

# A Quantitative Measure of Electrostatic Perturbation in Holo and Apo Enzymes Induced by Structural Changes

Sandeep Chakraborty\*

Department of Biological Sciences, Tata Institute of Fundamental Research, Mumbai, India

## Abstract

Biological pathways are subject to subtle manipulations that achieve a wide range of functional variation in differing physiological niches. In many instances, changes in the structure of an enzyme on ligand binding germinate electrostatic perturbations that form the basis of its changed catalytic or transcriptional efficiency. Computational methods that seek to gain insights into the electrostatic changes in enzymes require expertise to setup and computing prowess. In the current work, we present a fast, easy and reliable methodology to compute electrostatic perturbations induced by ligand binding (MEPP). The theoretical foundation of MEPP is the conserved electrostatic potential difference (EPD) in cognate pairs of active site residues in proteins with the same functionality. Previously, this invariance has been used to unravel promiscuous serine protease and metallo- $\beta$ -lactamase scaffolds in alkaline phosphatases. Given that a similarity in EPD is significant, we expect differences in the EPD to be significant too. MEPP identifies residues or domains that undergo significant electrostatic perturbations, and also enumerates residue pairs that undergo significant polarity change. The gain in a certain polarity of a residue with respect to neighboring residues, or the reversal of polarity between two residues might indicate a change in the preferred ligand. The methodology of MEPP has been demonstrated on several enzymes that employ varying mechanisms to perform their roles. For example, we have attributed the change in polarity in residue pairs to be responsible for the loss of metal ion binding in fructose 1,6-bisphosphatases, and corroborated the pre-organized state of the active site of the enzyme with respect to functionally relevant changes in electric fields in ketosteroid isomerases.

**Citation:** Chakraborty S (2013) A Quantitative Measure of Electrostatic Perturbation in Holo and Apo Enzymes Induced by Structural Changes. PLoS ONE 8(3): e59352. doi:10.1371/journal.pone.0059352

**Editor:** Freddie Salsbury Jr, Wake Forest University, United States of America

**Received:** October 12, 2012; **Accepted:** February 13, 2013; **Published:** March 14, 2013

**Copyright:** © 2013 Chakraborty. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was funded by the Tata Institute of Fundamental Research (Department of Atomic Energy) and the Department of Science and Technology (JC Bose Award Grant). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The author has declared that no competing interests exist.

\* E-mail: sanchak@gmail.com

## Introduction

The modulation of the catalytic rates of enzymes is a critical aspect for the regulation of flux in biological pathways [1]. Since the structure of an enzyme is a key determinant of its functionality, this modulation is often achieved through the allosteric binding of various cofactors that restructure the active site [2]. An alternate strategy makes use of various effector molecules that bind to activator or repressor proteins, and influence their transcriptional efficiency through subtle structural changes [3,4]. In many cases, the structural changes arising from mutations lead to depletion of metal ions, misfolding, aggregation and pathological phenotypes [5]. There are several such examples of sophisticated manipulations devised by evolution to effect a wide range of functional variation in differing physiological niches [5,6].

We have established the non-triviality of the spatial and electrostatic congruence in cognate pairs of residues in the active site in several enzymes with the same catalytic function (CLASP) [7]. CLASP was used to unravel a serine protease scaffold in alkaline phosphatases [7], and a scaffold recognizing a  $\beta$ -lactam (imipenem) in a cold-active *Vibrio* alkaline phosphatase [8,9]. Since the correlation in the electrostatic potential difference (EPD) between residues is established, the comparison of EPD in two different structures of the same protein should indicate residues/domains of the protein that undergo significant deviation in electrostatic properties. These perturbed domains are likely to be

the sites responsible for the different response of the protein to native and non-native substrates. CLASP analysis of the EPD in the catalytic residues in Class A  $\beta$ -lactamases identified the dichotomy in the proton abstraction mechanism [10], which led to a method for enumerating the possible pathways for proton abstraction in the active site [11].

Enzymes accelerate reactions by many orders of magnitude compared to the uncatalysed reaction in water [12,13]. Although it has been long known that enzymes achieve these phenomenal rate accelerations by binding the transition state better than the substrate and thereby lower the activation energy, the exact roles of various factors (steric effects, conformational fluctuations, dynamical effects, electrostatic pre-organization, entropy changes, reactant destabilization, quantum tunneling, etc.) in accomplishing this is highly debated [14,15]. Computational studies based on quantum, classical, and statistical techniques have evolved in the last few decades for studying the effect of structural changes on the chemical competency of the active site in enzymes. Typically these methods can access the free energy landscape for a chemical reaction through quantum mechanical (QM) calculations on model substrates [16], through molecular mechanical (MM) empirical functions [17–20] or through a hybrid quantum mechanical molecular mechanical (QM/MM) approach [21]. These computational approaches can be combined with accelerated conformational sampling methods (free energy perturbation

[22], umbrella sampling [23], etc.) to finally obtain the free energy profile for a chemical reaction of interest. Ultrahigh-resolution diffraction studies [24–27] and photoelectron spectroscopy [28] are complementary experimental methods usually applied to gain insights in the critical residues involved in catalysis in enzymes. Active site residues have  $pK_a$  values that differ considerably from their intrinsic values [29,30], and these deviations, computed using macroscopic [31], microscopic [32] or empirical models [33], have been used for predicting active site residues [34–36]. The  $pK_a$  values obtained from various methods have been recently benchmarked using experimental values, concluding that the most striking result of this blind test was that nobody performed significantly better than the rest' [37]. A "highly parallelized molecular dynamics code on a high-performance capability computer" has revealed the initial stages of misfolding caused by metal depletion in Cu-Zn superoxide dismutase [38]. Another work was guided by computationally derived  $pK_a$  shifts to search for, and eventually detect, a tyrosinate in the active site of a bacterial ketosteroid isomerase [39]. These methods require considerable expertise, in terms of both understanding the computational algorithm and the understanding of the biological system, for setting up the simulations and can be computationally intensive. A fast, simple and accurate method to probe the electrostatic perturbations resulting from structural changes would be quite useful for such studies.

In the current work, we present a methodology for measuring electrostatic perturbation in proteins by contrasting the holoenzyme to the apoenzyme - **M**easuring **e**lectrostatic **p**erturbations in **p**roteins (MEPP).

MEPP takes as input two protein structures (mutations are accepted, but not insertions or deletions) and extracts the electrostatic potentials using APBS [40]. The perturbation of each residue with respect to its neighboring residues (within a parameterized distance) in the two structures is computed. In the default mode all amino acids are included, whereas in a stricter mode we consider only polar residues. Residue pairs that undergo polarity reversal are enumerated and Pymol scripts are generated for the visualization of the perturbed residues/domains.

The methodology of MEPP has been demonstrated on several enzymes that employ varying mechanisms to perform their roles. We have attributed the change in polarity in key residue pairs to be responsible for the loss of metal ion binding in some enzymes [2]. Further, some regulatory enzymes modulate their functionality by inducing structural changes in a domain (which undergoes little change in the electrostatic profile) using larger electrostatic perturbations in another domain on binding of cofactors [3,4]. Some enzymes act through a motion induced domino effect in adjacent subunits, and have minimal electrostatic perturbations since structural changes suffice to produce the desired changes [6,41]. We also corroborated the pre-organized state of the active site of the enzyme with respect to functionally relevant changes in electric fields in ketosteroid isomerase, a prototype enzyme for gaining insights into the electrostatics of catalytic reactions [42,43]. Finally, certain pathological phenotypes resulting from mutations and metal-depletion are shown to undergo electrostatic perturbations in the critical residues [5,38]. Thus, MEPP provides a fast, easy and reliable methodology to compute electrostatic perturbations induced by ligand binding, and has invariably identified relevant residues in all the enzymes we have studied.

## Results and Discussion

We demonstrate the application of the MEPP methodology on different enzymes. These enzymes use varying mechanisms to

perform their respective roles. The choice of the radial distance that encompasses interacting residues needs to be made judiciously. A small radius will not include enough residues, while a large one will include irrelevant ones. The small runtime of MEPP allows fast iterations to help decide a reasonable value.

We have chosen a radial distance of 6Å to compute interacting residues in the examples shown hereafter, and have demonstrated the invariance of results based on the radius within reasonable limits.

### 1 Diphtheria toxin repressor

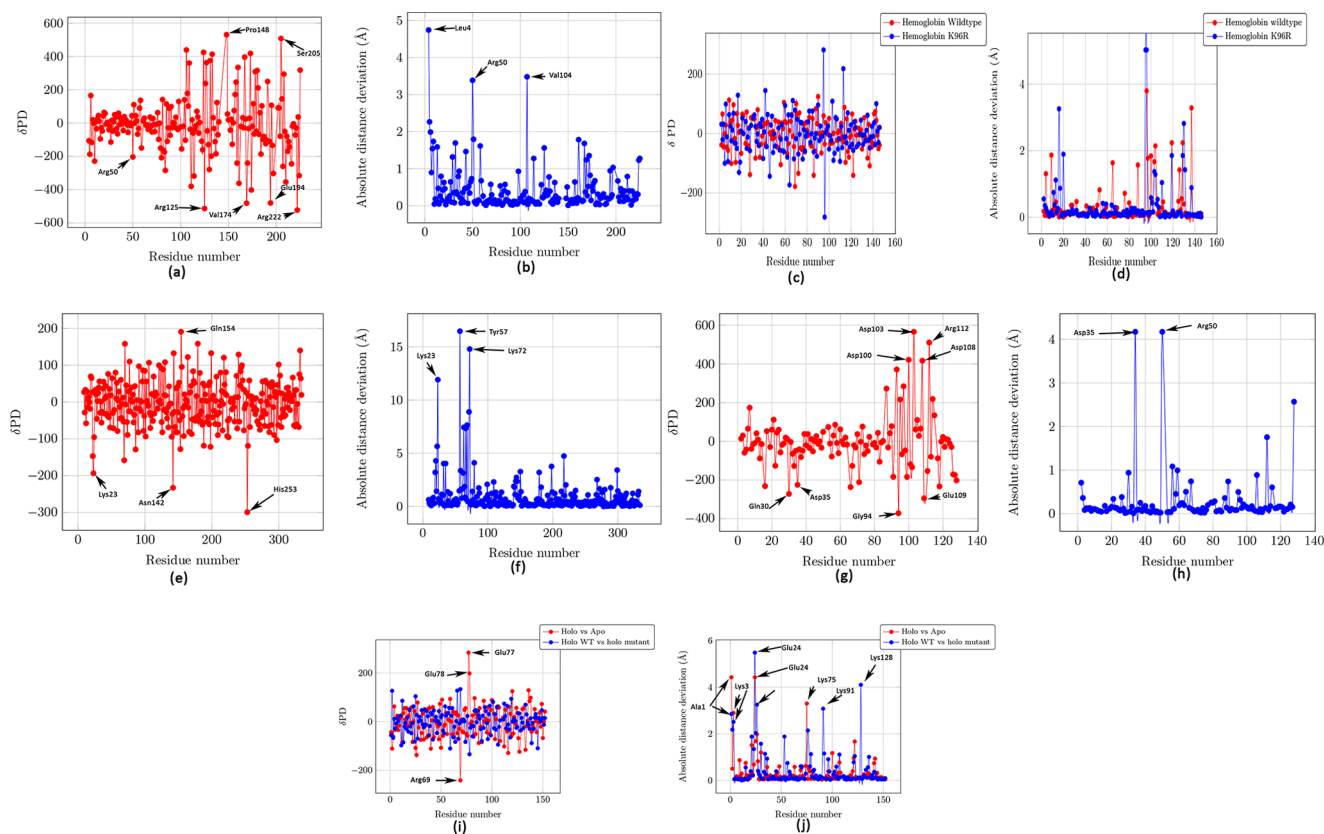
The diphtheria toxin repressor (DtxR) is a metal ion-activated repressor found in Gram-positive prokaryotes [4,44]. A 12- amino acid flexible tether links the N-terminal domain (residues 1–136) to the C-terminal. A mutation of either one of the residues in the primary metal binding site (Met10, Cys102, Glu105, or His106) to alanine results in complete inactivation of repressor activity [45]. We compared the holoenzyme (PDBid:1BI3) (which ligands a zinc ion) to the apoenzyme (PDBid:1BI2) of the DtxR from *Corynebacterium diphtheriae* [2]. We observed that the N-terminal undergoes minimal electrostatic perturbation (Fig. 1a), but larger conformational changes than the C-terminal (Fig. 1b), and the tether residues are where the residues having either low or large electrostatic perturbations are demarcated. Further, the significant polarity reversal in His106 and Glu105 is an explanation for their critical roles determined through mutational studies (Table 1) [45]. Other residue pairs with major polarity reversals are ones that bind the ancillary metal ion site (Glu83 and His79), and one of the residues that "contributes two ligands, Glu170 and Gln173, to the ancillary metal ion-binding site" (Gln173 and Asn169) [3]. Lastly, the Arg50 present in the HTH region has been shown to be essential for nucleic acid binding [44], and is seen to have both structural and electrostatic perturbation (Fig. 1a and b).

Fig. 2 shows the residues that have significant electrostatic perturbation, and clearly demonstrates the different responses of the N and C-terminals of the protein to the binding of the ligand. The Pymol script to visualize such perturbations is emitted during MEPP processing.

The N-terminal domain of DtxR that "undergoes a dynamic structural organization leading to homodimerization and target DNA binding" [45], but minor electrostatic changes can be explained by the fact that most repressors binds DNA in the apoenzyme loosely and intermittently. Hence, it is logical to assume that the electrostatics should be favorable to the configuration required for DNA binding. Electrostatic stimulation in the C-terminal transfers structural change to the N-terminal via a tether, enhancing the binding affinity. The invariance of electrostatic perturbation based on the radial distance from the active site used to choose interacting residues is seen in Fig. S1.

### 2 A Co-operative Dimeric Hemoglobin

Hemoglobins are proteins that transport oxygen in all vertebrates and in some invertebrates [6,41]. The ligand affinity of each subunit is influenced by the conformation changes in the neighboring subunits on ligation [6]. The origins of the cooperativity in hemoglobin through quantitative calculations of the free energy have been previously done by force field calculations [46] and by mathematically formulating the thermodynamic concept of order of free energy couplings [47,48]. These computations are much more rigorous than those used by MEPP, and consequently allow a more detailed analysis of the cooperativity in these enzymes. We compared the holoenzyme of the homodimeric hemoglobin HbI from *Scapharca inaequivalvis* which has bound CO and heme (PDBid: 3SDH) to the apoenzyme



**Figure 1. Electrostatic and spatial perturbations in different enzymes on ligand binding.** Perturbations for each residue is computed with respect to residues within a radial distance of 6Å. (a) and (b) Diphtheria toxin repressor (holoenzyme:1B13/apoenzyme:1B12), (c) and (d) Hemoglobin WT (holoenzyme:3SDH/apoenzyme:4SDH) Hemoglobin K96R (holoenzyme:3UHI/apoenzyme:3UHK), (e) and (f) Fructose 1,6-bisphosphatase (holoenzyme:1EYK/apoenzyme:1CNQ), (g) and (h) Ketosteroid isomerase (holoenzyme:3IPT/apoenzyme:1OPY), (i) and (j) Cu-Zn superoxide dismutase WT (holoenzyme:1HL5/apoenzyme:1HL4), and mutant A4V (holoenzyme:1UXM/apoenzyme:1HL5)  
doi:10.1371/journal.pone.0059352.g001

which is CO free (PDBid: 4SDH) [41]. Table 1 shows that Ser70 and His69 have a major polarity reversal, which corroborates “the strategic location of the distal histidine” next to the subunit junction [6].

However, we do not see significant electrostatic perturbation around Lys96 and Phe97 (Fig. 1c), although there is a reasonable spatial rearrangement for Lys96 (Fig. 1d). These residues “cover the heme binding pocket in the ligated state and swing open in the unligated state” [6]. The lack of electrostatic perturbation fits well with the proposed mechanism of cooperative ligand induced motion where there the dimer allows “slight relative rotation but preventing sliding” [6]. Lys96 is involved in the intersubunit coupling. Fig. 3 shows the proximity of His69 to the Lys96 from the neighboring subunit, reaffirming the MEPP identification of His69 as a critical residue for the mechanism of catalysis in hemoglobin. Interestingly, when we compared the holoenzyme (bound CO and heme, PDBid: 3UHI) to the apoenzyme (CO free, PDBid: 3UHK) for the K96R mutant, we observe a distinct electrostatic perturbation for Arg96, a meaningful polarity reversal between Glu95 and Arg96, and a conformational change for Arg96 as well (Fig. 1d).

### 3 Fructose 1,6-bisphosphatase

Fructose 1,6-bisphosphatase is a critical regulatory enzyme in gluconeogenesis that requires divalent cations for activity, and is inhibited allosterically by adenosine monophosphate (AMP) [49].

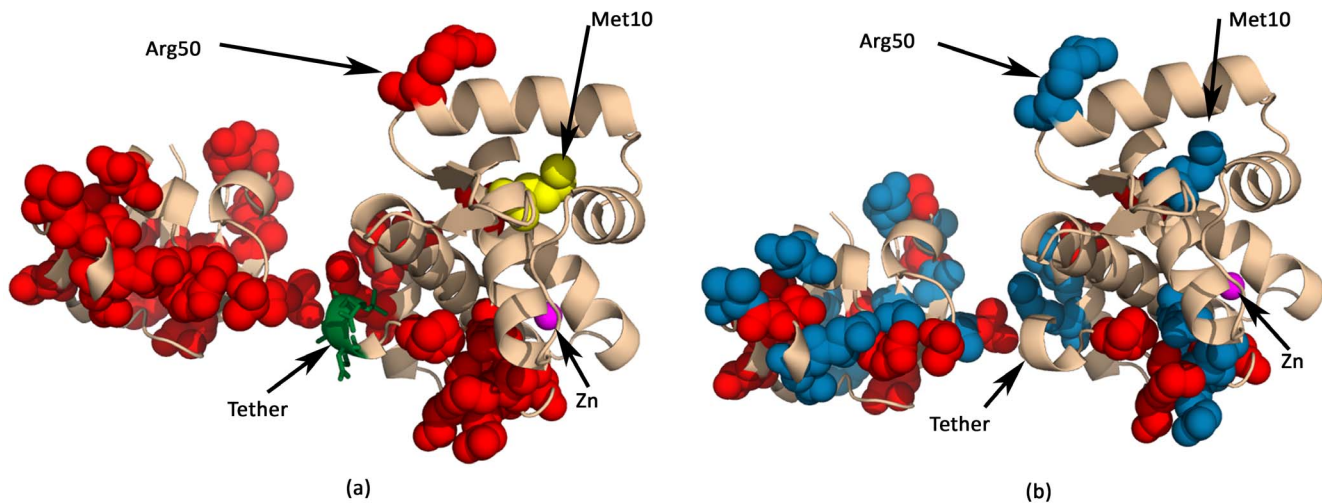
We compared the holoenzyme (PDBid:1EYK) (which ligands AMP) to the apoenzyme (PDBid:1CNQ) [2]. In the apoenzyme, zinc ions binds to all three metal sites and magnesium ions to two of three metal sites, while in the holoenzyme the divalent cations bind to a single site [2]. The residues liganding metal in the apoenzymes are M1 = (Glu97, Asp118, Leu120), M2 = (Asp118, Asp121, Glu280) and M3 = (Asp68, Glu97), while in the holoenzyme the single metal is liganded by residues (M2 = Asp118, Asp121, Gly28) (metal site numbering is non-standard). Table 1 shows the pairs of residues that have a significant reversal in their potential differences in the two structures, which includes the pair (Glu97-Glu98). Thus, it is seen that the reversal of the potential in Glu97, which ligands two metal ions, has resulted in its rejection of the metal ions in metal binding sites M1 and M3.

Another residue pair that is observed to have undergone polarity reversal is Gln154 and Asn142. Once again, Asn142 is known to be an important residues as the “major AMP site interacts with Asn142 from an adjacent tetramer in the crystal, while the minor site does not” [49]. Fig. 1e also shows that Asn142 gains significant negative potential in the holoenzyme when compared to neighboring residues in the apoenzyme. However, we did not find any reference in existing literature to the critical nature of His253 which is indicated to have both a polarity reversal and accrual of potential (Fig. 1f and Table 1). Fig. 1f shows that Tyr57 and Ly72 undergo a significant spatial rearrangement, a fact that corroborates previous studies where fluorescence from a tryptophan Tyr57 mutation showed an altered

**Table 1.** Residue pairs that undergo significant polarity reversal: PD1 and PD2 are the electrostatic potential difference between residues 'a' and 'b' in the holoenzyme and apoenzyme, respectively.

Apo PDB/Holo PDB	a	b	PD1	PD2	$\delta$ PD
Diphtheria toxin repressor 1BI3/1BI2	GLU83	HIS79	8.6	-150.7	159.2
	ASP88	ARG84	70.6	-117.2	187.8
	HIS106	GLU105	435.8	-18.2	454
	ASP122	ARG84	19.7	-182.3	202
	GLN173	ASN169	571.5	-126.3	697.8
Hemoglobin WT 3SDH/4SDH	GLU225	ARG114	284.5	-33.6	318.1
	SER70	HIS69	77.8	-143.2	221
Hemoglobin K96R 3UHI/3UHK	ASP89	ASN86	73.9	-79.1	153
	GLU95	ARG96	47	-233.9	280.9
Fructose 1,6-bisphosphatase 1EYK/1CNQ	THR103	ARG104	85.3	-75.9	161.1
	GLU97	GLU98	95.6	-66.5	162.1
	GLN154	ASN142	44.5	-188.4	232.9
	TYR164	HIS253	217.3	-82.1	299.4
	ASN179	ARG198	19.9	-165.8	185.7
Ketosteroid isomerase 3IPT/1OPY	TYR240	TYR226	61.3	-106.6	168
	HIS110	GLU109	59.4	-94.2	153.6
Cu-Zn superoxide dismutase 1HLS/1HL4	GLU77	ARG69	15.3	-267.8	283.1
	HIS120	HIS48	27.7	-138.4	166.1

$\delta$ PD = (PD1 - PD2). Electrostatic potential in dimensionless units of  $kT/e$  where  $k$  is Boltzmann's constant,  $T$  is the temperature in K and  $e$  is the charge of an electron. Threshold value for PD is 150 PD units.  
doi:10.1371/journal.pone.0059352.t001



**Figure 2. Visualizing the residues/domains that undergo significant electrostatic perturbation in the diphtheria toxin repressor (PDBid:1BI3).** Threshold value above which perturbation are considered significant is 200 potential (difference) units. It can be seen that the C-terminal undergoes more electrostatic changes than the N-terminal. (a) Residues undergoing electrostatic perturbation of magnitude greater than the threshold value are colored in red and yellow (Met10). Met10 ligands metal ions, and its mutation to alanine results in complete inactivation of repressor activity. (b) Residues undergoing positive and negative electrostatic perturbations greater than the threshold value are colored in red and blue respectively. As expected, they are interspersed in order to maintain electrostatic homogeneity.  
doi:10.1371/journal.pone.0059352.g002

conformational state for the loop 52–72 which plays a critical role in the mechanism of catalysis and allosteric inhibition of fructose 1,6-bisphosphatase [2].

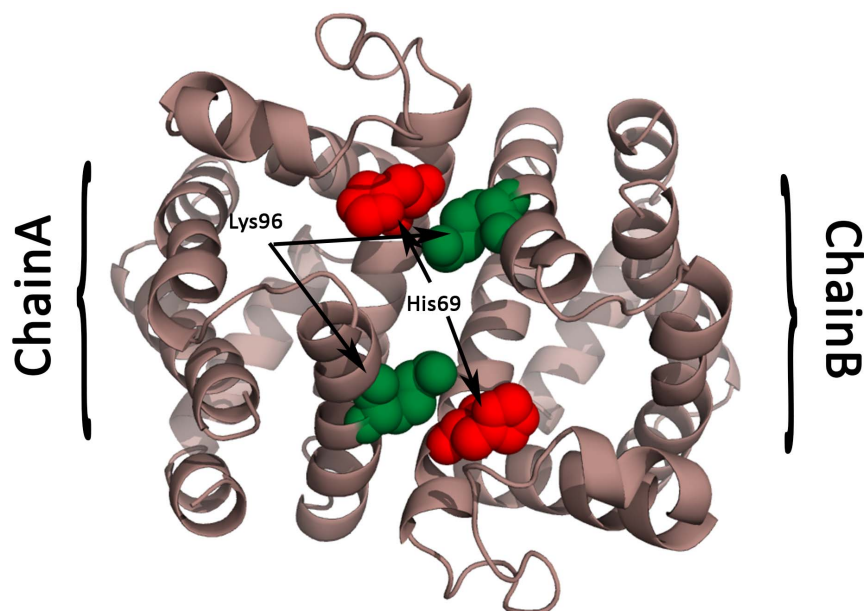
#### 4 Ketosteroid isomerase

Ketosteroid isomerase (KSI) catalyzes an allylic isomerization reaction at a diffusion-controlled rate, and is a prototype enzyme for gaining insights into the electrostatics of catalytic reactions [39,42,43,50]. The relative importance of the electrostatic preorganization in the active site to achieve the phenomenal acceleration is highly debated. While some groups, proponents of the electrostatic preorganization theory [46], emphasize that enzymes work by electrostatic preorganization' [42,51], others have claimed that KSI achieves its catalytic prowess through a combination of modest contributions from several mechanisms rather than from a single dominant contribution' [50]. We compared the holoenzyme (PDBid:3IPT,Y16S,D40N mutant) (which ligands equilenin, an intermediate analog) to the apoenzyme (PDBid:1OPY) from *Pseudomonas putida*. The mutation is intended to mimic the protonated D40 present in the intermediate complex. Fig. 1g shows that the active site undergoes minimal electrostatic perturbation. This corroborates the conclusion arrived by Jha et al. that the active site of the enzyme is pre-organized with respect to functionally relevant changes in electric fields since dipoles in the active site remained immobile with respect to the changing electric field [43].

#### 5 Cu-Zn superoxide dismutase

The mutant form of Cu-Zn superoxide dismutase (SOD1) is implicated in the hereditary form of the neurodegenerative disease Amyotrophic Lateral Sclerosis (ALS), commonly known as Lou Gehrig's disease, that effects neurons in the brain and spinal cord [52]. It has been shown that the metal-depleted SOD1 dimer gains toxic characteristics due to misfolding, unfolding, and aggregation [38]. We compared the holoenzyme (PDBid:1HL5, binds both Cu and Zn) to the apoenzyme (PDBid:1HL4, binds only Zn). The residues binding Zn are His63, His71, His80 and Asp83, while Cu is liganded by His48, His63, and His120. It is seen from Table 1





**Figure 3. Residues involved in intersubunit coupling in hemoglobin (PDBid:3SDH).** Lys96 (in green) in one subunit is in close proximity to His69 (in red) in the neighboring subunit. MEPP detects significant electrostatic perturbation in His69 and not in Lys96. On the other hand, Lys96 undergoes a distinct conformational change. doi:10.1371/journal.pone.0059352.g003

that His48 has a change in polarity with respect to His120. The spatial movement in this pair in the native structure has been previously noted [5]. Further, the other pair showing polarity reversal (Arg69-Glu77) was established as one of the 16 pairs that have long-range interactions with large interaction energies [52]. Also, Glu77 is a part of a non-native helix (Gly73-Glu77) that is adjacent to the zinc active site [53]. Fig. 4a shows that Glu77, which gains positive potential with respect to residues in the vicinity, is physically close to Glu77 from other subunits, and possibly generates a repulsive force that destabilizes the structure.

Next, we compared two holoenzymes (PDBid:1HL5 and 1UXM), where one of the holoenzymes (PDBid: 1UXM) has the mutation A4V [53]. This mutation has been implicated in generating a toxic phenotype. While Fig. 1i shows that there is little difference in the electrostatics in the protein, there are larger structural differences, especially at one of the electrostatic loop residues (Lys128) (Fig. 1j). In the apo enzyme, Asn86 is involved in H-bonds to residues Lys128-Gly130 of an adjacent SOD1 molecule [5,38]. A conformational change in Lys128 possibly disrupts these bonds between the closely placed Asn86 and Lys128 in neighboring subunits (Fig. 4b).

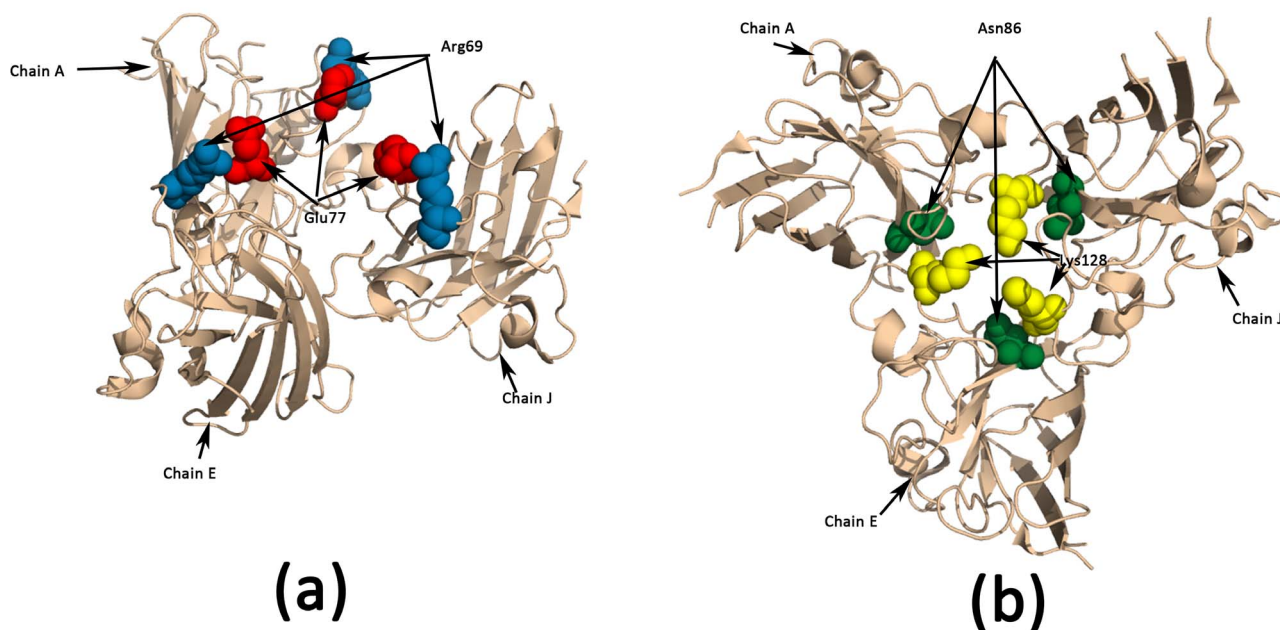
## 6 Limitations

The inability of simple continuum electrostatic models, such as those used MEPP, to always give a true representation of the electrostatic milieu of the active site is a well-known fact [54,55]. It has been argued that the electrostatic energies in proteins computed using such implicit models are sensitive to the value assigned to the dielectric constant [56]. However, MEPP computes the difference in electrostatic potentials rather than absolute values, and the (empirical) significance of the electrostatic potential difference in the active site has been established through several examples [7,9,11]. For example, although computational  $pK_a$  predictions for Tyr residues that form an extended hydrogen bond network in the oxyanion hole of the ketosteroid isomerase give a value more than 11, experimental results show the presence of

tyrosinate in the active site [39]. Computations assuming the ionization of these Tyr residues were able to better explain the observed frequency shifts than those obtained by presupposing the ionization of Asp103, substantiating the initial error in estimating the  $pK_a$  values of the Tyr residues [39].

In order to obtain the most accurate values for electrostatic energies, one has to solve the Schrödinger equation for the wavefunction of the interacting atoms [57]. Due to interactions between electrons, this equation does not have analytic solutions for atoms with more than one electron [58]. *Ab initio* methods use various approximations, for example the orbital approximation which assumes a linear combination of the wavefunctions of each electron, to obtain solutions for multi-electron atoms. One such technique, the Hartree-Fock self consistent field, starts with an approximate form of all orbitals and iteratively approaches convergence for the energies of the various orbitals [58]. However, these methods are computationally unfeasible even for moderately large systems. The QM/MM method has been proposed which uses quantum mechanical analysis for the reacting fragments, and classical analysis for the rest of the protein [21]. Such methods produce more authentic results, but require expertise and intensive computational resources. Such methods produce more authentic results, but require expertise and intensive computational resources, and are routinely used to gain insights into the significance of electrostatic effects in catalytic reactions [15,59]. It is to be noted that the current work is incapable of assigning the effects of the measured electrostatic perturbation.

The proper choice of the radial distance is another heuristic that needs to be varied keeping in the mind the differently shaped configurations of active sites in enzymes. The small runtimes of MEPP facilitates the iteration needed to derive such values. The variation of the perturbations in residues has been observed to be minimal (Fig. S1).



**Figure 4. Residues involved in intersubunit interactions in Cu-Zn superoxide dismutase (PDBid:1HL5).** (a) Glu77 (in red) and Arg69 (in blue) undergo significant perturbation. It can be seen that Glu77, which gains positive potential with respect to residues in the vicinity, is physically close to Glu77 from other subunits, and possibly generates a repulsive force that destabilizes the structure. (b) Asn86 (in green) and Lys128 (in yellow) from neighboring subunits are seen to be closely placed to each other, and are involved in H-bonds. Conformational changes in Lys128 possibly disrupt these bonds.

doi:10.1371/journal.pone.0059352.g004

## 7 Summary

The theoretical foundation of the current work lies in the conserved electrostatic potential difference (EPD) in cognate pairs of active site residues in proteins with the same functionality. This similarity is observed in structures solved independently over many years, and also holds true for convergently evolved proteins (for example, in the two major families of serine proteases, chymotrypsin and subtilisin [60]). This also speaks highly of the reliability of the APBS and PDB2PQR implementations [40,61].

Given that the EPD is significant, we expect differences in the EPD to be significant too. The EPD changes of one residue in two structures of the same protein with respect to residues in the vicinity should therefore indicate regions of the proteins that undergo significant electrostatic perturbation. The conformational changes accompanying the binding of a cofactor do not always imply a concomitant change in the electric field of significance. For instance, we found little electrostatic perturbation in hemoglobin, where the proposed model suggests that binding of the first ligand in one subunits induces a binding event in the other partner subunit [6].

Also, the electrostatic perturbation is not always uniform throughout the protein. The magnitude and location of these changes are primarily dictated by the mechanism through which the enzyme performs its role. For example, the N-terminal of the diphtheria toxin repressor (DtxR) undergoes minimal electrostatic perturbation in spite of large conformational changes. The situation is exactly reversed for the C-terminal. It is only logical that a regulatory enzyme, like the DtxR, should not undergo a dramatic change in its electrostatic profile, since such enzymes are still capable of binding substrate, albeit loosely, and intermittently perform their roles [3,4]. This low affinity is due to structural considerations, and the enzyme using the electrostatics of an adjacent domain effects the required structural modifications. A definite reversal of polarity is desired in cases where the

functionality changes require the gain/loss of a metal cofactor. Such is the case for fructose 1,6-bisphosphatase, where the binding of adenosine monophosphate results in the reversal of the polarity of residues that ligand two metal ions, resulting in the holoenzyme failing to bind these ions [2]. Thus, MEPP provides a fast, easy and reliable methodology to compute electrostatic perturbations induced by ligand binding, and has invariably identified relevant residues in all the enzyme we have probed.

## Methods

MEPP takes as an input two protein structures - the holo ( $P^h$ ) and the apo enzyme ( $P^a$ ). The number of residues in  $P^h$  and  $P^a$  are expected to be the same, although a (user defined) number of residue types are allowed to be different. In other words, mutations are accepted, but not insertions or deletions. APBS and PDB2PQR packages are used to calculate the potential of the atoms in the protein [40,60]. Let us denote the potential of the reactive atom of residue  $R_i$  in protein  $P^x$  as  $Pot(P^x, R_i)$ . The electrostatic potential difference (EPD) between two residues is given by Equation 1. For each residue  $Res_i$  in protein  $P^h$ , we compute the set of residues ( $\Phi(P^{h,neigh}_i)$ ) that are within a parameterized radius  $P_{rad}$  (Equation 2).

Next, we compute the summation of the EPDs of residue  $R_i$  with respect to residues in  $\Phi(P^{h,neigh}_i)$  for both  $P^h$  and  $P^a$  (Equations 3 and 4). Note that for residues in the apoenzyme, we consider residues that are within radius  $P_{rad}$  in the holoenzyme ( $\Phi(P^{h,neigh}_i)$ ), although structural changes might have changed this set if it were to be computed on the apoenzyme. This ensures that the number of residues being compared is the same. Finally, we define the perturbation of residue  $R_i$  ( $E_i^{perturb}$ ) as the difference in the summation of the EPDs of residue  $R_i$  with respect to residues in  $\Phi(P^{h,neigh}_i)$  (Equation 5).

MEPP also identifies the set of residue pairs that undergo a shift in polarity ( $\Phi(x, y)^{\text{PolarityReverse}}$ ). This is computed by choosing residue pairs that have a change in sign in their EPD (magnitude of the product  $PD(P^h, Res_i, Res_j) * PD(P^a, Res_i, Res_j)$  should be less than zero), and the absolute magnitude of their

difference should be larger than a user defined threshold value ( $P_{\text{thresh}}$ ) (Equation 6).

An optional feature restricts  $E^{\text{perturb}}$  computations to polar residues only (Equation 7). MEPP outputs a Pymol script for the visualization of the perturbed residues/domains which highlights the residues that have undergone perturbations more than a user defined threshold. The colored regions can be further differentiated based on the polarity of the perturbation. An analysis similar to the one done for pairwise EPD is done for pairwise distance deviations. More sophisticated methods exist for quantifying spatial changes [62].

The MEPP package is written in Perl on Ubuntu. Hardware requirements are modest - all results here are from a simple workstation (2GB ram) and runtimes were a few minutes at the most. The APBS parameters were set as described previously in [7]. APBS writes out the electrostatic potential in dimensionless units of  $kT/e$  where  $k$  is Boltzmann's constant,  $T$  is the temperature in K and  $e$  is the charge of an electron. We extensively integrated and used the freely available BioPerl [63] modules and Emboss [64] tools. All protein structures were rendered by PyMol (<http://www.pymol.org/>). The source code is made available at [www.sanchak.com/mepp](http://www.sanchak.com/mepp).

$$PD(P^x, res_i, res_j) = Pot(P^x, res_i) - Pot(P^x, res_j) \quad (1)$$

$$\phi(P^h)_i^{\text{neigh}} = \{ \forall Res_x \in P^h : (x \neq i) \wedge (\text{distance}(Res_i, Res_x) \leq P_{rad}) \} \quad (2)$$

$$PD_i^h = \sum_{\forall Res_x \in \phi(P^h)_i^{\text{neigh}}} (PD(P^h, res_i, res_x)) \quad (3)$$

$$PD_i^a = \sum_{\forall Res_x \in \phi(P^h)_i^{\text{neigh}}} (PD(P^a, Res_i, Res_x)) \quad (4)$$

$$E_i^{\text{perturb}} = PD_i^h - PD_i^a \quad (5)$$

$$\phi(x, y)^{\text{PolarityReverse}} = \{ \forall Res_x, Res_y \in P^h : x \neq y \wedge (PD(P^h, Res_i, Res_j) * PD(P^a, Res_i, Res_j) \leq 0) \} \quad (6)$$

$$\wedge (\text{abs}(PD(P^h, Res_i, Res_j) - PD(P^a, Res_i, Res_j)) > P_{\text{thresh}}) \}$$

$$\phi(P^h)_i^{\text{neigh}} = \{ \forall Res_x \in P^h : (x \neq i) \wedge (\text{distance}(Res_i, Res_x) \leq R) \wedge (\text{Polar}(Res_x)) = \text{true} \} \quad (7)$$

## Supporting Information

**Figure S1 Invariance of electrostatic perturbation based on radial distance from the active site used to choose interacting residues.** In the diphtheria toxin repressor from *Corynebacterium diphtheriae*, the C-terminal undergoes more electrostatic perturbation compared to the N-terminal, and this change is independent of the radial distance which defines interacting residues (PDF)

## Acknowledgments

I gratefully acknowledge B. J. Rao, Department of Biological Sciences, Tata Institute of Fundamental Research, for technical discussions and infrastructural support. I am grateful to Bjarni Asgeirsson from the Science Institute, Department of Biochemistry, University of Iceland for correcting the revised manuscript. I also thank Ravindra Venkatramani, Department of Chemical Sciences, Tata Institute of Fundamental Research for stimulating discussions on this subject.

## Author Contributions

Conceived and designed the experiments: SC. Performed the experiments: SC. Analyzed the data: SC. Contributed reagents/materials/analysis tools: SC. Wrote the paper: SC.

## References

- Suarez RK, Staples JF, Lighton JR, West TG (1997) Relationships between enzymatic flux capacities and metabolic flux rates: nonequilibrium reactions in muscle glycolysis. *Proc Natl Acad Sci USA* 94: 7065–7069.
- Choe JY, Fromm HJ, Honzatko RB (2000) Crystal structures of fructose 1,6-bisphosphatase: mechanism of catalysis and allosteric inhibition revealed in product complexes. *Biochemistry* 39: 8565–8574.
- Love JF, vanderSpek JC, Marin V, Guerrero L, Logan TM, et al. (2004) Genetic and biophysical studies of diphtheria toxin repressor (DtxR) and the hyperactive mutant DtxR(E175K) support a multistep model of activation. *Proc Natl Acad Sci USA* 101: 2506–2511.
- Pohl E, Holmes RK, Hol WG (1998) Motion of the DNA-binding domain with respect to the core of the diphtheria toxin repressor (DtxR) revealed in the crystal structures of apo- and holo-DtxR. *J Biol Chem* 273: 22420–22427.
- Strange RW, Antonyuk S, Hough MA, Doucette PA, Rodriguez JA, et al. (2003) The structure of holo and metal-deficient wild-type human Cu, Zn superoxide dismutase and its relevance to familial amyotrophic lateral sclerosis. *J Mol Biol* 328: 877–891.
- Ren Z, Srajer V, Knapp JE, Royer WE (2012) Cooperative macromolecular device revealed by meta-analysis of static and time-resolved structures. *Proc Natl Acad Sci USA* 109: 107–112.
- Chakraborty S, Minda R, Salaye L, Bhattacharjee SK, Rao BJ (2011) Active site detection by spatial conformity and electrostatic analysis-unravelling a proteolytic function in shrimp alkaline phosphatase. *PLoS ONE* 6: e28470.
- Helland R, Larsen RL, Asgeirsson B (2009) The 1.4 Å crystal structure of the large and cold-active *Vibrio sp.* alkaline phosphatase. *Biochim Biophys Acta* 1794: 297–308.
- Chakraborty S, Asgeirsson B, Minda R, Salaye L, Frere JM, et al. (2012) Inhibition of a cold-active alkaline phosphatase by imipenem revealed by *in silico* modeling of metallo-β-lactamase active sites. *FEBS Lett* 586: 3710–3715.
- Dambon C, Raquet X, Lian LY, Lamotte-Brasseur J, Fonze E, et al. (1996) The catalytic mechanism of β-lactamases: NMR titration of an active-site lysine residue of the TEM-1 enzyme. *Proc Natl Acad Sci USA* 93: 1747–1752.
- Chakraborty S (2012) Enumerating pathways of proton abstraction based on a spatial and electrostatic analysis of residues in the catalytic site. *PLoS ONE* 7: e39577.
- Wolfenden R, Snider MJ (2001) The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* 34: 938–945.
- Kraut J (1988) How do enzymes work? *Science* 242: 533–540.
- Kamerlin SC, Warshel A (2010) At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* 78: 1339–1375.

15. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, et al. (2006) Electrostatic basis for enzyme catalysis. *Chem Rev* 106: 3210–3235.
16. Chen P, Solomon EI (2004) Oxygen activation by the noncoupled binuclear copper site in peptidyl-glycine alpha-hydroxylating monooxygenase. Reaction mechanism and role of the noncoupled nature of the active site. *J Am Chem Soc* 126: 4991–5000.
17. MacKerell AD, Banavali N, Foloppe N (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* 56: 257–265.
18. van Gunsteren WF, Berendsen JC (1987) Groningen Molecular Simulation (GROMOS) Library Manual. Biomos, Nijenborgh 16, Groningen, NL.
19. Warshel A, Weiss RM (1981) Empirical valence bond calculations of enzyme catalysis. *Ann N Y Acad Sci* 367: 370–382.
20. Warshel A (2003) Computer simulations of enzyme catalysis: methods, progress, and insights. *Annu Rev Biophys Biomol Struct* 32: 425–443.
21. Warshel A, Levitt M (1976) Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103: 227–249.
22. Zwanzig RW (1954) High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* 22: 1420–1426.
23. Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* 23: 187–199.
24. Minasov G, Wang X, Shoichet BK (2002) An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation. *J Am Chem Soc* 124: 5333–5340.
25. Nukaga M, Mayama K, Hujer AM, Bonomo RA, Knox JR (2003) Ultrahigh resolution structure of a class A  $\beta$ -lactamase: on the mechanism and specificity of the extended-spectrum SHV-2 enzyme. *J Mol Biol* 328: 289–301.
26. Tomanicek SJ, Wang KK, Weiss KL, Blakeley MP, Cooper J, et al. (2011) The active site protonation states of perdeuterated Toho-1 -lactamase determined by neutron diffraction support a role for Glu166 as the general base in acylation. *FEBS Lett* 585: 364–368.
27. Fuhrmann CN, Daugherty MD, Agard DA (2006) Subangstrom crystallography reveals that short ionic hydrogen bonds, and not a His-Asp low-barrier hydrogen bond, stabilize the transition state in serine protease catalysis. *J Am Chem Soc* 128: 9086–9102.
28. Vonderach M, Ehrler OT, Matheis K, Karpuschkin T, Papalazarou E, et al. (2011) Probing electro-static interactions and structural changes in highly charged protein polyanions by conformer-selective photoelectron spectroscopy. *Phys Chem Chem Phys* 13: 15554–15558.
29. Warshel A (1981) Electrostatic basis of structure-function correlation in proteins. *Acc. Chem. Res.* 14: 284–290.
30. Forsyth WR, Antosiewicz JM, Robertson AD (2002) Empirical relationships between protein structure and carboxyl pKa values in proteins. *Proteins* 48: 388–403.
31. Bashford D, Karplus M (1990) pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* 29: 10219–10225.
32. Warshel A (1981) Calculations of enzymatic reactions: calculations of pKa, proton transfer reactions, and general acid catalysis reactions in enzymes. *Biochemistry* 20: 3167–3177.
33. Bas DC, Rogers DM, Jensen JH (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* 73: 765–783.
34. Elcock AH (2001) Prediction of functionally important residues based solely on the computed energetics of protein structure. *J Mol Biol* 312: 885–896.
35. Bate P, Warwicker J (2004) Enzyme/non-enzyme discrimination and prediction of enzyme active site location using charge-based methods. *J Mol Biol* 340: 263–276.
36. Wei Y, Ko J, Murga LF, Ondrechen MJ (2007) Selective prediction of interaction sites in protein structures with THEMATICCS. *BMC Bioinformatics* 8: 119.
37. Alexov E, Mehler EL, Baker N, Baptista AM, Huang Y, et al. (2011) Progress in the prediction of pKa values in proteins. *Proteins* 79: 3260–3275.
38. Strange RW, Antonyuk S, Hough MA, Doucette PA, Rodriguez JA, et al. (2003) The structure of holo and metal-deficient wild-type human Cu, Zn superoxide dismutase and its relevance to familial amyotrophic lateral sclerosis. *J Mol Biol* 328: 877–891.
39. Fafarman AT, Sigala PA, Schwans JP, Fenn TD, Herschlag D, et al. (2012) Quantitative, directional measurement of electric field heterogeneity in the active site of ketosteroid isomerase. *Proc Natl Acad Sci USA* 109: 299–308.
40. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98: 10037–10041.
41. Royer WE (1994) High-resolution crystallographic analysis of a co-operative dimeric hemoglobin. *J Mol Biol* 235: 657–681.
42. Kamerlin SC, Sharma PK, Chu ZT, Warshel A (2010) Ketosteroid isomerase provides further support for the idea that enzymes work by electrostatic preorganization. *Proc Natl Acad Sci USA* 107: 4075–4080.
43. Jha SK, Ji M, Gaffney KJ, Boxer SG (2011) Direct measurement of the protein response to an electrostatic perturbation that mimics the catalytic cycle in ketosteroid isomerase. *Proc Natl Acad Sci USA* 108: 16612–16617.
44. White A, Ding X, vanderSpek JC, Murphy JR, Ringe D (1998) Structure of the metal-ion-activated diphtheria toxin repressor/tox operator complex. *Nature* 394: 502–506.
45. Love JF, VanderSpek JC, Murphy JR (2003) The src homology 3-like domain of the diphtheria toxin repressor (DtxR) modulates repressor activation through interaction with the ancillary metal ion-binding site. *J Bacteriol* 185: 2251–2258.
46. Warshel A (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J Biol Chem* 273: 27035–27038.
47. Johnson ML (1986) Analysis of the order of free energy couplings between ligand binding and subunit assembly in human hemoglobin. *Biochemistry* 25: 791–797.
48. Weber G (1987) Free energy couplings between ligand binding and subunit association in hemoglobin are of first order. *Biochemistry* 26: 331–332.
49. Ke H, Thorpe CM, Seaton BA, Marcus F, Lipscomb WN (1989) Molecular structure of fructose-1,6-bisphosphatase at 2.8-Å resolution. *Proc Natl Acad Sci USA* 86: 1475–1479.
50. Kraut DA, Sigala PA, Pybus B, Liu CW, Ringe D, et al. (2006) Testing electrostatic complementarity in enzyme catalysis: hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biol* 4: e99.
51. Warshel A, Sharma PK, Chu ZT, Aqvist J (2007) Electrostatic contributions to binding of transition state analogues can be very different from the corresponding contributions to catalysis: phenolates binding to the oxyanion hole of ketosteroid isomerase. *Biochemistry* 46: 1466–1476.
52. Khare SD, Ding F, Dokholyan NV (2003) Folding of Cu, Zn superoxide dismutase and familial amyotrophic lateral sclerosis. *J Mol Biol* 334: 515–525.
53. Schmidlin T, Kennedy BK, Daggett V (2009) Structural changes to monomeric CuZn superoxide dismutase caused by the familial amyotrophic lateral sclerosis-associated mutation A4V. *Biophys J* 97: 1709–1718.
54. Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. *Science* 268: 1144–1149.
55. Grochowski P, Trylska J (2008) Continuum molecular electrostatics, salt effects, and counterion binding - a review of the Poisson-Boltzmann theory and its modifications. *Biopolymers* 89: 93–113.
56. Schutz CN, Warshel A (2001) What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* 44: 400–417.
57. Eisberg R, Resnick R (1985) *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles*. Wiley, second edition.
58. Atkins P, De Paula J (2006) *Physical Chemistry*. Oxford University Press, 8th edition.
59. Gao J, Ma S, Major DT, Nam K, Pu J, et al. (2006) Mechanisms and free energies of enzymatic reactions. *Chem Rev* 106: 3188–3209.
60. Rawlings ND, Barrett AJ (1993) Evolutionary families of peptidases. *Biochem J* 290 ( Pt 1): 205–218.
61. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32: W665–667.
62. Tuzmen C, Erman B (2011) Identification of ligand binding sites of proteins using the Gaussian Network Model. *PLoS ONE* 6: e16474.
63. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, et al. (2002) The bioperl toolkit: Perl modules for the life sciences. *Genome research* 12: 1611–1618.
64. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16: 276–277.