Conservation of Electrostatic Properties within Enzyme Families and Superfamilies†

Dennis R. Livesay,‡,§ Per Jambeck,|| Atipat Rojnuckarin,# and Shankar Subramaniam*,‡,||

Department of Chemistry, University of Illinois, Urbana, Illinois 61820, USA, Departments of Bioengineering and Chemistry and Biochemistry, University of California at San Diego, San Diego, California 92093-0412, USA, and Department of Chemical Engineering, University of Wisconsin, Madison, Wisconsin 53706, USA

Received September 27, 2002; Revised Manuscript Received January 30, 2003

ABSTRACT: Electrostatic interactions play a key role in enzyme catalytic function. At long range, electrostatics steer the incoming ligand/substrate to the active site, and at short distances, electrostatics provide the specific local interactions for catalysis. In cases in which electrostatics determine enzyme function, orthologs should share the electrostatic properties to maintain function. Often, electrostatic potential maps are employed to depict how conserved surface electrostatics preserve function. We expand on previous efforts to explain conservation of function, using novel electrostatic sequence and structure analyses of four enzyme families and one enzyme superfamily. We show that the spatial charge distribution is conserved within each family and superfamily. Conversely, phylogenetic analysis of key electrostatic residues provide the evolutionary origins of functionality.

The limited number of observed protein folds (∼1000) (*1*) has led to the grouping of similar enzymes into families (>50% pairwise sequence identity) and superfamilies (<50% pairwise sequence identity). Members of the same family or superfamily generally catalyze the same (or very similar) reactions on structurally related substrates. Recently, more diverse enzyme superfamilies have been identified (*2*) in which members catalyze very different overall reactions, while maintaining a common mechanistic strategy. These observations have led us to investigate a wide array of protein families and superfamilies to uncover what evolution has conserved to maintain function. We show that enzyme families and superfamilies across species often use key electrostatic properties as the *leit-motif* for maintaining similar functionality despite tolerating significant sequence and structural variability.

It is widely recognized that catalytic properties of an enzyme are impacted by electrostatics. This is shown experimentally, as well as computationally, by varying the pH and ionic strength of the solvent. Electrostatics are important in both intramolecular and intermolecular forces. The overall stability of a protein structure is considered a sum of contributions of intermolecular electrostatic interactions, hydrogen bonding, van der Waals, and hydrophobic interactions. In fact, it has been shown (*3*) that thermophiles often stabilize protein structures through enhanced electrostatic interactions. The same types of ionic side chain interactions that stabilize protein structures often mediate complex formation. At long ranges, electrostatics dominate enzyme−substrate encounters through steering forces. Through steering forces, the approaching substrate molecule is directed to the active site of the enzyme by the complementary electric field around the enzyme (*4*−*6*).

Recently, much work has focused on using electrostatic potential surfaces as tools for predicting function. Comparisons of electrostatics profiles has been employed to derive functional similarities across protein families. Exemplars where common functional regions in proteins have been characterized include cholinesterases (ChEs)[1] and a class of cell-adhesion proteins (*7*). Analysis of the electrostatic potentials of homology models of 104 members of the Pleckstrin homology (PH) domain (*8*) reveal that electrostatic properties are generally conserved, despite large amounts of sequence diversity. Electrostatic potentials have also been used to identify functionally important regions in RNA structures (*9*), despite the uniform distribution of negative charge. Analysis of the electrostatic potentials of RNA molecules reveals unusual electrostatic features that correspond to functionally important regions.

In this work, we show that conservation of various electrostatic properties is a common theme used by protein families and superfamilies to maintain function. We present here the results of four protein families: the CuZnSOD family, the c-type lysozyme family, the ferritin family, and the myoglobin family, and one structurally diverse super-

† This work was supported by National Science Foundation Grants DBI 96-04223 and MCB9873384, National Institutes of Health Grant R01 GM 46535, and a National Center for Supercomputing Applications metacenter computer allocation to S.S.

* Corresponding author: Shankar Subramaniam. Department of Bioengineering, University of California at San Diego, La Jolla, CA 92093. Telephone: 858-822-0986. E-mail: shankar@ucsd.edu.

‡ University of Illinois.

|| University of California at San Diego.

University of Wisconsin.

§ Present address: Department of Chemistry, California State Polytechnic University Pomona, Pomona, California, 91768, USA.

[1] Abbreviations: ChE, cholinesterase; PH, Pleckstrin homology; SCOP, structural classification of proteins; UHBD, University of Houston Brown Dynamics; LPBE, linearized Poisson−Boltzmann equation; NLPBE, nonlinear Poisson−Boltzmann equation; OPLS, optimized potentials for liquid systems; CuZnSOD, copper, zinc superoxide dismutase; PDF, probability density function; RMS, root-mean-square; BD, Brownian dynamics; MR, mandalate racemase; MLE, muconate lactonizing enzyme.

10.1021/bi026918f CCC: $25.00
Published on Web 03/04/2003

family: the enolase superfamily. The five examples are taken from the four SCOP classes (structural classification of proteins): α, *β*, α, and *β*, and α + *β*. The examples are chosen based on their structural and functional diversity, amount of structural information available, and amount of experimental functional data. In each case, key sequence/structure motifs have been conserved across the family as a way of maintaining function through electrostatics. The role electrostatics play in each family varies. In the CuZnSOD family, electrostatics (*6*) define the long-range steering forces. In the myoglobin family, electrostatics (*10*−*11*) provide for the surface potential that allows the enzyme to interact with complementary enzymes. In the ferritin family, electrostatics provide for monomer association (*12*−*13*), leading to its iron transport ability. And, in the c-type lysozyme family (*14*) and the enolase superfamily electrostatics provide the specific short-range acid/base interactions for catalysis (*15*). In each example, the overall pH profile of key electrostatic properties is qualitatively conserved within the family or superfamily.

METHODS

Continuum Electrostatic and Brownian Dynamics Calculations. The displayed electrostatic potential maps and residue pK_a values are calculated using the University of Houston Brownian Dynamics (UHBD) suite of programs (*16*). UHBD implements the method described in Gilson (*17*) and Antosiewicz et al. (*18*) to calculate pK_a values of titrating residues in a protein. Table 1 presents the PDB accession number and species of all structures studied. In each case, only one structure from each species listed in the family was studied. In both cases, the linearized Poisson−Boltzmann equation (LPBE) was solved using the Choleski preconditioned conjugate gradient method. The protein was centered on a 65 × 65 × 65 grid with each grid unit equaling 1.5 Å. Use of the more computationally expensive nonlinear Poisson−Boltzmann equation (NLPBE) does not result in qualitatively different electrostatic potential maps. Focusing is used around each titrating site with the grid spacing becoming 1.2, 0.75, and 0.25 Å. A solvent dielectric constant of 80 is used with a protein dielectric constant of 20 is used for all pK_a calculations. An interior dielectric of 4 is used for all electrostatic potential map calculations and Brownian dynamics simulations. Protein partial-charges are taken from the CHARMM parameter set (*19*) and radii from the Optimized Potentials for Liquid Systems (OPLS) (*20*). The ionic strength equals 150 mM and the temperature is 298 K. The calculations take between 1 and 5 h each on a single Silicon Graphics Origin2000 R10 000 processor at the National Center for Supercomputing Applications.

For the purposes of calculating Brownian dynamics encounter rate constants, the electrostatic forces were obtained using the NLPBE. To get higher accuracy, we used a larger grid (127 × 127 × 127 grid with each grid unit equaling 1.0 Å). All other electrostatic parameters are the same as above. The substrate is modeled as a single sphere with a hydrodynamic radius of 2.05 Å and an exclusion radius equal to that of the van der Waal's radius of an oxygen atom, 1.5 Å. The reaction criterion is set at 7.0 Å between the copper ion and diffusing substrate, and the b and q surfaces are set at 80 and 400 Å, respectively. Each simulation was run for 20 000 trajectories on a Silicon Graphics Origin2000 supercomputer at the National Center for Supercomputing Applications and took approximately 7 h on a single R10 000 processor.

Table 1: Species and PDB Identification Codes for Each Structure Studied

CuZnSOD family	
human	1SPD
bovine	2SOD
clawed frog	1XSO
spinach	1SRD
yeast	1SDY
P. leiognathi	1YAI
E. coli	1ESO
enolase superfamily	
enolase family	
S. cerevisiae	1EBG
S. cerevisiae	1EBH
S. cerevisiae	1ELS
S. cerevisiae	1NEL
S. cerevisiae	1ONE
S. cerevisiae	2ONE
H. vulgaris	1PDY
H. vulgaris	1PDZ
S. cerevisiae	3ENL
S. cerevisiae	4ENL
S. cerevisiae	5ENL
S. cerevisiae	6ENL
S. cerevisiae	7ENL
MLE-like family	
P. aeruginosa	1DTN
P. aeruginosa	1MDL
P. putida	1MDR
P. putida	1MNS
P. putida	1MRA
P. putida	2MNR
P. putida	1BKH
P. putida	1MUC
P. putida	2MUC
P. putida	3MUC
A. eutrophus	2CHR
P. putida	1BQG
C-type lysozyme family	
C. virginianus	1DKJ
hen egg white	1HEL
guinea fowl	1HHL
Japanese quail	2IHL
T. aculeatus	1JUG
rainbow trout	1LMN
hen egg white	1LZY
human	1REX
horse milk	2EQL
pheasant	1JHL
myoglobin family	
horse heart	1AZI
Asian elephant	1EMY
sea hare	1MBA
common seal	1MBS
pig	1MYG
yellowfin tuna	1MYT
human	2MM1
ferritin family	
E. caballus	1AEW
E. coli	1BCF
human	1FHA
R. catesbeiana	1RCD
D. vulgaris	1RYT

Sequence Alignments and Phylogenetic Analysis. The Pfam database (*21*) contains 143 copper, zinc superoxide dismutase

(CuZnSOD), and 98 enolase superfamily sequences. All evolutionary precursor and viral sequences are purged, and the remaining sequences are aligned using HMMER (*22*), a profile Hidden Markov model software. In each case, the seed alignment is the structural alignment of generated by the All-to-All Protein Structure Alignment server from San Diego Supercomputer Center (*23*). Phylogenetic trees are calculated from the entire HMMER sequence alignments and a subset of electrostatically important residues using CLUSTAL W (*24*) suite of programs.

The electrostatic subsequence used in the CuZnSOD example is taken from first and second electrostatic shell regions and the S−S subloop, which are regions identified by mutagenesis experiments (*25*) as being important to the shape of the electrostatic field about the enzyme (for further discussion see Results and Discussion). The electrostatic subsequence used in the enolase superfamily example was taken from the original alignment that includes conserved areas of significant calculated pK_a shifts from their aqueous values.

Charged Atom Probability Density Functions. The charged atom/target atom probability density functions (PDFs) are created by counting the occurrence of charged atoms within 16 Å sphere about the copper ion and then smoothed using an exponential smoothing function. In each case, only the structure used in the pH profile calculation is studied, except in the enolase superfamily where all solved structures are studied. The target atom varies across the five families: CuZnSOD family − catalytic copper ion, enolase superfamily − average substrate α-carbon position, c-type lysozyme family − average substrate cleavage site (oxygen atom), myoglobin family − heme iron ion, and ferritin family − structural iron ion in the center of the four-helix bundle.

RESULTS AND DISCUSSION

The Copper, Zinc Superoxide Dismutase Family. Copper, zinc superoxide dismutases (CuZnSODs) are metalloenzymes that protect cells from the oxidative damage of superoxide (O_2^-) free radicals, a toxic byproduct of aerobic metabolism. The enzyme is a homodimer with 153 amino acid residues (the human ortholog) with one copper and one zinc ion per subunit. Each monomer is composed of a slightly flattened Greek key β-barrel core. Dismutation of superoxide occurs by the alternate reduction and oxidation of the active site copper ion, yielding one molecule each of oxygen and hydrogen peroxide. CuZnSOD is extremely efficient in catalysis, with a rate constant that approaches the diffusion limit. However, the active site constitutes a very small portion of the enzyme surface and a uniform collision mechanism fails to predict such a high catalytic rate. The experimentally determined ionic strength dependence of the reaction rate indicates that the observed kinetics is driven by electrostatic guidance (*26*) where the anionic substrate is steered down the cationic potential toward the active site copper ion. Because the rate constant is so high, it is generally accepted that the reaction proceeds without major conformational changes in the enzyme.

Thus far, X-ray structures for seven wild-type CuZnSODs have been solved (*27*−*32*). All structures are very similar, particularly in the active site regions. However, there are key differences between eukaryotic and prokaryotic enzymes (*33*). In all structures, six histidine residues and a single aspartate residue coordinate the copper and zinc ions. The copper ion is coordinated by atoms His46 ND1, His48 NE2, His63 NE2, and His120 NE2 (henceforth, numbering will always represent human CuZnSOD). The zinc ion is coordinated by the remaining imidazole nitrogen of His63 (ND1), a carboxylate oxygen of Asp83 (OD2), and ND1 atoms of His71 and His80. Investigations of the active site have revealed a handful of residues responsible for the "cationic funnel" shape of the electric field about the active site (*6*, *25*, *34*−*36*). Arg143 (conserved across all known CuZnSOD sequences) is most certainly key to enzyme efficiency. The activity of R143K mutants is much lower than their wild-type counterparts (*4*), leading to the hypothesis that the role of Arg143 is more than just electrostatic. From the essentially conserved position of the side chain, it has been suggested that the residue also helps to orient the incoming superoxide free radical toward the active site copper ion.

Bordo et al. (*25*) have identified four residues at the upper rim of the active site that, along with Arg143, define the shape and strength of the electric field. The residues are at positions 132, 133, 136, and 137, and are designated as *first shell electrostatic residues*. They also identify four other residues that have smaller, but still significant, effects on the electrostatic field. These are designated as *second shell electrostatic residues*, and are at positions 121, 122, 131, and 135. The electrostatically active residues are not strictly evolutionarily conserved. The active site residues in prokaryotic enzymes are substantially rearranged as compared to that (*33*) from the eukaryotes. In fact, in the structural alignment of all available CuZnSOD monomers there are gaps in the prokaryotic sequences at positions 132, 133, and 136, all of which are electrostatic shell residues. The prokaryotic enzymes compensate for these deletions with the insertion of an electrostatically active dilysine loop (with signature sequence KDxK) proximal to the active site. Quaternary structure differences in the active form of the enzyme also exist between the prokaryotic and eukaryotic structures. In the eukaryotes, all enzymes are homodimers with very similar interfaces. *Photobacterium leiognathi* was the very first prokaryotic CuZnSOD structure determined (*37*−*38*) that revealed a novel dimer interface unlike that of the eukaryotic structures. Recently, the solved *E. coli* structure has revealed that the active enzyme is a monomer (*32*).

CuZnSOD has evolved to be one of the most functionally conserved enzymes in nature, yet the sequence, structure, and net charge vary considerably across species. Among the seven sequences with known structure, the identities range from 25.0 to 81.7%. Differences in structure occur mainly in differences of loop length and amino acid content, with the most obvious example being the added dilysine loop of prokaryotic enzymes. Structural differences also manifest themselves in α-carbon root-mean-square (RMS) differences up to 2.4 Å. The net charge of the enzymes also varies over a wide range. The net charge of the seven known structures ranges from −8 to +2. Despite the evolutionarily distance between CuZnSODs, key sequence/structure motifs must be conserved that maintain the efficiency of the enzymes near the diffusion limit. The conservation manifests itself as very similar electrostatic fields surrounding the enzyme. A cursory

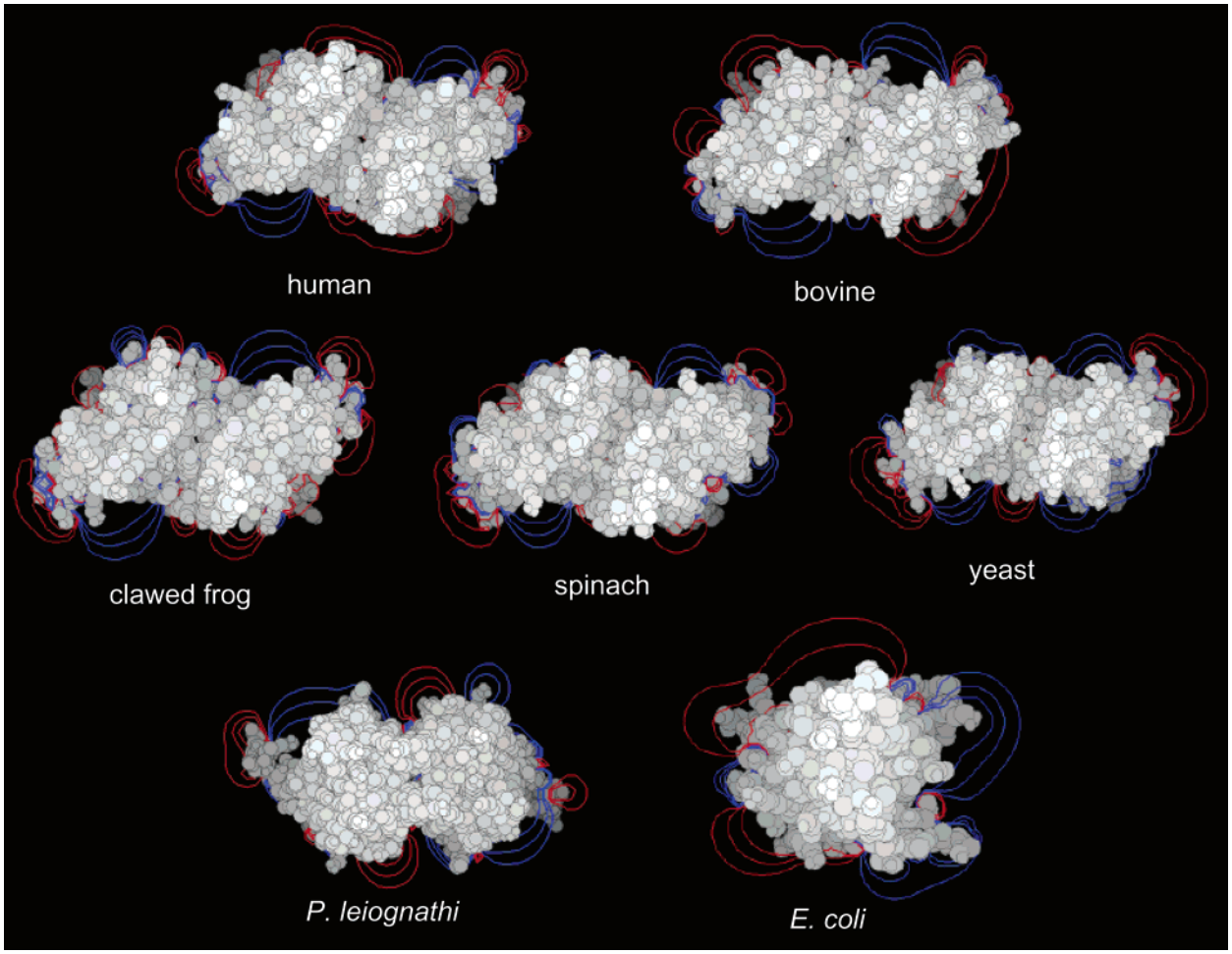

FIGURE 1: Electrostatic potential contours around the (A) human, (B) bovine, (C) clawed frog, (D) spinach, (E) yeast, (F) *P. leiognathi*, and (G) *E. coli* CuZnSOD enzymes. Red indicates the negative potential, while blue indicates the positive potential. The lines represent potential values at 2, 1, and 0.5 kcal/mol/e. The cross section is taken through the average *z* coordinate of the copper ions. The five eukaryotic structures are in the same orientation and the active sites are indicated by the cationic potential on opposite side of the homodimer. *P. leiognathi* has a different dimer interface and the location of its active sites is different from the eukaryotic structures; however, the cationic funnel of the active site is still quite obvious. The *E. coli* structure is a monomer, and its active site is indicated by the cationic funnel to the right of the structure shown.

examination of the electrostatic potential maps (Figure 1) reveals how the incoming superoxide ion is steered toward the active site through a conserved "cationic funnel".

Brownian dynamics (BD) calculated rate constants for the seven known structures (Table 2) are the quantitative compliment of electrostatic potential maps. BD simulations of all known CuZnSOD structures support the claim that the evolutionarily distant enzymes have conserved function with rate constants close to the diffusion limit. The calculated values presented here are in good agreement with experimental (*4*, *33*, *35*, *36*, *39*) and previously calculated results (*6*, *33*, *35*, *36*, *39*−*43*). The reaction rates range for the functional enzyme are conserved within the same order of magnitude ($\times 10^9$ 1/M s). Both bovine structures have the fastest calculated rates (6.25 and 5.92 $\times 10^9$ 1/M s), whereas the *E. coli* enzyme (which functions as a monomer) displays the slowest calculated rate (0.86 $\times 10^9$ 1/M s). In all cases, the calculated rate of the dimers is approximately the sum of the rates of its two constituent monomers chains. The differences of same species X-ray dimer pair ranges from 0.5 to 8.6%, which is much smaller than the variability across species. The key result of the calculations is that the premise that function has been conserved to preserve the catalytic rate near the diffusion limit is strongly supported by the calculated rate constants.

From the alignment of all 116 CuZnSOD sequences in the PFAM database, it is apparent that the primary structure conservation is not absolute. No electrostatic shell residues are absolutely conserved, and only four positions in the electrostatic loop are absolutely conserved (Asp124, Gly138, Gly141, and Arg143). This raises the question of "how does nature maintain the efficiency of the enzyme?" Three key electrostatic motifs have been discovered that provide insight into the nature of the underlying conservation. The conserved electrostatic motifs range from simple local conservation of charge to complex evolutionary origins of the CuZnSOD family.

The simplest of the conserved electrostatic motifs, first reported in 1994 by Bordo et al. (*25*), is very straightforward. The net charge carried on the electrostatic loop is conserved at or near −1 (Figure 2) through concerted mutations as seen in the complete 116-CuZnSOD sequence alignment with the HMMER program. When there has been a charge changing mutation at one point on the electrostatic loop, there is generally a charge compensating mutation elsewhere. It has been suggested that the conservation of charge at the

Table 2: Brownian Dynamics Calculated Rate Constants for CuZnSOD-Superoxide Encounter[a]

species	PDB	monomers	homodimers	charge[b]
human	1SPD	A 2.09 ± 0.16		
		B 3.18 ± 0.20	AB 5.04 ± 0.25	−4
bovine	2SOD	O 3.07 ± 0.20		
		B 2.65 ± 0.18	OB 6.25 ± 0.27	−2
		Y 3.76 ± 0.22		
		G 2.90 ± 0.19	YG 5.92 ± 0.27	−2
clawed frog	1XSO	A 2.13 ± 0.17		
		B 2.24 ± 0.17	AB 4.46 ± 0.23	−6
spinach	1SRD	A 2.31 ± 0.17		
		B 2.54 ± 0.18	AB 4.37 ± 0.23	−8
		C 2.61 ± 0.18		
		D 2.34 ± 0.17	CD 4.78 ± 0.24	−8
yeast	1SDY	A 2.18 ± 0.17		
		B 2.30 ± 0.17	AB 4.14 ± 0.23	0
		C 2.27 ± 0.17		
		D 2.93 ± 0.19	CD 4.12 ± 0.23	0
P. leiognathi	1YAI	A 3.58 ± 0.21	AA 6.83 ± 0.28	+2
		B 2.86 ± 0.19		
		C 3.27 ± 0.20	BC 5.57 ± 0.26	+2
E. coli[c]	1ESO	A 0.86 ± 0.11		−2

[a] Rate constants were calculated with the University of Houston Brown dynamics suite of programs using the full nonlinear Poisson−Boltzmann equation and are × 10^9 1/M s the reported value. [b] Mean pH 7.0 charge. [c] The *E. coli* CuZnSOD functions as a monomer.

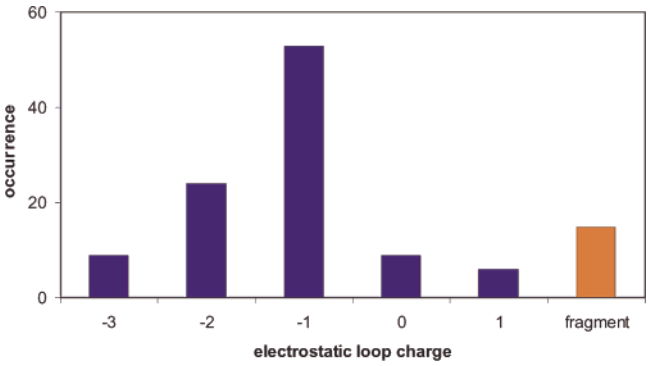

FIGURE 2: Histogram of net electrostatic loop charge. Charge is conserved at or near −1 through concerted mutations. Fifteen fragment sequences without terminal electrostatic loops have been included in the overall CuZnSOD alignment.

secondary structure level is doubly important: (i) for electrostatic steering purposes (*34*, *44*), and (ii) as an energetically stable way of maintaining the delicate charge balance on the loop (*45*).

The observation that prokaryotic orthologs (which lack several of the key electrostatic shell residues) qualitatively maintain the electric field about the enzyme reveals that the shape of the field depends on more than just the electrostatic loop charge. By comparing the charge of only the electrostatic loops across species, charge contributions from positions further away in sequence space (such as the dilysine subloop) are neglected. We have constructed charged atom/copper ion pairwise distance probability density functions (PDFs) for each of the seven structures to probe deeper into the sequence/structure/electrostatics relationship. The distance PDFs are analogous to the pair distributions in liquids, except that our sampling is from the aligned sequences of the SOD enzyme, and we count all charged atoms around a point (in this case the copper atom). The PDFs plot the occurrence of a charged atom occurring within 16 Å of the copper ion. This overcomes the awkwardness associated with a sequence position dependent description, i.e., the primary structure around an active site. From Figure 3A, it is evident that the qualitative shape of the PDFs is conserved across the seven structures. Conserved PDFs indicate a conserved spatial distribution of charge around the active site. The residues that lead to the observed PDFs are not entirely restricted to the electrostatic loop, indicating that function is maintained by a more global structural relationship.

The sequence-level evolutionary origins of the conserved active site charge distribution can be explained using phylogeny. Alignment of key electrostatic residues is sufficient to produce qualitatively equivalent phylogenetic trees for the CuZnSOD enzyme family (Figure 4). The electrostatic residues investigated are three short (five to 10 characters) contiguous sequences taken at the first and second electrostatic shells, as well as the prokaryotic KDxK motif. The alignment of the key electrostatic residues is less than 10 percent of the total alignment. The average composition of the subsequence is 4.2 acidic residues, 3.5 basic residues, 3.4 polar-neutral residues, 5.3 nonpolar residues, and 6.2 gaps. A consensus residue was identified for 10 of the 22 positions within the subsequence. In each case, the phylogenetic tree groups the enzyme family into seven groups: bacteria, extracellular/yeast, mammals/reptiles/fish, nematoda, antropoda, and two plant groups. The electrostatic subsequence tree appears to be made up of less sequences, but this is an illusion arising from the drastically reduced amount of information based on sequence variability across the protein family (i.e., the lines in the subsequence tree often represent more than one sequence). This result shows that despite little observable sequence conservation from a cursory examination, there exists a strong correlation between sequence and the functionality of the enzyme. This is strong support for our hypothesis that function (electrostatic steering) has driven the evolution of the CuZnSOD enzyme family.

The Enolase Superfamily. Members of the enolase superfamily catalyze different overall reactions while retaining a common mechanistic strategy. Members belong to the α/*β*-barrel structural class and are composed of two major domains and a third, smaller C-terminal domain (*15*). Each of the reactions catalyzed by members of the superfamily involves the formation of an enolic intermediate derived from the carboxylate substrate by abstraction of an α-proton (*46*). The energetically expensive reaction is made feasible by a combination of electrostatic stabilization and hydrogen bonding interactions with weakly acidic functional groups. Depending on the exact reaction, the intermediate is diverted to the observed products by different active site functional groups.

As is typical for α/*β*-barrel proteins, the active sites of members of the enolase superfamily are located at the carboxy-terminal ends of the barrel domains, with an independently conserved amino-terminal domain providing additional substrate binding interactions. The functional groups are located at the ends of most or all of the *β*-sheets or in the loops that connect the *β*-sheet with the following α-helix. This architecture provides a modular design such that functional groups can evolve independently to enable different types of stabilization of enolic intermediates. Recent evidence suggests that electrostatics largely contributes to this design strategy. The electrostatic properties of seven α/*β*-

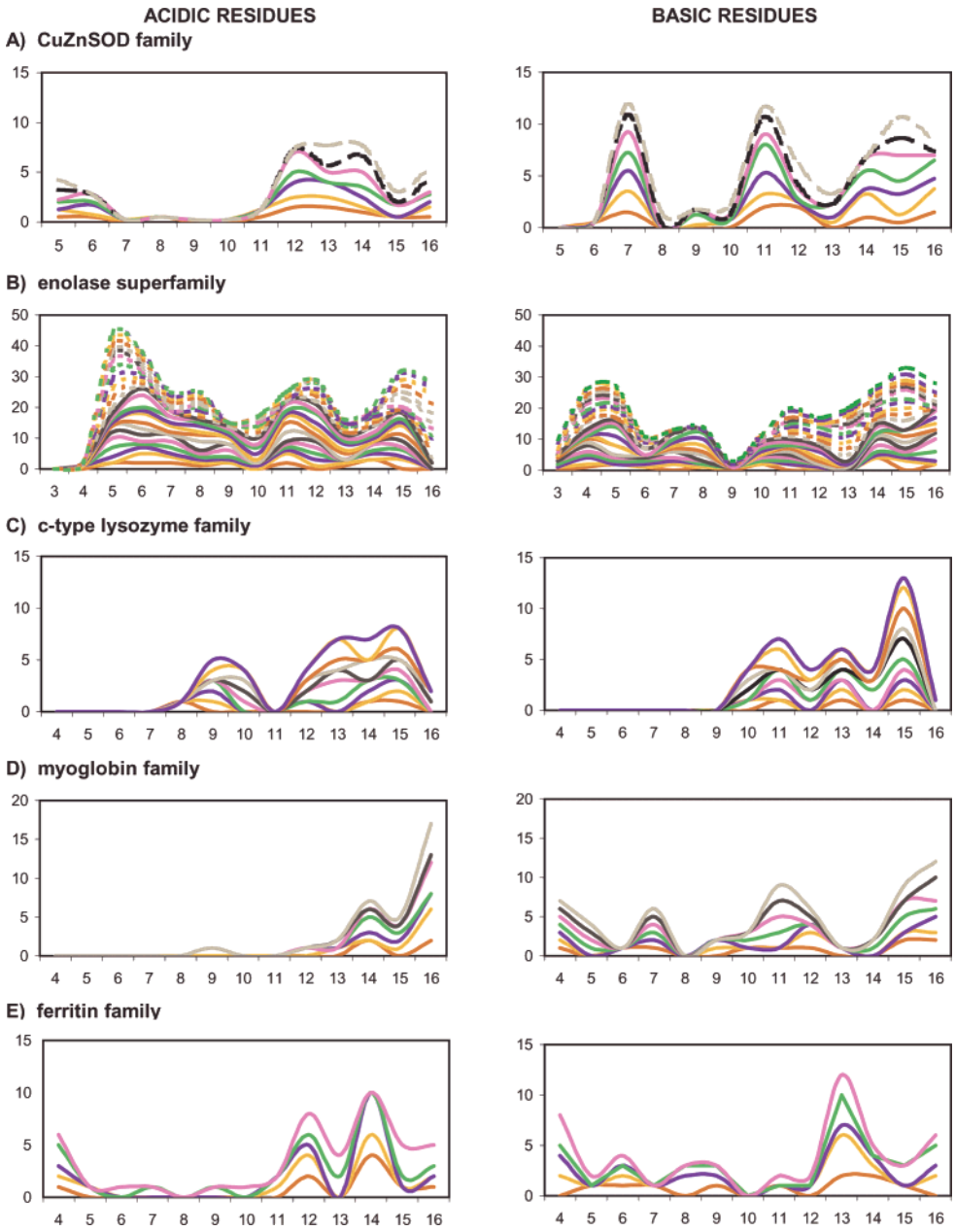

FIGURE 3: Stacked charged atom probability density functions (PDFs) plot acidic and basic residues occurring within 16 Å of some target atom, meaning the results for each subsequent protein are displayed in relation to the previous. (A) The target atom for the CuZnSOD family is the active site copper ion. The eukaryotic structures are shown in solid lines, while prokaryotic structures are shown in dashed lines. (B) The target atom for the enolase superfamily is the α-carbon of the carboxycylic acid substrate. The enolase family is shown in solid lines, while the mandelate racemase-muconate lactonizing enzyme family is shown in dashed lines. The target atoms for (C) the c-type lysozyme family, (D) the myoglobin family, and the (E) the ferritin family are the cleavage site of the cell wall substrate, the heme iron atom, and iron ion in the center of the four-helix bundle, respectively. The identity of each protein within the figure (top to bottom) is the same as the order in Table 1.

barrel proteins from different enzyme families (including mandelate racemase from the enolase superfamily) have been investigated (*47*). In each case, the backbone of the α/*β*-barrel results in the same electric field about the enzyme. Addition of side chains modulates the field pattern to focus on a specific area near the active site. This design strategy is used in other superfamilies based on α/*β*-barrel proteins and may explain why this fold represents nearly 10% of all structurally characterized proteins.

On the basis of sequence similarity, mandelate racemase (MR) has been shown to belong to the enolase superfamily (*47*). MR equilibrates the enantiomers of mandelate by a Mg^{2+}-dependent mechanism that involves the abstraction of the α-proton by a general base, Lys 166 or His 297, depending on the enantiomer. Muconate lactonizing enzyme (MLE) catalyzes a completely different reaction, but also belongs to the enolase superfamily. MLE generates its enolic intermediate through a base catalyzed proton abstraction via a Lys 166 homologue (Lys 169). Not all members of the enolase superfamily have exactly homologous catalytic base residues. In fact, when considering all members of the superfamily with known structures (enolase, MR, MLE, chloromuconate cycloisomerase, and galactonate dehydratase), in no case are both position and identity of the catalytic residue conserved. The position of catalytic atoms is, however, often conserved. This type of conservation

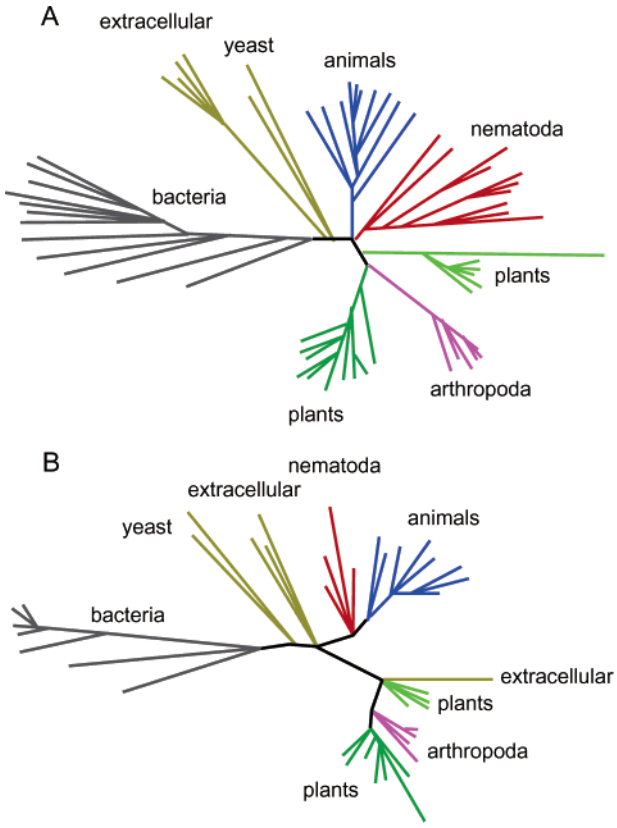

FIGURE 4: Phylogenetic trees built from (A) the complete CuZn-SOD alignment, and (B) a subsequence (<10%) taken from the original alignment that includes those positions deemed important by experimental mutagenesis experiments (Bordo et al., 1994) to the shape and magnitude of the electric field about the enzyme. The two phylogenetic trees are qualitatively the same, hinting at the evolutionary origins of the CuZnSOD family. Randomly generated test cases of same length fail to qualitatively cluster in the same groupings.

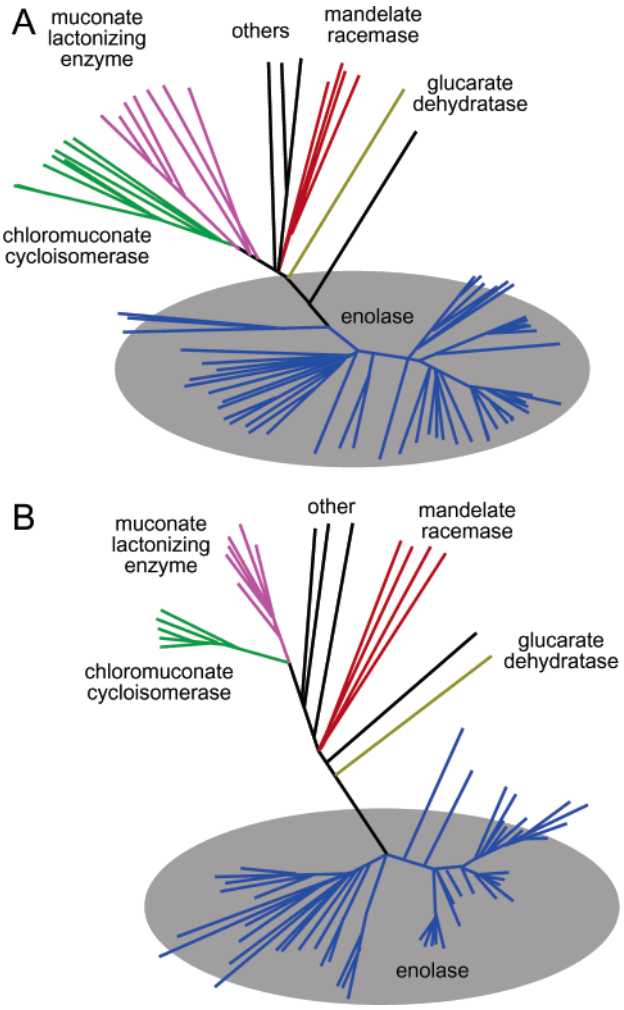

FIGURE 5: Phylogenetic trees built from (A) the complete enolase superfamily alignment, and (B) a subsequence (<13%) taken from the original alignment that includes those positions that display conserved areas of significant calculated pK_a shifts from their aqueous values (see Figure 6). The two phylogenetic trees are qualitatively the same. Randomly generated test cases of the same length fail to qualitatively cluster in the same groupings.

makes employing structural information a necessity for bioinformatic investigations (sequence based inferences) to be useful. To probe the structural conservation, we have constructed charged atom:substrate α-carbon pairwise distance probability density functions (PDFs) for each member of the enolase superfamily. The PDFs plot the occurrence of a charged atom occurring within 16 Å of the mean α-carbon position. From Figure 3B, it is evident that the qualitative shape of the PDFs is conserved across the entire enolase superfamily. The residues that lead to the PDFs are not strictly conserved. This result suggests that the structural placement of charged atoms is conserved despite differences in the amino acid sequence.

Phylogenetic analysis of key electrostatic residues shows that complex sequence correlations exist in the enolase superfamily (Figure 5). Owing to insufficient experimental evidence available to guide our subsequence selection, continuum electrostatics methods are used to calculate pK_a values of titrating residues on each of the five species. Conserved areas of significant pK_a shifts from aqueous values are used to identify key electrostatic residues. Figure 6 shows the four areas (which represent less than 13% of the overall sequence) that are extracted from the complete sequence alignment to create an electrostatically relevant subsequence. The average composition of the subsequence is 7.7 acidic residues, 6.9 basic residues, 12.6 polar-neutral residues, 33.4 nonpolar residues, and 18.3 gaps. A consensus residue was identified for 25 of the 79 positions within the subsequence. Phylogenetic analysis of the complete sequence alignment and the electrostatically relevant subsequence results in qualitatively equivalent trees. Each tree splits the enzyme superfamily into seven groups: enolases, MRs, MLEs, chrloromuconate cycloisomerases, glucarate dehydratases, and two groups of others. As is the case within the CuZnSOD family phylogenetic comparisons establish that electrostatics have driven the evolution of the enolase superfamily. Randomly generated test cases of the same length fail to qualitatively cluster in the same groupings.

Other Protein Families. A cursory sampling of other families reveals the conserved spatial distribution of charge to be a robust evolutionary mechanism. Charge−charge PDFs reveal the same types of electrostatic conservation in the c-type lysozyme, myoglobin, and ferritin families. These families compliment the previous two examples discussed. Taken together, the five examples are representative of the four main SCOP classes. The examples are chosen based on their structural and functional diversity, amount of structural information available, and amount of experimental

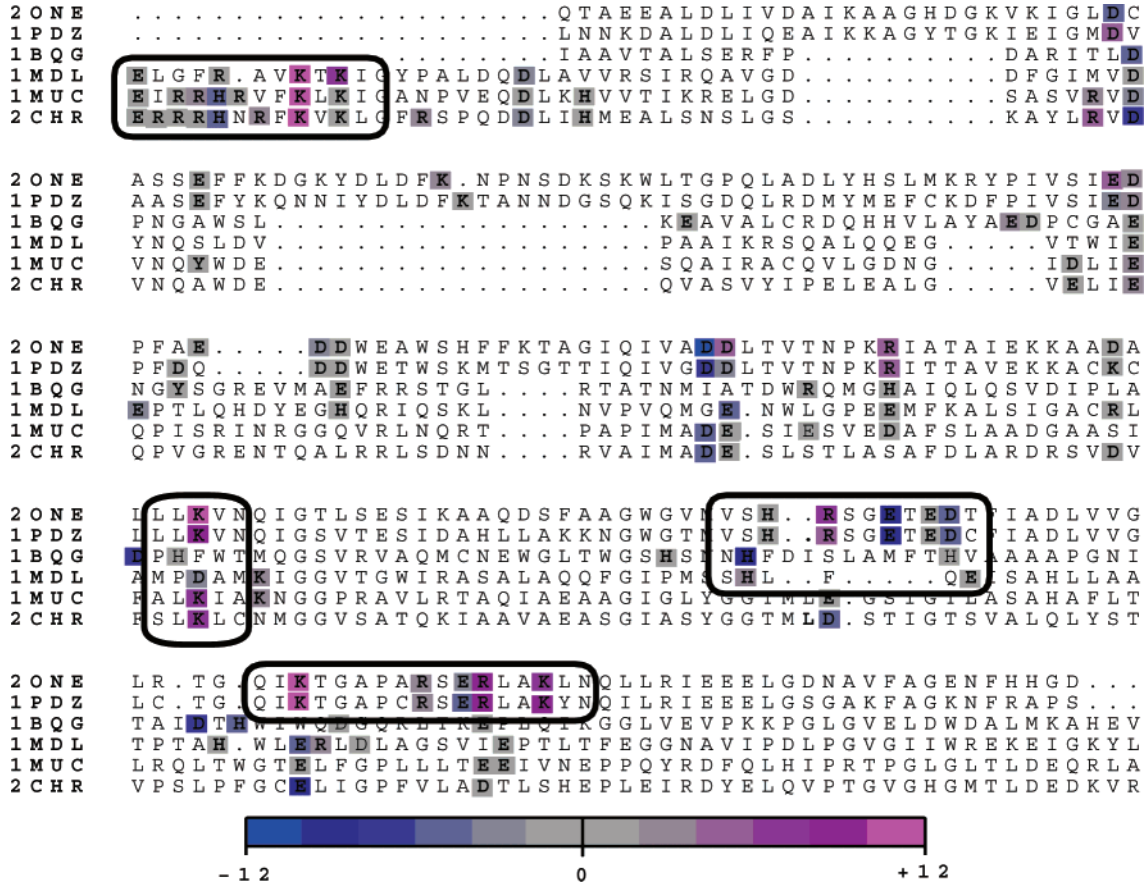

FIGURE 6: Multiple sequence alignment of each representative structure within the enolase superfamily colored by shifts in pK_a values from the aqueous values. The enolase family is represented by structures 2one and 1pdz, whereas the mandelate racemase-muconate lactonizing enzyme family is represented by the remaining structures. Conserved areas of significant pK_a shifts are used to identify an appropriate electrostatic sub-sequence. Two of the four conserved regions include the experimentally identified catalytic residues.

functional data. Additionally, the functional role of electrostatics varies between each family. Lysozyme is a small hydrolytic enzyme of ubiquitous distribution, found especially in tears and other animal secretions, and in hen egg whites. The enzyme dissolves certain types of bacteria by hydrolyzing 1,4-β linkages between *N*-acetylmuramic acid and *N*-acetyl-D-glucosamine in peptidoglycans of bacterial cell walls (*48*). As in the enolase superfamily electrostatics provide for the specific acid/base interactions of catalysis in the c-type lysozyme family (*15*). Beginning with its published 2.0 Å structure in 1960 (*49*), the structure/function relationships in myoglobin are arguably the most studied to date. The globular oxygen-carrying enzyme is extremely compact and almost entirely α-helical. The electrostatic surface potential of myoglobin provides the complementary surfaces for interaction with electron-transfer partners (*50*). Ferritin is an iron storage protein found in tissues. The protein is made up of 24 subunits arranged with octahedral symmetry (*12*−*13*). Each monomer is a four-helix bundle. The active protein has a large internal cavity (∼80 Å diameter) that can hold up to 4500 ferric ions. Regions of calculated charge density correlates well with previously identified portions of the proteins' active site (*12*). Further, theoretical studies have confirmed the importance of electrostatics in mediating the active ferritin assembly (*13*). Like the CuZnSOD family and the enolase superfamily, in each of these examples, the overall spatial organization of charge has been conserved (see Figure 3C,D). In each case, the conserved spatial distribution of charge has allowed for sequence variation across species, while maintaining functionality.

CONCLUSIONS

The basic tenet of evolution of protein families is the conservation of function. Function arises from the structural properties of the particular enzyme, which in turn arise from sequence. The subtle sequence/structure/function relationships are often very complex and difficult to decipher. We present here four protein families and one superfamily where function is maintained through conservation of key electrostatic properties. We demonstrate through phylogenetic analysis of only the electrostatically relevant residues, that electrostatics is a driving-force for evolution. Further, the importance of structure in preserving the electrostatic environment of the families is clearly presented through the charge probability density functions, electrostatic potential maps and association rate constants.

ACKNOWLEDGMENT

We wish to thank Dr. Andrew McCammon for providing the UHBD suite of programs for calculation of the rate constants.

REFERENCES

1. Chothia, C., and Murzin, A. G. (1993) New folds for all-beta proteins. *Structure 1*, 217−222.

2. Babbitt, P. C., and Gerlt, J. A. (1997) Understanding enzyme superfamilies. *J. Biol. Chem. 272*, 30591−30594.
3. Xiao, L., and Honig, B. (1999) Electrostatic contributions to the stability of hyperthermophilic proteins. *J. Mol. Biol. 289*, 1435−1444.
4. Fisher, C. L., Cabelli, D. E., Tainer, J. A., Hallewell, R. A., and Getzoff, E. D. (1994) The role of arginine 143 in the electrostatics and mechanism of Cu, Zn superoxide dismutase: computational and experimental evaluation by mutational analysis. *Proteins 19*, 24−34.
5. Kozack R. E., and Subramaniam, S. (1993) Brownian dynamics simulations of molecular recognition in an antibody−antigen system. *Protein Sci. 2*, 915−926.
6. Getzoff, E. D., Cabelli, D. E., Fisher, C. L., Parge, H. E., Viezzoli, M. S., Banci, L., and Hallewell, R. A. (1992) Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature 358*, 347−351.
7. Botti, S. A., Felder, C. E., Sussman, J. L., and Silman, I. (1998) Electrotactins: a class of adhesion proteins with conserved electrostatic and structural motifs. *Prot. Eng. 11*, 415−420.
8. Blomberg, N., Gabdoulline, R. R., Nilges, M., and Wade, R. C. (1999) Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins 37*, 379−387.
9. Chin, K., Sharp, K. A., Honig, B., and Pyle, A. M. (1999) Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nat. Struct. Biol. 6*, 1055−1061.
10. Muller, J. D., McMahon, B. H., Chien, E. Y., Sligar, S. G., and Nienhaus, G. U. (1999) Connection between the taoxonomic substates and protonation of histidines 64 and 97 in carbonmonoxy myoglobin. *Biophys. J. 72*, 1036−1051.
11. Ramos, C. H., Kay, M. S., and Baldwin, R. L. (1999) Putative interhelix ion pairs involved in the stability of myoglobin. *Biochemistry* 27, 9783−9790.
12. Douglas, T., and Ripoll, D. R. (1998) Calculated electrostatic gradients in recombinant human H-chain ferritin. *Protein Sci. 7*, 1083−1091.
13. Takahashi, T. (1997) Significant role of electrostatic interactions for stabilization of protein assemblies. *Adv. Biophys. 34*, 41−54.
14. Muraki, M., Harata, K., Hayashi, Y., Machida, M., and Jigami, Y. (1991) The importance of precise positioning of negatively charged carboxylate in the catalytic action of human lysozyme. *Biochim. Biophys. Acta 1079*, 229−337.
15. Zhang, E., Brewer, J. M., Minor, W., Carreira, L. A., and Lebioda, L. (1997) Mechanism of enolase: the crystal structure of asymmetric dimmer enolase-2-phospho-D-glycerate/enolase-phosphenolpyruvate at 2.0 Å resolution. *Biochemistry 36*, 12526−12534.
16. Madura, J. D., Briggs, J. M., Wade, R. C., Davis, M. E., Lutty, B. A., Ilin, A., Antosiewicz, J., Gilson, M. K., Gagheri, B., Scott, L. R., and McCammon, J. A. (1995) Electrostatics and diffusion of molecules in solution, simulations with the University of Houston Brownian dynamics program. *Comput. Phys. Commun. 91*, 57−95.
17. Gilson, M. K. (1993) Multiple-site titration and molecular modeling, two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins 15*, 266−282.
18. Antosiewicz, J., McCammon, J. A., and Gilson, M. K. (1994) Prediction of pH-dependent properties of proteins. *J. Mol. Biol. 238*, 415−436.
19. Brooks, R. B., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM, a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem. 4*, 187−217.
20. Jorgensen, W. L., and Tirado-Rives, J. (1988) The OPLS potential function for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc. 110*, 1657−1666.
21. Sonhammer, E. L., Eddy, S. R., Birney, D., Bateman, A., and Durbin, R. (1998) A comprehensive database of protein families based on seed alignments. *Proteins 28*, 405−420.
22. Eddy, S. R. (1996) Hidden Markov models. *Curr. Opin. Struct. Biol. 6*, 361−365.
23. Shindyalov, I. N., and Bourne, P. E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng. 11*, 739−747.
24. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Res. 22*, 4673−4680.
25. Bordo, D., Djinovic, K., and Bolognesi, M. (1994) Conserved patterns in the Cu,Zn superoxide dismutase family. *J. Mol. Biol. 238*, 366−386.
26. Lepock, J. R., Arnold, L. D., Torrie, B. H., Andrews, B., and Kruuv, J. (1985) Structural analyses of various Cu2+, Zn2+-superoxide dismutases by differential scanning calorimetry and Raman spectroscopy. *Arch. Biochem. Biophys. 241*, 243−251.
27. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S., and Richardson, D. C. (1982) Determination and analysis of the 2 Å structure of copper, zinc superoxide dismutase. *J. Mol. Biol. 160*, 181−217.
28. Parge, H. E., Getzoff, E. D., Scandell, C. S., Hallewell, R. A., and Tainer, J. A. (1986) Crystallographic characterization of recombinant human CuZn superoxide dismutase. *J. Biol. Chem. 261*, 16215−16218.
29. Redford, S. M., McRee, D. E., Getzoff, E. D., Steinman, H. M., and Tainer, J. A. (1990) Crystallographic characterization of a Cu, Zn superoxide dismutase from Photobacterium leiognathi. *J. Mol. Biol. 212*, 449−451.
30. Kitagawa, Y., Tanaka, N., Hata, Y., Kusunoki, M., Lee, G. P., Katsube, Y., Asada, K., Aibara, S., and Morita, Y. (1991) Three-dimensional structure of Cu, Zn-superoxide dismutase from spinach at 2.0 Å resolution. *J. Biochem. 109*, 477−485.
31. Djinovic, K., Gatti, G., Coda, A., Antolini, L., Pelosi, G., Desideri, A., Falconi, M., Marmocchi, F., Rotilio, G., and Bolognesi, M. (1992) Crystal structure of yeast Cu, Zn superoxide dismutase. Crystallographic refinement at 2.5 Å resolution. *J. Mol. Biol. 225*, 791−809.
32. Pesce, A., Capasso, C., Battistoni, A., Folcarelli, S., Rotilio, G., Desideri, A., and Bolognesi, M. (1997) Unique structural features of the monomeric Cu, Zn superoxide dismutase from *Escherichia coli*, revealed by X-ray crystallography. *J. Mol. Biol. 274*, 408−420.
33. Folcarelli, S., Battistoni, A., Falconi, M., O'Neill, P., Rotilio, G., and Desideri, A. (1998) Conserved enzyme−substrate electrostatic attraction in prokaryotic Cu, Zn superoxide dismutase. *Biochem. Biophys. Res. Comm. 244*, 908−911.
34. Desideri, A., Falconi, M., Polticelli, F., Bolognesi, M., Djinovic, K., and Rotilio, G. (1992) Evolutionary conservativeness of electric field in the Cu, Zn superoxide dismutase active site. *J. Mol. Biol. 223*, 337−342.
35. Polticelli, F., Battistoni, A., Bottaro, G. Carri, M. T., O'Neill, P., Desideri, A., and Rotilio, G. (1994) Mutation of Lys-120 and Lys-134 drastically reduces the catalytic rate of Cu, Zn superoxide dismutase. *FEBS Lett. 352*, 76−78.
36. Polticelli, F., Falconi, M., O'Neill, P., Petruzelli, R., Galtieri, A., Calabrese, L., Rotilio, G., and Desideri, A. (1994) Molecular modeling and electrostatic potential calculations on chemically modified Cu, Zn superoxide dismutases from *Bos taurus* and shark *Prionace glauca*: role of lys 134 in electrostatically steering the substrate to the active site. *Arch. Biochem. Biophys. 312*, 22−30.
37. Bourne, Y., Redford, S. M., Steinman, H. M., Lepock, J. R., Tainer, J. A., and Getzoff, E. D. (1996) Novel dimeric interface and electrostatic recognition in bacterial Cu, Zn superoxide dismutase. *Proc. Natl. Acad. Sci. U.S.A. 93*, 12777−12779.
38. Bordo, D., Matak, D., Djinovic-Carugo, K., Rosano, C., Pesce, A., Bolognesi, M., Stroppolo, M. E., Falconi, M., Battistoni, A., and Desideri, A. (1999) Evolutionary constraints for dimer formation in prokaryotic Cu, Zn superoxide dismutase. *J. Mol. Biol. 285*, 283−296.
39. Polticelli, F., Battistoni, A., O'Neill, P., Rotilio, G., and Desideri, A. (1996) Identification of the residues responsible for the akaline inhibition of Cu, Zn superoxide dismutase: a site-directed mutagenesis approach. *Protein Sci. 5*, 248−253.
40. Sharp, K., Fine, R., and Honig, B. (1987) Computer simulations of the diffusion of a substrate to an active site of an enzyme. *Science 236*, 1460−1463.
41. Sines, J. J., McCammon, J. A., and Allison, S. A. (1992) Kinetic effects of multiple charge modifications in enzyme−substrate reactions: Brownian dynamics simulations of Cu, Zn superoxide dismutases. *J. Comput. Chem. 13*, 66−69.
42. Sines, J. J., Allison, S. A., and McCammon, J. A. (1990) Point charge distributions and electrostatic steering in enzyme/substrate encounter: Brownian dynamics of modified copper/zinc superoxide dismutases. *Biochemistry 29*, 9403−9412.

43. Allison, S. A., Ganti, G., and McCammon, J. A. (1985) Simulation of the diffusion-controlled reaction between superoxide and superoxide dismutase. I. Simple models. *Biopolymers 24*, 1323−1336.
44. Getzoff, E. D., Tainer, J. A., Weiner, P. K., Kollman, P. A., Richardson, J. S., and Richardson, D. C. (1983) Electrostatic recognition between superoxide and copper, zinc superoxide dismutase. *Nature 306*, 287−290.
45. Banci, L., Cabelli, D. E., Getzoff, E. D., Hallewell, R. A., and Viezzoli, M. S. (1993) An essential role for the conserved Glu-133 in the anion interaction with superoxide dismutase. *J. Inorg. Biochem. 50*, 89−100.
46. Babbitt, P. C., Mrachko, G. T., Hasson, M. S., Huisman, G. W., Kolter, R., Ringe, D., Petsko, G. A., Kenyon, G. L., and Gerlt, J. A. (1995) A functionally diverse enzyme superfamily that abstracts the a protons of carboxylic acids. *Science 267*, 1159−1161.
47. Raychaudhuri, S., Younas, F., Karplus, P. A., Faerman, C. H., and Ripoll, D. R. (1997) Backbone makes significant contribution to the electrostatics of α/β-barrel proteins. *Protein Sci. 6*, 1849−1857.
48. Kirby, A. J. (1987) Mechanism and stereoselective effects in the lyozyme reaction. *CRC Crit. Rev. Biochem. 22*, 283−315.
49. Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. Structure of myoglobin. *Nature 185*, 422−427.
50. Ratnayake, C. K., and Regnier, F. E. (1996) Lateral interaction between electrostatically adsorbed and covalently immobilized proteins on the surface of cation-exchange sorbents. *J. Chromatogr. A 30*, 25−32.

BI026918F