Published in final edited form as:

Biotechnol J. 2009 April ; 4(4): 495–500. doi:10.1002/biot.200800299.

The Empirical Valence Bond as an Effective Strategy for Computer-Aided Enzyme Design

Alexandra Vardi-Kilshtain^{1,†}, Maite Roca^{1,2,†}, and Arie Warshel^{1,*}

¹ Department of Chemistry, University of Southern California, Los Angeles, California, 90089-1062. USA

² Departamento de Química Física y Analítica, Universidad Jaume I, 12071 Castellón, SPAIN

Abstract

The ability of the empirical valence bond (EVB) to be used in screening active site residues in enzyme design is explored in a preliminary way. This validation is done by comparing the ability of this approach to evaluate the catalytic contributions of various residues in chorismate mutase. It is demonstrated that the EVB model can serve as an accurate tool in the final stages of computer aided enzyme design (CAED). The ability of the model to predict quantitatively the catalytic power of enzymes should augment the capacity of current approaches for enzyme design.

Keywords

Enzyme design; Transition state stabilization; Preorganization effect

1. Introduction

Computer aided enzyme design (CAED) has become a subject of major attention and major activity in recent years (e.g [1–5]). Unfortunately the resulting constructs have been significantly less effective than the corresponding natural enzymes [1,6–8] and the reasons for this limited successes are not completely clear. One way to judge the effectiveness of different approaches for enzyme design is to use them in predicting the effect of different mutations on the catalytic rates of the corresponding enzymes.

Previous studies [9–18] have demonstrated that the empirical valence bond (EVB) method can give reliable prediction of enzyme activity and mutational effect. Although other quantum mechanics/molecular mechanics (QM/MM) approaches are also useful (e.g. [19,20]), the EVB is much less time demanding and thus allow for the expensive conformational search needed in enzyme design. Thus we focus here on the examination of the performance of the EVB method in reproducing observed mutational effects in different forms of chorismate mutase (CM). The results obtained in this preliminary study are quite promising, indicating that the EVB can be used in the final screening stage in computer aided enzyme design.

*To whom correspondence should be address. Arie Warshel, E-mail: warshel@usc.edu, Phone: +1 213 740 4114; Fax +1 213 740 2701.

†These authors contributed equally to this work

2. Methods

As stated above we explore here the performance of the EVB in computer aided enzyme design. The EVB has been described extensively in many works (e.g. [9,12,14]) and thus we provide here only key information about this method.

The EVB begins with the resonance states (or more precisely, diabatic states) corresponding to classical valence-bond structures. These basis states are mixed to describe the reacting system. The potential energies of the diabatic states (H_{11} and H_{22}) and the mixing term (H_{12}) are represented by the Hamiltonian matrix elements,

$$H_{ii} = \varepsilon_i = \alpha_{gas}^i + U_{intra}^i(R, Q) + U_{inter}^i(R, Q, r, q) + U_{solvent}^i(r, q) \quad (1)$$

$$H_{ij} = A \exp(-a|\Delta R'|)$$

Here $\mathbf{R}$ and $\mathbf{Q}$ represent the atomic coordinates and charges, respectively, of the reactants or products ("solute") in the diabatic states, and $\mathbf{r}$ and $\mathbf{q}$ are the coordinates and charges of the surrounding water or protein ("solvent"). α_{gas}^i is the energy of the i^{th} diabatic state in the gas-phase, where all the fragments are taken to be at infinity; $U_{intra}^i(R, Q)$ is the intramolecular potential of the solute system (relative to its minimum) in this state; $U_{inter}^i(R, Q, r, q)$ represents the interaction between the solute atoms and the surrounding solvent atoms; and $U_{solvent}^i(r, q)$ represents the potential energy of the solvent.

The adiabatic ground-state energy (E_g) and the corresponding eigenvector (C_g) are obtained by solving the secular equation

$$H_{EVB} C_g = E_g C_g \quad (2)$$

The simplicity of the EVB formulation makes it relatively straightforward to obtain analytical derivatives of the potential surface by using the Hellmann-Feynman theorem for the first derivatives of E_g , and thus to sample the EVB energy surface by molecular-dynamics (MD) simulations. This is done by a combined free energy perturbation (FEP) umbrella sampling (US) procedure that provides the free energy function ($\Delta G^\ddagger(x)$) that is needed to calculate the activation free energy ($\Delta G^\ddagger$). The details of the FEP/US mapping procedure used to evaluate the EVB free energy surface are described elsewhere [12], and here we review only essential points for the simple case of two diabatic states. In such a case we use a mapping potential of the form,

$$\varepsilon_m = (1 - \theta_m)\varepsilon_1 + \theta_m\varepsilon_2 \quad (3)$$

where θ_m changes from 0 to 1 in $n+1$ fixed increments ($\theta_m = 0/n, 1/n, 2/n, \dots, n/n$). The free energy ΔG_m associated with changing λ from 0 to m/n can be evaluated by a free-energy perturbation (FEP) procedure. The free energy functional that corresponds to the adiabatic ground state surface, E_g , is obtained by the FEP-umbrella sampling (FEP/US) method as

described elsewhere ([9,12,14]). The EVB has been used extensively by our group and has been implemented (sometimes under different names) by leading research groups (see [21]).

The EVB calculations were evaluated by using the MOLARIS simulation program [22] using the ENZY MIX force field. The EVB activation barriers were calculated at the configurations selected by using the same free energy perturbation umbrella sampling (FEP/US) approach used in all of our EVB studies. The simulation systems were solvated by the surface constrained all atom solvent (SCAAS) model [22] using a water sphere of 18 Å radius centered on the center of mass of the substrate, while long-range electrostatic effects were treated by the local reaction field (LRF) method [22]. The EVB region consisted of the entire substrate, a total of 24 atoms. The FEP mapping was evaluated by 21 frames of 20 ps each for moving along the reaction coordinate with our all atom surface constrained spherical model. All the simulations were done at 300 K with a time step of 1 fs. In order to obtain reliable results we repeated the simulations 5 times with different initial conditions (obtained from arbitrary points in the relaxation trajectory). Along the EVB simulation there was no positional constraint on the protein atoms and the EVB parameters used are the same to those used in the previous EVB study on CM [23]. The charges of the EVB states (reactant state and product state) were obtained by using the Gaussian03 package [24]. This was done by optimizing the structures of the reactant and product in the gas phase using B3LYP [25,26] with the 6-31+g(d) basis set [27,28] and solvating the optimized structures in the PCM solvent model [29,30].

To exploit the power of the EVB method in quantitative studies of enzymatic reactions, it is essential to fit the EVB free energy profile of the reference reaction in water, adjusting the parameters to reproduce the observed barrier for the reaction in solution. This was done in our previous studies [23,31]. The parameters optimized in the above procedure were used in the evaluation of the activation free energies in the different CM systems without any further adjustment (just using the EVB parameters of the reaction in aqueous solution).

3. Results

In this work we chose as a benchmark the ability to reproduce the catalytic activity of different mutants of different forms of enzyme chorismate mutase (CM) that catalyzes rearrangement of chorismate to prephenate (e.g see Gajewski JJ, Jurayj J, Kimbrough DR, Gande ME, Ganem B & Carpenter BK (1987) *J. Am. Chem. Soc.* 109: 1170–1186). The studied systems are the trimer *Bacillus subtilis* chorismate mutase (BsCM), [32] the homodimeric chorismate mutase from *Escherichia coli* (EcCM) [33] and the monomeric chorismate mutase mMjCM obtained by Hilvert and coworkers [34] by topological redesign of the thermostable EcCM homologue from *M. jannaschii* (MjCM). The coordinates for BsCM, EcCM (X-ray) and mMjCM (NMR) structures were obtained from the Protein Data Bank, Brookhaven National Laboratory, with PDB access codes 1COM, 1ECM and 2GTU, respectively.

In this work we considered; a) the native EcCM and the V35I, V35A mutants that have been used recently by Mayo and coworkers [35], b) the monomer mMjCM that has been designed by Hilvert and coworkers [36,37], and c) the native BsCM and the R90G-BsCM and R90Cit-BsCM mutations of refs [38,39]. In all of these cases the observed values of k_{cat}/K_m and k_{cat} are known. The position of the mutated residues is shown in Fig 1. Here we explore the performance of the EVB screening approach.

Our main strategy for calculations of activation energies, is the EVB method used by our group and others in quantitative studies of enzyme catalysis of many enzymes (e.g. [9]). The EVB has been described in great details elsewhere [12,13] and was used recently in a study of CM [23,31].

This approach can evaluate $\Delta g_{\text{cat}}^{\ddagger}$ (that corresponds to k_{cat}) rather than the $\Delta g_{\text{p}}^{\ddagger}$ considered in the previous section. Here we explored the catalytic power of the trimer (BsCM), dimer (EcCM) and monomer (mMjCM) by calculating the EVB surfaces for different mutants and native proteins considered in Table 1. In each of the systems considered we started by generating the given sequence from the native protein proceeding to 100 ps MD simulations in order to relax the given structures. Five structures were saved during each relaxation process and then used to generate the EVB surface and obtain the activation free energies and reorganization energies.

The calculated activation barriers of the different mutants are compared to the corresponding observed values in Table 1 and Fig. 2. As seen from the table and the figure the agreement between the calculated and observed results is excellent and can be considered to be a quantitative agreement. It should be noted in this respect that the performance of our calculations may not be fully appreciated by those who note the small deviations between the calculated and observed values in Fig. 2 while overlooking the fact that we actually reproduced quantitatively the absolute values of the activation free energies without any parameterization on the reaction in the enzyme. Obtaining such a quantitative prediction is very encouraging.

Thus the EVB approach can be used in the final screening stage of CAED approaches and perhaps can be defined as the “gold standard” of our approaches.

In order to assess the effectiveness of a given CAED approach it is important to have a clear idea on the computational resources needed to obtain the given result. Here we report the computer time needed for each of the approaches considered above, or in other words the price per performance ratio. The relevant estimates are summarized in Table 2. As seen from the table after appropriate relaxation it is possible to screen 9 mutations on 100 nodes each time considering that each node has two processors and then we could screen around 60 mutations overnight. While this may sound an investment of a major resource it might provide at present the most practical way of getting accurate screening.

The present study focused on the EVB approach and left for a subsequent studies the exploration of the performance of faster but less quantitative approaches, including the LRA, LIE and the PDL/D/S-LRA methods [40] that have been found to provide qualitative but physically reasonable trend (e.g. [41]), and even ground-state simulations may be useful in the screening process [42]. It is also important to point out that the case of CM is somewhat simpler than cases where there are several transition states with similar energy [43]. However, after identifying the highest energy TS one can use a small number of EVB calculations (usually three) to identify the free energy of the rate limiting TS. Obviously if we find that direct calculations of the TS binding free energy are sufficiently accurate we can use just a single step of such calculations focusing on the highest TS.

4. Concluding remarks

Despite recent impressive advances in enzyme design (e.g. [7,8]), the enzymes generated by current design efforts are still far less efficient than naturally evolved enzymes. Overall, it seems to us that a part of the problem in previous CAED has been the limited focus on modeling of the actual chemical step in the actual enzyme active site. Probably, the main problem in designing enzymes with native activity is related to the ability to predict the proper transition state (TS) stabilization. This difficulty is due in a large part to the difficulty of capturing the preorganization effect. That is, as shown in a large series of studies (see ref. [9] for a review) the major part of the catalytic effect (in any enzyme that was examined quantitatively) is due to the fact that the dipoles of the active site polar groups are already partially oriented toward the TS charges and thus do not have to pay for this polarization process. On the other hand, in the uncatalyzed reaction in water about half of the energy of interaction between the water dipoles and the TS is paid in the reorganization of the water molecules. Thus, the calculations

of the TS activation free energy are very different than the calculations of the interaction between the charged TS and the active site. This is probably demonstrated in the best way in the recent study of the catalytic effect of ketosteroid isomerase [44]. Obviously, attempts to evaluate the catalytic effect by using gas phase models or by looking at the electrostatic interaction between different residues and the TS are unlikely to reproduce the correct catalytic effect since it is impossible to assess the preorganization effect without including the protein and its reorganization in the simulations.

One clear but non trivial way to progress in enzyme design is to calculate the activation free energies in the proposed active sites, and this work explored the feasibility of such a strategy by using the EVB approach. Thus our work is not so much about effective early screening but mainly about the requirements of the final stage in the screening process. In this respect we believe that the EVB provides a very effective way for performing the final stage in the screening process. We are also developing an examining more qualitative approaches for early screening and these approaches will be reported in subsequent studies.

Acknowledgements

This work was supported by NIH grant GM24492 and the University of Southern California's High Performance Computing and Communication Center (HPCC) for computer time. We would like to thank Professor Don Hilvert for insightful discussion. M.R. thanks the Generalitat Valenciana from Spain for the postdoctoral fellowship.

Abbreviations

EVB	empirical valence bond
CAED	computer aided enzyme design
FEP/US	free energy perturbation/umbrella sampling
CM	chorismate mutase

References

1. Toscano MD, Woycechowsky KJ, Hilvert D. Minimalist active-site redesign: Teaching old enzymes new tricks. *Angew Chem Int Ed* 2007;46:4468–4470.
2. Seeling B, Szostak JW. Selection and evolution of enzyme from a partially randomized non-catalytic scaffold. *Nature* 2007;448:828–831. [PubMed: 17700701]
3. Varadarajan N, Gam J, Olsen MJ, Georgiou G, Iverson B. *Proc Natl Acad Sci USA* 2005;102:6855–6860. [PubMed: 15867160]
4. Bolon DN, Mayo SL. Enzyme-like proteins by computational design. *Proc Natl Acad Sci USA* 2001;98:14274–14279. [PubMed: 11724958]
5. Kaplan J, DeGrado WF. De novo design of catalytic proteins. *Proc Natl Acad Sci USA* 2004;101:11566–11570. [PubMed: 15292507]
6. Lippow SM, Tidor B. Progress in computational protein design. *Curr Opin Biotechnol* 2007;18:305–311. [PubMed: 17644370]
7. Rothlisberger D, et al. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* 2008;453:190–195. [PubMed: 18354394]
8. Jiang L, et al. De Novo Computational Design of Retro-Aldol Enzymes. *Science* 2008;319:1387–1391. [PubMed: 18323453]

9. Warshel A, Sharma PK, Kato M, Xiang Y, Liu HB, Olsson MHM. Electrostatic basis for enzyme catalysis. *Chem Rev* 2006;106:3210–3235. [PubMed: 16895325]
10. Warshel A, Sussman F. Toward computer-aided site-directed mutagenesis of enzymes. *Proc Natl Acad Sci USA* 1986;83:3806. [PubMed: 3520555]
11. Rao SN, Singh UC, Bash PA, Kollman PA. Free Energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin. *Nature* 1987;328:551–554. [PubMed: 3302725]
12. Warshel, A. Wiley Interscience; New York: 1991.
13. Aqvist J, Warshel A. Simulation of Enzyme-Reactions Using Valence-Bond Force-Fields and Other Hybrid Quantum-Classical Approaches. *Chem Rev* 1993;93:2523–2544.
14. Liu H, Warshel A. The Catalytic Effect of Dihydrofolate Reductase and Its Mutants Is Determined by Reorganization Energies. *Biochemistry* 2007;46:6011–6025. [PubMed: 17469852]
15. Hwang JK, Warshel A. Semiquantitative calculations of catalytic free energies in genetically modified enzymes. *Biochemistry* 1987;26:2669–2673. [PubMed: 3300766]
16. Hansson T, Nordlund P, Aqvist J. Energetics of Nucleophile Activation in a Protein Tyrosine Phosphatase. *J Mol Biol* 1997;265:118–127. [PubMed: 9020976]
17. Bjelic S, Aqvist J. Catalysis and Linear Free Energy Relationships in Aspartic Proteases. *Biochemistry* 2006;45:7709–7723. [PubMed: 16784222]
18. Mones L, Kulhanek P, Florian J, Simon I, Fuxreiter M. Probing the two-metal ion mechanism in the restriction endonuclease BamHI. *Biochemistry* 2007;46:14514–14523. [PubMed: 18020376]
19. Mulholland AJ. Computational enzymology: modelling the mechanisms of biological catalysts. *Biochem Soc Trans* 2008;36:22–26. [PubMed: 18208378]
20. Marti S, Andres J, Moliner V, Silla E, Tunon I, Bertran J. Predicting an Improvement of Secondary Catalytic Activity of Promiscuous Isochorismate Pyruvate Lyase by Computational Design. *J Am Chem Soc* 2008;130:2894–2895. [PubMed: 18281989]
21. Warshel, A.; Florian, J. *The Encyclopedia of Computational Chemistry*. Schleyer, PVR; Jorgensen, WL.; Schaefer, HFI.; Schreiner, PR.; Thiel Wand Glen, R., editors. John Wiley & Sons; Chichester, UK: 2004.
22. Lee FS, Chu ZT, Warshel A. Microscopic and Semimicroscopic Calculations of Electrostatic Energies in Proteins by the Polaris and Enzymix Programs. *J Comput Chem* 1993;14:161–185.
23. Strajbl M, Shurki A, Kato M, Warshel A. Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J Am Chem Soc* 2003;125:10228–10237. [PubMed: 12926945]
24. Frisch, MJ., et al. Gaussian 03, revision C 03. Inc, G., editor. Wallingford, CT: 2004.
25. Becke AD. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J Chem Phys* 1993;98:1372–1377.
26. Lee CT, Yang WT, Parr RG. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys Rev B* 1988;37:785–789.
27. Hehre WJ, Ditchfie R, Pople JA. Self-Consistent Molecular-Orbital Methods. 12. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular-Orbital Studies of Organic-Molecules. *J Chem Phys* 1972;56:2257.
28. Clark T, Chandrasekhar J, Spitznagel GW, Schleyer PV. Efficient Diffuse Function-Augmented Basis-Sets for Anion Calculations. The 3–21+G Basis Set for 1st-Row Elements, Li-F. *J Comput Chem* 1983;4:294–301.
29. Cossi M, Barone V, Cammi R, Tomasi J. Ab initio study of solvated molecules: A new implementation of the polarizable continuum model. *Chem Phys Lett* 1996;255:327–335.
30. Tomasi J, Cammi R, Mennucci B. Medium Effects on the properties of chemical systems: An overview of recent formulations in the polarizable continuum model (PCM). *Int J Quantum Chem* 1999;75:783–803.
31. Roca M, Messer B, Hilvert D, Warshel A. On the relationship between folding and chemical landscapes in enzyme catalysis. *Proc Natl Acad Sci USA* 2008;105:13877–13882. [PubMed: 18779576]
32. Chook YM, Gray JV, Ke H, Lipscomb WN. The monofunctional chorismate mutase from *Bacillus subtilis*. Structure determination of chorismate mutase and its complexes with a transition state analog

- and prephenate, and implications for the mechanism of the enzymatic reaction. *J Mol Biol* 1994;240:476–500. [PubMed: 8046752]
33. Lee AY, Karplus PA, Ganem B, Clardy J. Atomic-Structure of the Buried Catalytic Pocket of *Escherichia-Coli* Chorismate Mutase. *J Am Chem Soc* 1995;117:3627–3628.
 34. Vamvaca K, Vögeli B, Kast P, Pervushin K, Hilvert D. An enzymatic molten globule: Efficient coupling of folding and catalysis. *Proc Natl Acad Sci USA* 2004;101:12860–12864. [PubMed: 15322276]
 35. Lassila JK, Keefe JR, Kast P, Mayo SL. Exhaustive Mutagenesis of Six Secondary Active-Site Residues in *Escherichia coli* Chorismate Mutase Shows the Importance of Hydrophobic Side Chains and a Helix N-Capping Position for Stability and Catalysis. *Biochemistry* 2007;46:6883–6891. [PubMed: 17506527]
 36. Vamvaca K, Vogeli B, Kast P, Pervushin K, Hilvert D. An enzymatic molten globule: Efficient coupling of folding and catalysis. *Proceedings of the National Academy of Sciences of the United States of America* 2004;101:12860–12864. [PubMed: 15322276]
 37. Pervushin K, Vamvaca K, Vogeli B, Hilvert D. Structure and dynamics of a molten globular enzyme. *Nat Struct Mol Biol* 2007;14:1202–1206. [PubMed: 17994104]
 38. Kast P, Grisostomi C, Chen IA, Li S, Kregel U, Xue Y, Hilvert D. A Strategically Positioned Cation Is Crucial for Efficient Catalysis by Chorismate Mutase. *J Biol Chem* 2000;275:36832–36838. [PubMed: 10960481]
 39. Kienhofer A, Kast P, Hilvert D. Selective Stabilization of the Chorismate Mutase Transition State by a Positively Charged Hydrogen Bond Donor. *J Am Chem Soc* 2003;125:3206–3207. [PubMed: 12630863]
 40. Warshel A, Sharma PK, Kato M, Parson WW. Modeling electrostatic effects in proteins. *Biochim Biophys Acta, Proteins Proteomics* 2006;1764:1647–1676.
 41. Xiang Y, Oelschlaeger P, Florian J, Goodman MF, Warshel A. Simulating the effect of DNA polymerase mutations on transition-state energetics and fidelity: Evaluating amino acid group contribution and allosteric coupling for ionized residues in human pol beta. *Biochemistry* 2006;45:7036–7048. [PubMed: 16752894]
 42. Martínek V, Bren U, Goodman MF, Warshel A, Florian J. DNA polymerase β catalytic efficiency mirrors the Asn279-dCTP H-bonding strength. *FEBS Letters* 2007;581:775–780. [PubMed: 17286973]
 43. Florian J, Goodman MF, Warshel A. Computer simulations of protein functions: Searching for the molecular origin of the replication fidelity of DNA polymerases. *Proc Natl Acad Sci USA* 2005;102:6819–6824. [PubMed: 15863620]
 44. Warshel A, Sharma PK, Chu ZT, Aqvist J. 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* 2007;46:1466–1476. [PubMed: 17279612]

Figure 1.

Scheme of the Chorismate Mutase active site. The figure shows the structure of the transition state and its interactions with the neighboring residues, (a) for the dimer and monomer and (b) for the trimer.

Figure 2.
Correlation between the calculated and observed activation free energies.

Table 1Calculated and observed $\Delta g_{\text{cat}}^{\ddagger}$ ($\Delta \Delta g_{\text{cat}}^{\ddagger}$).^a

	$\Delta g_{\text{cat}}^{\ddagger}$ ($\Delta \Delta g_{\text{cat}}^{\ddagger}$) _{calc}	$\Delta g_{\text{cat}}^{\ddagger}$ ($\Delta \Delta g_{\text{cat}}^{\ddagger}$) _{obs}
EcCM	15.3 ± 1.5 (0.0)	15.3(0.0)
V35I-EcCM	13.3 ± 0.6 (−2.0)	15.0(−0.3)
V35A-EcCM	15.2 ± 0.9 (−0.1)	15.7(0.4)
mMjCM	16.2 ± 1.7 (0.9)	16.8(1.5)
BsCM	16.6 ± 1.6 (1.3)	15.3(0.0)
R90Cit-BsCM	23.7 ± 2.5 (8.4)	21.1(5.8)
R90G-BsCM	23.8 ± 2.2 (8.5)	22.5(7.2)

^aEnergies in kcal/mol.

Table 2

The performance times of the $\Delta g_{\text{cat}}^{\ddagger}$ calculation.^a

	Number of nodes	Number of processors	Computation time
MD Relaxation	1	1	3 hours
$\Delta g_{\text{cat}}^{\ddagger}$ using EVB	11	22	1,5 hours

^aThe calculations are done on the USC HPCC (High Performance Computing and Communication) Linux computer, using the Dual Intel P4 3.0 GHz 2GB Memory nodes. The time involves an average over the three forms of CM and also for calculating $\Delta g_{\text{cat}}^{\ddagger}$ is an average over five runs.