schneider TR. SHELXL: High-resolution refinement. Methods Enzymol. Volume 277; 1997. p 319-343.

33. Brünger AT. X-PLOR version 3.1 A system for x-ray crystallography and NMR. New Haven: Yale University Press; 1992. p 382.

34. Martz E. Protein Explorer: easy yet powerful macromolecular visualization. Trends Biochem Sci. Volume 27; 2002. p 107-109.

35. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng. Volume 11; 1998. p 739-747.

36. Krissinel E, Henrick K. Protein structure comparison in 3D based on secondary structure matching (SSM) followed by Ca alignment, scored by a new structural similarity function. In: Kungl PJ, editor. 5th International Conference on Molecular Structural Biology. Vienna; 2003. p 88.

Chapter 3

A structurally conserved water molecule in Rossmann dinucleotide-binding domains

(Adapted from Bottoms CA, Smith PE, Tanner JJ. A structurally conserved water molecule in Rossmann dinucleotide-binding domains. Protein Sci 2002;11(9):2125-2137.)

Chapter 3 Abstract

A computational comparison of 101 high-resolution (≤1.90 Å) enzyme-dinucleotide (NAD, NADP, FAD) complexes was performed to investigate the role of solvent in dinucleotide recognition by Rossmann fold domains. The typical binding site contains about 9-12 water molecules, and about thirty percent of the hydrogen bonds between the protein and the dinucleotide are water-mediated. Detailed inspection of the structures reveals a structurally conserved water molecule bridging dinucleotides with the well-known glycine-rich phosphate-binding loop. This water molecule displays a conserved hydrogen-bonding pattern. It forms hydrogen bonds to the dinucleotide pyrophosphate, two of the three conserved glycine residues of the phosphate-binding loop, and a residue at the C-terminus of strand four of the Rossmann fold. The conserved water molecule is also present in high-resolution structures of apo enzymes. However, the water molecule is not present in structures displaying significant deviations from the classic Rossmann fold motif, such as having nonstandard topology, containing a very short phosphate-binding loop, or having alpha-helix "A" oriented perpendicular to the beta-sheet. Thus, the conserved water molecule appears to be an inherent structural feature of the classic Rossmann dinucleotide-binding domain.

Keywords: NAD; NADP; FAD; dinucleotide-protein interactions; Rossmann fold; structurally conserved water molecule; molecular recognition.

Abbreviations: A, adenine; ACP, acyl carrier protein; ADH, alcohol dehydrogenase; CNS, Crystallography and NMR system; CoA, coenzyme A; dh, dehydrogenase; dTDP, deoxythymidine diphosphate; EBI-MSD, European Bioinformatics Institute Macromolecular Structure Database; FAD, flavin adenine dinucleotide; GAPDH, D-glyceraldehyde-3-phosphate dehydrogenase; GDP, Guanosine diphosphate; N, nicotinamide; NAD, nicotinamide adenine dinucleotide; NADP, nicotinamide adenine dinucleotide phosphate; NMN, nicotinamide adenine mononucleotide; P, pyrophosphate; PDB, Protein Data Bank; Ra, adenine ribose; rd, reductase; RMSD, root mean square deviation; Rn, nicotinamide ribose; UDP-galactose, uridine 5'-diphosphate galactose.

3.1 Introduction

Flavin adenine dinucleotide (FAD), nicotinamide adenine dinucleotide (NAD), and nicotinamide adenine dinucleotide phosphate (NADP) serve as enzyme cofactors in many essential biological processes, such as glycolysis (NAD), the citric acid cycle (FAD and NAD), and photosynthesis (NADP) (Figure 3.1).

Because of the central role played by adenine dinucleotides in biological redox chemistry, many dinucleotide-dependent enzymes are potential drug design targets for the treatment of cancer,[1-4] bacterial infections,[5] and trypanosomal diseases.[6-10] Knowledge of the structural and dynamic features that govern dinucleotide recognition by enzymes is important for understanding the many biochemical roles of dinucleotides and for designing analogues for therapeutic applications.

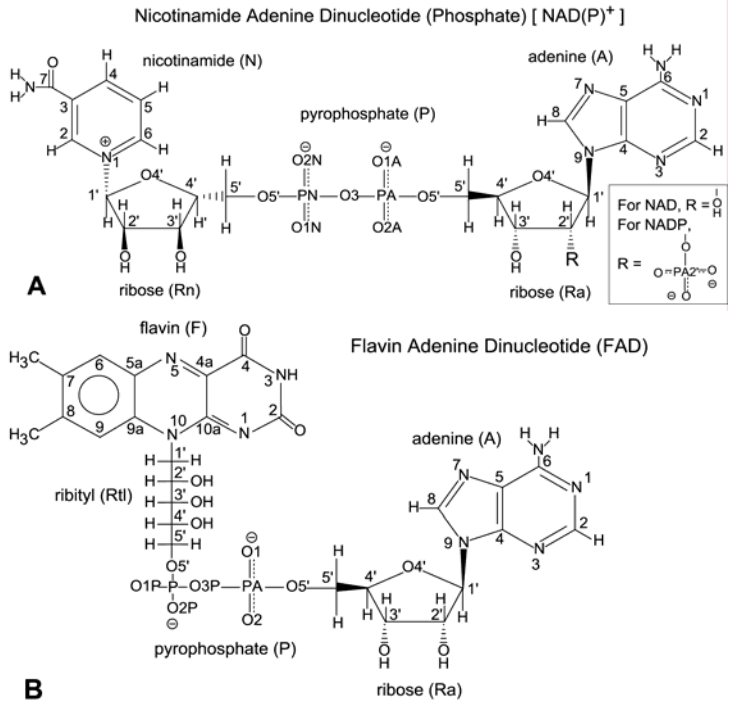


Figure 3.1. Chemical structures and nomenclature for (A) NAD(P)$^+$ and (B) FAD.

The Rossmann fold was first identified in dinucleotide-binding proteins[11] and remains one of the most thoroughly studied dinucleotide-binding folds. The Rossmann fold, or mononucleotide-binding motif, is a single βαβαβ motif that binds a mononucleotide. The

fold that binds NAD and NADP consists of two mononucleotide-binding motifs that are structurally related by a pseudo 2-fold rotation, with the most N-terminal strands adjacent to each other (Figure 3.2a). Together, the two βαβαβ motifs form a six-stranded parallel β-sheet flanked by α-helices, with relative strand order 321456. The fold that binds FAD typically contains one mononucleotide-binding motif with two additional parallel β-strands, giving a relative strand order of 32145 (Figure 3.2b).[12]

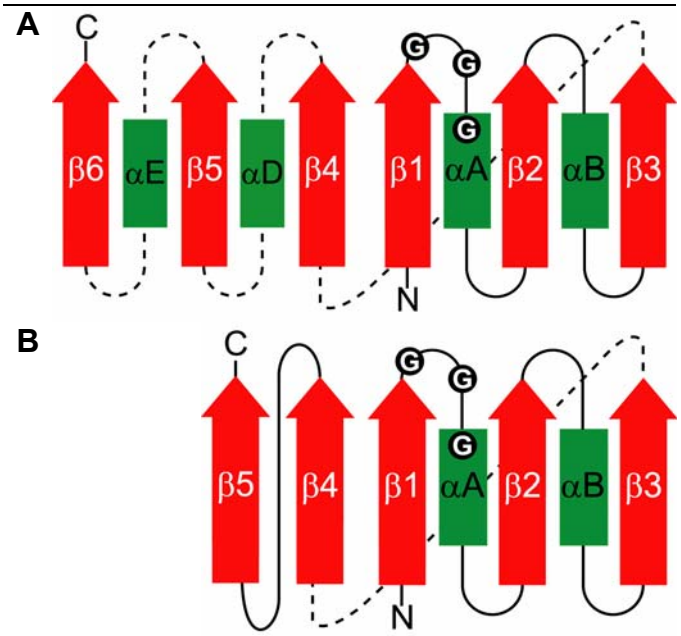


Figure 3.2. Classic Dinucleotide-binding Rossmann-fold topologies. Arrows designate β-strands and rectangles denote α-helices. Circles represent conserved glycine residues in the pyrophosphate-binding loop. (a) Topology of NAD(P)-binding proteins. (b) Topology of FAD-binding proteins.

Although dinucleotide-binding domains show very low overall sequence homology, large portions of their protein backbones superimpose very well.[13] However, there are two common sequence features. First, a glycine-rich phosphate-binding loop connects the C-terminus of β1 with the N-terminus of αA (Figure 3.2).[14] Typically, this loop contains three conserved glycine residues arranged in patterns such as GXGXXG. Mutations in the conserved glycine residues of the loop have been correlated with attenuation or elimination of enzyme activity[15-17] and also with disease.[18] The second feature is a side chain interaction with the adenine ribose group of the dinucleotide. In NAD- and FAD-binding Rossmann fold proteins, the carboxylate of Asp or Glu interacts

with the adenine ribose hydroxyls. In contrast, NADP-binding proteins typically have Arg interacting with the monophosphate at the O2' position of adenine ribose.[13]

The conserved characteristics of dinucleotide-binding proteins have been described in several studies;[11,19] however, the role that structural water plays in dinucleotide recognition has not been extensively analyzed. Five years ago, Carugo and Argos reported a study of NAD(P)-protein interactions,[20] using a data set of 32 NAD(P)-enzyme complexes representing 19 enzymes. The resolution of these structures ranged from 1.6 to 3.20 Å, averaging 2.30 Å, and only eight structures had 1.9 Å resolution or better. Since their work, many high-resolution (≤1.90 Å) structures have become available for study. Here we describe the results of a survey of 101 high-resolution enzyme/dinucleotide complexes that display the Rossmann dinucleotide-binding fold. These higher resolution structures allow an accurate description of the role of solvent in dinucleotide recognition. We find that bridging water molecules contribute significantly to the dinucleotide-binding interface in all of the enzymes studied. Moreover, we identified a structurally conserved water molecule that links, through hydrogen bonding, the glycine-rich loop and the dinucleotide pyrophosphate moiety. We assert that this conserved water molecule is an integral characteristic of dinucleotide-binding Rossmann fold domains, and thus it contributes significantly to dinucleotide recognition.

3.2 Results

3.2.1 Data set of structures analyzed

All crystal structures (as of Jan 2002) in the Protein Data Bank (PDB)[21] with NAD(P) or FAD (Figure. 3.1) bound to a Rossmann fold motif and having a resolution of 1.9 Å or better were selected for analysis. The resulting data set consists of 101 structures, and

represents 43 enzymes and 40 species (Tables 3.1, 3.2, and 3.3). The crystallographic resolution ranges from 1.15 to 1.90 Å, averaging 1.70 Å. Notwithstanding our conservative resolution cutoff, this study is one of the largest comparisons of dinucleotide-binding proteins performed to date.

Table 3.1. NAD-binding proteins

Enzyme	Species	PDB codes and references	Sequence pattern	Wat
			(G...X**G**XX**G)**	
alcohol dehydrogenase	*Equus caballus*	1EE2[22]; 1HET, 1HEU[23]; 2OHX[24]; 3BTO[25]	**G**...L**G**GV**G**	✓
3-dehydroquinate synthase	*Aspergillus nidulans*	1DQS[26]	**G**...G**G**VI**G**	
GAPDH	*Bacillus stearothermophilus*	1GD1[27]	**G**...F**G**RI**G**	✓
	Escherichia coli	1GAD[28]	**G**...F**G**RI**G**	✓
L-3-hydroxyacyl-CoA dh	*Homo sapiens*	1F0Y[29]	**G**...G**G**LM**G**	✓
D-2-hydroxyisocaproate dh	*Lactobacillus casei*	1DXY[30]	**G**...T**G**HI**G**	✓
lactate dehydrogenase	*Plasmodium falciparum*	1LDG[31]	**G**...S**G**MI**G**	✓
phenylalanine dehydrogenase	*Rhodococcus sp.* M4	1BW9[32]; 1C1D, 1C1X[33]	**G**...L**G**AV**G**	✓
			(G..XX**G**XX**G)**	
dTDP-glucose 4,6-dehydratase	*Escherichia coli*	1BXK [34]	**G**..GA**G**FI**G**	✓
malate dehydrogenase	*Escherichia coli*	1EMD [35]	**G**..AA**G**GI**G**	✓
	Thermus flavus	1BMD [36]	**G**..AA**G**QI**G**	✓
sulfolipid biosynthesis	*Arabidopsis thaliana*	1QRR [37]	**G**..GD**G**YC**G**	✓
UDP-galactose 4-epimerase	*Escherichia coli*	1NAH[38]; 1UDA, 1UDB, 1UDC[39]; 1XEL[40]; 2UDP[41]	**G**..GS**G**YI**G**	✓
	Homo sapiens	1EK5, 1EK6[42]; 1HZJ[43]	**G**..GA**G**YI**G**	✓
			(G..XXX**G**X**G)**	
7-α-hydroxysteroid dh	*Escherichia coli*	1FMC[44]	**G**..AGA**G**I**G**	✓
D-glucose-1-dehydrogenase	*Bacillus megaterium*	1GCO[45]	**G**..SST**G**L**G**	✓
L-3-hydroxyacyl-CoA dh	*Rattus norvegicus*	1E6W[46]	**G**..GAS**G**L**G**	✓
meso-2,3-butanediol dh	*Klebsiella pneumoniae*	1GEG[47]	**G**..AGQ**G**I**G**	✓
			(GXXXXXXX**G)**	
enoyl ACP reductase	*Brassica napus*	1D7O[48]; 1ENO[49]	**G**IADDNGY**G**	✓
	Escherichia coli	1QG6[50]; 1QSG[51]	**G**VASKLSIA	✓
NMN adenylyltransferase	*Methanobacterium thermoautotrophicum*	1EJ2[52]	**G**RMQPFHR**G**	

Tables 3.1, 3.2, and 3.3 show alignments of the glycine-rich pyrophosphate-binding sequences for the enzymes studied. The majority of enzymes in our data set display the pyrophosphate-binding loop sequence patterns GXGXXG, GXXGXXG, or GXXXGXG, where G is Gly and X is any amino acid residue. Two enzymes have the sequence pattern

GXXXXXG, and two have the pattern GXXXXXXXG. Note that sometimes Ala substitutes for the last Gly. And, in one case, dihydropyrimidine dehydrogenase, Ser substitutes for the last Gly. Note that all the FAD-binding proteins in our data set except one have the phosphate-binding loop sequence GXGXXG. Quinone reductase does not have a glycine-rich loop, although it is still classified as a Rossmann fold protein.

Table 3.2. NADP-binding proteins

Enzyme	Species	PDB codes and references	Sequence pattern	Wat
			(G...XGXXG)	
acetohydroxy acid isomeroreductase	*Spinacia oleracea*	1YVE[53]	**G...WG**SQA	✓
adrenodoxin reductase	*Bos taurus*	1E1M[54]	**G...QG**NVA	✓
glutathione reductase	*Homo sapiens*	1GRB[55]	**G...AG**YIA	✓
NADP(H) transhydrogenase.	*Bos taurus*	1D4O[56]	**G...YG**LCA	
			(G..XXGXXG)	
biliverdin-IX β reductase	*Homo sapiens*	1HDO, 1HE2, 1HE3, 1HE4, 1HE5[57]	**G..ATGQTG**	✓
coenzyme F420H2: $NADP^+$ oxidoreductase (Fno)	*Archaeoglobus fulgidus*	1JAY[58]	**G..GTGNLG**	✓
GDP 4-keto-6-deoxy-D-mannose epimerase reductase	*Escherichia coli*	1E6U[59]	**G..HRGMVG**	✓
L-lactate/malate dehydrogenase	*Methanococcus jannaschii*	1HYE[60]	**G..ASGRVG**	✓
			(G..XXXGXG)	
1,3,6,8-tetrahydroxy-naphthalene reductase	*Magnaporthe grisea*	1JA9[61]	**G..AGRGIG**	✓
carbonyl reductase	*Mus musculus*	1CYD[62]	**G..AGKGIG**	✓
mannitol dehydrogenase	*Agaricus bisporus*	1H5Q[63]	**G..GNRGIG**	✓
nitric-oxide synthase	*Rattus norvegicus*	1F20[64]	**G..PGTG**(IA)	
sepiapterin reductase	*Mus musculus*	1OAA[65]	**G..ASRGFG**	✓
tropinone reductase-II	*Datura stramonium*	2AE2[66]	**G..GSRGIG**	✓
			(G..XXXXXG)	
methylenetetrahydrofolate dh	*Homo sapiens*	1A4I[67]	**G..RSKIVG**	✓
pteridine reductase	*Leishmania major*	1E7W[68]	**G..AAKRLG**	✓
			Other	
dihydrofolate reductase	*Candida albicans*	1AI9, 1AOE[69]; 1IA1, 1IA2, 1IA3, 1IA4[70]	**G**RKT	
	Escherichia coli	1RA2, 1RA3, 1RC4, 1RX2, 1RX9[71]	**G**GGRV	
	Gallus gallus	8DFR[72]	**G**KKT	
	Lactobacillus casei	3DFR[73]	**G**RRT	
	Mycobacterium tuberculosis	1DF7, 1DG7[74]	**G**RRT	
	Pneumocystis carinii	1DYR[75]; 2CD2[76]	**G**RKT	

Table 3.3. FAD-binding proteins				
Enzyme	**Species**	**PDB codes and references**	**Sequence pattern**	**Wat**
			(G...XGXXG)	
adrenodoxin reductase	*Bos taurus*	1CJC[77]; 1E1M[54]	**G...SGPAG**	✓
alkyl hydroperoxide rd	*Escherichia coli*	1FL2[78]	**G...SGPAG**	✓
cholesterol oxidase	*Brevibacterium sterolicum*	1COY, 3COX [79]	**G...SGYGG**	✓
	Streptomyces sp.	1B4V[80]	**G...TGYGA**	✓
D-amino acid oxidase	*Rhodotorula gracilis*	1C0K, 1C0L, 1C0P[81]	**G...SGVIG**	✓
dihydropyrimidine dh	*Sus scrofa*	1H7W[82]	**G...AGPAS**	✓
flavocytochrome C3	*Shewanella frigidimarina*	1QJD[83]	**G...SGGAG**	✓
glucose oxidase	*Aspergillus niger*	1CF3[84]	**G...GGLTG**	✓
	Penicillium amagasakiense	1GPE[84]	**G...GGLTG**	✓
glutathione reductase	*Escherichia coli*	1GER[85]	**G...GGSGG**	✓
	Homo sapiens	1DNC, 1GSN[86]; 1GRB [55]; 3GRS [87]	**G...GGSGG**	✓
p-hydroxybenzoate hydroxylase	*Pseudomonas fluorescens*	1PBE[88]	**G...AGPSG**	✓
polyamine oxidase	*Zea mays*	1B37, 1B5Q[89]; 1H82, 1H83[90]	**G...AGMSG**	✓
sarcosine oxidase	*Bacillus sp.*	1EL5, 1EL7, 1EL8[91]	**G...AGSMG**	✓
trypanothione reductase	*Crithidia fasciculata*	1FEC[92]	**G...AGSGG**	✓
			Other	
quinone reductase	*Homo sapiens*	1D4A[93]; 1H69[2]; 1KBQ[94]	AHSERTSFNY	

3.2.2 Identification of a structurally conserved water molecule

All the structures in our database have several water molecules in the interface between the dinucleotide and the protein. On average, FAD-, NADP-, and NAD-binding proteins accommodate about 12, 11, and 9 interfacial (within 3.75Å of the protein and dinucleotide) water molecules per dinucleotide-binding site, respectively. As expected, many of these interfacial water molecules hydrogen bond to both the dinucleotide and the protein, thus forming water-mediated hydrogen bonds. Figure 3.3 shows the average numbers of direct and water-mediated hydrogen bonds formed between the protein and the five groups of each of the dinucleotides. Water-mediated hydrogen bonds comprise 29%, 32%, and 29% of the protein/dinucleotide hydrogen bonds for NAD, NADP, and FAD, respectively. The pyrophosphate group forms almost as many water-mediated hydrogen bonds (2.7 for NAD, 2.4 for NADP, and 3.5 for FAD) as direct hydrogen bonds

(2.8 for NAD, 3.2 for NADP, and 4.1 for FAD) to the protein. The remaining groups average one or fewer water-mediated hydrogen bonds, with the exception of adenine ribose in NADP. This group averages more than two water-mediated hydrogen bonds, a result due to its monophosphate. Thus, bridging water molecules are concentrated around the pyrophosphate and monophosphate groups, though they are commonly found associated with the rest of the dinucleotide as well.

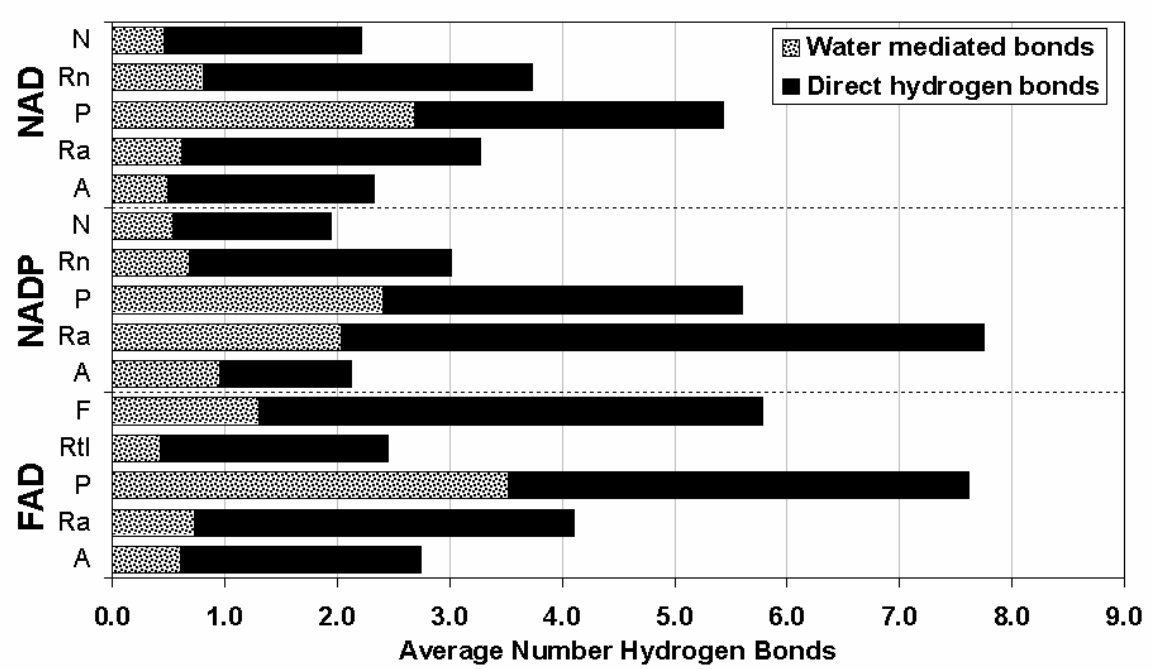


Figure 3.3. Protein/dinucleotide hydrogen bonds by groups. Hydrogen bonds to each group are divided into direct and water-mediated categories. Their sum represents the total number of hydrogen bonds between the protein and dinucleotide. The various groups of the three dinucleotides are abbreviated as follows: N, nicotinamide; Rn, nicotinamide ribose; P, pyrophosphate; Ra, adenine ribose; A, adenine; F, flavin isoalloxazine; Rtl, flavin ribityl side chain.

The presence of numerous water-mediated hydrogen bonds between protein and dinucleotides implicates water as a significant component of dinucleotide recognition. Superimposition of the structures allowed us to identify whether any of the bridging water

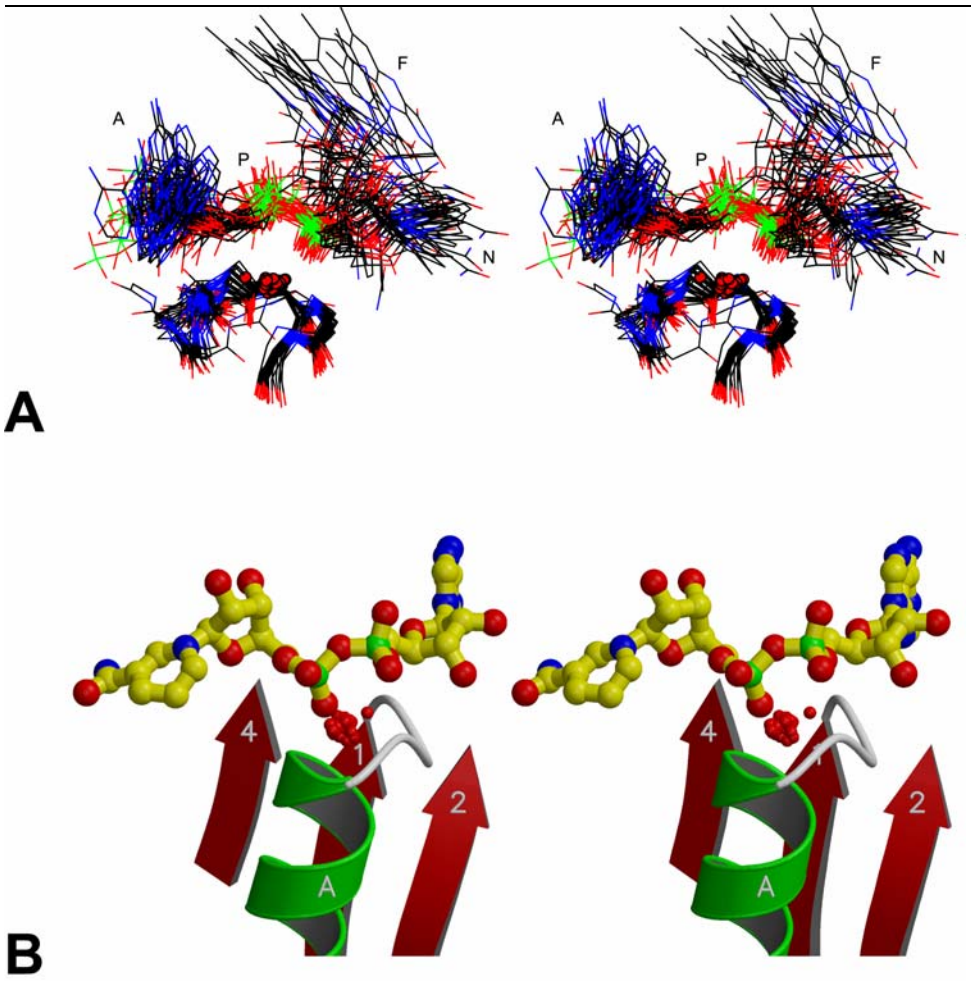


Figure 3.4. Superposition of protein structures. (A) Stereoimage showing the dinucleotides, phosphate-binding loops, and structurally conserved water molecules of 37 enzymes. (B) Stereoimage showing the structurally conserved water molecules of 77 structures superimposed according to their glycine-rich loops as described in Methods. The NADH and protein shown are from alcohol dehydrogenase (1HET). In both (A) and (B), the oxygen atoms of the water molecules are shown as red spheres with 1/5 van der Waals radii. Figures created using MOLSCRIPT [95] and Raster3D [96].

molecules bound in structurally conserved sites. The crystal structures were superimposed using the coordinates of only three alpha carbon atoms within the glycine-rich loop, as described in Methods. This analysis revealed a structurally conserved water molecule located at the N-terminus of the phosphate binding helix (i.e. αA) in 77 structures, which represents 37 of 43 enzymes studied. The structures displaying the conserved water molecule are indicated by the checkmark in Tables 3.1-3.3. Figure 3.4A

shows the superimposed phosphate binding loops, cofactors, and structurally conserved water molecules of these 37 enzymes. Note the almost perfect superimposition of the phosphate moiety, the glycine-rich loop, and the structurally conserved water molecule of these different structures. In contrast, adenine, nicotinamide, and flavin do not superimpose as well. Figure 3.4B shows the structurally conserved water molecule from 77 structures after superposition onto the alcohol dehydrogenase structure 1HET, as described in Methods. Also shown are secondary structurally elements and NADH of 1HET. These water molecules have an RMSD of 0.5 Å from a reference water molecule, which indicates that the water molecules occupy essentially the same location in their respective structures. As will be discussed below, the six enzymes that do not bind this water molecule show significant deviations from the classic Rossmann fold motif.

3.2.3 Hydrogen bonds formed by the structurally conserved water molecule

The structurally conserved water molecule typically makes four hydrogen bonds. Two of the four bonds are invariant and two vary according to the sequence pattern of the glycine-rich phosphate-binding loop. The hydrogen-bonding pattern of the conserved water molecule is shown schematically in figure 3.5A, and an example of this hydrogen bond coordination is shown in figure 3.5B. A table detailing the hydrogen bonding partners of the structurally conserved water molecule in each dinucleotide-binding site is provided in Table 3.4.

The two invariant hydrogen bonds formed by the conserved water molecule involve pyrophosphate and the amine of the last conserved Gly (Figure. 3.5). Thus, the water molecule always links the pyrophosphate to the glycine-rich loop. Moreover, the

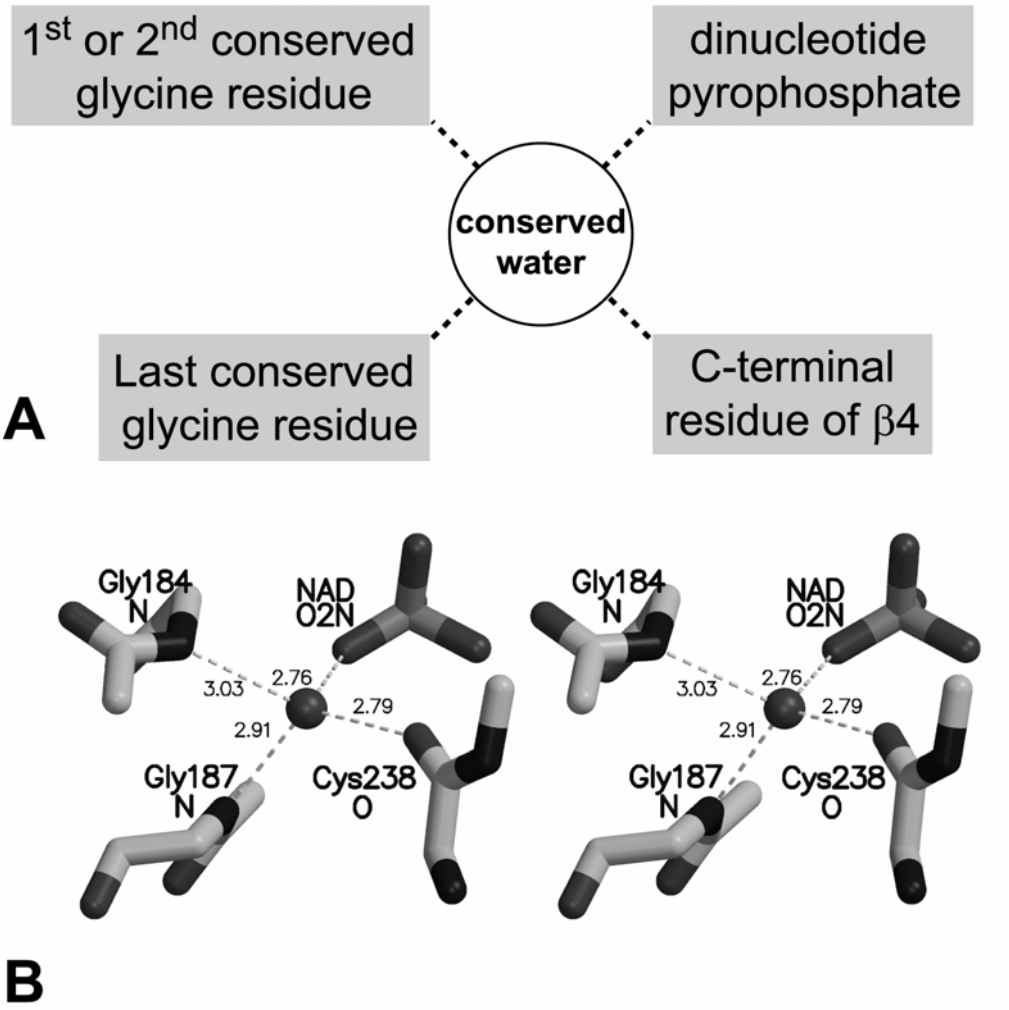


Figure 3.5. Hydrogen bonding patterns of the structurally conserved water molecule. (A) Schematic of the hydrogen bonds formed by the structurally conserved water molecule. The dotted lines denote hydrogen bonds. (B) An example of the structurally conserved water molecule's hydrogen bond coordination as seen in phenylalanine dehydrogenase of *Rhodococcus sp.* M4 (PDB code 1C1D). Gly184 and Gly187 donate hydrogen bonds via their amide nitrogens, while the carbonyl of Cys238 accepts a hydrogen bond. Only backbone atoms are shown. Distances are in angstroms.

interaction with the pyrophosphate is stereospecific. Though the pyrophosphate is not chiral, the oxygen atoms are distinct stereochemically (i.e. prochiral).[97] Almost without exception, the pyrophosphate atom that interacts with the structurally conserved water molecule is O2N in the case of NAD or NADP binding proteins. In FAD binding proteins, however, it is always O1P (i.e. equivalent to O1N of NAD(P)) that interacts with the structurally conserved water molecule. Interestingly, the only two structures in

our study that have both FAD and NADP bound are also the only ones in which NADP binds to the structurally conserved water molecule via its pyrophosphate atom O1N (see 1E1M and 1GRB).

The partners for the other two hydrogen bonds formed by the conserved water molecule depend upon the sequence of the glycine-rich loop. One sequence-dependent hydrogen bond involves either the first or second conserved Gly. This bond occurs in structures with the sequence patterns GXGXXG, GXXGXXG, and GXXXXXXXG. It also occurs in seven of the eight structures with the sequence pattern GXXXGXG, but does not occur in structures with the sequence pattern GXXXXXG. In structures with the sequence GXGXXG, the water molecule forms a hydrogen bond with the amino group of the second conserved Gly. An example of this hydrogen-bonding pattern is shown in figure 3.5B. In structures with the sequence patterns GXXGXXG, GXXXGXG, and GXXXXXXXG, the water molecule instead forms a hydrogen bond with the carbonyl group of the first conserved Gly.

The second sequence-dependent hydrogen bond involves a C-terminal residue of β4. In structures with the pattern GXGXXG and GXXGXXG, this hydrogen bond usually involves the backbone carbonyl of either a small residue or a hydrophobic residue (i.e. Ala, Cys, Gly, Leu, Phe, Ser, or Val for the structures studied). However, occasionally it will involve the hydroxyl of Ser or Thr. In structures with the sequence pattern GXXXGXG, the C-terminal residue of β4 usually donates a hydrogen bond via the side chain of an Asn. In structures with the sequence pattern GXXXXXXXG, the C-terminal residue of β4 donates a hydrogen bond via a Ser hydroxyl.

The hydrogen bonding patterns observed in structures with the sequence GXXXXXG are minor variations of the paradigm described above. There are two such structures in our data set, 1A4I (human methylenetetrahydrofolate dehydrogenase) and 1E7W (*Leishmania major* pteridine reductase). The water molecule of 1A4I does not superimpose as well as the others, as is evident from its 1.1 Å separation from the others in figure 3.4. The conserved water molecule of 1A4I displays a hydrogen-bonding pattern identical to that of GXGXXG structures, with Ser of its GRSKIVG sequence playing the role of the second conserved Gly. On the other hand, the hydrogen bond coordination in 1E7W is very similar to that of proteins with the sequence GXXXGXG. The only exception is that the first conserved glycine residue interacts with the structurally conserved water molecule via another water molecule inside the phosphate-binding loop. To our knowledge, this is the only case in which the structurally conserved water molecule forms a hydrogen bond with another water molecule.

Finally, the conserved water molecule sometimes forms a fifth hydrogen bond. For example, in structures with the sequence pattern GXXXGXG, the conserved water molecule forms a fifth hydrogen bond with the carbonyl of the third "X" residue. And, it is common to see the methylene of the second or third conserved Gly within hydrogen bonding distance of the structurally conserved water molecule. Several studies have mentioned the presence of CH...O hydrogen bonds in NAD(P) interfaces.[20,98] Though CH...O hydrogen bonds are somewhat unconventional, the methylene of glycine can theoretically form a significant hydrogen bond. Such a bond would have about half the energy of a typical hydrogen bond from water.[99]

3.2.4 Structures lacking the conserved water molecule

Six enzymes (3-dehydroquinate synthase, NMN adenylyltransferase, NADP(H) transhydrogenase, nitric-oxide synthase, dihydrofolate reductase, quinone reductase) lack the structurally conserved water molecule (Tables 3.1-3.3). These enzymes show significant deviations from the standard Rossmann fold motif in terms of their sequence and/or structure. Thus, the conserved water molecule is found only in classic Rossmann fold structures.

Some of the unusual features of these enzymes are obvious. For example, 3-dehydroquinate synthase has a nonstandard topology in which the phosphate-binding loop connects β4 and αD instead of β1 and αA. No other enzyme in our data set has this unusual topology. In NADP(H) transhydrogenase and dihydrofolate reductase, the long axis of αA runs perpendicular to the β-sheet rather than parallel to it as in all the other structures. NMN adenylyltransferase differs from the other enzymes in our data set because NAD is a product of catalysis rather than a redox cofactor. The pyrophosphate binds near the beginning of the phosphate-binding loop rather than at the end of the loop near the N-terminus of αA as occurs in the classic Rossmann fold structures. Thus, the pyrophosphate interacts with the Rossmann fold in a fundamentally unique way that is related to the nature of the reaction catalyzed.

3-dehydroquinate synthase, NADP(H) transhydrogenase, and nitric-oxide synthase show a subtle, but very important, deviation from the standard Rossmann fold structure concerning the hydrogen bonding potential of the last residue of their phosphate-binding sequences. In the classic Rossmann fold, this residue occurs at the beginning of αA (Figure 3.2) and its amino group donates one of the invariant hydrogen bonds to the

structurally conserved water molecule (Figure 3.5A). However, in 3-dehydroquinate synthase, NADP(H) transhydrogenase, and nitric-oxide synthase, this residue cannot hydrogen bond to a water molecule because it is the fifth or later residue of the helix, and its amino group must form an "n+4" α-helix hydrogen bond with the carbonyl four residues preceding it in the sequence.

Lastly, dihydrofolate reductase and quinone reductase deviate from the classic Rossmann fold because of the length and sequence, respectively, of their phosphate binding loops. The dihydrofolate reductases have short phosphate-binding loops. Their loops consist of 4-5 residues compared to 6-9 residues in the other structures (Tables 3.1-3.3). The resulting loop between β1 and αA is not wide enough to accommodate a water molecule. Quinone reductase, on the other hand, has no glycine residues in its phosphate-binding loop. This is a critical feature, because, in the classic Rossmann fold motif, the first Gly of the phosphate-binding loop always occupies the first quadrant (i.e. $\phi>0$, $\psi>0$) of the Ramachandran plot. This region of ϕ-ψ space is typically inaccessible to non-glycine residues. Thus, the phosphate-binding loop of quinone reductase cannot adopt the backbone conformations typically observed in classic Rossmann fold proteins.

3.3 Discussion

3.3.1 Bridging water molecules

NAD, NADP, and FAD have extraordinary hydrogen bonding potential. They have 19, 22, and 22 polar nitrogen and oxygen atoms, respectively. Therefore, it is not surprising that NAD, NADP, and FAD form about 16, 19, and 23 hydrogen bonds with the protein. However, the observation that water molecules mediate about 30% of these

hydrogen bonds is a novel result and indicates that bridging water is an important component of dinucleotide recognition.

More generally, our results demonstrate a pitfall of ignoring water molecules when analyzing protein crystal structures. In the present case, neglecting water molecules would have given the incorrect impression that Rossmann fold domains greatly underutilize the hydrogen bonding potential of dinucleotides. This issue is critical for ligand docking calculations and structure-based drug design and optimization.[100,101]

The result that water molecules are important in dinucleotide recognition is perhaps not surprising, because solvent is important in protein recognition of small molecules, DNA, and protein. For example, about 40% of all protein-DNA hydrogen bonds are water-mediated,[102] and protein-protein interfaces contain an average of about 22 water molecules and 11 water-mediated hydrogen bonds.[103]

3.3.2 Structurally conserved water molecule

We discovered a water molecule that bridges the glycine-rich loop and the pyrophosphate in 77 of the 101 structures studied. The structurally conserved water molecule binds at the N-terminus of αA and it displays a conserved hydrogen-bonding pattern (Figure 3.4). Structurally conserved water molecules have also been identified in fatty acid binding proteins[104] and serine proteases.[105] Water molecules of ligand-binding sites have also been found to be conserved between pairs of homologous proteins from different species.[106] Several authors have noted the presence of a water molecule that we refer to as the structurally conserved water molecule, but they did not realize its ubiquity in Rossmann fold domains nor its conserved hydrogen-bonding pattern.[22,30,42,49,63,107-109] In NAD-binding proteins, we found the water molecule to be more highly conserved than

the carboxylate interaction with the adenine-ribose diol. Not only do the structures in our data set represent a significant cross section of enzymes, but they also represent a variety of species and even biological kingdoms. Such structural conservation argues strongly for an important functional role.

The structurally conserved water molecule typically forms hydrogen bonds with two of the three conserved glycine residues, a C-terminal residue of β4, and the dinucleotide pyrophosphate. The water molecule always forms hydrogen bonds to both the pyrophosphate and the last conserved Gly, which is part of helix αA. This water-mediated hydrogen bond is significant, because

Previous studies have noted that direct hydrogen bonding between αA and the pyrophosphate is not optimal. Based on this observation, a favorable electrostatic interaction between the helix dipole of αA and the pyrophosphate has been considered an important factor in dinucleotide recognition.[14] This conclusion, however, was based on analyses of crystal structures that generally lacked water molecules; therefore, water-mediated hydrogen bonds could not be considered. Including water-mediated hydrogen bonds almost doubles the number of observed hydrogen bonds between the protein and pyrophosphate. Due to the high solvent content about the pyrophosphate and the several water-mediated hydrogen bonds, we wonder if the helix dipole would be an important contributing factor to pyrophosphate binding. Additional studies would be needed to resolve this issue.

Structures that exhibit the classic Rossmann fold motif bind the structurally conserved water molecule, whereas structures that have major deviations from the standard Rossmann fold do not. The following classic features are shared by all the

structures bearing the conserved water molecule. The β-sheet topology is 321456 for NAD(P)- and 32145 for FAD-binding domains, and αA is parallel to the β-sheet. The phosphate-binding loop is at least six residues long, and it contains Gly at the first position. The last residue of the phosphate-binding sequence is located at the beginning of αA where its amino group is available to hydrogen bond to the water.

The conserved water molecule appears to be a structural feature inherent to the classic Rossmann dinucleotide-binding fold itself, because this water molecule is also present in apo-structures. Examples of such apo-structures include the FAD utilizing enzyme L-aspartate oxidase, 1CHU,[110] the NAD enzyme malate dehydrogenase, 1MLD,[111] and the NADP enzyme secondary alcohol dehydrogenase, 1PED.[112] Some of the FAD-containing structures in our study also contained NADP-binding domains that lacked NADP. All of these apo-NADP domains also contained the structurally conserved water molecule.

There are several potential functional roles for the conserved water molecule in dinucleotide recognition. First, it could help maintain the unique conformation of the glycine-rich phosphate-binding loop. This role is suggested by the presence of the water molecule in apo structures of dinucleotide-binding enzymes, and by the fact that at least two residues of the glycine-rich loop hydrogen bond to the water molecule (Figure 3.5A). These residues interact with the water molecule through their backbone carbonyl or amide groups, thus these protein-water hydrogen bonds appear to stabilize the backbone conformation of the loop by maintaining these residues in specific phi and psi ranges. A second functional role for the water molecule is that it helps to maintain the cofactor in an extended conformation. Experimental[113] and computational studies[114,115] of NAD^+ in

various solvents suggest that NAD^+ is folded in aqueous solution such that the two bases stack against each other at a distance of about 4-5 Å. Thus, NAD^+ must unfold into an extended conformation during complexation to the enzyme. The extent to which the enzyme facilitates the unfolding of the cofactor is unknown; however, the protein presumably provides interactions that encourage and maintain the extended geometry. The hydrogen bonding and steric interactions provided by the conserved water molecule could help constrain the dihedral angles of the pyrophosphate group to the values observed in enzyme/dinucleotide complexes. A third possible functional role for the conserved water molecule is that the sequence independent hydrogen bond between the dinucleotide pyrophosphate and the water molecule (Figure 3.5) presumably provides a favorable enthalpic contribution to the free energy of binding. Future studies will explore these potential roles to evaluate their possible contributions to dinucleotide recognition and binding.

3.4 Methods

3.4.1 Selection and preparation of structures

Crystal structures of enzyme/dinucleotide complexes that contained at least one Rossmann fold were selected for analysis. This selection was restricted to structures having resolutions of 1.90 Å or better to ensure reliable solvent structure. Excluded were any structures having mutations other than for expression purposes and any structures having chemical modifications in the active site. 101 structures, representing 43 enzymes, were included in the study (Tables 3.1-3.3). Some of the proteins had representatives in multiple species.

To ensure an accurate accounting of interactions, including those involving solvent positions near subunit interfaces, the biologically relevant oligomeric forms of the enzyme were used. Typically, we used the "Likely Quaternary Structure" obtained from the European Bioinformatics Institute Macromolecular Structure Database (EBI-MSD) via the OCA-browser interface (http://oca.ebi.ac.uk/oca-bin/ocamain). Using symmetry operators, the EBI generates oligomeric structures from coordinates deposited in the PDB. Each structure downloaded from the EBI-MSD was superimposed onto its parent structure from the PDB using CNS[116] and visualized in O[117] to verify the accuracy of the symmetry expansion and to inspect the dinucleotide-binding site. In a few cases, the symmetry expansion was incorrect or incomplete, and CNS was used to create the correct quaternary structures from the corresponding PDB structures. In other cases, visual inspection revealed that a few subunits contained incompletely modeled NAD^+ molecules. These subunits were omitted from the subsequent analysis. Bound sulfate ions present due to the crystallization medium were removed.

To facilitate comparison, all the structures were superimposed onto a common structure using CNS. Each protein was superimposed onto the ADH structure of Meijers et al. (1HET) using the alpha-carbon coordinates of the first, third from last, and last residue of the phosphate-binding loop. The molecular visualization programs O and Protein Explorer[118] were used to examine the structures.

3.4.2 Hydrogen bonding calculations

Hydrogen bonds found in biological molecules typically have an acceptor-donor distance of 2.7-3.2 Å.[119] For the purposes of our study, the functional definition of a

hydrogen bond was the presence of an acceptor and a donor within 3.2 Å of each other. Only nitrogen and oxygen atoms were considered for our hydrogen bonding calculations (i.e. CH...O hydrogen bonds were not included in Figure 3.3). Due to the difficult nature of assigning hydrogen atom positions for water molecules and hydroxyl groups, as well as the absence of explicit hydrogen atoms in the crystal structures under study, an angle cutoff was not included in our definition of a hydrogen bond. Based on our functional definition, X-PLOR[120] was used to identify hydrogen bonds.

The data set of structures used in this study contains some redundancy; therefore, hydrogen-bonding calculations for each structure were weighted to avoid biasing the results toward enzymes heavily represented in our data set. Each enzyme contributed equally to the overall average by assigning each structure a weight equal to the inverse of the product of the number of sources per enzyme and the number of structures per source. For example, Table 3.1 lists three structures of UDP-galactose-4-epimerase from *H. sapiens* and six structures from *E. coli*. Thus, each *H. sapiens* UDP-galactose-4-epimerase structure would receive a weight of 1/6, while each *E. coli* UDP-galactose-4-epimerase structure would receive a weight of 1/12. Likewise, since there is only one source and one structure representing 7α-hydroxysteroid dehydrogenase (1FMC), it would receive a weight of 1. In addition, most of the structures used in this analysis are oligomers; therefore, the hydrogen bonding data were averaged over all the subunits within each structure. Each subunit contributed equally to the average for its structure.

Chapter 3 Acknowledgments

This project was partially supported by a Big 12 Faculty Fellowship Award to JJT.

Table 3.4. Atoms within 3.4Å of the structurally conserved water molecule.

For completeness, all atoms within 3.4Å of the structurally conserved water molecule are shown. As described in Methods, our functional definition of hydrogen bonds had a 3.2Å cutoff. Atoms in italics are greater than 3.2Å but less than 3.4Å from the structurally conserved water molecule. Therefore, all atoms in italics would not have been included in the calculation of hydrogen bonds to the structurally conserved water molecule. Carbon atoms were also excluded from the hydrogen bonding calculations. Atom notation is ATOMresidueNUMBER (e.g. Ngly201 = atom N of glycine residue 201).

Heading Abbreviations:
PDB ID, Protein Data Bank structure code; PDB wat, residue identification number of the structurally conserved water molecule of the corresponding PDB coordinate file; dinuc, dinucleotide; pyro, pyrophosphate; Gly3, last conserved glycine residue or equivalent; Gly1/Gly2, first and/or second conserved glycine or equivalent; C-term β4, atom from C-terminal residue of β-strand 4; Other N/O, nitrogen or oxygen atom not included in a previous column; dinuc C, dinucleotide carbon atom; Gly CA, alpha carbon of glycine residue; Other C, carbon atom not included in a previous column.

Atomic nomenclature:
C = backbone carbonyl carbon. CA = α carbon. CB = side chain β carbon. N = backbone nitrogen. ND2 = amide nitrogen of Asn side chain. O = backbone oxygen. OD1 = amide oxygen of Asn side chain. OG or OG1 = hydroxyl oxygen (i.e. of Ser or Thr). Note that ND2 and OD1 of Asn are essentially indistinguishable in the crystal structures studied. Atomic nomenclature for the dinucleotides conforms to figure 3.1.

PDB ID	PDB wat	dinuc	pyro	Å	Gly3	Å	Gly1/Gly2	Å	C-term β4	Å	Other N/O	Å	dinuc C	Å	Gly CA	Å	Other C	Å
1A4I	37	NDP	O2N	2.82	Ngly178	3.17	Nser174 OGser174	3.09 2.98	Oala215	2.81							*CBser174*	*3.33*
1B37	66	FAD	O1P O5'A	2.94 3.12	Ngly16	2.96	Ngly13	2.83	OGser265	2.67			*AC5**	*3.33*	CAgly13	3.12	*Cgly13* *CBser265*	*3.33* *3.36*
1B4V	631	FAD	O1P	2.69	Nala22	2.93	Ngly19	3.09	Ogly288	2.65	*Ngly21*	*3.34*			CAgly19	3.08	Cgly19 *CBala22*	3.06 *3.29*
1B5Q	39	FAD	O1P *O5'A*	2.87 *3.21*	Ngly16	2.90	Ngly13 *Ogly13*	2.81 *3.39*	OGser265	2.66	*OGser15*	*3.28*	*AC5**	*3.37*	CAgly13	3.10	*Cgly13*	*3.26*
1BMD	20	NAD	O2N	2.99	Ngly16	2.98	Ogly10	2.56							*CAgly13*	*3.21*	*Cgly13*	*3.29*
1BW9	33	NAD	O2N	2.68	Ngly187	2.79	Ngly184	3.01	Ocys238	2.87					*CAgly184*	*3.24*	*CAgly187*	*3.38*
1BXK	2	NAD	O2N	2.80	Ngly14	2.84	Ogly8	2.88							*CAgly11*	*3.39*		
1C0K	3002	FAD	O1P	2.72	Ngly1016	2.94	Ngly1013	2.87	Oala1178	2.76					*CAgly1013*	*3.31*		
1C0L	2002	FAD	O1P	2.61	Ngly1016	2.93	Ngly1013	2.78	Oala1178	2.84					*CAgly1013*	*3.21*		
1C0P	3003	FAD	O1P	2.70	Ngly1016	2.97	Ngly1013	2.90	Oala1178	2.76					*CAgly1013*	*3.32*		
1C1D	1163	NAD	O2N	2.77	Ngly187	2.92	Ngly184	3.04	Ocys238	2.80					*CAgly184*	*3.27*		
1C1X	1079	NAD	O2N	2.77	Ngly187	2.85	Ngly184	2.98	Ocys238	2.83					*CAgly184*	*3.29*		
1CF3	203	FAD	O1P	2.88	Ngly31	2.88	Ngly28	3.05	Oala288	2.75	*OG1thr30*	*3.37*			CAgly28	3.12	*Cgly28*	*3.27*
1CJC	16	FAD	O1P	2.64	Ngly18	2.90	Ngly15 *Ogly15*	2.73 *3.26*	Oser101	2.75					CAgly15	3.11	*Cgly15*	*3.26*

PDB ID	PDB wat	dinuc	pyro	Å	Gly3	Å	Gly1/Gly2	Å	C-term β4	Å	Other N/O	Å	dinuc C	Å	Gly CA	Å	Other C	Å
1COY	524	FAD	O1P	2.75	Ngly23	2.96	Ngly20	2.99	Oala288	2.68					*CAgly20*	*3.23*	Cgly20	3.20
1CYD	10	NAP	O2N *O5'A*	2.78 *3.24*	Ngly20	2.81	Ogly14	3.12	OD1asn83	2.71	Olys17	2.84						
1D7O	38	NAD	*O2N* O2A O5'A	*3.26* 3.07 3.12	Ngly33	3.00	Ogly25	2.86	*OGser136*	*3.27*					*CAgly33*	*3.40*		
1DNC	504	FAD	O1P	2.74	Ngly32	2.81	Ngly29 *Ogly29*	2.96 *3.37*	Oala155	2.79					*CAgly29*	*3.32*	*Cgly29* *CBala155*	*3.31* *3.36*
1DXY	341	NAD	O2N	2.78	Ngly157	2.98	Ngly154 *Ogly152*	2.81 *3.29*	Ohis204	2.81					*CAgly154*	*3.26*		
1E1M	9	FAD	O1P	2.60	Ngly18	2.82	Ngly15 Ogly15	2.81 3.16	Oser101	2.85					CAgly15	3.11	Cgly15	3.14
1E1M	192	NAP	O1N	2.44	*Nala157*	*3.24*	Ngly154	3.09	OGser328 Oser328	2.72 3.16					*CAgly154*	*3.21*		
1E6U	9	NAP	O2N	2.70	Ngly16	2.81	Ogly10	3.29	Oala62	2.82					*CAgly13*	*3.32*		
1E6W	3	NAD	O2N O5'A	2.83 3.20	Ngly23	2.88	Ogly17	3.01			Oser20	2.95	*NC5**	*3.23*				
1E7W	9	NDP	O2N	2.72	Ngly19	2.97	Ohoh6	3.12	ND2asn109	3.08	Olys16	3.14	*AC5**	*3.22*				
1EE2	6	NAD	O2N	2.78	Ngly203	3.03	Ngly200	2.93	Oval267	3.04					*CAgly200*	*3.32*		
1EK5	601	NAD	O2N O5'A	3.16 3.13	Ngly15	2.66	Ogly9	2.84							*CAgly12* *CAgly15*	*3.23* *3.27*	*Cgly12*	*3.31*
1EK6	1110	NADH	O2N O5'A	2.93 3.05	Ngly15	2.78	Ogly9	2.81							*CAgly12* *CAgly15*	*3.29* *3.28*	*Cgly12*	*3.33*
1EL5	11	FAD	O1P	2.68	Ngly15	2.89	Ngly12 Ogly10	3.04 3.11	Oser200	2.87								
1EL7	11	FAD	O1P *O5'A*	2.74 *3.34*	Ngly15	2.94	Ngly12 Ogly10	2.81 3.13	Oser200	3.03					CAgly12	3.15	*Cgly12*	*3.28*
1EL8	11	FAD	O1P	2.78	Ngly15	2.91	Ngly12 Ogly10	2.82 3.05	Oser200	2.87					*CAgly12*	*3.21*	*Cgly12*	*3.31*
1EMD	316	NAD	O2N *O5'A*	2.84 *3.29*	Ngly13	3.01	Ogly7	2.98	*Oser76*	*3.26*					*CAgly10*	*3.29*		
1ENO	45	NAD	NO5*	2.70	Ngly33	3.13	Ogly25	2.96	OGser136	3.08								
1F0Y	818	NAD	O2N *O5'A*	2.75 *3.37*	Ngly27	2.91	Ngly24	2.88	Oala107	2.75					*CAgly24*	*3.29*		
1FEC	52	FAD	O1P	2.61	Ngly15	2.82	Ngly12	3.05	Oala158	2.94					*CAgly12*	*3.28*	*Cgly12* *CBala158*	*3.37* *3.38*

PDB ID	PDB wat	dinuc	pyro	Å	Gly3	Å	Gly1/Gly2	Å	C-term β4	Å	Other N/O	Å	dinuc C	Å	Gly CA	Å	Other C	Å
1FL2	528	FAD	O1P	2.65	Ngly224	2.84	Ngly221 *Ogly221*	2.76 *3.34*	Oala321	2.70					*CAgly221*	*3.27*	*Cgly221*	*3.36*
1FMC	196	NAD	O2N *O5'A*	2.70 *3.22*	Ngly24	2.82	Ogly18	3.04	ND2asn95	2.85	Oala21	2.73	*NC5**	*3.39*				
1GAD	356	NAD	O2N	2.82	Ngly12	3.01	Ngly9 *Ogly7*	2.98 *3.37*	Oala95	3.10					CAgly9	3.11	*Cgly9*	*3.31*
1GD1	356	NAD	O2N	2.75	Ngly12	2.93	Ngly9 Ogly7	3.13 3.18	Oser95	2.85					*CAgly9*	*3.30*		
1GCO	1	NAD	O2N O5'A	2.69 2.84	Ngly20	2.87	Ogly14	3.17	OD1asn92	2.96	Othr17 *Nleu19*	2.78 *3.36*						
1GEG	530	NAD	O2N	2.69	Ngly15	2.83	Ogly9	2.90	ND2asn86	2.80	Ogln12	3.06	*NC5**	*3.35*				
1GER	11	FAD	O1P	2.75	Ngly16	2.93	Ngly13 *Ogly11*	2.90 *3.37*	Oala138	2.91					CAgly13	3.19	*Cgly13*	*3.24*
1GPE	21	FAD	O1P	2.83	Ngly36	2.97	Ngly33	2.93	Oala292	2.58	*OG1thr35*	*3.26*			CAgly33	3.04	*Cgly33*	*3.22*
1GRB	490	FAD	O1P	2.68	Ngly32	2.87	Ngly29 *Ogly29*	2.86 *3.38*	Oala155	2.89					CAgly29	3.10	Cgly29	3.16
1GRB	524	NDP	O1N	2.74	Nala199	3.10	Ngly196	2.88	Oala288	2.57					*CAgly196*	*3.24*	*Cgly196* CBala199	*3.35* 3.13
1GSN	504	FAD	O1P	2.66	Ngly32	2.95	Ngly29 *Ogly29*	2.83 *3.38*	Oala155	2.83					CAgly29	3.12	*Cgly29*	*3.21*
1H5Q	1125	NAP	O2N	2.76	Ngly24	3.11					Oarg21 Nile23	2.82 3.00	NC5*	3.19	CAgly22	3.35	*Cgly22*	*3.40*
1H7W	1171	FAD	O1P	2.57	Nser199 *OGser199*	3.01 *3.38*	Ngly196 Ogly196	2.92 3.17	Ogly282	2.73					CAgly196	3.10	Cgly196	3.14
1H82	232	FAD	O1P O5'A	2.77 3.09	Ngly16	2.91	Ngly13 *Ogly13*	2.92 *3.29*	OGser265	2.79	*OGser15*	*3.34*	*AC5**	*3.34*	CAgly13	3.14	*Cgly13*	*3.22*
1H83	122	FAD	O1P O5'A	2.85 3.08	Ngly16	3.01	Ngly13	2.86	OGser265	2.63			*AC5**	*3.35*	CAgly13	3.12	*Cgly13* *CBser265*	*3.35* *3.35*
1HDO	32	NAP	O2N *O3*	2.84 *3.31*	Ngly16	2.84	Ogly10	3.03	Oleu74	3.19					*CAgly13*	*3.23*	*Cgly13*	*3.32*
1HE2	27	NAP	O2N	2.90	Ngly16	2.84	Ogly10	3.04	*Oleu74*	*3.23*					*CAgly13*	*3.23*	*Cgly13*	*3.33*
1HE3	29	NAP	O2N	2.91	Ngly16	2.79	Ogly10	3.00	*Oleu74*	*3.26*					*CAgly13*	*3.21*	*Cgly13*	*3.30*
1HE4	29	NAP	O2N *O3*	2.86 *3.40*	Ngly16	2.81	Ogly10	3.08	Oleu74	3.20					*CAgly13*	*3.25*	*Cgly13*	*3.34*
1HE5	27	NAP	O2N	2.90	Ngly16	2.83	Ogly10	3.01	*Oleu74*	*3.23*					CAgly13	3.19	*Cgly13*	*3.33*

PDB ID	PDB wat	dinuc	pyro	Å	Gly3	Å	Gly1/Gly2	Å	C-term β4	Å	Other N/O	Å	dinuc C	Å	Gly CA	Å	Other C	Å
1HET	762	NAD	O2N	2.79	Ngly204	3.00	Ngly201	3.00	Oval268	3.04					*CAgly201*	*3.37*		
1HEU	629	NAD	O2N	2.79	Ngly204	3.00	Ngly201	3.00	Oval268	3.06					*CAgly201*	*3.36*		
1HYE	1007	NAP	O2N	3.06	Ngly13	2.87	Ogly7	2.86	OG1thr81	2.89							*CG1val12*	*3.23*
1HZJ	1101	NAD	O2N O5'A	2.82 3.13	Ngly15	2.77	Ogly9	2.87							*CAgly12* *CAgly15*	*3.33* *3.34*	*Cgly12*	*3.37*
1JA9	630	NDP	O2N	2.76	Ngly42	2.93	Ogly36	3.06	ND2asn114	2.98	Oarg39	2.78	AC4*	3.40				
1JAY	226	NAP	O2N	2.75	Ngly13	2.85	*Ngly10* Ogly7	*3.35* 3.12	*Othr71*	*3.32*					CAgly10	3.17	Cgly10	3.11
1LDG	13	NAD	O2N	2.87	Ngly32	2.95	Ngly29	2.77	OG1thr97	2.92					*CAgly29*	*3.21*	*Cgly29*	*3.40*
1NAH	15	NAD	O2N O5'A	2.71 3.17	Ngly13	2.77	Ogly7	2.85							*CAgly13*	*3.39*		
1OAA	12	NAP	O2N	2.70	Ngly21	2.97	Ogly15	3.05	ND2asn101	2.90	Oarg18 *Nphe20*	2.78 *3.32*						
1PBE	419	FAD	O1P	2.82	Ngly14	2.97	Ngly11	2.96	Ocys158	2.48					CAgly11	3.15	*Cgly11*	*3.32*
1QG6	7	NAD	O2N	2.70	Nala21	3.03	Ogly13	2.94	OGser91	2.84			*NC5**	*3.29*			*CBala21*	*3.33*
1QJD	908	FAD	O1P	2.71	Ngly138	2.93	Ngly135	2.94	Oala312	2.66					CAgly135	3.14	*Cgly135*	*3.28*
1QRR	444	NAD	O2N *O5'A*	2.71 *3.24*	Ngly14	2.71	Ogly8	2.80							*CAgly14*	*3.23*		
1QSG	74	NAD	O2N	2.88	Nala21	2.99	Ogly13	2.95	OGser91	2.86								
1UDA	509	NAD	O2N O5'A	2.80 3.10	Ngly13	2.84	Ogly7	2.80										
1UDB	523	NAD	O2N O5'A	2.77 3.18	Ngly13	2.71	Ogly7	2.90							*CAgly13*	*3.39*		
1UDC	514	NAD	O2N O5'A	2.71 3.15	Ngly13	2.79	Ogly7	2.89										
1XEL	1	NAD	O2N O5'A	2.82 3.14	Ngly13	2.79	Ogly7	2.81										
1YVE	314	NDP	O2N	2.73	Ngly134	2.84	Nala137	2.95	Oleu199	2.81					*CAgly134*	*3.40*	CBala137	3.09
2AE2	114	NAP	O2N O5'A	2.76 3.03	Ngly22	2.83	*Ogly16*	*3.25*	ND2asn94	2.89	Oarg19	2.81	*AC5** *AC4**	*3.31* *3.34*				
2OHX	77	NAD	O2N	2.73	Ngly204	3.07	Ngly201	2.91	Oval268	3.19					*CAgly201*	*3.28*		
2UDP	11	NAD	O2N *O5'A*	2.79 *3.23*	Ngly13	2.67	Ogly7	2.69							*CAgly13*	*3.22*		

PDB ID	PDB wat	dinuc	pyro	Å	Gly3	Å	Gly1/Gly2	Å	C-term β4	Å	Other N/O	Å	dinuc C	Å	Gly CA	Å	Other C	Å
3BTO	64	NAD	O2N	2.73	Ngly201	3.02	Ngly204	2.94	Oval268	3.07								
3COX	524	FAD	O1P	2.70	Ngly23	3.08	Ngly20	2.97	Oala288	2.70					CAgly20	3.15	Cgly20	3.16
3GRS	45	FAD	O1P	2.66	Ngly32	2.96	Ngly29	2.88	Oala155	2.88					CAgly29	3.17	Cgly29	3.19

Chapter 3 References

1. Franchetti P, Cappellacci L, Perlini P, Jayaram HN, Butler A, Schneider BP, Collart FR, Huberman E, Grifantini M. Isosteric analogues of nicotinamide adenine dinucleotide derived from furanfurin, thiophenfurin, and selenophenfurin as mammalian inosine monophosphate dehydrogenase (type I and II) inhibitors. J Med Chem 1998;41(10):1702-1707.

2. Faig M, Bianchet MA, Winski S, Hargreaves R, Moody CJ, Hudnott AR, Ross D, Amzel LM. Structure-based development of anticancer drugs: complexes of NAD(P)H:quinone oxidoreductase 1 with chemotherapeutic quinones. Structure (Camb) 2001;9(8):659-667.

3. Nagai M, Natsumeda Y, Konno Y, Hoffman R, Irino S, Weber G. Selective up-regulation of type II inosine 5'-monophosphate dehydrogenase messenger RNA expression in human leukemias. Cancer Res 1991;51(15):3886-3890.

4. Konno Y, Natsumeda Y, Nagai M, Yamaji Y, Ohno S, Suzuki K, Weber G. Expression of human IMP dehydrogenase types I and II in *Escherichia coli* and distribution in human normal lymphocytes and leukemic cell lines. J Biol Chem 1991;266(1):506-509.

5. Zhang R-g, Evans G, Rotella FJ, Westbrook EM, Beno D, Huberman E, Joachimiak A, Collart FR. Characteristics and crystal structure of bacterial inosine-5'-monophosphate dehydrogenase. Biochemistry 1999;38:4691-4700.

6. Aronov AM, Verlinde CL, Hol WG, Gelb MH. Selective tight binding inhibitors of trypanosomal glyceraldehyde-3-phosphate dehydrogenase via structure-based drug design. J Med Chem 1998;41(24):4790-4799.

7. Bressi JC, Verlinde CL, Aronov AM, Shaw ML, Shin SS, Nguyen LN, Suresh S, Buckner FS, Van Voorhis WC, Kuntz ID, Hol WG, Gelb MH. Adenosine analogues as selective inhibitors of glyceraldehyde-3-phosphate dehydrogenase of Trypanosomatidae via structure-based drug design. J Med Chem 2001;44(13):2080-2093.

8. Van Calenbergh S, Verlinde CL, Soenens J, De Bruyn A, Callens M, Blaton NM, Peeters OM, Rozenski J, Hol WG, Herdewijn P. Synthesis and structure-activity relationships of analogs of 2'-deoxy-2'-(3-methoxybenzamido)adenosine, a selective inhibitor of trypanosomal glycosomal glyceraldehyde-3-phosphate dehydrogenase. J Med Chem 1995;38(19):3838-3849.

9. Verlinde CL, Callens M, Van Calenbergh S, Van Aerschot A, Herdewijn P, Hannaert V, Michels PA, Opperdoes FR, Hol WG. Selective inhibition of trypanosomal glyceraldehyde-3-phosphate dehydrogenase by protein structure-based design: toward new drugs for the treatment of sleeping sickness. J Med Chem 1994;37(21):3605-3613.

10. Verlinde CLMJ, Hol WGJ. Structure-based drug design: Progress, results and challenges. Structure 1994;2:577-587.

11. Rossmann MG, Moras D, Olsen KW. Chemical and biological evolution of a nucleotide-binding protein. Nature 1974;250(463):194-199.

12. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247(4):536-540.

13. Branden C, Tooze J. Chapter 10: Enzymes that bind nucleotides. Introduction to Protein Structure. First ed. New York: Garland Publishing; 1991. p 141-159.

14. Wierenga RK, De Maeyer MCH, Hol WGJ. Interaction of pyrophosphate moieties with α-helixes in dinucleotide binding proteins. Biochemistry 1985;24:1346-1357.

15. Eschenbrenner M, Chlumsky LJ, Khanna P, Strasser F, Jorns MS. Organization of the multiple coenzymes and subunits and role of the covalent flavin link in the complex heterotetrameric sarcosine oxidase. Biochemistry 2001;40(18):5352-5367.

16. Nishiya Y, Imanaka T. Analysis of interaction between the *Arthrobacter* sarcosine oxidase and the coenzyme flavin adenine dinucleotide by site-directed mutagenesis. Appl Environ Microbiol 1996;62(7):2405-2410.

17. Rescigno M, Perham RN. Structure of the NADPH-binding motif of glutathione reductase: efficiency determined by evolution. Biochemistry 1994;33(19):5721-5727.

18. van Grunsven EG, van Berkel E, Ijlst L, Vreken P, de Klerk JB, Adamski J, Lemonde H, Clayton PT, Cuebas DA, Wanders RJ. Peroxisomal D-hydroxyacyl-CoA dehydrogenase deficiency: resolution of the enzyme defect and its molecular basis in bifunctional protein deficiency. Proc Natl Acad Sci 1998;95(5):2128-2133.

19. Lesk AM. NAD-binding domains of dehydrogenases. Curr Opin Struct Biol 1995;5(6):775-783.

20. Carugo O, Argos P. NADP-dependent enzymes. I: Conserved stereochemistry of cofactor binding. Proteins 1997;28(1):10-28.

21. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank (http://www.rcsb.org/). Nucleic Acids Res 2000;28(1):235-242.

22. Adolph HW, Zwart P, Meijers R, Hubatsch I, Kiefer M, Lamzin V, Cedergren-Zeppezauer E. Structural basis for substrate specificity differences of horse liver alcohol dehydrogenase isozymes. Biochemistry 2000;39(42):12885-12897.

23. Meijers R, Morris RJ, Adolph HW, Merli A, Lamzin VS, Cedergren-Zeppezauer ES. On the enzymatic activation of NADH. J Biol Chem 2001;276(12):9316-9321.

24. Al-Karadaghi S, Cedergren-Zeppezauer ES, Petratos K, Hovmoeller S, Terry H, Dauter Z, Wilson KS. Refined crystal structure of liver alcohol dehydrogenase-NADH complex at 1.8 Å resolution. Acta Crystallog sect D 1994;50:793-807.

25. Cho H, Ramaswamy S, Plapp BV. Flexibility of liver alcohol dehydrogenase in stereoselective binding of 3-butylthiolane 1-oxides. Biochemistry 1997;36(2):382-389.

26. Carpenter EP, Hawkins AR, Frost JW, Brown KA. Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. Nature 1998;394(6690):299-302.

27. Skarzynski T, Moody PC, Wonacott AJ. Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. J Mol Biol 1987;193(1):171-187.

28. Duee E, Olivier-Deyris L, Fanchon E, Corbier C, Branlant G, Dideberg O. Comparison of the structures of wild-type and a N313T mutant of *Escherichia coli* glyceraldehyde 3-phosphate dehydrogenases: implication for NAD binding and cooperativity. J Mol Biol 1996;257(4):814-838.

29. Barycki JJ, O'Brien LK, Strauss AW, Banaszak LJ. Sequestration of the active site by interdomain shifting. Crystallographic and spectroscopic evidence for distinct conformations of L-3-hydroxyacyl-CoA dehydrogenase. J Biol Chem 2000;275(35):27186-27196.

30. Dengler U, Niefind K, Kiess M, Schomburg D. Crystal structure of a ternary complex of D-2-hydroxyisocaproate dehydrogenase from *Lactobacillus casei*, NAD^+ and 2-oxoisocaproate at 1.9 Å resolution. J Mol Biol 1997;267(3):640-660.

31. Dunn CR, Banfield MJ, Barker JJ, Higham CW, Moreton KM, Turgut-Balik D, Brady RL, Holbrook JJ. The structure of lactate dehydrogenase from *Plasmodium falciparum* reveals a new target for anti-malarial design. Nature structural biology 1996;3(11):912-915.

32. Vanhooke JL, Thoden JB, Brunhuber NM, Blanchard JS, Holden HM. Phenylalanine dehydrogenase from *Rhodococcus sp.* M4: high-resolution X-ray analyses of inhibitory ternary complexes reveal key features in the oxidative deamination mechanism. Biochemistry 1999;38(8):2326-2339.

33. Brunhuber NM, Thoden JB, Blanchard JS, Vanhooke JL. *Rhodococcus* L-phenylalanine dehydrogenase: kinetics, mechanism, and structural basis for catalytic specificity. Biochemistry 2000;39(31):9174-9187.

34. Hegeman AD, Gross JW, Frey PA. Probing catalysis by Escherichia coli dTDP-glucose-4,6-dehydratase: identification and preliminary characterization of functional amino acid residues at the active site. Biochemistry 2001;40(22):6598-6610.

35. Hall MD, Banaszak LJ. Crystal structure of a ternary complex of *Escherichia coli* malate dehydrogenase citrate and NAD at 1.9 Å resolution. J Mol Biol 1993;232(1):213-222.

36. Kelly CA, Nishiyama M, Ohnishi Y, Beppu T, Birktoft JJ. Determinants of protein thermostability observed in the 1.9-Å crystal structure of malate dehydrogenase from the thermophilic bacterium *Thermus flavus*. Biochemistry 1993;32(15):3913-3922.

37. Mulichak AM, Theisen MJ, Essigmann B, Benning C, Garavito RM. Crystal structure of SQD1, an enzyme involved in the biosynthesis of the plant sulfolipid headgroup donor UDP-sulfoquinovose. Proc Natl Acad Sci 1999;96(23):13097-13102.

38. Thoden JB, Frey PA, Holden HM. Crystal structures of the oxidized and reduced forms of UDP-galactose 4-epimerase isolated from *Escherichia coli*. Biochemistry 1996;35(8):2557-2566.

39. Thoden JB, Hegeman AD, Wesenberg G, Chapeau MC, Frey PA, Holden HM. Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from *Escherichia coli*. Biochemistry 1997;36(21):6294-6304.

40. Thoden JB, Frey PA, Holden HM. Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from *Escherichia coli*: implications for the catalytic mechanism. Biochemistry 1996;35(16):5137-5144.

41. Thoden JB, Frey PA, Holden HM. High-resolution X-ray structure of UDP-galactose 4-epimerase complexed with UDP-phenol. Protein Science 1996;5(11):2149-2161.

42. Thoden JB, Wohlers TM, Fridovich-Keil JL, Holden HM. Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase. Biochemistry 2000;39(19):5691-5701.

43. Thoden JB, Wohlers TM, Fridovich-Keil JL, Holden HM. Human UDP-galactose 4-epimerase. Accommodation of UDP-N-acetylglucosamine within the active site. J Biol Chem 2001;276(18):15131-15136.

44. Tanaka N, Nonaka T, Tanabe T, Yoshimoto T, Tsuru D, Mitsui Y. Crystal structures of the binary and ternary complexes of 7α-hydroxysteroid dehydrogenase from *Escherichia coli*. Biochemistry 1996;35(24):7715-7730.

45. Yamamoto K, Kurisu G, Kusunoki M, Tabata S, Urabe I, Osaki S. Crystal structure of glucose dehydrogenase from *Bacillus megaterium* IWG3 at 1.7 Å resolution. J Biochem (Tokyo) 2001;129(2):303-312.

46. Powell AJ, Read JA, Banfield MJ, Gunn-Moore F, Yan SD, Lustbader J, Stern AR, Stern DM, Brady RL. Recognition of structurally diverse substrates by type II 3-hydroxyacyl-CoA dehydrogenase (HADH II)/amyloid-β binding alcohol dehydrogenase (ABAD). J Mol Biol 2000;303(2):311-327.

47. Otagiri M, Kurisu G, Ui S, Takusagawa Y, Ohkuma M, Kudo T, Kusunoki M. Crystal structure of meso-2,3-butanediol dehydrogenase in a complex with NAD^+ and inhibitor mercaptoethanol at 1.7 Å resolution for understanding of chiral substrate recognition mechanisms. J Biochem (Tokyo) 2001;129(2):205-208.

48. Roujeinikova A, Levy CW, Rowsell S, Sedelnikova S, Baker PJ, Minshull CA, Mistry A, Colls JG, Camble R, Stuitje AR, Slabas AR, Rafferty JB, Pauptit RA, Viner R, Rice DW. Crystallographic analysis of triclosan bound to enoyl reductase. J Mol Biol 1999;294(2):527-535.

49. Rafferty JB, Simon JW, Baldock C, Artymiuk PJ, Baker PJ, Stuitje AR, Slabas AR, Rice DW. Common themes in redox chemistry emerge from the X-ray structure of oilseed rape (*Brassica napus*) enoyl acyl carrier protein reductase. Structure 1995;3(9):927-938.

50. Ward WH, Holdgate GA, Rowsell S, McLean EG, Pauptit RA, Clayton E, Nichols WW, Colls JG, Minshull CA, Jude DA, Mistry A, Timms D, Camble R, Hales NJ, Britton CJ, Taylor IW. Kinetic and structural characteristics of the inhibition of enoyl (acyl carrier protein) reductase by triclosan. Biochemistry 1999;38(38):12514-12525.

51. Stewart MJ, Parikh S, Xiao G, Tonge PJ, Kisker C. Structural basis and mechanism of enoyl reductase inhibition by triclosan. J Mol Biol 1999;290(4):859-865.

52. Saridakis V, Christendat D, Kimber MS, Dharamsi A, Edwards AM, Pai EF. Insights into ligand binding and catalysis of a central step in NAD^+ synthesis: structures of *Methanobacterium thermoautotrophicum* NMN adenylyltransferase complexes. J Biol Chem 2001;276(10):7225-7232.

53. Biou V, Dumas R, Cohen-Addad C, Douce R, Job D, Pebay-Peyroula E. The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a

herbicidal transition state analog determined at 1.65 Å resolution. EMBO J 1997;16(12):3405-3415.

54. Ziegler GA, Schulz GE. Crystal structures of adrenodoxin reductase in complex with $NADP^+$ and NADPH suggesting a mechanism for the electron transfer of an enzyme family. Biochemistry 2000;39(36):10986-10995.

55. Karplus PA, Schulz GE. Substrate binding and catalysis by glutathione reductase as derived from refined enzyme: substrate crystal structures at 2 Å resolution. J Mol Biol 1989;210(1):163-180.

56. Prasad GS, Sridhar V, Yamaguchi M, Hatefi Y, Stout CD. Crystal structure of transhydrogenase domain III at 1.2 Å resolution. Nature structural biology 1999;6(12):1126-1131.

57. Pereira PJ, Macedo-Ribeiro S, Parraga A, Perez-Luque R, Cunningham O, Darcy K, Mantle TJ, Coll M. Structure of human biliverdin IXβ reductase, an early fetal bilirubin IXβ producing enzyme. Nature structural biology 2001;8(3):215-220.

58. Warkentin E, Mamat B, Sordel-Klippert M, Wicke M, Thauer RK, Iwata M, Iwata S, Ermler U, Shima S. Structures of F420H2:$NADP^+$ oxidoreductase with and without its substrates bound. EMBO J 2001;20(23):6561-6569.

59. Rosano C, Bisso A, Izzo G, Tonetti M, Sturla L, De Flora A, Bolognesi M. Probing the catalytic mechanism of GDP-4-keto-6-deoxy-d-mannose Epimerase/Reductase by kinetic and crystallographic characterization of site-specific mutants. J Mol Biol 2000;303(1):77-91.

60. Lee BI, Chang C, Cho SJ, Eom SH, Kim KK, Yu YG, Suh SW. Crystal structure of the MJ0490 gene product of the hyperthermophilic archaebacterium *Methanococcus jannaschii*, a novel member of the lactate/malate family of dehydrogenases. Journal of molecular biology 2001;307(5):1351-1362.

61. Liao DI, Thompson JE, Fahnestock S, Valent B, Jordan DB. A structural account of substrate and inhibitor specificity differences between two naphthol reductases. Biochemistry 2001;40(30):8696-8704.

62. Tanaka N, Nonaka T, Nakanishi M, Deyashiki Y, Hara A, Mitsui Y. Crystal structure of the ternary complex of mouse lung carbonyl reductase at 1.8 Å resolution: the structural origin of coenzyme specificity in the short-chain dehydrogenase/reductase family. Structure 1996;4(1):33-45.

63. Horer S, Stoop J, Mooibroek H, Baumann U, Sassoon J. The crystallographic structure of the mannitol 2-dehydrogenase $NADP^+$ binary complex from *Agaricus bisporus*. J Biol Chem 2001;276(29):27555-27561.

64. Zhang J, Martasek P, Paschke R, Shea T, Siler Masters BS, Kim JJ. Crystal structure of the FAD/NADPH-binding domain of rat neuronal nitric-oxide synthase. Comparisons with NADPH-cytochrome P450 oxidoreductase. J Biol Chem 2001;276(40):37506-37513.

65. Auerbach G, Herrmann A, Gutlich M, Fischer M, Jacob U, Bacher A, Huber R. The 1.25 Å crystal structure of sepiapterin reductase reveals its binding mode to pterins and brain neurotransmitters. EMBO J 1997;16(24):7219-7230.

66. Yamashita A, Kato H, Wakatsuki S, Tomizaki T, Nakatsu T, Nakajima K, Hashimoto T, Yamada Y, Oda J. Structure of tropinone reductase-II complexed with $NADP^+$ and pseudotropine at 1.9 Å

resolution: implication for stereospecific substrate binding and catalysis. Biochemistry 1999;38(24):7630-7637.

67. Allaire M, Li Y, MacKenzie RE, Cygler M. The 3-D structure of a folate-dependent dehydrogenase/cyclohydrolase bifunctional enzyme at 1.5 Å resolution. Structure 1998;6(2):173-182.

68. Gourley DG, Schuttelkopf AW, Leonard GA, Luba J, Hardy LW, Beverley SM, Hunter WN. Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites. Nat Struct Biol 2001;8(6):521-525.

69. Whitlow M, Howard AJ, Stewart D, Hardman KD, Kuyper LF, Baccanari DP, Fling ME, Tansik RL. X-ray crystallographic studies of *Candida albicans* dihydrofolate reductase. High resolution structures of the holoenzyme and an inhibited ternary complex. J Biol Chem 1997;272(48):30289-30298.

70. Whitlow M, Howard AJ, Stewart D, Hardman KD, Chan JH, Baccanari DP, Tansik RL, Hong JS, Kuyper LF. X-Ray crystal structures of *Candida albicans* dihydrofolate reductase: high resolution ternary complexes in which the dihydronicotinamide moiety of NADPH is displaced by an inhibitor. J Med Chem 2001;44(18):2928-2932.

71. Sawaya MR, Kraut J. Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. Biochemistry 1997;36(3):586-603.

72. Matthews DA, Bolin JT, Burridge JM, Filman DJ, Volz KW, Kaufman BT, Beddell CR, Champness JN, Stammers DK, Kraut J. Refined crystal structures of *Escherichia coli* and chicken liver dihydrofolate reductase containing bound trimethoprim. J Biol Chem 1985;260(1):381-391.

73. Bolin JT, Filman DJ, Matthews DA, Hamlin RC, Kraut J. Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution. I. General features and binding of methotrexate. J Biol Chem 1982;257(22):13650-13662.

74. Li R, Sirawaraporn R, Chitnumsub P, Sirawaraporn W, Wooden J, Athappilly F, Turley S, Hol WG. Three-dimensional structure of *M. tuberculosis* dihydrofolate reductase reveals opportunities for the design of novel tuberculosis drugs. J Mol Biol 2000;295(2):307-323.

75. Champness JN, Achari A, Ballantine SP, Bryant PK, Delves CJ, Stammers DK. The structure of *Pneumocystis carinii* dihydrofolate reductase to 1.9 Å resolution. Structure 1994;2(10):915-924.

76. Cody V, Galitsky N, Rak D, Luft JR, Pangborn W, Queener SF. Ligand-induced conformational changes in the crystal structures of *Pneumocystis carinii* dihydrofolate reductase complexes with folate and $NADP^+$. Biochemistry 1999;38(14):4303-4312.

77. Ziegler GA, Vonrhein C, Hanukoglu I, Schulz GE. The structure of adrenodoxin reductase of mitochondrial P450 systems: electron transfer for steroid biosynthesis. J Mol Biol 1999;289(4):981-990.

78. Bieger B, Essen LO. Crystal structure of the catalytic core component of the alkylhydroperoxide reductase AhpF from *Escherichia coli*. J Mol Biol 2001;307(1):1-8.

79. Li J, Vrielink A, Brick P, Blow DM. Crystal structure of cholesterol oxidase complexed with a steroid substrate: implications for flavin adenine dinucleotide dependent alcohol oxidases. Biochemistry 1993;32(43):11507-11515.

80. Yue QK, Kass IJ, Sampson NS, Vrielink A. Crystal structure determination of cholesterol oxidase from *Streptomyces* and structural characterization of key active site mutants. Biochemistry 1999;38(14):4277-4286.

81. Umhau S, Pollegioni L, Molla G, Diederichs K, Welte W, Pilone MS, Ghisla S. The x-ray structure of D-amino acid oxidase at very high resolution identifies the chemical mechanism of flavin-dependent substrate dehydrogenation. Proc Natl Acad Sci 2000;97(23):12463-12468.

82. Dobritzsch D, Schneider G, Schnackerz KD, Lindqvist Y. Crystal structure of dihydropyrimidine dehydrogenase, a major determinant of the pharmacokinetics of the anti-cancer drug 5-fluorouracil. EMBO J 2001;20(4):650-660.

83. Taylor P, Pealing SL, Reid GA, Chapman SK, Walkinshaw MD. Structural and mechanistic mapping of a unique fumarate reductase. Nature structural biology 1999;6(12):1108-1112.

84. Wohlfahrt G, Witt S, Hendle J, Schomburg D, Kalisz HM, Hecht HJ. 1.8 and 1.9 Å resolution structures of the *Penicillium amagasakiense* and *Aspergillus niger* glucose oxidases as a basis for modelling substrate complexes. Acta Crystallog sect D 1999;55(5):969-977.

85. Mittl PR, Schulz GE. Structure of glutathione reductase from *Escherichia coli* at 1.86 Å resolution: comparison with the enzyme from human erythrocytes. Protein science : a publication of the Protein Society 1994;3(5):799-809.

86. Becker K, Savvides SN, Keese M, Schirmer RH, Karplus PA. Enzyme inactivation through sulfhydryl oxidation by physiologic NO-carriers. Nature structural biology 1998;5(4):267-271.

87. Karplus PA, Schulz GE. Refined structure of glutathione reductase at 1.54 Å resolution. J Mol Biol 1987;195(3):701-729.

88. Schreuder HA, Prick PA, Wierenga RK, Vriend G, Wilson KS, Hol WG, Drenth J. Crystal structure of the p-hydroxybenzoate hydroxylase-substrate complex refined at 1.9 Å resolution. Analysis of the enzyme-substrate and enzyme-product complexes. J Mol Biol 1989;208(4):679-696.

89. Binda C, Coda A, Angelini R, Federico R, Ascenzi P, Mattevi A. A 30 Å long U-shaped catalytic tunnel in the crystal structure of polyamine oxidase. Structure Fold Des 1999;7(3):265-276.

90. Binda C, Angelini R, Federico R, Ascenzi P, Mattevi A. Structural bases for inhibitor binding and catalysis in polyamine oxidase. Biochemistry 2001;40(9):2766-2776.

91. Wagner MA, Trickey P, Chen ZW, Mathews FS, Jorns MS. Monomeric sarcosine oxidase: 1. Flavin reactivity and active site binding determinants. Biochemistry 2000;39(30):8813-8824.

92. Strickland CL, Karplus PA. Overexpression of *Crithidia fasciculata* trypanothione reductase and crystallization using a novel geometry. Acta Crystallog sect D 1995;51:337-341.

93. Faig M, Bianchet MA, Talalay P, Chen S, Winski S, Ross D, Amzel LM. Structures of recombinant human and mouse NAD(P)H:quinone oxidoreductases: species comparison and structural changes with substrate binding and release. Proc Natl Acad Sci 2000;97(7):3177-3182.

94. Winski SL, Faig M, Bianchet MA, Siegel D, Swann E, Fung K, Duncan MW, Moody CJ, Amzel LM, Ross D. Characterization of a mechanism-based inhibitor of NAD(P)H:quinone

oxidoreductase 1 by biochemical, X-ray crystallographic, and mass spectrometric approaches. Biochemistry 2001;40(50):15135-15142.

95. Kraulis PJ. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. Journal of Applied Crystallography 1991;24:946-950.

96. Merritt EA, Bacon DJ. Raster3D Photorealistic Molecular Graphics. Methods Enzymol 1997;277:505-524.

97. Schultze P, Feigon J. Chirality errors in nucleic acid structures. Nature 1997;387(6634):668.

98. Chu PY, Hwang MJ. New insights for dinucleotide backbone binding in conserved C5'-H . . . O hydrogen bonds. J Mol Biol 1998;279(4):695-701.

99. Scheiner S, Kar T, Gu Y. Strength of the $C^{\alpha}H..O$ hydrogen bond of amino acid residues. J Biol Chem 2001;276(13):9832-9837.

100. Raymer ML, Sanschagrin PC, Punch WF, Venkataraman S, Goodman ED, Kuhn LA. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. J Mol Biol 1997;265(4):445-464.

101. Marrone TJ, Briggs JM, McCammon JA. Structure-based drug design: computational advances. Annu Rev Pharmacol Toxicol 1997;37:71-90.

102. Luscombe NM, Laskowski RA, Thornton JM. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res 2001;29(13):2860-2874.

103. Janin J. Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. Structure Fold Des 1999;7(12):R277-279.

104. Likic VA, Juranic N, Macura S, Prendergast FG. A "structural" water molecule in the family of fatty acid binding proteins. Protein science 2000;9(3):497-504.

105. Sreenivasan U, Axelsen PH. Buried water in homologous serine proteases. Biochemistry 1992;31(51):12785-12791.

106. Poornima CS, Dean PM. Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins. J Comput Aided Mol Des 1995;9(6):521-531.

107. Dessen A, Quemard A, Blanchard JS, Jacobs WR, Jr., Sacchettini JC. Crystal structure and function of the isoniazid target of *Mycobacterium tuberculosis*. Science 1995;267(5204):1638-1641.

108. Lamzin VS, Dauter Z, Popov VO, Harutyunyan EH, Wilson KS. High resolution structures of holo and apo formate dehydrogenase. J Mol Biol 1994;236(3):759-785.

109. Pantano S, Alber F, Lamba D, Carloni P. NADH interactions with WT- and S94A-acyl carrier protein reductase from *Mycobacterium tuberculosis*: An ab initio study. Proteins 2002;47(1):62-68.

110. Mattevi A, Tedeschi G, Bacchella L, Coda A, Negri A, Ronchi S. Structure of L-aspartate oxidase: implications for the succinate dehydrogenase/fumarate reductase oxidoreductase family. Structure Fold Des 1999;7(7):745-756.

111. Gleason WB, Fu Z, Birktoft J, Banaszak L. Refined crystal structure of mitochondrial malate dehydrogenase from porcine heart and the consensus structure for dicarboxylic acid oxidoreductases. Biochemistry 1994;33(8):2078-2088.

112. Korkhin Y, Kalb AJ, Peretz M, Bogin O, Burstein Y, Frolow F. NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of *Clostridium beijerinckii* and *Thermoanaerobacter brockii*. J Mol Biol 1998;278(5):967-981.

113. Zens AP, Bryson TA, Dunlap RB, Fisher RR, Ellis PD. Nuclear magnetic resonance studies of pyridine dinucleotides.7.1 The solution conformational dynamics of the adenosine portion of nicotinamide adenine dinucleotide and other related purine containing compounds. J Am Chem Soc 1976;98(24):7559-7564.

114. Smith PE, Tanner JJ. Molecular dynamics simulations of NAD^+ in solution. J Am Chem Soc 1999;121:8637-8644.

115. Smith PE, Tanner JJ. Conformations of nicotinamide adenine dinucleotide (NAD^+) in various environments. J Mol Recognit 2000;13(1):27-34.

116. Brunger 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, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta crystallographica Section D, Biological crystallography 1998;54 (Pt 5):905-921.

117. Jones TA, Zou JY, Cowan SW, Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallog sect A 1991;47 (Pt 2):110-119.

118. Martz E. "Protein Explorer Software", http://proteinexplorer.org. For a list of further references, see http://molvis.sdsc.edu/protexpl/pe_lit.htm. 2001.

119. Jeffrey GA. An Introduction to Hydrogen Bonding. New York: Oxford University Press; 1997.

120. Brunger AT. X-PLOR, Version 3.1:A system for x-ray crystallography and NMR. New Haven: Yale University Press; 1992.

Vita

Christopher A. Bottoms was born in Oklahoma City, Oklahoma in 1975. He graduated from Piedmont High School in Piedmont, Oklahoma in 1993. He served as a missionary in Spain from 1994 to 1996 with The Church of Jesus Christ of Latter-day Saints. He graduated with a B.S. in Biochemistry and Biology from Oklahoma City University in 1999. He received his Ph.D. in Genetics from the University of Missouri-Columbia in 2005. In 2005 he was awarded a National Library of Medicine Postdoctoral Informatics Fellowship in the Health Management and Informatics Department, University of Missouri-Columbia.