

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Computational studies of antibody and enzyme catalysis

Permalink

<https://escholarship.org/uc/item/0ch5c8tr>

Author

Chong, Lillian T.

Publication Date

2002

Peer reviewed|Thesis/dissertation

Computational Studies of Antibody and Enzyme Catalysis

by

Lillian T. Chong

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

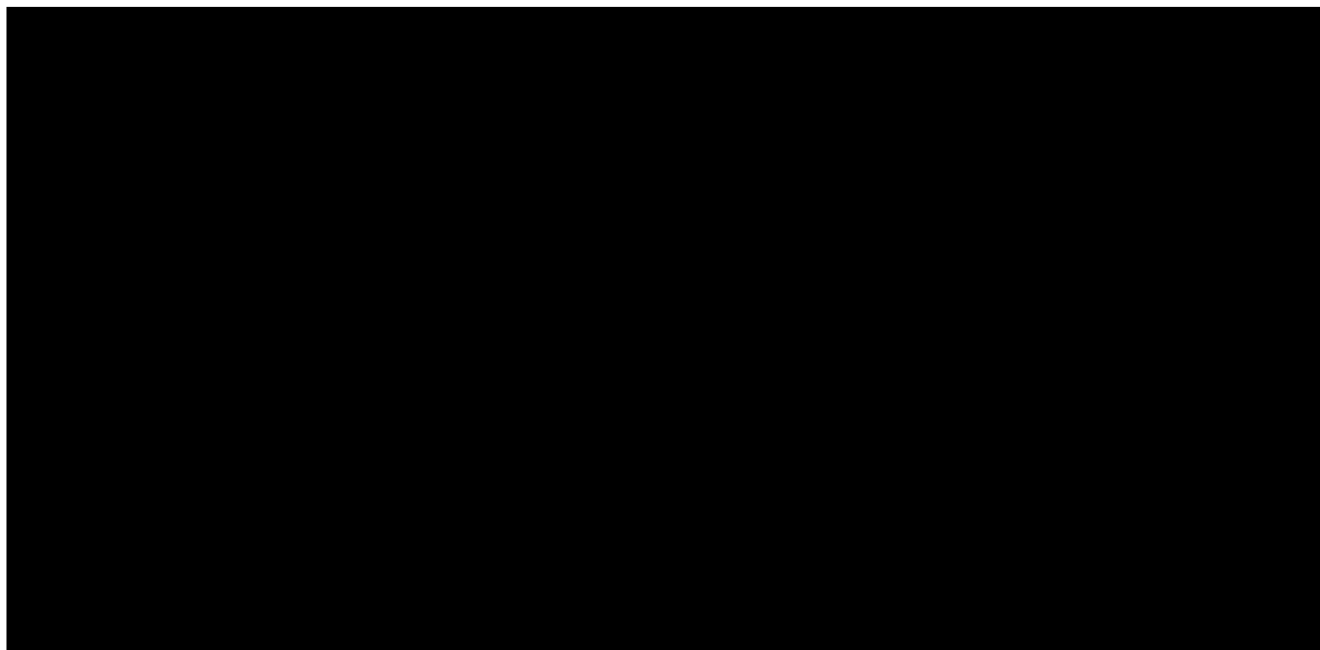
Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO



Copyright 2002

by

Lillian T. Chong

Dedication

This work is dedicated to the late Peter Kollman for his guidance and contagious enthusiasm that continue to inspire me in my life.

Acknowledgements

My most sincere thanks to Tack Kuntz for being a wonderful mentor to me throughout my graduate years, particularly after Peter's death last spring. I also thank my dear friends and family for their love and support.

Two chapters contained in this dissertation have been previously published.

Chapter 2 was originally published in *Proc. Natl. Acad. Sci. USA* in 1999.

Chapter 4 was published in *J. Am. Chem. Soc.* in 2000.

Abstract:

This dissertation examines the structural and energetic aspects of antibody and enzyme catalysis through the application of molecular dynamics simulations, quantum mechanical calculations, and free energy analyses. Chapter 1 contains a brief background on antibody and enzyme catalysis as well as an introduction to the computational methods used in the upcoming chapters. Chapter 2 investigates the optimization of hapten binding affinity during affinity maturation of 48G7. In Chapter 3, the hapten-programmed mechanism of catalysis is explored for the highly efficient esterase antibody, 17E8, and the unique amidase antibody, 43C9, which hydrolyzes esters as well. In Chapter 4, the work is extended to analysis of a proposed mechanism of catalysis for orotidine 5'-phosphate decarboxylase, the most proficient enzyme known. The concluding chapter contains a perspective on the importance of the insight gained from this doctoral work in guiding the pursuit for the design of improved and novel catalysts.



Table of Contents:

Chapter 1.

Introduction.....	1
Thesis overview.....	2
Computational approaches.....	6
References.....	10

Chapter 2. Molecular Dynamics and Free Energy Calculations Applied to Affinity

Maturation in Antibody 48G7.....	13
Abstract.....	14
Introduction.....	15
Methods.....	16
Results and Discussion.....	20
Approximations and Future Directions.....	32
Conclusions.....	35
Acknowledgements.....	37
References.....	38
Appendix.....	41

Chapter 3. Direct Hydroxide Attack: A Mechanism of Antibody Catalysis for

Esterase 17E8 as well as the Unique Amidase 43C9.....	43
Abstract.....	44

Introduction.....	45
Methods.....	49
Results.....	54
Discussion and Conclusions.....	74
Acknowledgements.....	77
References.....	78
Appendix.....	82

Chapter 4. An Alternative Explanation for the Catalytic Proficiency of Orotidine

5'-Phosphate Decarboxylase.....	84
Abstract.....	85
Introduction.....	86
Methods.....	90
Results.....	99
Discussion and Conclusions.....	116
Acknowledgements.....	128
References.....	129
Appendix.....	134

Chapter 5. Conclusion..... 136

References.....	141
-----------------	-----

List of Tables:

Chapter 2:

Table 1. Energetic analysis of 48G7 antibody-hapten complex formations.....	23
Table 2. Energetic analysis of mutant 48G7 antibody-hapten complex formations	26
Table 3. Electrostatic interactions between hapten and antibody.....	28
Table 4. Electrostatic analysis of Tyr H33 in 48G7 antibody-hapten complex formations.....	30
Table A. Partial atomic charges of the hapten.....	41
Table B. Additional force field parameters used for the hapten.....	42

Chapter 3:

Table 1. Relative energies for hydroxide attack on model ester and amide substrates for 43C9.....	57
Table 2. Relative energies for hydroxide attack on model ester and amide substrate for 17E8.....	60
Table 3. Average energetic contributions to hapten complex formations of 43C9 and 17E8.....	68
Table 4. Computed average binding free energies for TS complexes from <i>re</i> and <i>si</i> attack by hydroxide ion.....	69
Table 5. Average energetic contributions to GS and TS complex formations of 43C9.....	70

Table 6. Average energetic contributions to GS and TS complex formations of 17E8.....	71
Table A. Additional force field parameters used for ligands.....	82

Chapter 4:

Table 1. Average MM free energy differences for the C5-protonated intermediate relative to the ground state.....	103
Table 2. Statistical data on distances between N ζ Lys72 and C5 of the ligand	111
Table 3. Average energetic contributions to enzyme complex formations with the 6-aza-UMP inhibitor.....	112
Table 4. Average energetic contributions to GS and TS complex formations...	113
Table 5. Average energetic contributions to enzyme complex formations with 2-thio-OMP and 4-thioOMP.....	115
Table A. Atom types and partial charges of ligands.....	134
Table B. Additional force field parameters used for ligands.....	135

List of Figures:

Chapter 2:

- Figure 1. RMS fluctuations of the germ line and mature antibody-hapten complexes.....22
- Figure 2. Orientations of Tyr H33 and hapten in the 48G7 germ line and mature antibody-hapten complexes.....32

Chapter 3:

- Figure 1. Base-catalyzed hydrolysis of an aryl ester or anilide.....45
- Figure 2. Optimized geometries of reactants and transition states for hydroxide attack on model ester and amide substrates for 43C9.....56
- Figure 3. Optimized geometries of reactants and transition states for hydroxide attack on model ester and amide substrates for 17E8.....60
- Figure 4. Schematic depiction of hydrogen bonds formed with the phosphonate moiety of the bound haptens in 43C9 and 17E8.....63
- Figure 5. Computed activation barriers for hydrolysis *via* direct hydroxide attack.....73

Chapter 4:

Figure 1. Ligands used in this study.....	90
Figure 2. Final model system used in the quantum calculations.....	91
Figure 3. Energy profiles for C5- and C6-protonations of orotate by an ammonium ion with a formic acid hydrogen bonded to the orotate carboxyl group.....	99
Figure 4. Energy profiles for C5-protonation of orotate by ammonium ion with geometries taken from Figure 3.....	100
Figure 5. Energy profile for the C5-protonation of orotate by a methyammonium ion.....	101
Figure 6. Partial charges for the QM atoms in the ground and C5-protonated intermediate states used in the QM-FE calculations.....	104
Figure 7. Calculated gas-phase proton affinity for different sites of orotate and deprotonated uracil.....	105
Figure 8. Calculated gas-phase proton affinity for different derivations of orotate.....	105
Figure 9. Energy profiles for the decarboxylation of orotate and of the C5-protonated intermediate.....	106
Figure 10. Qualitative energy profile illustrating the plausibility of the proposed mechanism of catalysis by ODCase.....	108
Figures 11. Stereoviews of final structures from MD trajectories.....	110
Figure 12. Reaction mechanisms considered in this study.....	126

Thesis Overview:

Natural enzymes are remarkable catalysts. The fastest enzymes, which include superoxide dismutase and triosephosphate isomerase, are limited by the rate at which they encounter substrate while those that have not achieved this diffusion-controlled limit ($k_{\text{cat}}/K_M \approx 10^9 \text{ M}^{-1} \text{ s}^{-1}$) can have apparent bimolecular rate constants between 10^6 and $10^8 \text{ M}^{-1} \text{ s}^{-1}$ irrespective of the rate of the corresponding uncatalyzed reaction [1]. The subsequent rate accelerations over the uncatalyzed reaction are very high, in the range of 10^6 to 10^{12} , and even reaching 10^{17} in the case of orotidine 5'-phosphate decarboxylase [1].

How do enzymes achieve these large rate accelerations? Factors that have been found to contribute to enzymatic catalysis include nucleophilic attack on the substrate (*i.e.* by serine hydroxyl or cysteine thiol), enhancement of the nucleophilicity of attacking groups by a general base (*i.e.* imidazole of histidine), stabilization of a negative charge with a positively charged metal ion (*i.e.* Zn^{2+} or Mg^{2+}), and preorganization of the reacting groups upon binding such that less entropy is lost during the course of catalysis. Another important element, which is not usually found in simple reactions in solution, is the role of binding energy between the enzyme and its substrate.

In the late 1940s, Pauling suggested that enzymatic catalysis is dominated by the ability of the enzyme to bind transition states more tightly than ground states [2]. About twenty years later, Jencks suggested that stable molecules resembling the transition state of a reaction might be used as haptens to elicit antibodies with tailored catalytic activities and selectivities that cannot be obtained by standard chemical methods or do not exist among natural enzymes [3]. Implementation of this clever idea was made possible over

the last two decades by the development of monoclonal antibodies, the synthesis of transition-state analogs (called haptens), and versatile screening assays. More than 100 reactions have been successfully accelerated by antibodies [4, 5], which is a nice proof of principle that transition-state stabilization is a strategy used by enzymes to promote catalysis [6].

Although catalytic antibodies are primitive compared to enzymes with rate accelerations in the range of 10^3 - 10^5 , they provide an alternative vantage point for examining the relationship between binding energy and catalysis. Instead of working backwards from a fully evolved enzyme, one can start with an antibody catalyst as a simple model system that has been developed to illuminate specific mechanistic questions. In addition, functional analysis of the antibody intermediates that arise during affinity maturation [7-9] can shed light on the evolution of catalytic function.

This dissertation examines the structural and energetic aspects of antibody and enzyme catalysis using molecular dynamics, quantum mechanical calculations, and free energy analyses. An understanding of these aspects is important for the development of improved catalysts and complements the insight gained from crystallography and biochemical studies.

The research described begins with a detailed energetic analysis of the affinity maturation in catalytic antibody 48G7 (Chapter 2). The first structural insights into the amazing process by which the immune system optimizes the binding affinity of antibodies for a foreign particle (*i.e.* hapten) were provided by Schultz and co-workers in 1997 [7, 8], who obtained high-resolution crystal structures of both the germ line and mature forms of 48G7 in its hapten-bound and unbound states. The structures revealed

that the dramatic increase in affinity ($>10^4$ -fold) of the mature form for hapten is due to nine somatic mutations that are distant from the binding site [10]. How do these mutations lead to the increased binding affinity? With the timely availability of the MM-PBSA methodology described below, we were able to explore the energetics involved in hapten binding for both the germ line and mature forms of 48G7. This was the first energetic analysis of affinity maturation in antibodies, which provides a glimpse into how transition-state stabilization by enzymes might be optimized.

Chapter 3 extends analysis of the role of binding energy in catalytic antibodies to their mechanisms of catalysis. Specifically, the hapten-programmed mechanism of direct hydroxide attack on the scissile carbonyl of the substrate is investigated for the highly efficient esterase antibody, 17E8, and the unique amidase antibody, 43C9, which hydrolyzes amides as well. This mechanism has been suggested by Hilvert, based on examination of the crystal structures of seven independently generated esterase antibodies, which have active-site features that stabilize the transition-states for base-catalyzed ester hydrolysis [11, 12]. A detailed analysis of the mechanism for these antibodies has not been pursued. In this study, quantum mechanical and free energy calculations are applied to explore the plausibility of the proposed mechanism.

Finally, Chapter 4 describes a novel mechanism of catalysis by orotidine 5'-phosphate decarboxylase (ODCase), the most proficient enzyme known. Because of its high catalytic efficiency, the mechanism of ODCase has been of considerable interest since the first measurement of its activity in 1995 [1]. Even before structures of the enzyme were available, various mechanisms were proposed [13-16]. The publications of crystal structures solved for the enzyme from four different species and four independent

research groups in the February issue of *Proceedings of the National Academy of Sciences of the United States of America* in 2000 [17-21] were together, an exciting step towards understanding the mechanism of ODCase. A theoretical study of the enzyme by Gao and co-workers was published in the same issue as the structural studies, proposing a ground-state destabilization mechanism for the enzyme [21]. This mechanism was widely accepted such that ODCase was viewed as a paradigm for the ground-state destabilization mechanism, which has not yet been proved to be operative for any enzyme. We have found this mechanism to be counterintuitive for reasons described in Chapter 4 and demonstrate the plausibility of an alternative mechanism through the use of molecular dynamics simulations, quantum-mechanical calculations, and free energy analyses.

Computational approaches:

Two recently developed approaches have allowed us to gain valuable insights into mechanisms of enzyme catalysis that are not accessible by structural or biochemical studies. One is the MM-PBSA free energy approach [22], which can be used to examine noncovalent aspects of catalysis such as the role of binding energy in the differential stabilization of the transition states and ground states of the reaction. The other is the QM-FE approach [23, 24], which can be applied to steps that involve covalent catalysis.

MM-PBSA. The MM-PBSA approach is a hybrid method combining molecular mechanics and continuum solvent calculations [22]. The cornerstone of this approach is its success in the accurate determination of relative binding and conformational free energies for a variety of macromolecular systems [25]. Free energies (ΔG) are computed with the following equation:

$$\Delta G = \Delta E_{MM} + \Delta G_{\text{solv}} - T\Delta S.$$

These energies are computed for each configuration of the system from an ensemble obtained by molecular dynamics simulation with explicit solvent. First, a molecular-mechanical energy term (ΔE_{MM}) is computed for each configuration stripped of solvent molecules and counterions. This “gas-phase” energy term is evaluated with the potential energy function used in the Cornell *et al.* force field [26] of the AMBER molecular dynamics simulation package [27]. The all-atom potential energy function [$E(\mathbf{r}^N)$], which is a function of the positions (\mathbf{r}) of N atoms, is defined as a summation of pairwise bond, angle, dihedral, van der Waals, and electrostatic terms:

$$E(\mathbf{r}^N) = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 \\ + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

The conformational entropy of the each solute configuration ($-\text{T}\Delta\text{S}$) is approximated as the sum of translational, rotational, and vibrational terms. The translational and rotational terms are computed semi-classically while the vibrational term is estimated by normal mode analysis. The free energy of solvation (ΔG_{solv}) is approximated as the sum of electrostatic and nonpolar contributions. The electrostatic solvation contribution is computed with the Poisson-Boltzmann (PB) approach while the nonpolar contribution is estimated by a surface-area (SA) dependent term, hence the name MM-PBSA.

In the PB approach, the solvent is represented by a continuum with a dielectric constant of 80 (for room-temperature water) while the solute is treated as low dielectric region with a constant of 1 (for vacuum). Higher dielectric values (*i.e.* 2 or 4) are used for the solute interior in some cases to include the capacity for electronic polarization due to the reaction field from the solvent. The solute and solvent, with associated charges and dielectric boundaries, are mapped onto a cubic lattice grid. The electrostatic potential at each grid point is determined by using iterative numerical techniques to obtain a finite difference solution to the Poisson equation:

$$\nabla \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) + 4\pi \rho(\mathbf{r}) = 0.$$

The total electrostatic solvation energy (ΔG_{PB}) is then defined as the change in electrostatic energy from transferring the solute from a low ($\epsilon = 1$) to high dielectric ($\epsilon = 80$) medium:

$$\Delta G_{PB} = \frac{1}{2} \sum_i q_i (\phi_i^{80} - \phi_i^1)$$

The nonpolar contribution to solvation (ΔG_{SA}) is approximated as linearly dependent on the solvent accessible surface area (SASA):

$$\Delta G_{SA} = \gamma (\Delta SASA) + \beta,$$

where $\gamma = 0.00542$ and $\beta = 0.92$ kcal/mol as parameterized for the PARSE radii in the DelPhi2.0 software package [28]. The solvent assessable surface area is computed with Sanner's MSMS software [29].

QM-FE. The QM-FE approach [23, 24] is a hybrid of quantum mechanical and classical free energy perturbation calculations that has provided accurate estimates of the activation energies for enzyme-catalyzed reactions such as those of trypsin [30] and catechol *O*-methyl transferase [31]. In this approach, the change in quantum-mechanical energy (ΔE_{QM}) for a step involving bond cleavage or formation is computed using the minimal portions of the enzyme and substrate where bonding is changing. Use of a minimal QM model for the reaction step enables the use of high-level *ab initio* (*i.e.* MP2) or density functional methods (*i.e.* B3LYP) that include electron correlation. As a

tentative strategy for taking into account the bonding between the atoms of the QM model and the rest of the protein, “link atoms” are included where appropriate in the model. Atom charges for the QM model are derived using the RESP method [32] in the AMBER software package [27] with Lagrangian constraints on the link atoms. These charges are internally consistent with the rest of the protein. To incorporate environmental effects on the reaction step, the QM model along with these charges are spliced into the framework of the rest of the protein, which is represented by an accurate molecular mechanical model. The change in interaction free energy (ΔG_{FE}) of the QM region with the rest of the system is then computed with classical free energy perturbation calculations for the reaction step. The overall free energy change (ΔG^*) for the reaction step is thus defined as

$$\Delta G^* = \Delta E_{QM} + \Delta G_{FE}.$$

References:

1. Radzicka, A. and R. Wolfenden, *A proficient enzyme*. Science, 1995. **267**(5194): p. 90-93.
2. Pauling, L., Am. Sci., 1948. **36**: p. 51-58.
3. Jencks, W.P., *Catalysis in Chemistry and Enzymology*. 1969, New York: McGraw-Hill.
4. Pollack, S.J., J.W. Jacobs, and P.G. Schultz, *Selective chemical catalysis by an antibody*. Science, 1986. **234**: p. 1570-73.
5. Tramontano, A., K.D. Janda, and R.A. Lerner, *Catalytic antibodies*. Science, 1986. **234**: p. 1566-70.
6. Lerner, R.A., S.J. Benkovic, and P.G. Schultz, *At the Crossroads of Chemistry and Immunology: Catalytic Antibodies*. Science, 1991. **252**: p. 659-667.
7. Wedemayer, G.J., et al., *Structural insights into the evolution of an antibody combining site*. Science, 1997. **276**(5319): p. 1665-9.
8. Wedemayer, G.J., et al., *Crystal structures of the free and liganded form of an esterolytic catalytic antibody*. J Mol Biol, 1997. **268**(2): p. 390-400.
9. Romesberg, F.E., et al., *Immunological origins of binding and catalysis in a Diels-Alderase antibody*. Science, 1998. **279**(5358): p. 1929-33.
10. Patten, P.A., et al., *The immunological evolution of catalysis*. Science, 1996. **271**(5252): p. 1086-91.
11. MacBeath, G. and D. Hilvert, *Hydrolytic antibodies: variations on a theme*. Chem Biol, 1996. **3**(6): p. 433-45.
12. Hilvert, D., *Critical analysis of antibody catalysis*. Annu Rev Biochem, 2000. **69**: p. 751-93.
13. Beak, P. and B. Siegel, *Mechanism of Decarboxylation of 1,3-Dimethylorotic Acid. A Possible Role for Orotate Decarboxylase*. J. Am. Chem. Soc., 1976. **98**: p. 3601-3606.
14. Silverman, R. and M. Groziak, J. Am. Chem. Soc., 1982. **104**: p. 6434-6439.
15. Lee, J.K. and K.N. Houk, *A proficient enzyme revisited: the predicted mechanism for orotidine monophosphate decarboxylase*. Science, 1997. **276**(5314): p. 942-5.

16. Ehrlich, J.I., et al., *Evidence for a stepwise mechanism of OMP decarboxylase*. Journal of the American Chemical Society, 1999. 121(29): p. 6966-6967.
17. Appleby, T.C., et al., *The crystal structure and mechanism of orotidine 5'-monophosphate decarboxylase*. Proc Natl Acad Sci U S A, 2000. 97(5): p. 2005-10.
18. Harris, P., et al., *Structural basis for the catalytic mechanism of a proficient enzyme: orotidine 5'-monophosphate decarboxylase*. Biochemistry, 2000. 39(15): p. 4217-24.
19. Feng, W.Y., et al., *The mechanism of orotidine 5'-monophosphate decarboxylase: catalysis by destabilization of the substrate*. Biochemistry, 2000. 39(7): p. 1778-83.
20. Miller, B.G., et al., *Anatomy of a proficient enzyme: the structure of orotidine 5'-monophosphate decarboxylase in the presence and absence of a potential transition state analog*. Proc Natl Acad Sci U S A, 2000. 97(5): p. 2011-6.
21. Wu, N., et al., *Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase*. Proc Natl Acad Sci U S A, 2000. 97(5): p. 2017-22.
22. Srinivasan, J., et al., *Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices*. Journal of the American Chemical Society, 1998. 120(37): p. 9401-9409.
23. Kuhn, B. and P.A. Kollman, *QM-FE and molecular dynamics calculations on catechol O-methyltransferase: Free energy of activation in the enzyme and in aqueous solution and regioselectivity of the enzyme-catalyzed reaction*. Journal of the American Chemical Society, 2000. 122(11): p. 2586-2596.
24. Stanton, R.V., et al., *Combined ab initio and free energy calculations to study reactions in enzymes and solution: Amide hydrolysis in trypsin and aqueous solution*. Journal of the American Chemical Society, 1998. 120(14): p. 3448-3457.
25. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models*. Accounts of Chemical Research, 2000. 33(12): p. 889-897.
26. Cornell, W.D., et al., *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules*. Journal of the American Chemical Society, 1995. 117(19): p. 5179-5197.
27. Case, D.A., et al., *Amber 5.0*. 1997, San Francisco, CA: University of California.

28. Sharp, K.A. and B. Honig, *Electrostatic interactions in macromolecules: theory and applications*. *Annu Rev Biophys Biophys Chem*, 1990. **19**: p. 301-32.
29. Sanner, M.F., A.J. Olson, and J.C. Spehner, *Reduced surface: an efficient way to compute molecular surfaces*. *Biopolymers*, 1996. **38**(3): p. 305-20.
30. Stanton, R.V., et al., *Combined ab initio and Free Energy Calculations to Study Reactions in Enzymes and Solution: Amide Hydrolysis in Trypsin and Aqueous Solution*. *J. Am. Chem. Soc.*, 1998. **120**: p. 3448-3457.
31. Kuhn, B. and P.A. Kollman, *QM-FE and Molecular Dynamics Calculations on Catechol O-Methyltransferase: Free Energy of Activation in the Enzyme and in Aqueous Solution and Regioselectivity of the Enzyme-Catalyzed Reaction*. *J. Am. Chem. Soc.*, 2000. **122**: p. 2586-2596.
32. Bayly, C.I., et al., *A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model*. *Journal of Physical Chemistry*, 1993. **97**(40): p. 10269-10280.

**Chapter 2. Molecular Dynamics and Free Energy Calculations Applied to Affinity
Maturation in Antibody 48G7**

Lillian T. Chong, Yong Duan, Lu Wang, Irina Massova, and Peter A. Kollman

Graduate Group in Biophysics

University of California at San Francisco

San Francisco, CA 94143-0446

Reproduced with permission from *Proc. Natl. Acad. Sci. USA* (1999) **96**: 14330-14335.

Copyright 1999 National Academy of Sciences, U.S.A.

Abstract:

We investigated the relative free energies of hapten binding to the germ line and mature forms of the 48G7 antibody Fab fragments by applying a continuum model to structures sampled from molecular dynamics simulations in explicit solvent. Reasonable absolute and very good relative free energies were obtained. As a result of nine somatic mutations that do not contact the hapten, the affinity-matured antibody binds the hapten $>10^4$ tighter than the germ line antibody. Energetic analysis reveals that van der Waals interactions and nonpolar contributions to solvation are similar and drive the formations of both the germ line and mature antibody-hapten complexes. Affinity maturation of the 48G7 antibody therefore appears to occur through reorganization of the combining site geometry in a manner that optimizes the balance of gaining favorable electrostatic interactions with the hapten and losing those with solvent during the binding process. As reflected by lower RMS fluctuations in the antibody-hapten complex, the mature complex undergoes more restricted fluctuations than the germ line complex. The dramatically increased affinity of the 48G7 antibody over its germ line precursor is thus made possible by electrostatic optimization.

Introduction:

The immune system is capable of producing antibodies that can bind virtually any molecule with high affinity through the combinatorial association of variable, joining, and diversity genes with subsequent affinity maturation [1]. Affinity maturation involves repeated cycles of somatic mutations in the variable regions, preferentially stimulating high-affinity antibodies in the resulting diverse pool of antibodies until an optimal affinity is achieved [2, 3]. Indeed, nature has solved many of the complex problems associated with molecular recognition by generating and screening a large library of proteins over the course of a few weeks during an immune response. By unraveling the mechanism behind molecular recognition, we will gain a deeper understanding of the principles that lead to high-affinity and selective inhibitors for a target receptor in the area of drug design.

Over the past decade, chemists have tapped the enormous diversity of the immune system to obtain antibody catalysts for chemical reactions of interest by immunizing with transition-state analog haptens. Research involving catalytic antibodies has provided us with structural insights into the process of affinity maturation. Specifically, high-resolution x-ray crystal structures for the germ line and mature Fab fragments of the esterolytic antibody 48G7, by themselves and in complex with the *p*-nitrophenyl phosphonate hapten, have been attained [4, 5]. As a result of nine somatic mutations that are distant from the combining site, the mature antibody has a $>10^4$ greater affinity for the hapten with a dissociation constant (K_d) of 10 nM, compared with a K_d of 135 μ M for the germ line antibody [6]. Conformational changes occur in the variable region of the germ

line antibody in response to binding hapten; these changes become preorganized in the combining site of the mature antibody, which undergoes little change in structure on binding hapten. Structural aspects that lead to the dramatic difference in binding affinities between the germ line and mature antibody-hapten complexes are not well understood. The availability of high-resolution structural information on the 48G7 antibody presents the opportunity to perform a detailed energetic analysis of the system. A practical approach to calculating free energies is one developed by Srinivasan *et al.* in 1998 [7], which involves applying a continuum model to solute configurations sampled as “snapshots” from a molecular dynamics (MD) simulation by using explicit solvent. For each solute configuration, a molecular mechanical energy is determined. Free energies of solvation are estimated by using finite-difference Poisson-Boltzmann (PB) calculations for the electrostatics contribution and a surface-area-dependent term for the nonelectrostatic contribution to solvation. Solute entropic contributions are estimated from a harmonic analysis. This approach has been used to investigate nucleic acid conformational problems, such as the relative stability of the “A” and “B” forms of nucleic acid duplexes [7, 8]; it has also been used for a protein-RNA complex (C. Reyes and P.A.K., unpublished observations) and for different mutants of a peptide-protein complex [9]. Here the approach of Srinivasan *et al.* [7] is applied to a small molecule-protein complex problem: the relative binding free energies of the 48G7 germ line and mature antibody-hapten complexes.

Methods:

Preparation. Coordinates were extracted from the x-ray crystal structures of the 48G7 germ line and mature Fab fragments complexed with the p-nitrophenyl phosphonate hapten (1aj7 and 1 gaf for the germ line and mature Fab-hapten complexes, respectively, in the Protein Data Bank [4, 5]). All crystallographic waters were included (151 waters for the mature Fab-hapten complex). Hydrogen atoms were added to the heavy-atom positions of the crystal structures with the LeAP module of the AMBER 5.0 package [10]. His L189, His L198, and His L55 (in the mature Fab-hapten complex only) were protonated at the ϵ -nitrogen; His H35, His H164, and His H200 were protonated at the δ -nitrogen [light chain (L) and heavy chain (H)]. Charges for the hapten were obtained by restrained electrostatic potential fitting,[11] and the electrostatic potentials were produced by single-point quantum mechanical calculations at the Hartree-Fock level with a 6-31G* basis set. The resulting partial atomic charges of the hapten along with bond, angle, and dihedral parameters for the hapten not included in the Cornell *et al.* force field [12] are listed in the *Appendix*.

Simulation. MD simulations were performed by using the Cornell *et al.* force field [12] and the AMBER 5.0 suite of programs [10]. Long-range nonbonded interactions were truncated by using a 12-Å residue-based cutoff. The SHAKE algorithm was applied to constrain all bonds to their equilibrium values, thus removing high-frequency vibrations [13]. Each system was solvated by placing a spherical cap of TIP3P water molecules [14] with a radius of 25 Å from the geometric center of the hapten. Unfavorable interactions within the structures were relieved with steepest descent followed by conjugate gradient energy minimization until the RMS of the elements in the gradient vector was less than 10^{-4} kcal/(mol·Å). Solvent in the systems (including

crystallographic waters) was then equilibrated for 25 ps while raising the temperature from 0 K to 300 K. A second equilibration phase of 100 ps at 300 K was applied with the “belly” option in which only those residues within 12 Å of the hapten geometric center were allowed to move. Constant temperature was maintained by the Berendsen coupling algorithm with separate solute-solvent and solvent-solvent coupling [15]. Sampling of reasonable configurations for the given stable states of the antibody-hapten complex structures was conducted by running a 500-ps belly simulation with a 2-fs time step at 300 K.

Energetic Analysis. The general strategy used here involves calculating energies for “snapshot” configurations taken from the MD trajectories of the antibody-hapten complexes and then averaging the values [7]. Unbound antibody and hapten snapshots were also taken from the antibody-hapten complex trajectories. Fifty snapshots taken at 10-ps intervals from each antibody-hapten complex trajectory were used for analysis. For consistency, all waters were removed from each snapshot before energy calculations were performed. Removal of crystallographic waters was not a concern, because none of the waters lie within the combining site.

Total molecular-mechanical energies (E_{MM}) and internal energy (E_{int} where $int =$ bonds, angles, and dihedrals), van der Waals (E_{vdw}), and electrostatic (E_{elec}) components were determined by using the *anal* module of the AMBER 5.0 package [10] with the same force field [12] used in the MD simulations, with the inclusion of solute-solute pairwise interactions. No cutoff was used for the evaluation of nonbonded interactions.

Solute entropic contributions were estimated for the 50th snapshots taken from the trajectories by using the *nmode* module of the AMBER 5.0 package [10], which involves

a harmonic approximation to the normal modes and standard (quantum) formulas. Residues lying outside of the belly region in the simulations were removed before minimizing and proceeding with normal mode analysis. All minimizations and normal mode calculations were performed with a distance-dependent dielectric ($\epsilon = 4 r$, where r = interatomic distance in Å) to mimic solvent screening. Steepest descent followed by conjugate gradient minimizations was carried out with a nonbonded cutoff of 12 Å until the RMS of the elements in the gradient vector was less than 10^{-4} kcal/(mol·Å). The structures were further minimized with no cutoff for nonbonded interactions by using conjugate gradient and then Newton-Raphson minimizations until the RMS of the elements in the gradient vector was less than 10^{-5} kcal/(mol·Å). Normal mode calculations were then carried out with no cutoff for nonbonded interactions. The electrostatic contribution to the solvation free energy (G_{PB}) was determined with the PB approach.[16] This approach involves using a continuum solvent model, which represents the solute as a low dielectric medium ($\epsilon = 1$) with embedded charges and the solvent as a high dielectric medium ($\epsilon = 80$) with no salt. Atomic charge values were taken from the Cornell *et al.* force field [12] as consistent with the molecular-mechanical energy calculations. However, atomic radii were taken from the PARSE parameter set [17] instead of the Cornell *et al.* force field [12] because of the small size of hydrogens in the latter. The dielectric boundary is the contact surface between the radii of the solute and the radius (1.4 Å) of a water probe molecule. The DELPHI 2.0 computer program package was used to numerically solve the linearized PB equation on a cubic lattice by using iterative finite-difference methods [16]. One thousand iterations were performed for each calculation. The cubic lattice, which had a grid spacing of 0.5 Å, was scaled

such that its dimensions were 80% larger than the longest dimension of the solute; the points on the boundary of the grid were set to the sum of Debye-Huckel potentials [18].

The nonpolar contribution to the solvation free energy (G_{SA}) was determined by using the solvent-accessible surface area algorithm (SASA) of Sanner [19] in a linear parameterization, where $G_{SA} = \gamma (\text{SASA}) + \beta$ and the constants γ and β are 0.00542 kcal/Å² and 0.92 kcal/mol, respectively, for use with PARSE atomic radii values [17]. The solvent probe radius was set to 1.4 Å.

Computational Mutagenesis. Desired alanine and glycine mutations were made by editing the snapshots from the trajectories before energetic analysis by using an approach developed by Massova and Kollman [9]. In this approach, a residue is mutated to an alanine by truncating the residue at C γ and replacing C γ with a hydrogen atom at a 1.09-Å distance from C β along the C γ -C β bond; likewise, a glycine mutation involves truncating the residue at C β and replacing C β with a hydrogen atom. The partial charges of the residue were then changed to those appropriate for alanine or glycine by simply changing the topology files. Energetic analysis of the mutants was performed as described above for the wild-type structures.

Results and Discussion:

Structure and Dynamics of the Antibody-Hapten Complexes. MD simulations of the 48G7 germ line and mature antibody-hapten complexes were performed with a spherical cap of explicit waters centered at the hapten and a “belly,” in which only atoms within a defined distance from the hapten were allowed to move during

the simulation. Stable trajectories of the complexes resulted, as demonstrated by steady RMS fluctuations (see Figure 1) and RMS deviations from their minimized crystal structures (data not shown). As shown in Figure 1, the belly atoms in the germ line complex experienced greater RMS fluctuations (0.4 to 0.6 Å) than those in the mature complex (0.3 to 0.5 Å). RMS deviations of the belly atoms in the complexes from their minimized crystal structures were 1 Å and 0.7 Å, as averaged over the last 100 ps of the simulations for the germ line and mature complexes, respectively. These results suggest that the germ line complex has greater flexibility than the mature complex. The importance of this greater flexibility is underscored by two well-discussed notions of affinity maturation [4, 6]: (i) the germ line antibody must readily adapt to a variety of antigens to efficiently protect the host from disease; and (ii) the plasticity of the germ line antibody facilitates its remodeling by somatic mutation into a “mature” conformation that has a tighter fit for the repeatedly introduced antigen.

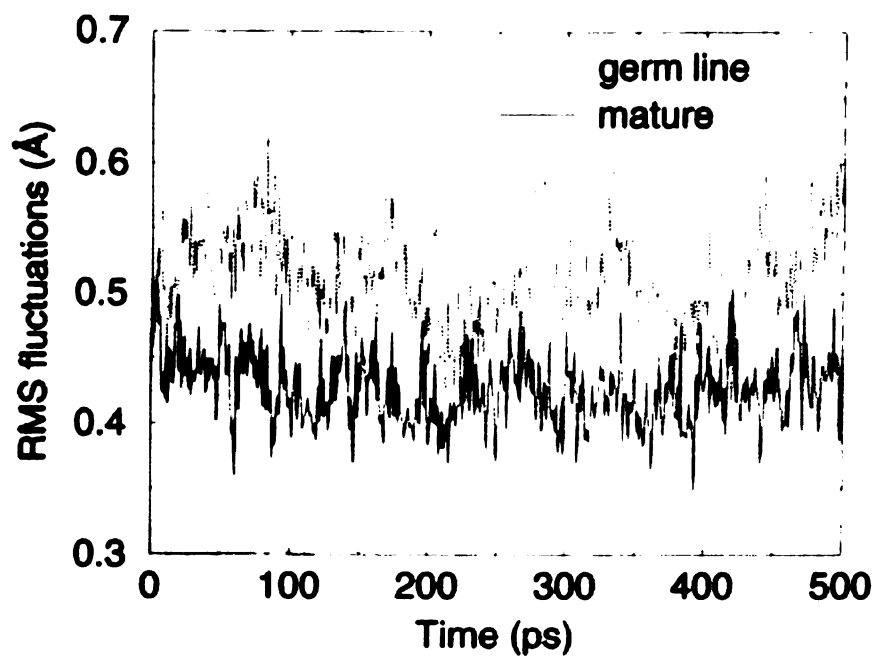


Figure 1. RMS fluctuations of the germ line and mature antibody-hapten complexes. RMS fluctuations are defined as RMS deviations of the structure at a given time from the average structure of the MD simulation. Only mobile atoms of the simulation, or “belly” atoms (see *Methods*), were considered for the RMS fluctuations.

Table 1. Energetic analysis of 48G7 antibody-hapten complex formations

	Germ line		Mature	
$\langle \Delta E_{es} \rangle$	-304.9	(2.5)	-349.1	(1.5)
$\langle \Delta E_{vdw} \rangle$	-35.9	(0.5)	-35.5	(0.4)
$\langle \Delta E_{MM} \rangle$	-340.8	(2.4)	-384.6	(1.4)
$\langle \Delta G_{SA} \rangle$	-4.4	(0.0)	-4.2	(0.1)
$\langle \Delta G_{PB} \rangle$	317.6	(2.2)	354.0	(1.2)
$\langle \Delta G_{solv} \rangle$	313.3	(2.2)	349.6	(1.2)
$\langle \Delta G_{es,tot} \rangle$	12.7	(0.9)	4.8	(0.5)
$\langle \Delta G_{tot} \rangle$	-27.5	(0.6)	-35.0	(0.5)
$-T\Delta S^b$	18.4		19.3	
ΔG_{bind}	-9.1		-15.7	

Values in parentheses are SEM.

Results are mean energies (in kcal/mol) and SEM from 50 equally spaced "snapshot" configurations of a 500-ps MD simulation. Each snapshot contains only the solute, which consists of all antibody and hapten atoms, including hydrogens. Definitions of the energetic components are as follows: ΔE_{es} = electrostatic molecular-mechanical energy; ΔE_{vdw} = van der Waals molecular-mechanical energy; ΔE_{MM} = total molecular-mechanical energy ($\Delta E_{es} + \Delta E_{vdw} + \Delta E_{mm}$); ΔG_{SA} = nonpolar contribution to the solvation energy; ΔG_{PB} = electrostatic contribution to the solvation energy calculated by the PB approach; ΔG_{solv} = total solvation energy ($\Delta G_{PB} + \Delta G_{SA}$), $\Delta G_{es,tot}$ = total electrostatic energy ($\Delta E_{es} + \Delta G_{PB}$); ΔG_{tot} = total energy without solute entropic contribution ($\Delta E_{MM} + \Delta G_{solv}$); $-T\Delta S$ = solute entropic contribution, where T = temperature and S = sum of translational, rotational, and vibrational entropies; and ΔG_{bind} = total energy with solute entropic contribution ($\Delta G_{tot} - T\Delta S$).

Energetic Analysis of the Antibody-Hapten Complexes. The results from energetic analysis of 50 equally spaced snapshots taken from each of the two MD simulations are summarized in Table 1. All energetic analysis was done for only a single MD trajectory of the desired antibody-hapten complex with unbound antibody and hapten snapshots taken from snapshots of that trajectory. The average binding free energy ($\langle \Delta G_{bind} \rangle$) for the complex was estimated by the following expression:

$$\langle \Delta G_{\text{bind}} \rangle = \langle \Delta E_{\text{MM}} \rangle + \langle \Delta G_{\text{solv}} \rangle - T\Delta S,$$

where $\langle \Delta E_{\text{MM}} \rangle$ is the molecular-mechanical energy, $\langle \Delta G_{\text{solv}} \rangle$ is the solvation energy, and $-T\Delta S$ is the solute entropic contribution. The molecular-mechanical energy consists of internal ($\langle \Delta E_{\text{int}} \rangle$), van der Waals ($\langle \Delta E_{\text{vdw}} \rangle$), and electrostatic ($\langle \Delta E_{\text{cs}} \rangle$) components. Note that, because the structures of each antibody in its unbound and bound states were the same, the internal component of the molecular-mechanical energy has zero contribution to the binding free energy ($\langle \Delta E_{\text{int}} = 0 \rangle$). The solvation energy consists of an electrostatic contribution ($\langle \Delta G_{\text{PB}} \rangle$) and a nonpolar contribution ($\langle \Delta G_{\text{SA}} \rangle$). Solute entropic contributions were determined only for the last snapshots of the MD trajectories. The calculated free energies of germ line complex formation ($\langle \Delta G_{\text{bind}} \rangle = -9.1$ kcal/mol) and of mature complex formation ($\langle \Delta G_{\text{bind}} \rangle = -15.7$ kcal/mol) agree fairly well with those from experiment (-5.3 kcal/mol and -10.9 kcal/mol, respectively, for the germ line and mature complexes) [6]. The favorable formations of these complexes are driven by the van der Waals contributions ($\langle \Delta E_{\text{vdw}} \rangle$ values) and the nonpolar contributions to solvation ($\langle \Delta G_{\text{SA}} \rangle$ values). These nonelectrostatic components are very similar for both the germ line and mature complex formations ($\langle \Delta E_{\text{vdw}} \rangle = -35.9$ kcal/mol and $\langle \Delta G_{\text{SA}} \rangle = -4.4$ kcal/mol for the germ line complex; $\langle \Delta E_{\text{vdw}} \rangle = -35.5$ kcal/mol and $\langle \Delta G_{\text{SA}} \rangle = -4.2$ kcal/mol for the mature complex). The solute entropic contributions ($-T\Delta S$ values) are similar as well for the germ line and mature complex formations (18.4 kcal/mol and 19.3 kcal/mol, respectively).

It is important to consider the electrostatic component of the molecular-mechanical energy ($\langle \Delta E_{es} \rangle$) together with the electrostatic contribution to solvation ($\langle \Delta G_{PB} \rangle$) when examining the role of electrostatics in the antibody-hapten complex formations. As demonstrated by numerous studies [20-26], electrostatics generally disfavor the docking of ligand and receptor molecules because the unfavorable change in the electrostatics of solvation is mostly, but not fully, compensated by the favorable electrostatics within the resulting ligand-receptor complex. Indeed, the total electrostatic energy contributions ($\langle \Delta G_{es, tot} \rangle$ values) for both the germ line and mature antibody-hapten complex formations are unfavorable, with values of 12.7 kcal/mol and 4.8 kcal/mol, respectively. The mature complex formation is less unfavorable than the germ line complex formation because of a less positive total electrostatic term in which the penalty paid by the electrostatics of solvation is better compensated by favorable electrostatic interactions within the complex. Thus, even though electrostatics destabilize antibody-hapten complex formation, it is the optimized balance of opposing electrostatic contributions that leads to tighter binding in the mature complex.

On the affinity maturation of the 48G7 antibody, changes in either its conformation or amino acid sequence may account for the optimization of electrostatics. In other words, the nine somatic mutations may increase the binding affinity of the antibody for hapten directly through long-range electrostatic interactions or indirectly by inducing a conformational change that leads to improved electrostatic interactions within the immediate binding vicinity. To isolate the effects of conformational change from those of sequence change, we produced identical amino acid sequences for the germ line and mature antibodies while preserving their respective original conformations; this was

done by replacing selected residues at the positions of the nine somatic mutations in one or both antibodies to alanines (or glycines, if the residues were glycines in either antibody) with no subsequent minimization. Details of this computational mutagenesis approach are described in *Methods*. We then proceeded to conduct an energetic analysis of the resulting “mutant” antibodies.

Table 2. Energetic analysis of mutant 48G7 antibody-complex formations

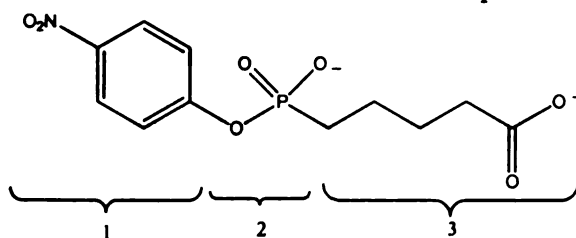
	kcal/mol	
	Germ line mutant	Mature mutant
$\langle \Delta E_{es} \rangle$	-374.0 (2.5)	-370.0 (1.3)
$\langle \Delta E_{vdW} \rangle$	-35.8 (0.5)	-35.5 (0.4)
$\langle \Delta E_{MM} \rangle$	-409.8 (2.4)	-405.4 (1.2)
$\langle \Delta G_{SA} \rangle$	-4.3 (0.0)	-4.4 (0.0)
$\langle \Delta G_{PB} \rangle$	386.5 (2.2)	374.6 (1.0)
$\langle \Delta G_{solv} \rangle$	382.1 (2.2)	370.3 (1.0)
$\langle \Delta G_{es,tot} \rangle$	12.5 (0.9)	4.6 (0.5)
$\langle \Delta G_{tot} \rangle$	-27.7 (0.7)	-35.2 (0.5)

In order to produce mutant germ line and mature antibodies with identical amino acid sequences, the following mutations were made: Ser L30, Asp L55, Glu H42, Asn H56, Asn H76 → Ala and Ser L34 → Gly in the germ line antibody; Asn L30, His L55, Lys H42, Asp H56, Lys H76, Thr H78 → Ala and Val H55, Asp H65 → Gly in the mature antibody. See Table 1 legend for definitions of the energy components.

Results from our energetic analysis of the mutant germ line and mature antibodies are shown in Table 2. The total free energies ($\langle \Delta G_{tot} \rangle$ values) for the mutant germ line and mature antibody-hapten complex formations are different, with values of -27.7 kcal/mol and -35.2 kcal/mol, respectively. These values are essentially unchanged from those of the wild-type germ line and mature antibody-hapten complex formations, which are -27.5 kcal/mol and -35.0 kcal/mol, respectively (see Table 1). Conformational differences are therefore key to the affinity maturation of the 48G7 antibody.

This result is not too surprising, because the somatic mutations are distant from the hapten (the closest one being 6 Å away). Nonetheless, the similarity of the results for the wild-type and mutant antibody-hapten complexes suggests that long-range electrostatic interactions have little, if any, effect on the formation of the complexes. The difference in the binding affinities of the germ line and mature antibodies for the hapten must therefore lie in the short-range electrostatics, which can be largely influenced by reorganization of the combining site. Electrostatic interactions between the hapten and the residues in the germ line and mature antibodies were thus explored, dividing the hapten into three moieties: a nitroaryl “head” moiety, a phosphonate moiety, and an alkyl “tail” moiety.

Table 3. Electrostatic interactions between hapten and antibody



	kcal/mol	
	Germ line	Mature
Head (1)	18.4 (0.3)	2.8 (0.3)
Phosphonate (2)	-187.6 (0.9)	-234.4 (0.8)
Tail (3)	-135.8 (2.1)	-117.5 (1.4)
Total	-304.9 (2.5)	-349.1 (1.5)

Results are mean interaction energies (in kcal/mol) and SEM from 50 configurations. Note that the values in the row labeled "totals" correspond to the $\langle \Delta E_{es} \rangle$ and $\langle \Delta E_{vdw} \rangle$ values shown in Table 1.

As shown in Table 3, the phosphonate moiety of the hapten forms the bulk of the favorable electrostatic interactions between the hapten and the residues in both the germ line and mature antibodies. In fact, the net sum of electrostatic interactions between the antibody and hapten ($\langle \Delta E_{es} \rangle$) is more favorable for the mature complex ($\langle \Delta E_{es} \rangle = -349.1$ kcal/mol) than for the germ line complex ($\langle \Delta E_{es} \rangle = -304.9$ kcal/mol) largely because of a more negative electrostatic term for the phosphonate moiety of the hapten (-234.4 kcal/mol for the mature complex vs. -187.6 kcal/mol for the germ line complex). Note that the electrostatic interactions of the hapten tail are more favorable in the germ line complex (-135.8 kcal/mol) than in the mature complex (-117.5 kcal/mol). However, the differences in the electrostatic interactions of the tail are not as great as those of the phosphonate moiety. The phosphonate moiety is thus a major binding determinant, as suggested earlier by crystallographic studies based on the formation of hydrogen bonds between the phosphonyl oxygens and the combining site residues, Arg L96 and His H35, in both the germ line and mature complexes; in the mature complex, the phosphonyl

oxygens also form an additional hydrogen bond with Tyr H33 [4, 5]. These crystallographic studies also indicate that the tail of the hapten may be disordered and highly flexible because of its weak electron density relative to that of the phosphonate moiety. This flexibility may introduce error into our calculations. Nonetheless, our results are based on MD simulations that show low RMS fluctuations ($<1 \text{ \AA}$) for the tail region of the hapten in both the germ line and mature complexes.

The interaction between the phosphonate moiety of the hapten and Tyr H33 is key to the more favorable network of electrostatic interactions between the phosphonate moiety of the hapten and the mature antibody. This interaction, which is worth about -17 kcal/mol , is much more attractive than it is in the germ line complex, where it amounts to about -5 kcal/mol . With such a favorable electrostatic interaction between just the Tyr H33 and the phosphonate moiety of the hapten, the electrostatic interactions formed by Tyr H33 may actually compensate, or more than compensate, for the penalty paid in the electrostatics of solvation on antibody-hapten complex formation.

These possibilities were addressed by developing a strategy for determining the total electrostatic contribution of an individual residue (*i.e.*, Tyr H33), which includes both its favorable electrostatic interactions with the hapten and its unfavorable loss of electrostatic interactions with solvent. This strategy involves creating a “discharged” version of the residue by setting all partial atomic charges within the residue to zero and using the resulting modified topology file for each snapshot taken from the MD simulation of the antibody-hapten complex. One can thus estimate the total electrostatic contribution of the residue relative to its discharged hydrophobic version. Electrostatic

contributions were calculated for the resulting “mutants” as described in *Methods* for the wild-type antibody-hapten complexes.

Table 4. Electrostatic analysis of Tyr H33 in 48G7 antibody-hapten complex formations

	kcal/mol	
	Germ line	Mature
$\langle \Delta \Delta G_{PB} \rangle$	12.4 (0.8)	8.5 (0.1)
$\langle \Delta \Delta E_{cs} \rangle$	-17.7 (1.2)	-16.5 (0.3)
$\langle \Delta \Delta G_{cs,tot} \rangle$	-5.3 (0.5)	-8.0 (0.3)

"Mutant" corresponds to the antibody-hapten complex system with "discharged" Tyr H33 (all partial atomic charges in Tyr H33 set to zero). See Table 1 legend for definition of electrostatic energetic terms.

Results for the discharged Tyr H33 mutants of both the germ line and mature antibodies in complexes with the hapten are shown in Table 4 along with the corresponding values for the wild-type antibody-hapten complexes. As predicted, the total electrostatic contributions of Tyr H33 to the germ line and mature complex formations are neutral, if not favorable, relative to the discharged version of the residue ($\langle \Delta \Delta G_{cs,tot} \rangle = -5.4$ kcal/mol for the germ line complex formation and $\langle \Delta \Delta G_{cs,tot} \rangle = -8.1$ kcal/mol for the mature complex formation). Interestingly, the net electrostatic interactions formed by Tyr H33 in the complex are about the same in the germ line and mature complexes; however, they are distributed differently as a result of the different orientations of the hapten in the two complexes. As shown in Figure 2, the phosphonate moiety of the hapten is located in similar positions for both complexes, whereas the head and the tail of the hapten are in rather different positions. In the germ line complex, the hydroxyl of Tyr H33 forms a strong hydrogen bond (worth -13 kcal/mol) with the carboxyl end of the hapten. This hydrogen bond is absent in the mature complex,

because the tail of the hapten has oriented itself away from Tyr H33; however, this missing hydrogen bond is compensated by a strong hydrogen bond (worth -17 kcal/mol) between the hydroxyl of a newly oriented Tyr H33 and the phosphonate moiety of the hapten. The net $\langle \Delta \Delta E_{es} \rangle$ (Table 4) is actually about 1 kcal/mol more favorable in the germ line antibody, but this is compensated by $\langle \Delta \Delta G_{PB} \rangle$, the solvation contribution, which is less favorable than that of the mature complex by about 4 kcal/mol. This more destabilizing electrostatic contribution to solvation is because of greater burial of Tyr H33 on binding of the germ line antibody to hapten with a solvent-accessible surface area of 45.2 \AA^2 in the unbound state and 25.1 \AA^2 in the bound state, compared with the values for the mature antibody, which are 22.4 \AA^2 and 12.1 \AA^2 for the unbound and bound states, respectively.

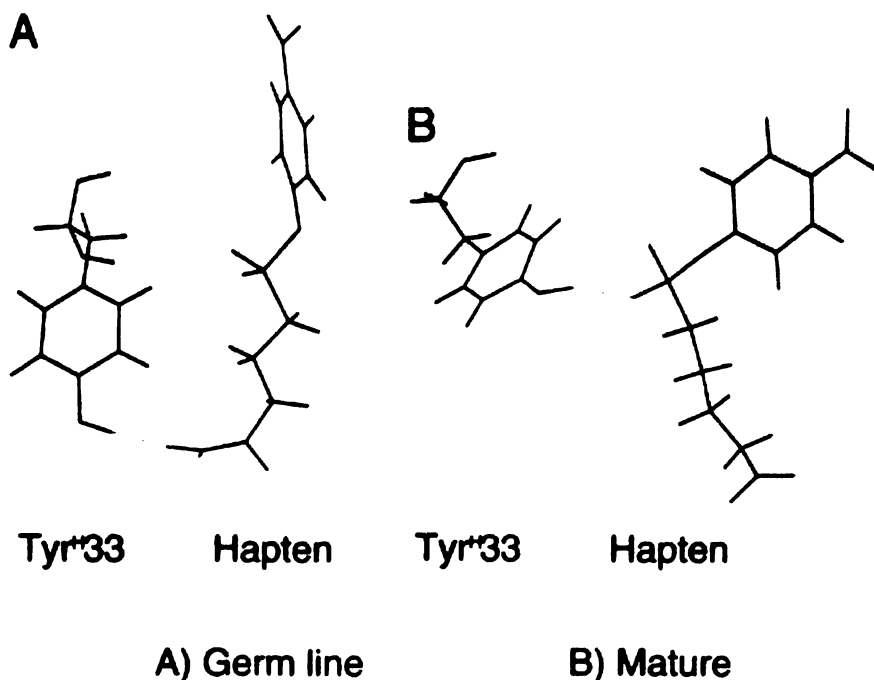


Figure 2. Orientations of Tyr H33 and hapten in the 48G7 germ line and mature antibody-hapten complexes. Orientations are taken from representative structures of the appropriate antibody-hapten complex trajectories. Hydrogen bonds are indicated by dashed lines. (A) Germ line antibody-hapten complex. The carboxyl end of the hapten is ideally oriented to form a hydrogen bond with the hydroxyl of Tyr H33. (B) Mature antibody-hapten complex. Hydrogen bonding is not possible between Tyr H33 and the newly positioned carboxyl end of the hapten. However, Tyr H33 has shifted toward the phosphonate moiety to enable the formation of a strong hydrogen bond between its hydroxyl and one of the phosphonyl oxygens. Produced with the MIDASPLUS graphics program [33].

Approximations and Future Directions:

Our approach involving the combination of MD simulations in explicit solvent with the use of continuum solvent models gives reasonable absolute and relative estimates of free energies for the formations of the 48G7 germ line and mature antibody-hapten complexes. These results encourage using this approach as an alternative to Aqvist's linear interaction energy method [27, 28], which also calculates free energies of binding from MD simulations.

A few advantages exist in our approach vs. the linear interaction energy method. One advantage is that our approach requires only one MD simulation, which is for the ligand-receptor complex. The linear interaction energy method requires the generation of two simulations, one for the unbound ligand and one for the ligand-receptor complex. Second, our approach also has the advantage of using the PB treatment of electrostatics, which tends to be more accurate than the linear response approximation for electrostatics used in Aqvist's method. Accurate treatment of electrostatics is particularly important for highly charged systems, such as the 48G7 germ line and mature antibody-hapten complexes, which involve a net charge of -2 on the hapten and net charges of $+3$ and $+5$ for the germ line and mature antibodies, respectively. Thirdly, our approach has the advantage of requiring no empirical parameters that are specific for a given system, whereas Aqvist's method involves an empirical expression for nonpolar effects, which contains parameter values that vary between systems [29].

There are some approximations in our study. First, we have used the snapshots from the MD trajectory of each antibody-hapten complex as representative structures of the unbound antibody and hapten states. Approximating the unbound and bound states of the antibody or of the hapten as the same would be problematic if major conformational changes occur on complex formation. However, only small conformational changes in the binding regions take place in the 48G7 germ line and mature complexes [4]. Furthermore, it has been shown for a peptide-protein complex [9] and a protein-RNA complex (C. Reyes and P.A.K., unpublished observations) that, if the true unbound states of the ligand and receptor of interest are used instead of being taken from the complex MD simulations, the resulting absolute binding free energies are of similar magnitudes as

when all the free energies are extracted from the single antibody-hapten complex simulation.

Another uncertainty is the harmonic approximation to the solute entropy and its estimation by only a single normal mode calculation. The harmonic approximation involved in normal mode analysis is generally a crude one and, in addition, because of the very shallow potential surface with explicit water, we have removed the waters and done the normal mode analysis with a distance-dependent dielectric ($\epsilon = 4 r$, where r = interatomic distance in Å). We used a distance-dependent dielectric, rather than setting the dielectric equal to one, because minimization with fully charged residues, in the latter case without the waters, is very unrealistic and is sure to move the structure far from the x-ray-determined position. We emphasize that we have used normal mode analysis only to determine approximate estimates of the change in solute entropy on antibody-hapten complex formation.

According to our results, the affinity maturation of the 48G7 antibody involves the optimization of electrostatics; van der Waals interactions appear optimal to begin with in the germ line antibody. These conclusions are consistent with the findings for the Diels-Alderase antibody 39-A11 [30], in which the mature form binds a hydrophobic hapten 40-fold tighter than the germ line form. In this situation, the increase in binding affinity is not as dramatic as that for the 48G7 antibody, because the germ line antibody has the ability to form nearly optimal interactions with the hapten. Thus, only two somatic mutations are necessary for the affinity maturation of 39-A11 [30].

As is evident in our energetic analysis of the germ line and mature antibody-hapten complexes, there is a delicate balance between the favorable gas-phase

electrostatics term and the unfavorable change in electrostatic contribution to the solvation. In our electrostatic analysis of Tyr H33, desolvation of the residue on binding of the antibody to the hapten plays a key role in determining the different extents of stabilization provided by the electrostatics of the residue toward complex formation. This type of analysis appears to be useful in isolating the electrostatic effects of individual residues and is likely to uncover more details about the optimization of the 48G7 antibody when applied to other residues in the combining site vicinity.

Conclusions:

Using a combination of molecular mechanical energies derived from MD simulations in explicit solvent, solvation free energies derived from a continuum solvent model, and solute entropic contributions derived from normal mode analysis, we have obtained reasonable absolute and very good relative free energies for the 48G7 germ line and mature antibody-hapten complex formations. Energetic analysis reveals that van der Waals interactions and nonpolar contributions to solvation provide the basis for the favorable absolute free energies of the germ line and mature complexes and are optimal even before affinity maturation. The key to the $>10^4$ times tighter binding in the mature complex lies in the electrostatics. By counteracting the favorable electrostatic interactions that form between the antibody and the hapten, the desolvation of the antibody plays an important role in determining the effect of electrostatics, as a whole, on the formation of the antibody-hapten complex. Alanine/glycine scanning reveals that the interactions of the charged residues involved in somatic mutations are not changed by affinity maturation. Rather, the combining site geometry appears to be reorganized in a

manner that optimizes the electrostatics. Increasing the rigidity of the antibody structure further optimizes the binding affinity of the antibody for the hapten. This increased rigidity is apparent in the lower RMS fluctuations of the mature complex relative to the germ line complex.

The results of this study have provided us with insight into the process of affinity maturation. First, the immune system must be able to produce high-affinity and specific antibodies rapidly, before the antigen has time to damage the host. In the case of 48G7, this rapid response appears to be achieved through the likely ability of the germ line antibody to adapt and bind to a variety of antigens, because it is more flexible than the mature antibody. Through already optimal hydrophobic interactions but suboptimal electrostatic interactions, the germ line antibody may bind with reasonable affinity to each of these antigens. Greater affinity and specificity are then accomplished through the optimization of electrostatic interactions in the binding site and increased rigidity of the antibody during affinity maturation of the 48G7 antibody.

In the design of an ideal ligand for a given receptor, we must consider the electrostatics and van der Waals interactions as well as the specific arrangement of hydrogen bonds. Optimization, as has been done by the 48G7 antibody, can involve minimizing the desolvation penalty for favorable electrostatic interactions. Such an optimization strategy requires less precision and a more limited set of residues than fine-tuning the precise arrangement of hydrogen bonds and packing of van der Waals interactions. Thus, it is likely that molecular design efforts will benefit from effective strategies for optimizing electrostatics, including desolvation [31, 32], as well as from

developing reliable methods for creating ideal hydrogen bonding or van der Waals interactions.

Acknowledgements:

We thank P. Schultz and R. Stevens for providing the crystal structures of the 48G7 antibody system; B. Honig for the DELPHI program; M. Sanner for the solvent-accessible surface area algorithm; and members of our research group, particularly O. Donini, J. Pitera, and C. Reyes, for their helpful suggestions. This work was supported in part by a National Institutes of Health (NIH) Grant GM-29072 to P.A.K. and by a National Science Foundation Fellowship to L.T.C. Graphics were provided by the Computer Graphics Laboratory, University of California, San Francisco (T. Ferrin, principal investigator, NIH P41 Grant RR-01081).

References:

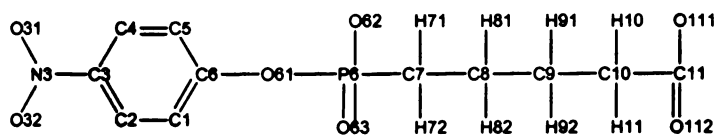
1. Tonegawa, S., *Somatic generation of antibody diversity*. Nature, 1983. **302**(5909): p. 575-81.
2. Siskind, G.W. and B. Benacerraf, *Cell selection by antigen in the immune response*. Adv Immunol, 1969. **10**: p. 1-50.
3. Kim, S., et al., *Antibody diversity: somatic hypermutation of rearranged VH genes*. Cell, 1981. **27**(3 Pt 2): p. 573-81.
4. Wedemayer, G.J., et al., *Structural insights into the evolution of an antibody combining site*. Science, 1997. **276**(5319): p. 1665-9.
5. Wedemayer, G.J., et al., *Crystal structures of the free and liganded form of an esterolytic catalytic antibody*. J Mol Biol, 1997. **268**(2): p. 390-400.
6. Patten, P.A., et al., *The immunological evolution of catalysis*. Science, 1996. **271**(5252): p. 1086-91.
7. Srinivasan, J., et al., *Continuum solvent studies of the stability of RNA hairpin loops and helices*. J Biomol Struct Dyn, 1998. **16**(3): p. 671-82.
8. Cheatham, T.E., 3rd, et al., *Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution*. J Biomol Struct Dyn, 1998. **16**(2): p. 265-80.
9. Massova, I. and P.A. Kollman, *Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies*. Journal of the American Chemical Society, 1999. **121**(36): p. 8133-8143.
10. Case, D.A., et al., *Amber 5.0*. 1997, San Francisco, CA: University of California.
11. Bayly, C.I., et al., *A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model*. Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
12. Cornell, W.D., et al., *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules*. Journal of the American Chemical Society, 1995. **117**(19): p. 5179-5197.
13. Ryckaert, J., G. Ciccotti, and H. Berendsen, J. Comput. Phys., 1977. **23**: p. 327-341.
14. Jorgensen, W., et al., J. Chem. Phys., 1983. **79**: p. 926-935.

15. Berendsen, H., et al., *J. Comput. Phys.*, 1984. **81**: p. 3684-3690.
16. Sharp, K.A. and B. Honig, *Electrostatic interactions in macromolecules: theory and applications*. *Annu Rev Biophys Biophys Chem*, 1990. **19**: p. 301-32.
17. Sitkoff, D., K.A. Sharp, and B. Honig, *Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models*. *Journal of Physical Chemistry*, 1994. **98**(7): p. 1978-1988.
18. Gilson, M.K. and B.H. Honig, *Calculation of electrostatic potentials in an enzyme active site*. *Nature*, 1987. **330**(6143): p. 84-6.
19. Sanner, M.F., A.J. Olson, and J.C. Spehner, *Reduced surface: an efficient way to compute molecular surfaces*. *Biopolymers*, 1996. **38**(3): p. 305-20.
20. Novotny, J. and K. Sharp, *Electrostatic fields in antibodies and antibody/antigen complexes*. *Prog Biophys Mol Biol*, 1992. **58**(3): p. 203-24.
21. Novotny, J., et al., *Empirical free energy calculations: a blind test and further improvements to the method*. *J Mol Biol*, 1997. **268**(2): p. 401-11.
22. Misra, V.K., et al., *Salt effects on ligand-DNA binding. Minor groove binding antibiotics*. *J Mol Biol*, 1994. **238**(2): p. 245-63.
23. Misra, V.K., et al., *Salt effects on protein-DNA interactions. The lambda ci repressor and EcoRI endonuclease*. *J Mol Biol*, 1994. **238**(2): p. 264-80.
24. Sharp, K.A., *Electrostatic interactions in hirudin-thrombin binding*. *Biophys Chem*, 1996. **61**(1): p. 37-49.
25. Shen, J. and J. Wendoloski, *Electrostatic Binding Energy Calculation Using the Finite Difference Solution to the Linearized Poisson-Boltzmann Equation - Assessment of Its Accuracy*. *Journal of Computational Chemistry*, 1996. **17**(3): p. 350-357.
26. Brucoleri, R.E., J. Novotny, and M.E. Davis, *Finite difference Poisson-Boltzmann electrostatic calculations: Increased accuracy achieved by harmonic dielectric smoothing and charge antialiasing*. *Journal of Computational Chemistry*, 1997. **18**(2): p. 268-276.
27. Aqvist, J., C. Medina, and J.E. Samuelsson, *A new method for predicting binding affinity in computer-aided drug design*. *Protein Eng*, 1994. **7**(3): p. 385-91.
28. Hansson, T., J. Marelus, and J. Aqvist, *Ligand binding affinity prediction by linear interaction energy methods*. *J Comput Aided Mol Des*, 1998. **12**(1): p. 27-35.

29. Wang, W., J. Wang, and P.A. Kollman, *What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations?* Proteins, 1999. **34**(3): p. 395-402.
30. Romesberg, F.E., et al., *Immunological origins of binding and catalysis in a Diels-Alderase antibody.* Science, 1998. **279**(5358): p. 1929-33.
31. Lee, L.P. and B. Tidor, *Optimization of electrostatic binding free energy.* Journal of Chemical Physics, 1997. **106**(21): p. 8681-8690.
32. Kangas, E. and B. Tidor, *Optimizing electrostatic affinity in ligand-receptor binding: Theory, computation, and ligand properties.* Journal of Chemical Physics, 1998. **109**(17): p. 7522-7545.
33. Ferrin, T., et al., *The MIDAS display system.* J. Mol. Graphics, 1988. **6**: p. 13-27.

Appendix:

Table A. Partial atomic charges of the hapten.



atom name	atom type	charge
C1	CA	-0.146860
H1	HA	0.179086
C2	CA	-0.224971
H2	HA	0.173445
C3	CB	-0.053011
N3	N*	0.785319
O31	O2	-0.517593
O32	O2	-0.517593
C4	CA	-0.185167
H4	HA	0.165068
C5	CA	-0.263732
H5	HA	0.134685
C6	CA	0.459948
O61	OS	-0.476414
P6	P	0.918815
O62	O2	-0.758979
O63	O2	-0.749969
C7	CT	-0.089054
H71	HC	0.036985
H72	HC	0.036985
C8	CT	0.036069
H81	HC	0.016225
H82	HC	0.016225
C9	CT	0.013449
H91	HC	-0.004153
H92	HC	-0.004153
C10	CT	-0.0066101
H10	HC	-0.024982
H11	HC	-0.024982
C11	C	0.797904
O111	O2	-0.836066
O112	O2	-0.826428

Table B. Additional force field parameters used for the hapten.

BOND	K_r (kcal mol ⁻¹ Å ⁻²)	r_{eq} (Å)			
O2-N*	750.0	1.210	= r_{eq} from AM1 geometry		
CA-OS	656.0	1.250	=C-O2		
P-CT	227.0	1.810	= r_{eq} from AM1 geometry		
ANGLE	K_θ (kcal mol ⁻¹ rad ⁻²)	θ_{eq} (deg)			
CA-CA-O2	70.0	120.0	=CT-C-O2		
CA-CB-N*	70.0	120.0	=CT-C-O2		
CA-CB-CA	70.0	120.0	=CT-C-O2		
CB-N*-O2	70.0	120.0	=CT-C-O2		
O2-N*-O2	70.0	120.0	=CT-C-O2		
HA-CA-HA	35.0	120.0	=CA-CA-HA		
CA-OS-P	90.0	120.0	= θ_{eq} from AM1 geometry		
OS-P-CT	90.0	120.0	= θ_{eq} from AM1 geometry		
P-CT-CT	40.0	109.50	=CT-CT-N*		
P-CT-HC	30.0	109.50	=CT-CT-N*		
DIHEDRAL	idivf*	$V_n/2$ (kcal mol ⁻¹)	γ (deg)	n	
CA-CA-OS-P	1	1.20	0.0	2	=X-CT-O2-X
X-P-CT-X	9	1.40	0.0	3	=X-CT-CT-X

*Factor by which torsional barrier is divided.

**Chapter 3. Direct hydroxide attack is a plausible mechanism for amidase antibody
43C9**

**Lillian T. Chong, Pradipta Bandyopadhyay, Thomas S. Scanlan, Irwin D. Kuntz,
Peter A. Kollman**

Graduate Group in Biophysics

University of California at San Francisco

San Francisco, CA 94143-0446

Reproduced in part with permission from *J. Am. Chem. Soc.*, submitted for publication.

Unpublished work copyright 2002 American Chemical Society.

Abstract:

Direct hydroxide attack on the scissile carbonyl of the substrate has been suggested as a likely mechanism for esterase antibodies elicited by phosphonate haptens, which mimic the transition states for the alkaline hydrolysis of esters [1]. The unique amidase activity of esterase antibody 43C9 has been attributed to nucleophilic attack by an active-site histidine residue on the substrate [2]. Yet, the combining site of 43C9 is strikingly similar to those of other esterase antibodies, particularly 17E8. Common features of the combining sites include a deep hydrophobic pocket for burial of the aryl leaving group and several hydrogen bond donating residues that stabilize the phosphonate moiety of the hapten. We have carried out *ab initio* quantum mechanical calculations, molecular dynamics simulations, and free energy calculations to assess the mechanism involving direct hydroxide attack for 43C9. Results support this mechanism and suggest that the mechanism is operative for other antiphosphonate antibodies that catalyze the hydrolysis of (*p*-nitro) phenyl esters.

Introduction:

Many biological processes depend on enzymes that promote the selective hydrolysis of esters and amides. To understand the basic principles underlying enzyme-catalyzed ester and amide hydrolysis, substantial effort has been made to mimic these enzymes. One such approach is the development of antibodies with tailored hydrolytic properties obtained by challenging the immune system with haptens, or stable analogs of the putative transition states. Of the more than 100 catalytic antibodies that have been developed, over half accelerate the hydrolysis of aryl ester substrates. The mechanism of hydrolysis for aryl esters as well as their corresponding activated amides, or anilides, is known to involve the formation of a transient, high energy intermediate. As shown in Figure 1, the intermediate as well as its flanking addition (TS1) and elimination (TS2) transition states can be mimicked by phosphonates or phosphoramidates, which elicit antibodies that hydrolyze aryl esters or anilides.

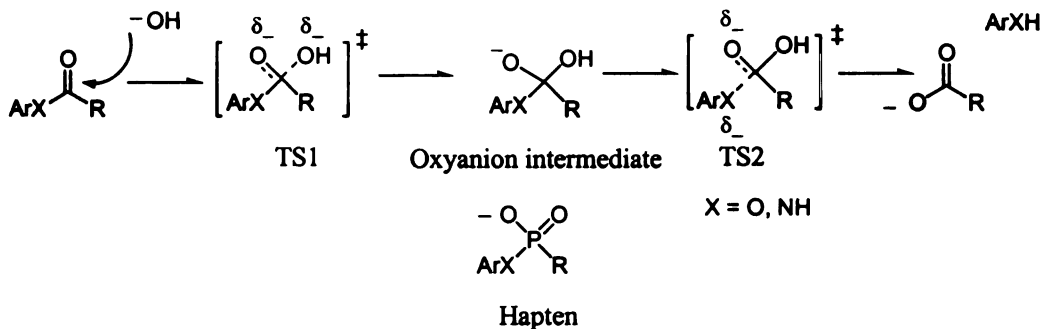


Figure 1. Base-catalyzed hydrolysis of an aryl ester or anilide. The corresponding phosphonate (X=O) or phosphoramidate (X=NH) transition state analog mimics the oxyanion intermediate as well as the addition (TS1) and elimination (TS2) transition states.

The crystal structures of seven independently generated antiphosphonate antibodies, which include two catalysts for the cleavage of phenyl esters (17E8 [4] and 29G11 [5]), two for the cleavage of *p*-nitrophenyl esters (CNJ206 [3] and 48G7 [6, 7]), and three (D2.3, D2.4, D2.5 [8, 9]) for the hydrolysis of *p*-nitrobenzyl esters, reveal combining sites that are strikingly similar. These antibodies appear to have in common a deep hydrophobic pocket for burial of the aryl/benzyl leaving group and several hydrogen bond donating residues that stabilize the phosphonate moiety [1, 10]. This motif has been recently referred to by Tantillo and Houk as a canonical binding array that recognizes the “transition state epitope” for the base-catalyzed hydrolysis of esters and amides [11]. Esterase, catalytic antibodies produced by phosphonate haptens are therefore likely to function *via* hydroxide attack on the corresponding substrate as programmed by their haptens [1].

One antibody that has an unexpected, different mechanism is 43C9, which is reported to have the unique ability to promote the hydrolysis of activated amides as well as esters. This antibody was generated against a phosphoramidate rather than a phosphonate hapten. The amidase activity of 43C9 has a maximum k_{cat} of 0.08 min^{-1} with a rate acceleration ($k_{\text{cat}}/k_{\text{uncat}}$) of 2.5×10^5 at pH 9.0 [12] while its esterase activity is among the highest reported with a maximum k_{cat} of 1500 min^{-1} and a rate acceleration of 2.7×10^4 at pH 9.3 [13]. Note that the reported k_{cat} is only an estimate from the pH-rate profile since product inhibition prevented its accurate measurement. Based on biochemical evidence, the mechanism of catalysis by 43C9 is thought to involve the formation of a covalent acyl-antibody intermediate by nucleophilic attack of the substrate by an active-site histidine residue, His L91 (L = light chain), with subsequent deacylation

by hydroxide ion [14, 15]. The role of His L91 as a nucleophile was suggested by the effects of mutating the residue to Gln in which a >50-fold decrease in catalytic efficiency resulted with little effect on ligand binding for the ester hydrolysis reaction [16]. Kinetic parameters of the mutant were, however, obtained using the *p*-chlorophenyl ester substrate instead of the *p*-nitrophenyl substrate to avoid problems with product inhibition. Nonetheless, nucleophilic attack by His L91 is supported by detection of the acyl intermediate by electrospray mass spectrometry at pH 5.9 and the fact that this intermediate was not observed under the same conditions with the His L91 → Gln mutant [17]. The crystal structures of free 43C9 and its complex with the ester hydrolysis product, *p*-nitrophenol, were recently solved [2]. Only small conformational changes occur in the unbound antibody upon binding product with an RMSD of 0.36 Å for the non-hydrogen atoms in the active site. The largest displacements by heavy atoms in this region are 1.33 Å and 1.12 Å by the terminal nitrogen atoms of the Arg L96 guanidinium group, which reorients through a 150° rotation about the C^δ-N^ε bond. Docking experiments with the product complex structure have suggested an orientation of the substrate in which its scissile carbonyl is in a reasonable position for nucleophilic attack by the imidazole side chain of His L91 [2]. This active-site histidine is therefore suggested to be the feature that distinguishes 43C9 from other esterase antibodies.

Although the crystal structure of the hapten complex of 43C9 is unavailable, superposition of the *p*-nitrophenyl moiety of the hapten onto the bound product and replacement of two crystallographic water molecules with the phosphonate oxygens reveals a similar pattern of residues involved in hapten recognition as the esterase antibodies described above, particularly 17E8 [2]. Based on sequence comparison of

active site residues, seven out of ten residues in the vicinity of the arylphosphonate moiety of the hapten in 17E8 are identical in 43C9 and another is altered conservatively (Leu L89 → Gly) [1]. Two of the identical residues are His H35 and Arg L96 (L = light chain, H = heavy chain), which form hydrogen bonds with the phosphonate oxygens of their corresponding haptens; the importance of these residues for transition-state stabilization is supported by site-directed mutagenesis experiments [16].

Antibody 17E8 is a highly efficient esterase that has a maximum k_{cat} of 223 min^{-1} with a rate acceleration of 8.3×10^3 at pH 9.5 [18] with no product inhibition. Based on the fact that the active site is similar to that of natural triad hydrolases [4] and on initial mechanistic characterizations [18], the mechanism of catalysis by 17E8 was originally thought to involve the rate-determining formation of an acyl intermediate by nucleophilic attack of an active-site serine residue. However, full wild-type activity of the Ser H99→Ala 17E8 mutant casts doubt on this possibility, suggesting that 17E8 is likely to function through a direct hydroxide attack mechanism [19]. Likewise, given that 43C9 contains the canonical binding array that is common to all the other esterase antibodies, it is surprising that its mechanism of catalysis is so different from that programmed by its hapten.

Here we explore the possibility of a direct hydroxide attack mechanism for 43C9 as well as 17E8 through *ab initio* quantum mechanical (QM) calculations, molecular dynamics (MD) simulations, and MM-PBSA free energy calculations [20]. Even though 17E8 has no detectable amidase activity, we examined the potential hydrolysis reactions of both the ester and amide substrates of each antibody. Results support the plausibility of hydroxide attack on the scissile carbonyl of the substrate as a mechanism

of catalysis for both 43C9 and 17E8 and suggest that this mechanism is likely for other hydrolytic antibodies with the same canonical binding array.

Methods:

QM Calculations. Phenyl acetate and acetanilide were used as minimal representations of the phenyl ester and amide substrates of 17E8, respectively. Likewise, *p*-nitrophenyl acetate and *p*-nitroacetanilide represented the *p*-nitrophenyl ester and *p*-nitroanilide substrates of 43C9, respectively. Calculations were performed both at the Hartree-Fock and hybrid density functional theory levels. Reported energies for stationary points along the reaction path are from calculations carried out at the HF/6-31+G**/HF-6-31+G* level or B3LYP/6-31+G**/B3LYP/6-31+G* level with zero point energy corrections (scaled by 0.9153 or 0.9806, respectively) [21] derived from frequency calculations at the corresponding level. Stationary points were characterized by calculating the Hessian matrix of energy second derivatives and by determining the number of imaginary frequencies. Intrinsic reaction coordinate (IRC) calculations were used to confirm the minima connected by each saddle point. Conductor-like polarized continuum model (COSMO) [22] self-consistent reaction field (SCRF) single point calculations at the B3LYP/6-31+G* level were performed at selected stationary points to estimate environmental effects. A dielectric constant value of 78.4 was used to simulate an aqueous environment while a value of 4 was used to approximate the protein environment [23]. Geometries in which the hydroxide ion was positioned near the amide hydrogen of *p*-nitroacetanilide led to deprotonation to form a water molecule and an

ionized substrate that was inert to hydrolysis. This side reaction was avoided in an optimized geometry where the hydroxide ion was positioned ~ 1 Å further away from the amide hydrogen. All *ab initio* calculations were performed with the GAUSSIAN 98 program [24].

MD Simulations. Models of each complex were prepared using the LeAP module of the AMBER 5.0 package [25]. The crystal structure of the single-chain variable fragment (scFv) of 43C9 in complex with the *p*-nitrophenol ester hydrolysis product (43CA in the Protein Data Bank) [2] was used as a starting point for building a model of the 43C9-hapten complex. As reported by Thayer *et al.* [2], superposition of the hapten's *p*-nitrophenyl group onto the bound *p*-nitrophenol and rotation about the C-N-P-C bond positions the hapten's two phosphamidate oxygen atoms close to the crystallographic waters 481 and 584. All crystallographic waters except for waters 481 and 584 were included in the starting model. Coordinates for the 17E8-hapten complex were taken from the 2.5-Å-resolution crystal structure of the antibody binding (Fab) fragment of 17E8 in complex with its norleucine phosphonate hapten (1EAP in the Protein Data Bank) [4]. No crystallographic waters were present in this crystal structure. Since the variable region of each antibody is sufficient for full antibody activity, [26] only the first 113 residues of the light chain and the first 117 residues of the heavy chain were included to represent the antibody. *N*-methyl capping groups were added to the C-terminal truncated ends while the N-terminal ends were treated as charged. Ionization states present in neutral solution were used for all charged residues. Tautomeric states of histidine residues (HID = protonation at the δ -nitrogen, HIE = protonation at the ϵ -nitrogen) were determined by analysis of hydrogen bonds; in ambiguous cases, the

residues were assigned to be HIE, which is the predominant tautomer for free histidines in solution [27]. Both His H35 and His L91 were assigned to be HIE for the ground-state (GS) complexes of 43C9 and HID in the transition-state (TS) complexes. In the TS complexes of 17E8, His L38, His L49, His H32, and His H35 were assigned to be HID, HIE, HID, and HIE, respectively; the same tautomeric forms were used for the GS complexes except that His H35 was assigned to be HID. Coordinates for the GS and TS complexes (transition states were approximated by tetrahedral intermediates) were obtained by modifying the bound hapten in the antibody-hapten complex structures. Charges for all ligands were determined by the restrained electrostatic potential (RESP) fitting method [28]. Electrostatic potentials were derived from HF/6-31G**//HF/6-31G* quantum mechanical calculations using the Gaussian 98 package [24]. The geometry of the tetrahedral intermediate from the ester hydrolysis catalyzed by 43C9 was optimized with the scissile ester bond constrained to 1.53 Å (the length of the bond in the ester hydrolysis intermediate of 17E8) since optimization without constraints yielded the products. All ligand parameters not present in the Parm99 parameter set [29] are reported in the *Appendix*.

In each MD simulation, we concentrated our attention on the binding site region of the antibody, permitting only residues (including ions, water molecules, and the ligand) that lie within 18 Å of the C ζ atom of Arg L96 to move. This mobile set of residues will be referred to as the “belly” region. Each antibody-ligand complex was solvated by placing a sphere of TIP3P water molecules [30] with a radius of 31 Å centered at the C ζ atom of Arg L96 and then simulated using an identical protocol and the AMBER 5.0 suite of programs [25]. To fix any errors introduced in the model-

building process, the molecular mechanics model described above was first subjected to 100 steps of steepest descent minimization followed by 1000 steps of conjugate gradient minimization. After this minimization, the system was equilibrated in two stages. At the beginning of each stage, the temperature is raised from 0 to 300 K during the first 2 ps after which the temperature remained at 300 K. In the first stage, only the ligand, ions in the belly region, and all waters were equilibrated for 40 ps. In the second stage, the entire belly region was equilibrated for 60 ps. Sampling of reasonable configurations for the resulting stable state of the antibody-ligand structure was obtained by running a 600-ps belly simulation with a 2-fs time step at 300 K. For each simulation, Berendsen temperature coupling [31] was used to maintain the system at its assigned temperature, while the SHAKE algorithm constrained the length of all bonds to their equilibrium values [32]. To minimize computational expense, long-ranged nonbonded interactions were only calculated out to a 12-Å residue-based cutoff distance. To allow for sufficient equilibration, only the latter 500 ps of each 600-ps simulation were used for subsequent analysis of the dynamics.

MM-PBSA Calculations. To compute the average binding free energy of each antibody-ligand complex, we selected 50 equally spaced “snapshot” configurations of the unbound antibody, unbound ligand, and antibody-ligand complex from the 101-600 ps interval of the 600 ps complex trajectory and performed energy calculations on the snapshots. Prior to the energy calculations, all waters and ions were removed from each snapshots. Free energy calculations were performed using the MM-PBSA approach, described in detail by Srinivasan *et al.* [33]. The binding free energy, ΔG_{bind} , is defined as follows:

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{cs}} + \Delta E_{\text{vdw}} + \Delta E_{\text{int}}$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{PB}} + \Delta G_{\text{SA}}$$

where ΔE_{MM} is the change in the total MM energy of the solute with an electrostatic component (ΔE_{cs}), a van der Waals component (ΔE_{vdw}), and an internal component (ΔE_{int}) consisting of bond, angle, and torsional energies; ΔG_{solv} is the solvation energy difference, which consists of an electrostatic contribution (ΔG_{PB}) determined by the Poisson-Boltzmann approach using the DelPhi2.0 software package [34] and a nonelectrostatic contribution (ΔG_{SA}) that is linearly dependent on the solvent accessible surface area [$\Delta G_{\text{SA}} = \gamma (\Delta \text{SASA}) + \beta$ where $\gamma = 0.00542$ and $\beta = 0.92$ kcal/mol]; and $-T\Delta S$ is the solute entropic contribution to the binding free energy. Solute entropic contributions were not calculated in this study since they are only crudely estimated by normal-mode analysis and likely to be similar for all the antibody-ligand complexes. To estimate the absolute binding free energies, a value of +20 kcal/mol was used, based on published results for the 48G7-hapten complex [35]. Because the same antibody and ligand configurations are used for their respective unbound and bound states, ΔE_{int} is equal to zero. Calculations of the electrostatic component to the MM energy and the electrostatic contribution to the solvation energy were performed using a solute, or “interior”, dielectric (ϵ_{int}) value of 1, consistent with our use of a nonpolarizable molecular mechanics force field. In this free energy approach, the solute response to charge fluctuations is estimated through explicit averaging of conformations taken from

an MD simulation rather than using an interior dielectric constant that is greater than unity for a single, static conformation.

The pK_a values of histidine residues in the vicinity of the bound hapten were computed using the DelPhi2.0 software package [34] and the continuum electrostatics approach described by Honig and co-workers [36]. The pK_a of a selected residue is defined as follows:

$$pK_a = pK_a^{sol} + \Delta\Delta G^{deprot}/RT \ln 10$$

where pK_a^{sol} is the pK_a of the residue free in neutral solution, $\Delta\Delta G^{deprot}$ is the energy of deprotonating the residue in the enzyme-substrate complex ($\Delta G^{complex}$) relative to that of deprotonating the residue when it is free in solution (ΔG^{sol}), R is the ideal gas constant, and T is the temperature. Coordinates for the antibody-hapten complexes were taken from the minimized structures. *N*-methyl and acetyl capping groups were added to histidine for our model of the residue free in solution. These calculations were performed with an interior dielectric constant (ϵ_{int}) of 20, which has been shown to predict protein pK_a 's more accurately than other values [37].

Results:

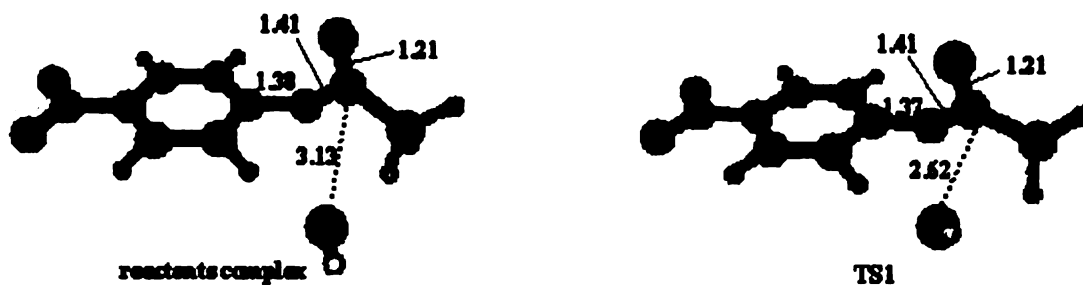
QM calculations:

Activation barriers for direct hydroxide attack on the scissile carbonyl of the substrate were estimated for the ester and amide hydrolysis reactions of 43C9 and 17E8

using *ab initio* QM calculations at the HF/6-31+G*//HF/6-31+G* and B3LYP/6-31+G*//B3LYP/6-31+G* levels of theory in the gas-phase. Environmental effects were incorporated using the COSMO continuum solvent model in single point energy calculations with B3LYP/6-31+G* geometries. *P*-nitrophenyl acetate and phenyl acetate were used to model the ester substrates for 43C9 and 17E8, respectively; the corresponding amides were used to model the amide substrates.

Hydrolysis of 43C9 model substrates. Optimized geometries at the B3LYP/6-31+G* level for the reactants and transition states involved in the hydrolysis reactions of the model ester and amide substrates for 43C9 are shown in Figure 2. For each of these reactions, a reactants complex between hydroxide ion and the model substrate was located. The ester hydrolysis reaction was concerted with the location of only one transition state, consistent with isotopic labeling studies of hydroxide attack on *p*-nitrophenyl acetate [38]. The amide hydrolysis reaction was stepwise, involving an addition transition state (TS1) and an elimination transition state (TS2). The angle of attack by hydroxide ion in the reactants complex is 112°, shifting to 104° in TS1 for the ester hydrolysis reaction; the angle of attack is 133°, shifting to 112° in TS1 for the amide hydrolysis reaction. Optimized geometries shown in Figure 2 are similar to those determined at the HF/6-31+G* level except for the distances between the oxygen of the hydroxide ion and the carbonyl carbon in the TS1 structures that are shorter by 0.3 Å. Geometries of the reactants and transition state at the HF/6-31+G* level for the ester hydrolysis reaction are the same as those reported by Tantillo and Houk at the same level of theory in a recent QM study on the fidelity of haptens for esterase antibodies [39].

Ester hydrolysis



Amide hydrolysis

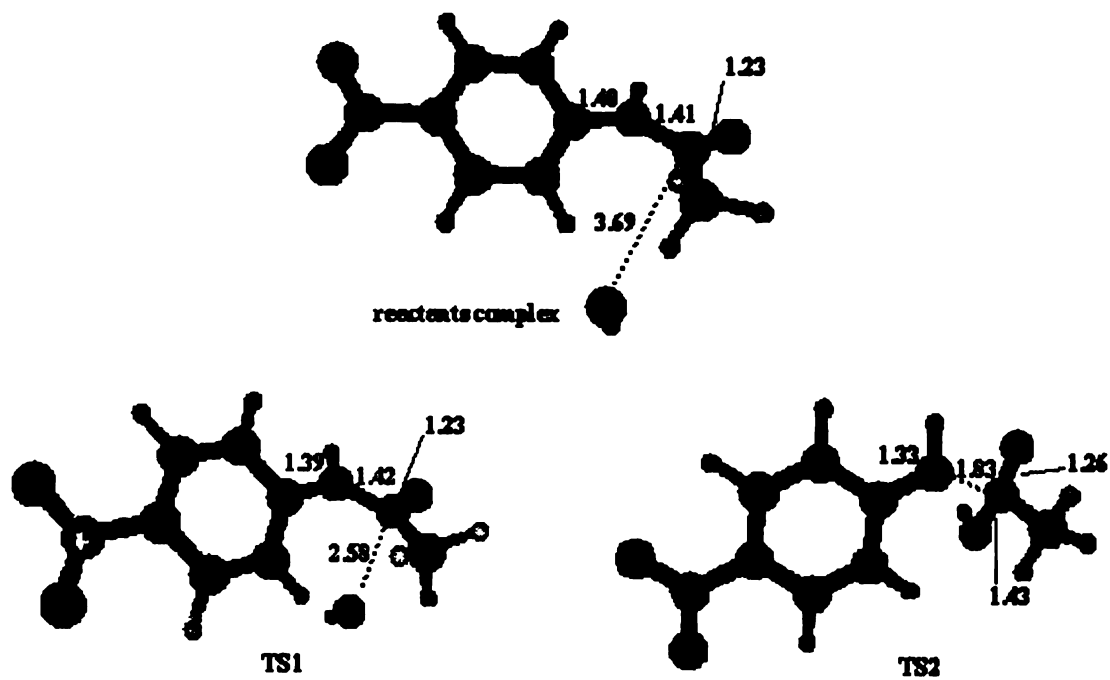


Figure 2. B3LYP/6-31+G* geometries of reactants and transition states for hydroxide attack on model ester and amide substrates for 43C9. Bond distances shown are in Å.

Table 1. Relative energies (kcal/mol) for hydroxide attack on model ester and amide substrates for 43C9 with rate-limiting barriers underlined. Gas-phase energies include zero point energy corrections calculated at the B3LYP/6-31+G* level. The relative energy of TS2 is not included for calculations of the barrier for ester hydrolysis since the reaction is concerted instead of stepwise. See text for definitions of the reactants and transition states.

		reactants			
	level of theory ^a	isolated	complex	TS1	TS2
ester hydrolysis	B3LYP/6-31+G*	40.1	0.0	<u>2.0</u>	--
	B3LYP/6-31+G*/COSMO($\epsilon=4$)	0.0	4.2	<u>4.4</u>	--
	B3LYP/6-31+G*/COSMO($\epsilon=78.4$)	0.0	18.3	<u>18.7</u>	--
amide hydrolysis	B3LYP/6-31+G*	38.0	0.0	<u>6.1</u>	-12.7
	B3LYP/6-31+G*/COSMO($\epsilon=4$)	0.0	2.0	<u>7.1</u>	-3.3
	B3LYP/6-31+G*/COSMO($\epsilon=78.4$)	0.0	14.5	<u>19.3</u>	13.5

^aGeometries optimized at the B3LYP/6-31+G* level.

Relative energies of the reactants and transition states involved in the hydrolysis reactions of the model ester and amide substrates for 43C9 are shown in Table 1. Only B3LYP/6-31+G* level results will be discussed here. Computations in the gas-phase reveal that the reactants “ion-molecule” complexes are 40.1 kcal/mol and 38.0 kcal/mol more stable than the isolated reactants for the ester and amide hydrolysis reactions, respectively. The activation barrier for the concerted ester hydrolysis reaction was computed to be 2.0 kcal/mol. The rate-limiting step for the stepwise amide hydrolysis reaction involves the addition of hydroxide ion, which leads to a barrier of 6.1 kcal/mol. This barrier is ~4 kcal/mol higher than that of the ester hydrolysis reaction, consistent with the experimental difference of ~6 kcal/mol. The latter was estimated using transition-state theory and the reported k_{cat} values for the antibody-catalyzed reactions at high pH where 15.5 kcal/mol ($k_{\text{cat}} \approx 1500 \text{ min}^{-1}$) [13] and 21.4 kcal/mol ($k_{\text{cat}} = 0.08 \text{ min}^{-1}$) [12] are the activation barriers for the hydrolyses of the actual ester and amide substrates of 43C9, respectively. As noted in the *Introduction*, the reported k_{cat} for

the ester hydrolysis reaction is not a measured value, but an estimate based on the pH-rate profile. A difference of ~ 7 kcal/mol is estimated from the measured k_{uncat} values for the reactions in solution at high pH where 21.6 kcal/mol ($k_{\text{uncat}} = 0.056 \text{ min}^{-1}$) [13] and 28.7 kcal/mol ($k_{\text{uncat}} = 3.4 \times 10^{-7} \text{ min}^{-1}$) [12] are the activation barriers for the hydrolyses of the ester and amide substrates, respectively. These estimated activation barriers are much greater than those computed for the model hydrolysis reactions in the gas-phase.

Environmental effects on the reaction barriers were explored using single point calculations at the B3LYP/6-31+G* level with the COSMO continuum solvent model. The proper choice of the dielectric constant to represent the protein environment is not obvious. We have chosen a dielectric constant of 4, which accounts for the response of the dipolar groups of the polypeptide backbone [23], but higher values may well be appropriate for regions of the protein surface, especially near charged residues [40]. A dielectric constant of 78.4 was used to represent an aqueous solution environment. With the incorporation of a polarizable environment, the reactants complex is raised higher in energy than the isolated reactants thus making the complex irrelevant. At a dielectric constant of 4, the barriers for the model ester and amide hydrolysis reactions increase to 4.4 kcal/mol and 7.1 kcal/mol, respectively. At a dielectric constant of 78.4, the barriers for the ester and amide hydrolysis reactions increase to 18.7 kcal/mol and 19.3 kcal/mol, respectively. Even though the relative barriers for the ester and amide hydrolysis reactions in solution ($\epsilon = 78.4$) were not reproduced, the computed barriers have been shown to be sensitive to the polarizability of the surrounding environment, leading to values that are in range of the estimated experimental barriers. Such calculations

involving a continuum model have been appropriately criticized [41-43] and are considered to yield qualitative results.

Hydrolysis of 17E8 model substrates. Figure 3 shows the optimized geometries at the B3LYP/6-31+G* level for the reactants and transition states involved in the hydrolysis reactions of the model ester and amide substrates for 17E8. As is the case for the hydrolysis reactions of the model 43C9 substrates, reactants “ion-molecule” complexes were located for the reactions of the model 17E8 substrates. The hydrolysis of phenyl acetate was concerted with TS1 being the only relevant transition state. When electron correlation was not included at the HF/6-31+G* level, however, a second, elimination transition state (TS2) was located. At both the B3LYP/6-31+G* and HF/6-31+G* levels of theory, the hydrolysis of phenyl acetanilide is stepwise, proceeding through TS1 and TS2. The angles of attack by hydroxide ion on phenyl acetate and phenyl acetanilide are similar to those for *p*-nitrophenyl acetate and *p*-nitrophenyl acetanilide reported above. The structures shown in Figure 3 are similar to those optimized at the HF/6-31+G* level except for the distances between the oxygen of the hydroxide ion and the carbon of the substrate carbonyl that are 0.4 Å shorter. Geometries at the HF/6-31+G* level for the ester hydrolysis reaction are the same as those reported by Tantillo and Houk at the same level of theory [39].

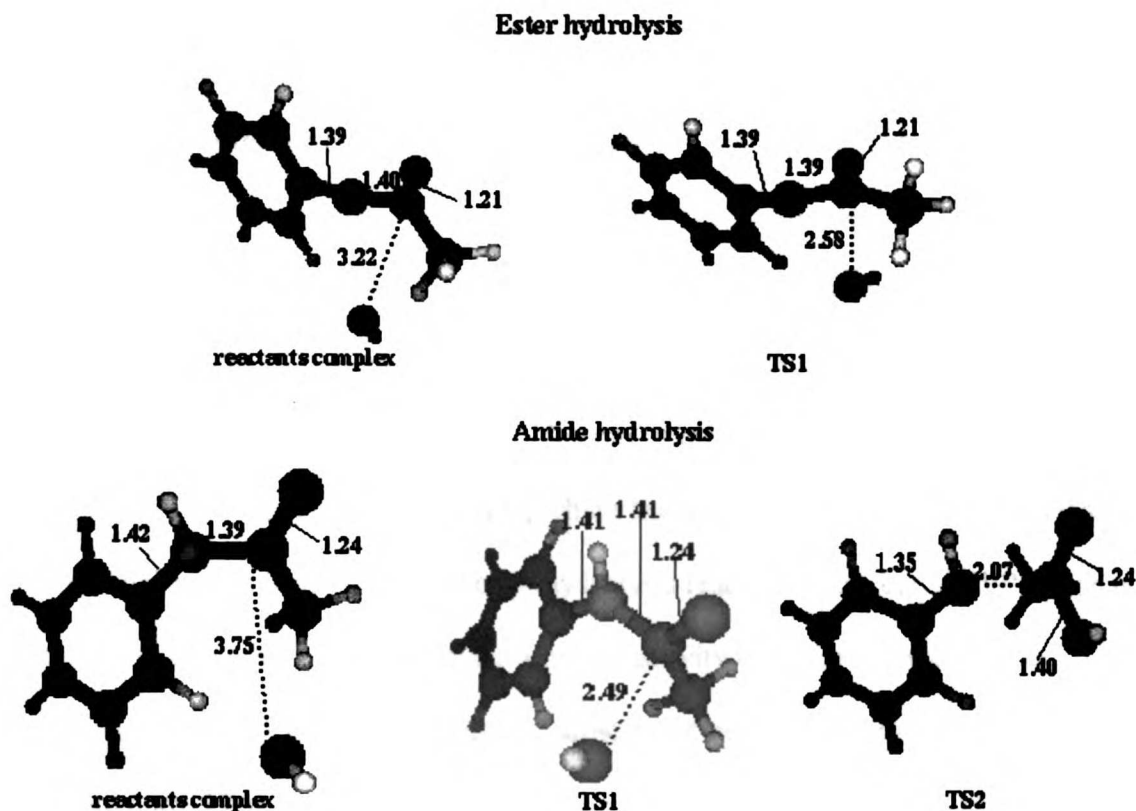


Figure 3. B3LYP/6-31+G* geometries of reactants and transition states for hydroxide attack on model ester and amide substrates for antibody 17E8. Bond distances shown are in Å.

Table 2. Relative energies (kcal/mol) for hydroxide attack on model ester and amide substrates for 17E8 with rate-limiting barriers underlined. Gas-phase energies include zero point energy corrections calculated at the B3LYP/6-31+G* level. The relative energy of TS2 is not included for calculations of the barriers for ester hydrolysis since the reaction is concerted. See text for definitions of the reactants and transition states.

		reactants			
	level of theory ^a	isolated	complex	TS1	TS2
ester hydrolysis	B3LYP/6-31+G*	25.0	0.0	<u>2.8</u>	--
	B3LYP/6-31+G*/COSMO($\epsilon=4$)	0.0	6.5	<u>9.5</u>	--
	B3LYP/6-31+G*/COSMO($\epsilon=78.4$)	0.0	15.3	<u>18.4</u>	--
amide hydrolysis	B3LYP/6-31+G*	24.9	0.0	<u>7.8</u>	-1.0
	B3LYP/6-31+G*/COSMO($\epsilon=4$)	0.0	6.4	15.9	<u>20.9</u>
	B3LYP/6-31+G*/COSMO($\epsilon=78.4$)	0.0	15.6	25.9	<u>37.0</u>

^aGeometries optimized at the B3LYP/6-31+G* level.

Relative energies of the reactants and transition states involved in the hydrolysis reactions of model 17E8 substrates are shown in Table 2. Computations in the gas-phase reveal that the reactants complexes are 25.0 kcal/mol and 24.9 kcal/mol more stable than the isolated reactants for the ester and amide hydrolysis reactions, respectively. The activation barrier for the concerted ester hydrolysis reaction was computed to be 2.8 kcal/mol, which is ~ 1 kcal/mol higher than that for the hydrolysis of the 43C9 model ester substrate (*p*-nitrophenyl acetate). The rate-limiting step for the amide hydrolysis reaction is the addition of hydroxide ion, which leads to a barrier of 7.8 kcal/mol that is ~ 5 kcal/mol higher than that for the ester hydrolysis reaction. This barrier difference is not as great as might be expected since amidase activity by 17E8 is undetectable.

However, COSMO calculations dramatically increase this difference to ~ 11 kcal/mol and ~ 19 kcal/mol when dielectric constants of 4 and 78.4 are used, respectively. These differences are much greater than those found for the model reactions of 43C9, consistent with the fact that only esterase activity is observed for 17E8 while 43C9 has both amidase and esterase activity. These significantly greater differences might be explained by the fact that aniline is a much poorer leaving group than *p*-nitrophenyl aniline due to the absence of the electron-withdrawing nitro group (the pK_a values of aniline and *p*-nitroaniline are 27 and 18.5, respectively) while the leaving group abilities of phenol and *p*-nitrophenol are more similar (pK_a values of 10 and 7, respectively). As is the case for the model reactions for 43C9, incorporation of a polarizable environment makes the reactants complex irrelevant. Interestingly, the rate-limiting step for the amide hydrolysis reaction changes from the step involving TS1 to that involving TS2 when dielectric constant values of 4 and 78.4 are used. Computations with a dielectric constant

of 4 increase the activation barriers for ester and amide hydrolysis to 9.5 kcal/mol and 20.9 kcal/mol, respectively, while computations with a dielectric constant of 78.4 increase the barriers even more to 18.4 kcal/mol and 37.0 kcal/mol, respectively. The values for the ester hydrolysis reaction are in range of the estimated experimental activation barriers of 16.7 kcal/mol ($k_{\text{cat}} = 223 \text{ min}^{-1}$) and 22.0 kcal/mol ($k_{\text{uncat}} = 0.027 \text{ min}^{-1}$) for the antibody-catalyzed and base-catalyzed reactions, respectively [18].

MD simulations:

A model of the 43C9-hapten complex was constructed using the same general strategy by Thayer *et al.* [2] in which the *p*-nitrophenyl moiety of the hapten was superimposed onto the *p*-nitrophenol product in the crystal structure of the antibody-product complex and the phosphonate oxygens of the haptens were positioned to replace two crystallographic waters. This strategy is plausible provided that the structures of the product-bound and hapten-bound forms of the antibody are similar as is the case for the esterase antibody D2.3 [8, 9]. Use of the *p*-nitrophenyl product as an ‘anchor’ for the hapten is supported by the observation that the *p*-nitrophenyl moiety of the hapten stayed in the binding pocket of the product with a maximum RMS fluctuation of 0.2 Å during the MD simulation of the 43C9-hapten complex. The MD simulation of the 43C9-hapten complex was stable with an RMS deviation of 0.6 Å from the minimized starting coordinates for the mobile, or ‘belly’ (see *Methods*), residues. A stable simulation of the 17E8-hapten complex was also obtained with an RMS deviation of 0.6 to 0.8 Å from the minimized crystal structure for the belly residues. RMS fluctuations of the belly residues in the 43C9- and 17E8-hapten complexes were 0.4 Å and 0.5 Å, respectively.

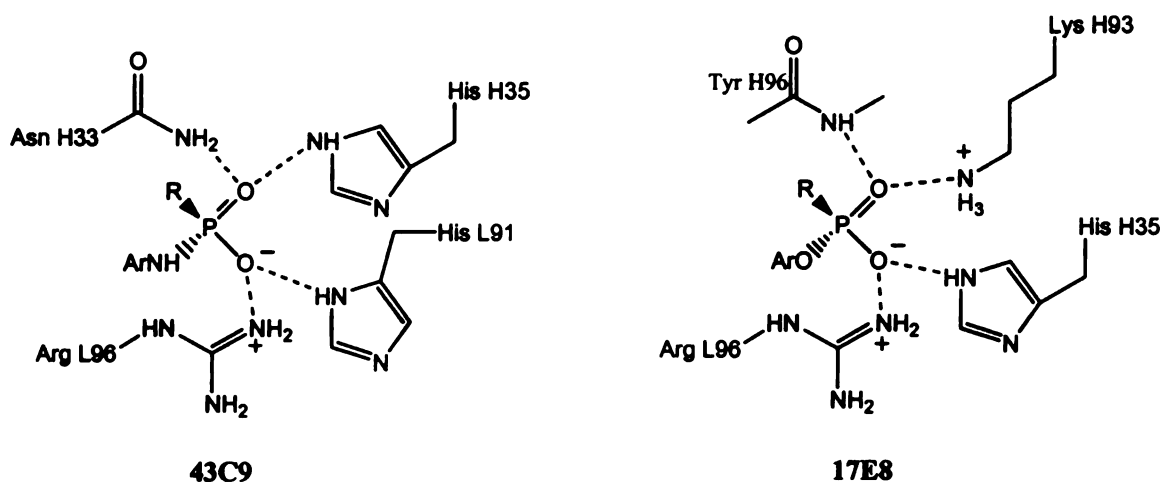


Figure 4. Schematic depiction of hydrogen bonds formed with the phosphonate moiety of the bound haptens in 43C9 and 17E8.

Throughout the MD simulations of the hapten complexes, it is apparent that antibodies 43C9 and 17E8 utilize similar general strategies for hapten recognition. Each antibody has a deep, hydrophobic pocket that is occupied by the aryl moiety of the hapten and forms hydrogen bonds with the phosphonate moiety of the hapten at the mouth of the binding site. Figure 4 shows the four hydrogen bonds that are formed by the active site residues of each antibody with the phosphonate moiety of the hapten; two of these residues, Arg L96 and His H35, are present in both antibodies. Like other esterase antibodies such as CNJ206 [3] and 48G7 [6, 7], 17E8 recognizes the phosphonate oxygens predominately through heavy chain residues – His H35, Lys H93, Tyr H97 – with the exception of Arg L96. Phosphonate recognition by 43C9, on the other hand, involves an equal number of heavy and light chain residues; the hydrogen bond formed between the pro-R oxygen of the hapten and the light chain residue, His L91, is unique among known esterase antibodies. In the crystal structure of the 17E8-hapten complex, no hydrogen bond is apparent between His H35 and the hapten. However, during the

equilibration stage of the MD simulation of the complex, the His H35 imidazole side chain rotated to its 'flipped' orientation, resulting in hydrogen bond formations with Trp H47 as well as with the pro-R oxygen of the hapten. In a recent study involving the docking of transition states to various esterase antibodies, it was found that this rotamer of His H35 in 17E8 produced a combining site with improved stabilization of the transition state [44]. Furthermore, the hydrogen bond between the side chain of the flipped rotamer of His H35 and Trp H47 is conserved among all other esterase antibody structures, including that of 43C9 [3, 6-9, 45]. Due to the negative charge of the hapten phosphonate moiety, one might expect the pK_a values of the histidine residues in the vicinity to be raised. To assess the possibility of charged histidine residues in the active site, the pK_a 's of these residues were computed using the continuum electrostatics approach described in *Methods*. The computed pK_a 's for His H35 in 17E8, His H35 in 43C9, and His L91 in 43C9 were found to be lowered from the intrinsic pK_a of 6.5 with values of 4.0, 3.5, and 4.8, respectively, while the pK_a of a distant, exposed histidine residue in 17E8, His L49, was computed to be close to the intrinsic pK_a with a value of 6.2. The histidine residues near the hapten phosphonate moiety were therefore assigned to be neutral.

Models of the ground-state and transition-state complexes for each hydrolysis reaction were constructed by modifying the hapten in the antibody-hapten complex to the substrate and tetrahedral intermediate, respectively. The hydroxide ion was not included in the ground-state complex since the ion is likely present at only a catalytic and not stoichiometric amount during the antibody-catalyzed reaction. Tetrahedral intermediates were used to approximate the transition states due to their similar geometries and charge

distributions. Stable trajectories were obtained for each ground-state and transition-state complex with RMS fluctuations ranging from 0.4 to 0.6 Å for the belly residues.

Dynamics of the ground-state complexes in our MD simulations of the complexes suggest that the direct hydroxide attack mechanism is plausible for 43C9 and 17E8. In each ground-state complex simulation for 43C9, one to two water molecules are found within 4 Å of the substrate carbonyl carbon in 41% of the 500 sampled configurations from each simulation. No water molecules were found within 4 Å of the substrate carbonyl carbon in the ground-state complex simulations of 17E8. However, the absence of water molecules in this vicinity does not rule out direct hydroxide attack since extremely long simulations would be required to allow for diffusion of water molecules into the binding site. We also note that, unlike the starting structures of the 43C9 simulations, the starting structures of 17E8 had no resolved crystallographic water molecules. Finally, in each ground-state simulation of 17E8, 99% of the configurations contain 1-4 water molecules in the first solvation shell (within 3.4 Å) of His H35, which can abstract a proton from a water molecule to form a hydroxide ion that may then diffuse near the substrate carbonyl.

Interestingly, nucleophilic attack by His L91 (protonated at the ϵ -nitrogen position) of 43C9 on its ester and amide substrates is also plausible from examination of the corresponding ground-state complex simulations. In the simulation of the ground-state for amide hydrolysis, the distance between the nucleophilic nitrogen of His L91 (protonated at the ϵ -nitrogen) and the carbonyl carbon of the substrate ranges between 3.7 to 5.5 Å. In the simulation of the ground-state for ester hydrolysis, the distance ranges between 3.9 and 4.6 Å.

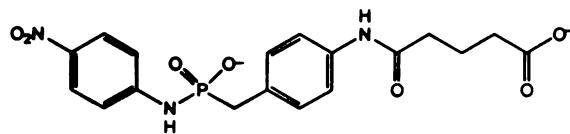
For each transition-state complex, both enantiomers (from *re* and *si* attack by hydroxide ion) of the tetrahedral intermediate were considered. As shown in Table 4, the intermediates resulting from *si* attack lead to more favorable computed transition-state complex formations for the hydrolysis reactions of 43C9 while those resulting from *re* attack lead to more favorable transition-state complex formations for the hydrolysis reactions of 17E8. These results were obtained *via* the MM-PBSA approach described in the section below. In each of the MD simulations of the more favorable transition-state complex formations, His H35 (protonated at the δ -nitrogen) forms a hydrogen bond with the proton of the attacking hydroxide ion while the other active-site residues involved in phosphonate recognition form hydrogen bonds with the analogous oxygens of the tetrahedral intermediate. The presence of a hydrogen bond acceptor to stabilize the transition state is not 'pre-programmed' by anionic phosphonate haptens, but has been predicted by Tantillo and Houk to be characteristic of the best catalysts in their recent transition state docking study involving various esterase antibodies [44]. Among the findings in their study was that *re* attack led to a more favorable transition-state complex for the ester hydrolysis reaction of 17E8, consistent with our results.

Free energy calculations:

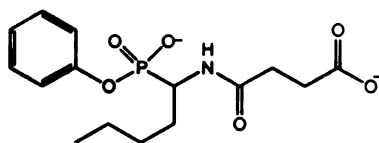
To compute the relative binding energies for the hapten complexes of 43C9 and 17E8, we used the MM-PBSA free energy approach [33], which has given accurate relative binding free energies in a number of studies [46], including one in which the relative binding free energies of the germ line and mature hapten complexes of esterase antibody 48G7 were computed [35]. This approach calculates free energies for solute

structures sampled from an MD simulation in explicit solvent using a combination of molecular mechanics and continuum solvent models. For each antibody-ligand complex, we used the bound antibody and ligand configurations from the simulation of the complex as representative structures of the unbound antibody and ligand. Thus, it was assumed that only small conformational changes occur upon formation of the antibody-ligand complexes, which is apparent for several catalytic antibodies that have preformed binding sites [6-9].

Table 3. Average energetic contributions to hapten complex formations of antibodies 43C9 and 17E8 in kcal/mol with standard errors of the mean in parentheses^a. See *Methods* for definitions of the energetic contributions.



Hapten for 43C9



Hapten for 17E8

	43C9	17E8
$\langle \Delta E_{\text{vdW}} \rangle$	-52.2 (0.3)	-41.4 (0.5)
$\langle \Delta E_{\text{es}} \rangle$	-202.8 (0.4)	-207.8 (0.7)
$\langle \Delta G_{\text{PB}} \rangle$	215.8 (0.4)	216.6 (0.7)
$\langle \Delta G_{\text{SA}} \rangle$	-5.5 (0.0)	-4.9 (0.0)
$-T\Delta S^{\text{b}}$	20	20
$\langle \Delta G_{\text{bind}} \rangle$	-24.7 (0.3)	-17.5 (0.4)
$\Delta G_{\text{bind}}(\text{exp})$	-12.5	-8.6

^aAverage energies from 50 equally spaced configurations from the 101-600 ps interval of the MD simulation. ^bEstimated value based on published results for similarly sized ligands.

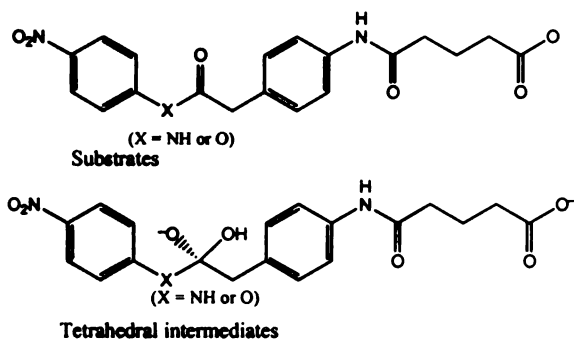
Table 4. Computed average binding free energies (kcal/mol) for transition-state (TS) complexes from *re* and *si* attack by hydroxide ion with standard errors of the mean in parentheses^a

	<i>re</i> attack	<i>si</i> attack
43C9 ester TS complex	16.3 (1.0)	-1.0 (0.8)
43C9 amide TS complex	2.2 (0.5)	-2.2 (0.6)
17E8 ester TS complex	-5.2 (0.5)	0.9 (0.7)
17E8 amide TS complex	-4.3 (0.6)	3.9 (0.7)

^aAverage energies from 50 equally spaced configurations from the 101-600 ps interval of the MD simulation including estimated solute entropic contribution of +20 kcal/mol.

Energetic contributions to the hapten binding free energies of 43C9 and 17E8 are shown in Table 3. Even though the computed binding free energies for the hapten complexes of 43C9 (-24.7 kcal/mol) and 17E8 (-17.5 kcal/mol) are much more favorable than the experimental values (-12.5 and -8.6 kcal/mol, respectively), the tighter binding of the 43C9 complex relative to that of the 17E8 complex has been qualitatively reproduced with reasonable quantitative agreement with the measured difference in binding free energies. The more favorable van der Waals contribution of -52.2 kcal/mol for the 43C9 complex compared to -41.4 kcal/mol for the 17E8 complex is consistent with the presence of an additional aryl moiety in the hapten of 43C9. Tighter binding in the 43C9 complex is due to more favorable nonpolar interactions with solvent ($\langle \Delta G_{SA} \rangle$) and within the solute ($\langle \Delta E_{vdW} \rangle$), compensating for the less favorable balance of electrostatic interactions ($\langle \Delta E_{es} \rangle + \langle \Delta G_{PB} \rangle$).

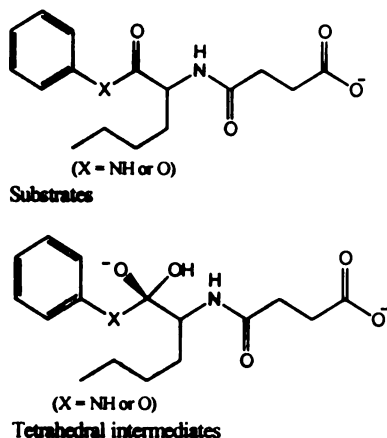
Table 5. Average energetic contributions to ground-state (GS) and transition-state (TS) complex formations of antibody 43C9 in kcal/mol with standard errors of the mean in parentheses^a. See *Methods* for definitions of the energetic contributions.



	ester hydrolysis		amide hydrolysis	
	GS complex	TS complex	GS complex	TS complex
$\langle \Delta E_{\text{vdW}} \rangle$	-44.5 (0.4)	-47.3 (0.5)	-45.6 (0.2)	-49.7 (0.5)
$\langle \Delta E_{\text{es}} \rangle$	-171.7 (1.4)	-391.6 (2.1)	-183.5 (1.2)	-381.5 (1.9)
$\langle \Delta G_{\text{PB}} \rangle$	205.6 (1.6)	423.3 (1.7)	217.4 (1.2)	414.5 (1.6)
$\langle \Delta G_{\text{SA}} \rangle$	-5.1 (0.0)	-5.4 (0.0)	-5.0 (0.0)	-5.4 (0.0)
$-T\Delta S^{\text{b}}$	20	20	20	20
$\langle \Delta G_{\text{bind}} \rangle$	4.3 (0.6)	-1.0 (0.8)	3.3 (0.4)	-2.2 (0.6)
$\Delta G_{\text{bind}}(\text{exp})$	-5.8	-11.9	-4.4	-11.7

^aAverage energies from 50 equally spaced configurations from the 101-600 ps interval of the MD simulation. ^bEstimated value based on published results for similarly sized ligands.

Table 6. Average energetic contributions to ground-state (GS) and transition-state (TS) complex formations of antibody 17E8 in kcal/mol with standard errors of the mean in parentheses^a. See Methods for definitions of the energetic contributions.



	ester hydrolysis		amide hydrolysis	
	GS complex	TS complex	GS complex	TS complex
$\langle \Delta E_{vdW} \rangle$	-43.5 (0.3)	-45.2 (0.4)	-43.6 (0.4)	-43.1 (0.4)
$\langle \Delta E_{es} \rangle$	-193.2 (1.3)	-402.9 (1.5)	-198.8 (1.1)	-394.4 (2.0)
$\langle \Delta G_{PB} \rangle$	223.0 (1.1)	427.8 (1.3)	229.2 (1.0)	418.2 (1.7)
$\langle \Delta G_{SA} \rangle$	-4.9 (0.0)	-5.0 (0.0)	-5.0 (0.0)	-5.0 (0.0)
$-T\Delta S^b$	20	20	20	20
$\langle \Delta G_{bind} \rangle$	1.4 (0.5)	-5.2 (0.5)	1.9 (0.5)	-4.3 (0.6)
$\Delta G_{bind}(exp)$	-4.6	-9.9	--	--

^aAverage energies from 50 equally spaced configurations from the 101-600 ps interval of the MD simulation. ^bEstimated value based on published results for similarly sized ligands.

The MM-PBSA free approach can also be used to estimate the relative ground-state and transition-state stabilizations of each hydrolysis reaction by computing the binding free energies of the ground-state and transition-state complexes simulated above. Only the transition-state complex simulations consisting of the more favorable enantiomers of the tetrahedral intermediates were used to estimate the transition-state stabilizations. Assuming that the antibody-catalyzed reaction has the same mechanism as

the non-catalyzed reaction, the ground-state stabilization can be estimated from the measured K_M for the catalyzed reaction and the transition-state stabilization, from incorporating the activation barriers for the catalyzed and uncatalyzed reactions (as estimated from the k_{cat} and k_{uncat} values, respectively) into the reaction profile. The estimated ground-state stabilizations for the ester (-5.8 kcal/mol) and amide (-4.4 kcal/mol) hydrolysis reactions of 43C9 are within 1 kcal/mol of each other as are the transition-state stabilizations (-11.9 kcal/mol and -11.7 kcal/mol for ester and amide hydrolysis, respectively) [12, 13]. As shown in Table 5, our free energy calculations reproduce these similarities as well as the relative stabilizations of the transition-state to the ground-state for each hydrolysis reaction of 43C9. The ground-state and transition-state stabilizations are also computed to be similar for the ester and amide hydrolysis reactions of 17E8 (see Table 6). Since no kinetic parameters were measurable for the amide hydrolysis reaction of 17E8, there is no experimental validation for the latter computed results. Nonetheless, for the ester hydrolysis reaction of 17E8, the computed relative stabilizations of the transition-state to the ground-state of -6.1 kcal/mol is in good agreement with the estimated experimental value of -5.3 kcal/mol (-9.9 kcal/mol and -4.6 kcal/mol for transition-state and ground-state stabilization, respectively) [18].

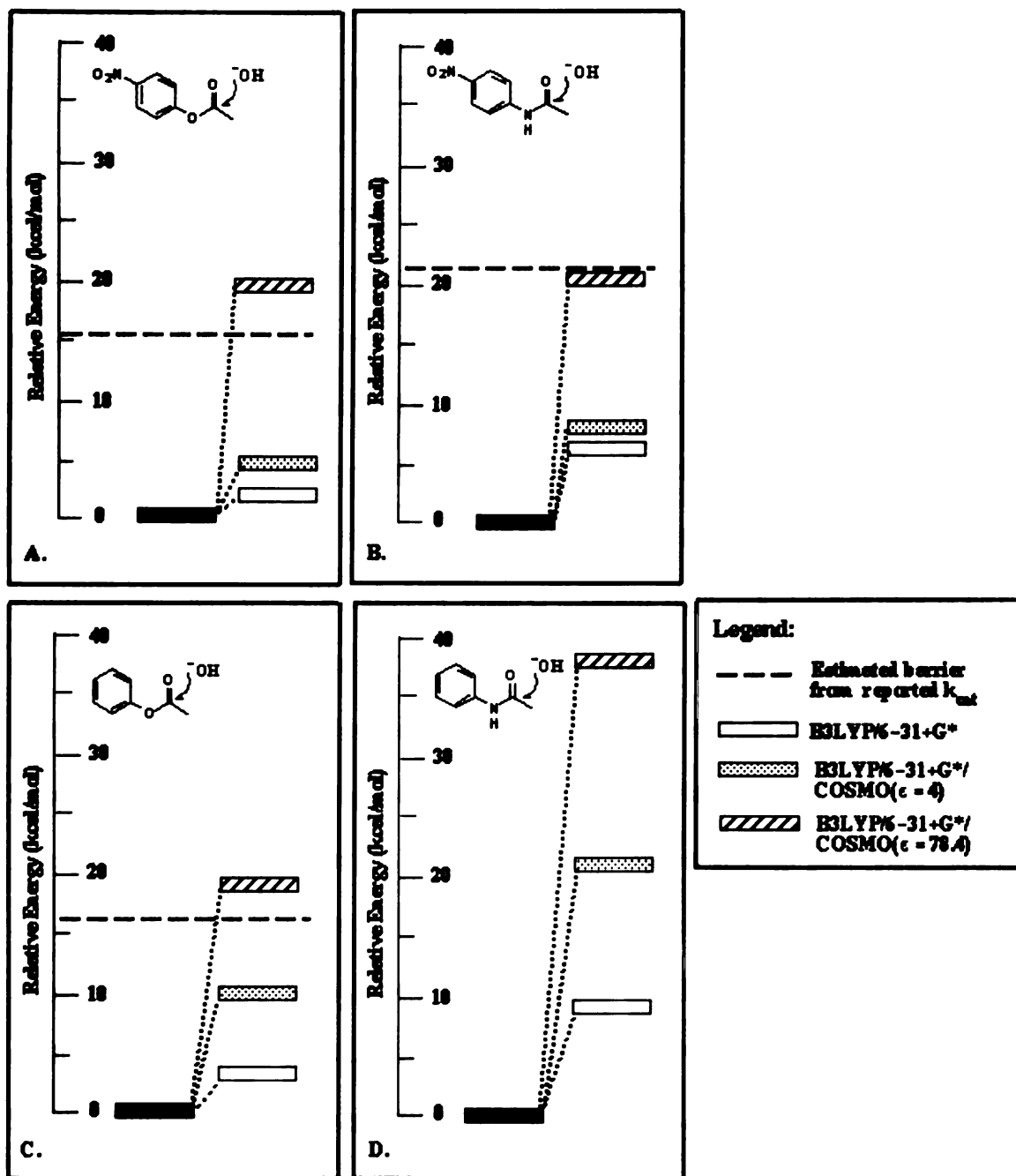


Figure 5. Computed activation barriers for hydrolysis *via* direct hydroxide attack of model ester (A) and amide substrates (B) for 43C9, and model ester (C) and amide (D) substrates for 17E8. No experimental barrier for the reaction in (D) is shown since amidase activity was not detectable for 17E8.

Discussion and Conclusions:

The computed activation barriers for the model reactions considered in this study are summarized in the energy profile diagrams of Figure 5. As discussed earlier in *Results*, the incorporation of environmental effects with the COSMO continuum model raises the gas-phase activation barriers for hydroxide attack on the model substrates to values in range of estimated, experimental activation barriers. Combined with our molecular dynamics simulations of the ground states that support the possibility of hydroxide attack and our free energy calculations that indicate no differential stabilization of the transition states (or ground states) for the potential ester and amide hydrolysis reactions of 43C9 and 17E8, these QM results suggest that direct hydroxide attack is a plausible mechanism of catalysis for both antibodies. In the case of 43C9, it is likely that His L91 acts as a general base, abstracting a proton from a nearby water to form hydroxide ion. This role is consistent with the decreased ester hydrolysis activity, but unchanged substrate affinity of the His L91→Gln mutant of 43C9 [16]. Assuming that the *p*-chlorophenyl ester substrate used in the kinetics experiments with the mutant is a reasonable substitute for the *p*-nitrophenyl ester substrate (see *Introduction*), these results for the mutant support our proposed mechanism. Likewise, the catalytic residue in 17E8, His H35, may act as a general base to produce hydroxide ion. The absence of detectable amidase activity by 17E8 might be, in part, rationalized by the fact that its substrates are less activated than the *p*-nitrophenyl substrates of 43C9 (see *Results* under *QM calculations*).

The possibility of histidine attack by 43C9 cannot, however, be ruled out based on the detection of the acyl-antibody intermediate by mass spectrometry [17] and our MD

simulations of the antibody-substrate complexes in which His L91 is positioned sufficiently close for nucleophilic attack on the substrates. Nonetheless, our results suggest that this more complex mechanism, which could not result directly from hapten recognition, provides no catalytic advantage over the more simple hydroxide attack mechanism that would result directly from hapten recognition. Detailed mechanistic inferences should be tentative for 43C9 at this time since a high-resolution structure of the 43C9-hapten complex is not known. Furthermore, our results are only semi-quantitative due to the limitations discussed below.

The first limitation is the use of a continuum model to incorporate environmental effects on the activation barriers in the QM calculations. Results from such calculations have indicated that the activation barriers for the model reactions are sensitive to the polarizability of the environment, increasing with the value of the dielectric constant. Quantitative interpretation of these results is limited. However, values obtained with a dielectric constant of 4 and 78.4 are in range of the barriers estimated from experimental rate constants. More accurate computations of the barriers would involve explicit incorporation of the protein environment as has been done in various hybrid QM and molecular mechanics approaches [49-51]. Given the large computational demands of such approaches, convergence problems that often emerge from their use, and the uncertainty of the 43C9-hapten complex model, these approaches were not applied in this study.

Other limitations exist with the free energy calculations used to estimate the ground-state and transition-state stabilizations for each antibody-catalyzed hydrolysis reaction. Errors may exist in the structural models of the ground-state and transition-state

complexes since these models were based on the corresponding hapten complex structures. In addition, the transition states of the reactions were approximated with the use of the corresponding tetrahedral intermediates. Uncertainties in the structures of these models may explain the fact that the computed binding free energies are not as favorable as those predicted from experimental data when the antibody-catalyzed mechanism is assumed to be the same as in solution. Furthermore, accurate calculation of solute entropic contributions may lead to more reasonable absolute binding free energies. However, such accuracy is not obtainable with current methods such as quasi-harmonic and normal mode analysis. Due to the computational demands of these methods, the solute entropic contribution was not computed in this study and estimated as described in *Methods*. Despite these limitations, the MD simulations of the complexes are stable with low RMS fluctuations and the relative binding free energies for the ground-state and transition-state complexes are in good agreement with those predicted by experiment.

In conclusion, *ab initio* QM calculations combined with MD simulations and free energy calculations support the proposed mechanism of direct hydroxide attack on the ester and amide substrates for antibodies 43C9 and 17E8. The computed activation barriers at the B3LYP/6-31+G**/B3LYP/6-31+G* level of theory are in range of experimental values when environmental effects are incorporated with the COSMO continuum model. Dynamics of the ground-state complexes support the possibility of hydroxide attack on the substrates while those of the transition-state complexes revealed the formation of hydrogen bonds with the “transition state epitope” as programmed by the haptens that mimic the transition states for hydroxide attack. Simulations of each

transition-state complex also revealed that an active-site histidine residue adopted the tautomeric form that enabled the residue to act as an unprogrammed hydrogen bond acceptor to the oxygen of the attacking hydroxide ion. Finally, free energy calculations predict that the ground-state and transition-state stabilizations are similar, in agreement with the relative stabilizations estimated from experimental kinetic parameters. The proposed mechanism may be generalized to all antibodies that have been elicited by arylphosphonate or arylphosphamidate haptens. Thus, direct hydroxide attack, as programmed by their haptens, is a likely mechanism for antibodies 48G7, 29G11, and CNJ206 as well as 43C9 and 17E8.

Acknowledgements

The authors thank K. Houk and D. Tantillo for helpful suggestions on improving the manuscript. This work was supported in part by National Science Foundation and Burroughs-Wellcome Fellowships to L.T.C. and a National Institutes of Health (NIH) Grant (GM29072) to P.A.K. Graphics were provided by the Computer Graphics Laboratory, University of California, San Francisco (T. Ferrin, P.I., NIH P41 Grant RR-01081).

References:

1. MacBeath, G. and D. Hilvert, *Hydrolytic antibodies: variations on a theme*. Chem Biol, 1996. 3(6): p. 433-45.
2. Thayer, M.M., et al., *Structural basis for amide hydrolysis catalyzed by the 43C9 antibody*. J Mol Biol, 1999. 291(2): p. 329-45.
3. Charbonnier, J.B., et al., *Crystal structure of the complex of a catalytic antibody Fab fragment with a transition state analog: structural similarities in esterase-like catalytic antibodies*. Proc Natl Acad Sci U S A, 1995. 92(25): p. 11721-5.
4. Zhou, G.W., et al., *Crystal structure of a catalytic antibody with a serine protease active site*. Science, 1994. 265(5175): p. 1059-64.
5. Guo, J., et al., *Mechanistically different catalytic antibodies obtained from immunization with a single transition-state analog*. Proc Natl Acad Sci U S A, 1995. 92: p. 1694-1698.
6. Wedemayer, G.J., et al., *Structural insights into the evolution of an antibody combining site*. Science, 1997. 276(5319): p. 1665-9.
7. Wedemayer, G.J., et al., *Crystal structures of the free and liganded form of an esterolytic catalytic antibody*. J Mol Biol, 1997. 268(2): p. 390-400.
8. Charbonnier, J.B., et al., *Structural convergence in the active sites of a family of catalytic antibodies*. Science, 1997. 275: p. 1140-1142.
9. Gigant, B., et al., *X-ray structures of a hydrolytic antibody and of complexes elucidate catalytic pathway from substrate binding and transition state stabilization through water attack and product release*. Proc Natl Acad Sci U S A, 1997. 94: p. 7857-7861.
10. Hilvert, D., *Critical analysis of antibody catalysis*. Annu Rev Biochem, 2000. 69: p. 751-93.
11. Tantillo, D.J. and K.N. Houk, *Canonical binding arrays as molecular recognition elements in the immune system: tetrahedral anions and the ester hydrolysis transition state*. Chem Biol, 2001. 8(6): p. 535-45.
12. Janda, K.D., et al., *Induction of an antibody that catalyzes the hydrolysis of an amide bond*. Science, 1988. 241(4870): p. 1188-91.
13. Gibbs, R.A., et al., *Substituent Effects on an Antibody-Catalyzed Hydrolysis of Phenyl Esters - Further Evidence for an Acyl-Antibody Intermediate*. Journal of the American Chemical Society, 1992. 114(9): p. 3528-3534.

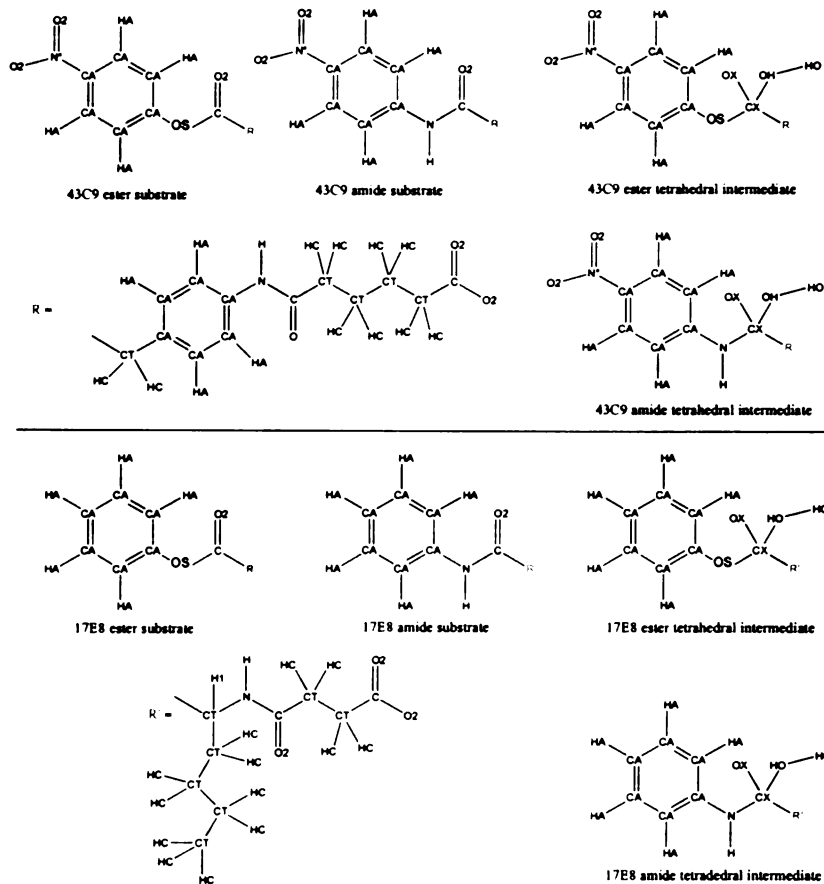
14. Stewart, J.D., et al., *Dissection of an antibody-catalyzed reaction*. Proc Natl Acad Sci U S A, 1994. **91**(16): p. 7404-9.
15. Benkovic, S.J., et al., *The enzymic nature of antibody catalysis: development of multistep kinetic processing*. Science, 1990. **250**(4984): p. 1135-9.
16. Stewart, J.D., et al., *Site-directed mutagenesis of a catalytic antibody: an arginine and a histidine residue play key roles*. Biochemistry, 1994. **33**(8): p. 1994-2003.
17. Krebs, J.F., et al., *Detection of a catalytic antibody species acylated at the active site by electrospray mass spectrometry*. Biochemistry, 1995. **34**(3): p. 720-3.
18. Guo, J.C., W. Huang, and T.S. Scanlan, *Kinetic and Mechanistic Characterization of an Efficient Hydrolytic Antibody - Evidence for the Formation of an Acyl Intermediate*. Journal of the American Chemical Society, 1994. **116**(14): p. 6062-6069.
19. Baca, M., et al., *Phage display of a catalytic antibody to optimize affinity for transition-state analog binding*. Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10063-8.
20. Srinivasan, J., et al., *Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices*. Journal of the American Chemical Society, 1998. **120**(37): p. 9401-9409.
21. Scott, A.P. and L. Radom, *Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Moller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors*. J. Phys. Chem., 1996. **100**(41): p. 16502-16513.
22. Barone, V., M. Cossi, and J. Tomasi, *Geometry optimization of molecular structures in solution by the polarizable continuum model*. Journal of Computational Chemistry, 1998. **19**(4): p. 404-417.
23. Gilson, M. and B. Honig, *The dielectric constant of a folded protein*. Biopolymers, 1986. **25**: p. 2097-2119.
24. Frisch, M.J., et al., *Gaussian 98*. revision A.6. 1998, Pittsburgh, PA: Gaussian, Inc.
25. Case, D.A., et al., *Amber 5.0*. 1997, San Francisco, CA: University of California.
26. Gibbs, R.A., et al., *Construction and characterization of a single-chain catalytic antibody*. Proc Natl Acad Sci U S A, 1991. **88**: p. 4001-4004.
27. Tanokura, M., M. Tasumi, and T. Miyazawa, *Biopolymers*, 1976. **15**: p. 393-401.

28. Bayly, C.I., et al., *A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model*. Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
29. Wang, J., Cieplak, P., and Kollman, P.A., *How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules?* J. Comput. Chem., 2000. **21**:1049-1074.
30. Jorgensen, W., et al., J. Chem. Phys., 1983. **79**: p. 926-935.
31. Berendsen, H., et al., J. Comput. Phys., 1984. **81**: p. 3684-3690.
32. Ryckaert, J., G. Ciccotti, and H. Berendsen, J. Comput. Phys., 1977. **23**: p. 327-341.
33. Srinivasan, J., et al., *Continuum solvent studies of the stability of RNA hairpin loops and helices*. J Biomol Struct Dyn, 1998. **16**(3): p. 671-82.
34. Sharp, K.A. and B. Honig, *Electrostatic interactions in macromolecules: theory and applications*. Annu Rev Biophys Biophys Chem, 1990. **19**: p. 301-32.
35. Chong, L.T., et al., *Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7*. Proc Natl Acad Sci U S A, 1999. **96**(25): p. 14330-5.
36. Yang, A.S., et al., *On the Calculation of Pk(a)S in Proteins*. Proteins-Structure Function and Genetics, 1993. **15**(3): p. 252-265.
37. Antosiewicz, J., J.A. McCammon, and M.K. Gilson, *The Determinants of Pk(a)S in Proteins*. Biochemistry, 1996. **35**(24): p. 7819-7833.
38. Hengge, A.C. and R.A. Hess, J. Am. Chem. Soc., 1994. **116**: p. 11256-11263.
39. Tantillo, D.J. and K.N. Houk, *Fidelity of Hapten Design: How Analogous Are Phosphonate Haptens to the Transition States for Alkaline Hydrolyses of Aryl Esters?* J. Org. Chem., 1999. **64**: p. 3066-3076.
40. Sheinerman, F., R. Norel, and B. Honig, *Electrostatic aspects of protein-protein interactions*. Current Opinion in Structural Biology, 2000. **10**: p. 153-159.
41. Warshel, A., et al., *Remarkable rate enhancement of orotidine 5'-monophosphate decarboxylase is due to transition-state stabilization rather than to ground-state destabilization*. Biochemistry, 2000. **39**(48): p. 14728-38.

42. Warshel, A., *Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites*. J Biol Chem, 1998. **273**(42): p. 27035-8.
43. Warshel, A. and J. Florian, *Computer simulations of enzyme catalysis: finding out what has been optimized by evolution*. Proc Natl Acad Sci U S A, 1998. **95**(11): p. 5950-5.
44. Tantillo, D.J. and K.N. Houk, *Transition State Docking: A Probe for Noncovalent Catalysis in Biological Systems. Application to Antibody-Catalyzed Ester Hydrolysis*. Journal of Computational Chemistry, 2002. **23**(1): p. 84-95.
45. Buchbinder, J.L., et al., *A Comparison of the Crystallographic Structures of Two Catalytic Antibodies with Esterase Activity*. J. Mol. Biol., 1998. **282**: p. 1033-1041.
46. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models*. Accounts of Chemical Research, 2000. **33**(12): p. 889-897.
47. Jencks, W.P., *Catalysis in Chemistry and Enzymology*. 1969, New York: McGraw-Hill.
48. Kirsch, J.F. and W.P. Jencks, *Nonlinear Structure-Reactivity Correlations. The Imidazole-Catalyzed Hydrolysis of Esters*. J. Am. Chem. Soc., 1964. **86**: p. 837-846.
49. Kuhn, B. and P.A. Kollman, *QM-FE and molecular dynamics calculations on catechol O-methyltransferase: Free energy of activation in the enzyme and in aqueous solution and regioselectivity of the enzyme-catalyzed reaction*. Journal of the American Chemical Society, 2000. **122**(11): p. 2586-2596.
50. Stanton, R.V., et al., *Combined ab initio and free energy calculations to study reactions in enzymes and solution: Amide hydrolysis in trypsin and aqueous solution*. Journal of the American Chemical Society, 1998. **120**(14): p. 3448-3457.
51. Warshel, A., *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, ed. Wiley. 1991, New York.

Appendix:

Table A. Additional force field parameters used for ligands



MASS	m (amu)		
CX	12.01		
OX	16.00		
BOND	K_r (kcal mol ⁻¹ Å ⁻²)	r_{eq} (Å)	
CA-OS	300.0	1.400	*
CA-N	490.0	1.335	=C-N
N*-O2	750.0	1.210	**
N-CX	337.0	1.530	=HF/6-31+G* geometry, K_r for N*-CT
CX-OX	320.0	1.260	=HF/6-31+G* geometry, K_r for CT-OS
CX-OH	320.0	1.420	=HF/6-31+G* geometry, K_r for CT-OS
CX-OS	320.0	1.530	=HF/6-31+G* geometry, K_r for CT-OS
CX-CT	310.0	1.526	=CT-CT

ANGLE	K_{θ} (kcal mol ⁻¹ rad ⁻²)	θ_{eq} (deg)			
CA-CA-OS	70.0	120.00	*		
CA-N-C	50.0	123.20	=CA-N2-CT		
CA-CA-N	70.0	120.00	=CA-CA-OS		
CA-CB-N*	70.0	120.00	=CT-C-O2		
CA-CB-CA	70.0	120.00	=CT-C-O2		
CB-N*-O2	70.0	120.00	=CT-C-O2		
O2-N*-O2	140.0	119.90	=O2-P-O2		
CA-N-H	35.0	120.00	=CA-N2-H		
CA-OS-C	60.0	117.00	=C-OS-CT		
C-CT-CA	63.0	110.10	=C-CT-N		
N-CX-OH	100.0	108.23	=O2-P-OS		
N-CX-OX	100.0	108.23	=O2-P-OS		
N-CX-CT	80.0	109.70	=N-CT-CT		
OH-CX-OX	140.0	119.90	=O2-P-O2		
CA-N-CX	50.0	123.20	=CA-N2-CT		
CX-N-H	50.0	118.04	=CT-N-H		
CX-OH-HO	55.0	108.50	=CT-OH-HO		
CX-CT-HC	50.0	109.50	=CT-CT-HC		
CX-CT-CA	63.0	114.00	=CT-CT-CA		
OH-CX-CT	50.0	109.50	=OH-CT-CT		
OX-CX-CT	50.0	109.50	=OH-CT-CT		
CA-OS-CX	60.0	117.00	=CM-OS-CT		
OS-CX-OH	100.0	108.23	=OS-P-OS		
OS-CX-OX	100.0	108.23	=OS-P-OS		
OS-CX-CT	50.0	109.50	=OS-CT-CT		
CT-CT-CX	40.0	109.50	=CT-CT-CT		
N-CT-CX	80.0	109.50	=N-CT-CT		
CX-CT-HI	50.0	109.50	=CT-CT-HI		
DIHEDRAL	idv [†]	$V_n/2$ (kcal/mol ⁻¹)	γ (deg)	n	
X-CA-OS-X	2	2.60	180.0	-2.	*
X-CA-OS-X	2	0.325	180.0	4.	*
X-CA-N-X	4	3.45	180.0	2.	
X-N-CX-X	6	0.00	0.0	2.	=X-N-CT-X
X-CX-OH-X	3	0.50	0.0	3.	=X-CT-OH-X
X-CX-CT-X	9	1.40	0.0	3.	=X-CT-CT-X
X-OS-CX-X	3	1.15	0.0	3.	=X-OS-CT-X
NONBONDED	σ (Å)	ϵ (kcal/mol)			
CX	1.9080	0.1094	=CT		
OX	1.7210	0.2104	=OH		

* Spellmeyer, D.C.; Grootenhuis, P.D.J.; Miller, M.D.; Kuyper, L.F.; and Kollman, P.A. *J. Phys. Chem.* **1990**, *94*, 4483-4491

** Chong, L.T.; Duan, Y.; Wang, L.; Massova, I.; Kollman P.A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14330-14335

† Factor by which torsional barrier is divided.

**Chapter 4. An Alternative Explanation for the Catalytic Proficiency of Orotidine
5'-Phosphate Decarboxylase**

Tai-Sung Lee*, Lillian T. Chong*, John D. Chodera, and Peter A. Kollman

*** These authors contributed equally to the work.**

Graduate Group in Biophysics

University of California at San Francisco

San Francisco, CA 94143-0446

Reproduced with permission from *J. Am. Chem. Soc.* (2000) **123**: 12837-12848.

Copyright 2000 American Chemical Society.

Abstract:

Orotidine 5'-phosphate decarboxylase (ODCase) is the most proficient enzyme known, enhancing the rate of decarboxylation of orotidine 5'-phosphate (OMP) by a factor of 10^{17} , which corresponds to a $\Delta\Delta G^\ddagger$ of ~ 24 kcal/mol. Ground-state destabilization through local electrostatic stress has been recently proposed as the basis of catalytic rate enhancement for a mechanism that is the same as in solution. We have carried out gas-phase *ab initio* quantum mechanical calculations combined with a free energy method, a continuum solvent model, and molecular dynamics simulations to assess an alternative mechanism. Although we are not able to reproduce the experimentally observed $\Delta\Delta G^\ddagger$ quantitatively, we present evidence that this $\Delta\Delta G^\ddagger$ is very large, in the range found experimentally. We thus conclude that the preferred mechanism may well be different from that in solution, involving an equilibrium pre-protonation of OMP C5 by a catalytic lysine residue that greatly reduces the barrier to subsequent decarboxylation.

Introduction:

Orotidine 5'-phosphate decarboxylase (ODCase) catalyzes the decarboxylation of orotidine 5'-phosphate (OMP), the final step in the *de novo* biosynthesis of uridine 5'-phosphate (UMP). This dimeric enzyme is unusually proficient, enhancing the rate of decarboxylation by a factor of 10^{17} , which corresponds to a $\Delta\Delta G^\ddagger$ of ~ 24 kcal/mol in neutral solution at room temperature [1]. No cofactors or metals are involved in the catalysis. To explain the enzyme's proficiency, numerous mechanisms have been proposed based on model systems, kinetic isotope experiments, and, more recently, crystal structures.

The first set of mechanisms was proposed prior to the availability of the enzyme crystal structures. Beak and Siegel suggested a zwitterion mechanism where O2-protonation of OMP yields an ylide intermediate, which acts as an electronic sink [2]. Silverman and Groziak proposed a nucleophilic addition to C5 of the substrate by an active site residue, which is then expelled upon decarboxylation [3]. On the basis of quantum mechanical calculations, Lee and Houk proposed a mechanism involving O4-protonation concerted with decarboxylation via a carbene intermediate [4]. Their results indicated that O4-protonation lowered the activation free energy for decarboxylation from a computed value of 42 kcal/mol in solution to 18 kcal/mol in the enzyme environment, which was modeled as a bulk medium with dielectric constant. A fourth proposal involved a modified version of this mechanism by Ehrlich *et al.*, who conducted multiple solvent deuterium and ^{13}C kinetic isotope effect experiments on catalysis by ODCase [5]. Their results indicate that catalysis involves a stepwise mechanism in which

a protonation event occurs prior to decarboxylation. This led them to propose O4-protonation followed by decarboxylation.

Kinetic isotope studies have provided evidence against mechanisms involving either nucleophilic addition to C5 of the substrate or O4-protonation. When the substrate was synthesized with deuterium at C5, Acheson *et al.* observed no secondary deuterium isotope effect, thereby providing no support for mechanisms involving nucleophilic addition to C5 of the substrate [6]. Furthermore, the observation by Smiley *et al.* of a large ^{13}C isotope effect in the enzymatic reaction has established that decarboxylation is the rate-determining step, thus suggesting that no covalent step, unless it is fast (*e.g.*, a proton-transfer event), occurs before decarboxylation [7]. The O4-protonation mechanisms are unlikely due to the finding by Shostak and Jones of a large effect on the reaction rate (greater than 10 000-fold reduction in k_{cat}) with an O2 \rightarrow S2 substitution and only a small effect on the rate (50% reduction in k_{cat}) with an O4 \rightarrow S4 substitution [8].

Recent crystal structures of ODCase from four different microbial sources in complexes with various inhibitors [9-13] have provided evidence against all of the mechanisms described above. The structures reveal no polar or charged residues near O2 or O4 of the substrate, thereby ruling out O2-protonation and O4-protonation mechanisms. Mechanisms involving nucleophilic addition to C5 are also unlikely due to the absence of a nucleophile near C5 of the substrate. In the vicinity of C5, however, is Lys72 (*Methanobacterium autotrophicum* numbering [13]), which has been revealed by site-directed mutagenesis to be an essential proton-donating catalytic residue that does not appear to be critical for substrate binding [14]. This catalytic residue is part of a unique, conserved Lys42-Asp70-Lys72-Asp75B charged array that lines the pocket for

the substrate carboxylate group (Asp75B is from the adjacent monomer). Various mechanisms have been proposed based on the crystal structures [15]. Wu *et al.* [13] and Appleby *et al.* [9] have interpreted the crystal structures as evidence for ground-state destabilization by the severe repulsion between the carboxylate of the substrate and the anionic form of Asp70. In the transition state, this charge repulsion would be diminished due to a decreased anionic character of the substrate carboxylate group. Wu *et al.* have suggested that ground-state destabilization drives unimolecular cleavage of the bond between C6 and the carboxylate carbon of the substrate, which is followed by protonation of C6 by Lys72 [13]. Using hybrid quantum mechanical/molecular mechanics (QM/MM) calculations, they find that the “electrostatic stress” free energy of 18 kcal/mol between the substrate carboxylate and Asp70 is compensated by the strong binding of the phosphoribosyl region, computed to be -26 kcal/mol. Appleby *et al.* have proposed an electrophilic substitution mechanism in which C6-protonation and decarboxylation occur in a concerted manner [9]. An interesting alternative to the ground-state destabilization mechanism has been proposed by Harris *et al.*, who suggest that a very short hydrogen bond may form between Asp70 and the substrate carboxylate [10].

Warshel *et al.* have recently used a combination of *ab initio* quantum mechanical calculations, empirical valence bond (EVB) simulations, and free-energy perturbation (FEP) calculations to demonstrate that the catalytic effect of ODCase is due to transition-state stabilization rather than ground-state destabilization [16]. They argue that the Wu *et al.* proposal [13], which involves ground-state destabilization due to repulsion between the substrate carboxylate and the anionic form of Asp70, is problematic since the destabilized substrate (or Asp70) will accept a proton from the solvent and form a new

stable ground state. Furthermore, they alert us to the fact that the extremely large binding energy (-26 kcal/mol) that was computed by Wu *et al.* [13] for the phosphoribosyl region of the substrate has no precedent and is inconsistent with recent experiments by Miller *et al.* [12, 17], who found that the region contributed only ~ 5 kcal/mol to stabilizing the transition-state analogue, 6-hydroxyuridine 5'-phosphate (BMP). Warshel *et al.* studied the reaction with the substrate carboxylate in its ionized form and with the neighboring protein groups in their ionized forms [16]. They observed that this configuration does not lead to ground-state destabilization. Their results reveal that the transition state is preferentially stabilized relative to the ground state, which is found to be stabilized to a small extent rather than destabilized. Upon transfer from the ground state to the transition state, there is a significant increase in the dipole moment of the reacting system since the charge separation is greater between the positively charged Lys72 and the negative charge on C6 of the substrate than between Lys72 and the substrate carboxylate. They conclude that the transition-state stabilization arises from the fact that the enzyme has preorganized the active site environment of the ground state in a configuration that contains electrostatic stress between Asp70 and Asp75B, but is complementary to the significantly larger dipole moment of the transition-state ion pair [16, 18].

We also find the notion of “ground-state destabilization” counterintuitive. Here, we present an alternative mechanism that explains the remarkable catalytic efficiency of ODCase. The mechanism involves equilibrium pre-protonation of C5 of OMP by Lys72 followed by decarboxylation, where Asp70 is in its neutral form. This protonation state of Asp70 is in contrast to that reported by Warshel *et al.* [16], who claimed that their pK_a calculations find Asp70 to be anionic, but did not report the computed value. Our

proposed mechanism is supported by QM-FE (quantum mechanical-free energy) calculations on the enzymatic reaction [19, 20], molecular dynamics (MD) simulations of the enzyme in complexes with various ligands, pictured in Figure 1, pK_a calculations, and free energy calculations using the MM-PBSA approach of Srinivasan *et al.* [21]. The mechanism is also consistent with available kinetic isotope effect experimental data on this enzyme.

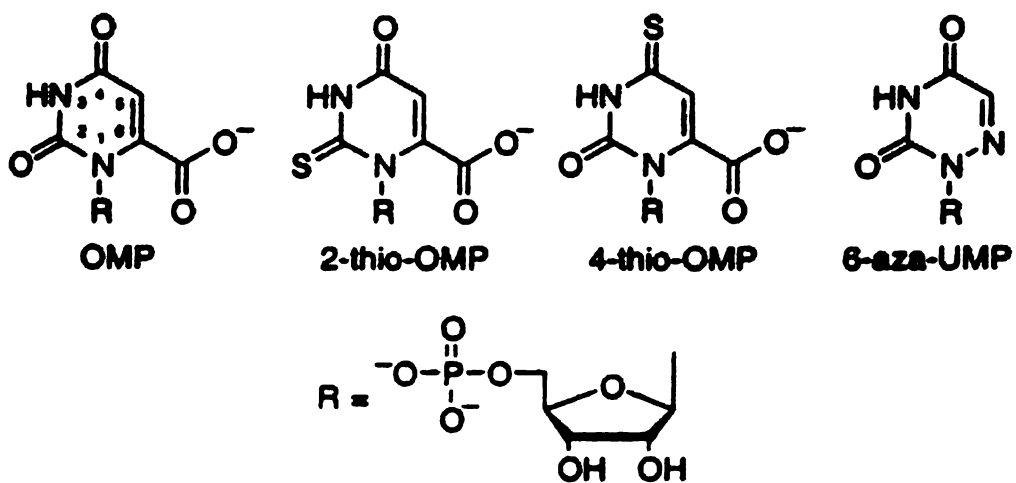


Figure 1. Ligands used in this study.

Methods:

QM Calculations on Model Systems:

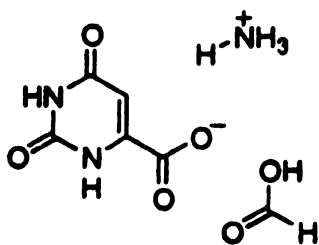


Figure 2. Final model system used in the quantum calculations.

Reaction energy profiles for C5- and C6-protonation of the substrate by Lys72 were calculated using a model system consisting of three species: orotate, ammonium ion, and a neutral formic acid (Figure 2). The coordinates for these species were taken from the corresponding molecules in the crystal structure. Justification for representing Asp70 in its neutral form as opposed to its anionic form is presented under *Discussion and Conclusions*. Each reaction energy profile was obtained by constraining the distance between the nearest hydrogen atom of the ammonium ion and the C5 atom or the C6 atom of the orotate ring. The constraint was done through fixing the corresponding internal coordinate in the Gaussian 98 input file, which is equivalent to a harmonic restraining force with infinite force constant. For each point along the reaction coordinate, all geometric parameters except for the constrained distance were optimized at the HF/6-31+G* level; single-point energies were calculated at both the HF/6-31+G* and MP2/6-31+G* levels. Reaction energy profiles for the decarboxylations of orotate and the C5-protonated intermediate were calculated at the same levels of theory. Single-point energies for the C5-protonation reaction profile as well as that for decarboxylation of the C5-protonated intermediate were also calculated at the MP2/cc-pVDZ level. In some cases, solvation effects were calculated using the PCM model within the Gaussian

98 package [22]. Gas-phase proton affinities (PAs) were computed by calculating the energy difference between the protonated and unprotonated species at the MP2/6-31+G**//HF/6-31+G* level. Equilibrium isotope effects (EIEs) for the C5-protonation step were determined by obtaining the ratio of fractionation factors for the orotate and the corresponding C5-protonated intermediate (orotate/intermediate). Fractionation factors at 298 K were computed by using the Bigeleisen equation [23] implementation via the QUIVER program [24], which employs the Cartesian force constant matrices for the respective states as input. The matrices were obtained using the Gaussian 98 package [22] at the HF/6-31+G**//HF/6-31+G* level and scaled by the default value of 0.8929. Unless otherwise noted, all of the above calculations were performed using the Gaussian 98 package [22] on the Origin 2000 cluster at the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign and the Cray SV1 cluster at the National Cancer Institute (NCI).

QM-FE Calculations:

Combined quantum mechanical and free energy (QM-FE) [19, 20] calculations were performed to incorporate the effect of the enzyme environment on the reaction energy profile for C5-protonation. This approach has also been used by Jorgensen on many organic reactions [25-27]. The overall free energy change, ΔG^* , for C5-protonation in the enzyme is approximated as

$$\Delta G^* = \Delta E_{\text{QM}} + \Delta G_{\text{FE}},$$

where ΔE_{QM} is the quantum mechanical (QM) energy difference and ΔG_{FE} the molecular mechanical (MM) free energy difference between the reactants and the C5-protonated intermediate. The QM region was defined as the orotate ring and the methylammonium ion (representing Lys72), the only parts of the system that undergo bond cleavage and formation. The MM region was defined as the rest of the substrate and protein. Link atoms consisted of the C1 atom of orotate and the C δ atom of Lys72, which were treated as hydrogens in the QM calculations, but as carbons in the MM free energy calculations. It is critical that parts of the system included in the QM region are not included in the MM region to avoid double counting. The neighboring charged residues of Lys72 (Lys42, Asp70, Asp75B) were assigned to the MM region due to the fact that including these residues in the QM *in vacuo* calculation would have greatly exaggerated their effect on the C5-protonation barrier compared to their influence in the higher dielectric environment of the enzyme. QM energies were taken from calculations on model systems (see QM results section, below). Charges for the QM region were obtained by a restrained electrostatic potential fitting method (RESP) [28] with the following modifications: the C1 atom charge, which was near zero and fluctuated slightly over different stages of the free energy calculation, was set to zero to avoid unphysical effects on the overall free energy change. The charge of the C δ atom of Lys72 was taken from that of the corresponding hydrogen in the QM region. The thermodynamic integration (TI) [29] method with 101 equally spaced windows for each direction was employed for the free energy calculations, using the MD protocol described above with different nonbonded cutoffs to study the influence of the nonbonded cutoff on the computed ΔG_{FE} . Various total simulation lengths ranging from 10 to 80 ps were performed. During each

simulation, one hydrogen attached to the Lys72 nitrogen “disappeared” gradually and became a dummy atom, while a dummy atom attached to the orotate C5 atom slowly became a hydrogen atom. This manner of performing the simulation obviates the need to consider the pathway for proton transfer from Lys72 to the orotate ring. Since other energy terms are already included in the QM energies, the MM free energies consist of only van der Waals and electrostatic interactions between the QM and MM regions that are extracted from the simulation, as has been done by Stanton *et al.* [19] and Kuhn *et al.* [20]

Molecular Dynamics (MD) Simulations:

Coordinates for the ODCase dimer were extracted from monomers C and D in the 1.5-Å-resolution crystal structure of the *Methanobacterium thermoautotrophicum* enzyme in complex with the 6-azauridine 5'-phosphate (6-aza-UMP) inhibitor (1DVJ in the Protein Data Bank).[13] The ligand in the active site of monomer C was included in the model, while the active site of monomer D was left empty. Even though the active species of the enzyme is a dimer, it has been shown that the monomers are catalytically independent, such that the E_2S_2 complex can be treated as E_2S . [30] All residues (9-222) of monomers C and D were included in the model with the exception of the cloning artifact and missing residues distant from the active site. Acetyl and *N*-methyl capping groups were added to the N-terminal and C-terminal truncated ends, respectively, using the LEaP module in the AMBER 5.0 package.[31] In the same module, hydrogen atoms were added, and all histidines were protonated by default at the δ -nitrogen. As described above, Asp70 in the ligand-bound active site was assigned to be neutral for all the

enzyme-ligand complexes unless otherwise noted. Ionization states present in neutral solution were used for all other charged residues. All crystallographic waters within 35 Å of the N ζ atom of Lys72 in monomer C were included, yielding a total of 316 waters. Each system was solvated by placing a spherical cap of TIP3P water molecules [32] with a radius of 35 Å centered at the N ζ atom of Lys72 in monomer C. Neutralizing counterions were added within the LEaP module. Coordinates for the substrate (OMP) and its thio-substituted analogues (2-thio-OMP and 4-thio-OMP) were obtained by modifying the 6-aza-UMP ligand in the LEaP module of the AMBER 5.0 package.[31] Based on the free energy reaction profile computed by Wu *et al.*, [13] a “transition-state” ligand was defined as having the same geometry as the substrate, but with a bond of 2.4 Å between C6 and the carboxyl carbon. This structure was optimized at the 6-31+G* level with the C6-CO₂- bond constrained to 2.4 Å. Ligand bond, angle, and dihedral parameters not present in the Cornell *et al.* force field [33] are listed in the *Appendix*. Charges for the ligands and transition-state model were obtained by the RESP method.[28] Phosphoribosyl and base fragments were defined by splitting the ligand at the C1*-N1 bond and capping with formamide and methyl groups, respectively. Electrostatic potentials for these fragments were then derived from HF/6-31+G*//HF/6-31+G* QM calculations using the Gaussian 98 package.[22] Charges were obtained by fitting each fragment independently such that the net charge of the phosphoribosyl fragment was -2 and that of the base fragment was -1 (for OMP, 2-thio-OMP, and 4-thio-OMP) or 0 (for 6-aza-UMP); the capping groups were constrained to have zero net charge. For OMP, good agreement was found between the charges obtained in this manner and those obtained by using the entire molecule for deriving the electrostatic

potentials. Independent treatment of the phosphoribosyl and base fragments is therefore valid for all the ligands in this study.

Molecular dynamics simulations were performed using the Cornell *et al.* force field [33] and the AMBER 5.0 suite of programs.[31] The SHAKE algorithm was applied to constrain all bonds to their equilibrium values.[34] A 12-Å residue-based cutoff was used for nonbonded interactions. Only residues (including ions, water molecules, and the ligand) that lie in the “belly”, the region within 18 Å of the N atom of Lys72, were allowed to move. Minimization and equilibration were performed in two stages. In the first stage, ions in the belly and all water molecules were minimized for 10 cycles of steepest descent followed by 990 cycles of conjugate gradient, and then equilibrated for 10 ps while the temperature was raised from 0 to 300 K. In the second stage, the entire belly region was minimized in the same manner, and then equilibrated for 60 ps while the temperature was raised from 0 to 300 K. Sampling of reasonable configurations for the given stable state of the enzyme-ligand structure was conducted by running a 300-ps belly simulation with a 2-fs time step at 300 K. Constant temperature was maintained by the Berendsen coupling algorithm with separate solute-solvent and solvent-solvent coupling.[35] The equilibrated coordinates and velocities of the enzyme-OMP complex were used as the starting points for the simulations of the enzyme-(2-thio-OMP) and the enzyme-(4-thio-OMP) complexes.

MM-PBSA Calculations:

To compute the average binding free energy of each enzyme-ligand complex, we sampled 15 equally spaced “snapshot” configurations of the unbound enzyme, unbound

ligand, and enzyme-ligand complex from the last 150 ps of the complex trajectory and performed energy calculations on the snapshots. Prior to the energy calculations, all waters and ions were removed from each snapshot. Free energy calculations were performed using the MM-PBSA approach, described in detail by Srinivasan *et al.* [21]

The binding free energy, ΔG_{bind} , is defined as follows:

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{es}} + \Delta E_{\text{vdW}} + \Delta E_{\text{int}}$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{PB}} + \Delta G_{\text{SA}}$$

where ΔE_{MM} is the change in the total MM energy of the solute with an electrostatic component (ΔE_{es}), a van der Waals component (ΔE_{vdW}), and an internal component (ΔE_{int}) consisting of bond, angle, and torsional energies; ΔG_{solv} is the solvation energy difference, which consists of an electrostatic contribution (ΔG_{PB}) determined by the Poisson-Boltzmann approach using the DelPhi2.0 software package [36] and a nonelectrostatic contribution (ΔG_{SA}) that is linearly dependent on the surface area; and $-T\Delta S$ is the solute entropic contribution to the binding free energy. Solute entropic contributions were not calculated in this study since they are only crudely estimated by normal-mode analysis and likely to be similar for all the enzyme-ligand complexes. Because the same enzyme and ligand configurations are used for their respective unbound and bound states, ΔE_{int} is equal to zero. Calculations of both the electrostatic component to the MM energy and the electrostatic contribution to the solvation energy were performed using two different solute, or “interior”, dielectric (ϵ_{int}) values, 1 and 4.

The pK_a value of Asp70 in the enzyme-substrate complex was calculated using a new methodology developed by B. Kuhn in our laboratory (B. Kuhn, unpublished results). This method employs the MM-GBSA approach in which generalized Born/surface area (GBSA) calculations are performed instead of the Poisson-Boltzmann and surface area calculations that are described above. The pK_a of a selected residue is defined as follows:

$$pK_a = pK_a^{sol} + \Delta\Delta G^{deprot}/RT \ln 10$$

where pK_a^{sol} is the pK_a of the residue free in neutral solution, $\Delta\Delta G^{deprot}$ is the energy of deprotonating the residue in the enzyme-substrate complex ($\Delta G^{complex}$) relative to that of deprotonating the residue when it is free in solution (ΔG^{sol}), R is the ideal gas constant, and T is the temperature. The alternative protonation state of Asp70 was modeled into 20 selected configurations from the trajectory of the enzyme-substrate complex (or Asp70 free in solution capped by acetyl and *N*-methyl groups). Free energies of the anionic and neutral Asp were obtained by minimizing Asp70 in the relevant environment and calculating the GBSA energy [37] using parameters from Jayaram *et al.* [38] and a nonbonded cutoff of 25 Å in the AMBER 6.0 suite of programs.[39] Both tautomers of the neutral Asp were taken into consideration, and the preferred protonation site was chosen on the basis of the lower free energy. All GBSA calculations were done using $\epsilon_{int} = 4$.

Results:

QM and QM-FE Calculations:

Energy Profile of C5 Protonation and C6 Protonation. As shown in Figure 3, the barrier for C5-protonation is significantly lower than that of C6-protonation when the energy profiles are calculated at either the HF/6-31+G*//HF/6-31+G* or the MP2/6-31+G*//HF/6-31+G* level. Furthermore, C5-protonation appeared more consistent with the experimental evidence for a stepwise mechanism [5] since it yielded an intermediate while C6-protonation proceeded directly to decarboxylation. We therefore focused on the C5-protonation mechanism.

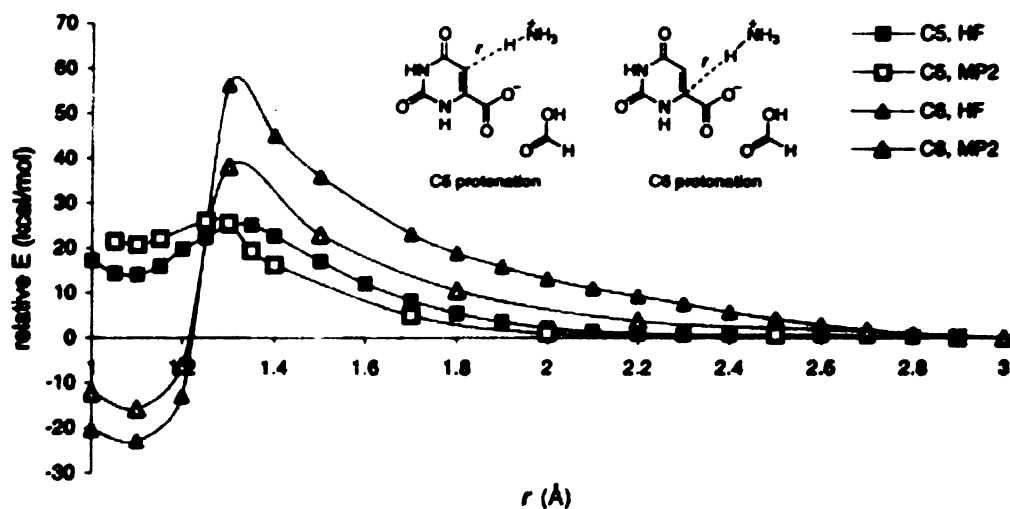


Figure 3. Energy profiles for the C5- and C6-protonations of orotate by an ammonium ion with a formic acid hydrogen bonded to the orotate carboxyl group at the HF/6-31+G*//HF/6-31+G* and MP2/6-31+G*//HF/6-31+G* levels.

To obtain geometries for the ammonium-orotate attack that are less distorted by the attraction between the ammonium and the substrate carboxylate, the C5-protonation energy profile was computed without formic acid, using the geometries optimized in the

presence of formic acid. The results of this calculation are shown in Figure 4. From this result and the result in Figure 3, the formic acid appears to keep the orotate carboxyl group from being distorted due to hydrogen bonding with the ammonium ion, rather than stabilizing the reaction through electronic effects. To more accurately represent Lys72 in the enzyme, the ammonium hydrogen corresponding to the C ξ position of Lys72 was replaced with a methyl group, important because methylamine has a greater proton affinity than NH₃. The energy barrier to C5-protonation by methylamine is ~18 kcal/mol, while the energy difference between the reactants and the intermediate is ~10 kcal/mol (Figure 4). Use of a larger basis set lowers these values to ~13 and ~9 kcal/mol, respectively, as determined at the MP2/cc-pVDZ level of theory. This substitutions at the 2 and 4 positions on the orotate ring caused changes of less than 1 kcal/mol on both energy barriers and relative energies (data not shown).

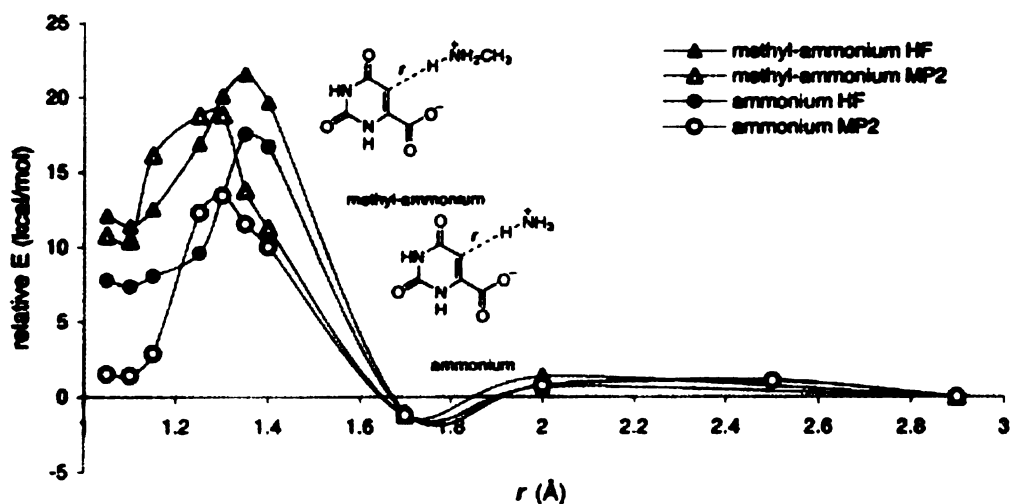


Figure 4. Energy profiles for C5-protonation of orotate by ammonium ion with geometries taken from Figure 3 and energies calculated at the HF/6-31+G* and MP2/6-31+G* levels. Profiles for C5-protonation by methylammonium ion are also shown.

The effect of the enzyme environment on the C5-protonation is challenging to estimate since there are several charged groups in the active site region. To examine the sensitivity of the protonation barrier to the polarizability of the enzyme environment, the PCM continuum solvent model was used with a range of dielectric constant values (ϵ) of 2, 4, and 80. Results are shown in Figure 5. The effect of the continuum model is to increase both the energy barrier and the relative energy. With $\epsilon = 2$, the energy barrier becomes ~ 26 kcal/mol and the relative energy ~ 17 kcal/mol. With $\epsilon = 4$ or 80, the corresponding barriers are even larger. However, the use of continuum models in this way has been appropriately criticized [16, 40, 41]. Such models, when used without explicit representation of the many charged residues near the orotate moiety, are unlikely to accurately reproduce the energetics of proton transfer.

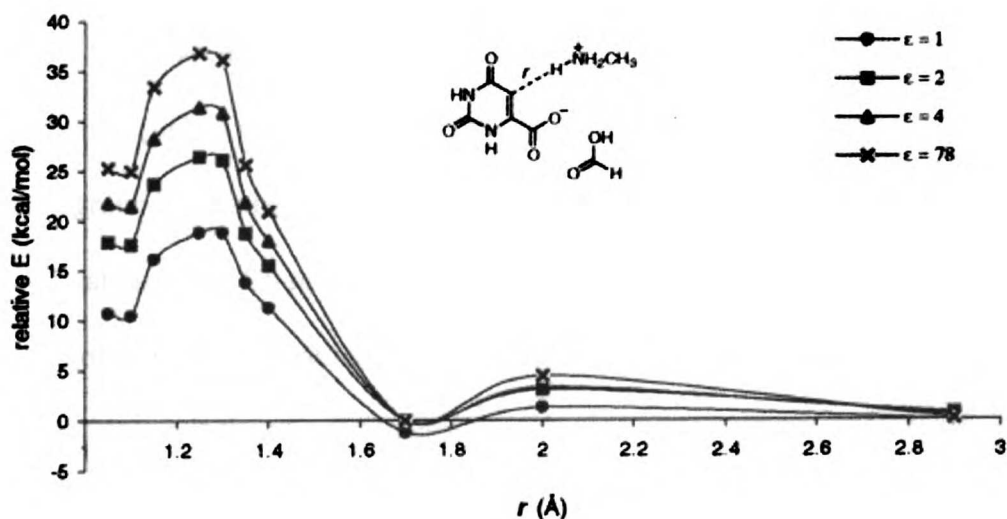


Figure 5 Energy profile for the C5-protonation of orotate by a methylammonium ion with geometries taken from Figure 3. The geometry of the methyl group was optimized at the HF/6-31+G* level with the COSMO continuum solvent model using different dielectric constants. Energies were calculated at the HF/6-31+G* and MP2/6-31+G* levels.

To more realistically account for the enzyme environment, QM-FE calculations [19, 20] were performed on this system to estimate the energy of the C5-protonated intermediate relative to the ground state. This methodology, which involves treating only the minimal part of the system with *ab initio* quantum mechanics and representing the rest with molecular mechanics, has been reviewed by Kollman *et al.* [42] and has accurately determined the ΔG^\ddagger for three other enzymes [19, 20, 43]. Warshel has had success with a similar approach involving EVB simulations [43]. As described in the *Methods* section, the overall free energy change (ΔG^*) for C5-protonation is approximated as the sum of the difference in QM energy (ΔE_{QM}) and the difference in the free energy of interaction (ΔG_{FE}). The above calculations have shown that ΔE_{QM} is ~ 10 kcal/mol. To be consistent with the experimental k_{cat} and the fact that decarboxylation is the rate-limiting step would require the stabilizing effect of the enzyme environment (ΔG_{FE}) to be less than -5 kcal/mol. The average differences in the free energy of interaction (ΔG_{FE}) are presented in Table 1, and the charge distributions of the QM atoms in the ground and intermediate states are shown in Figure 6. Although the computed stabilization of the C5-protonated intermediate is unrealistically more negative than -5 kcal/mol and still fluctuating as a function of nonbonded cutoff and simulation length, the qualitative result shows consistently that the enzyme stabilizes the intermediate relative to the reactants. Given the large uncertainty in the electrostatic energy with the numerous charged groups in the substrate and nearby residues, free energy calculations with particle mesh Ewald [44] would need to be performed to accurately calculate the energy of the intermediate relative to the ground state. However, this system is large for full periodic

box MD-FE simulations and, with the large number of charges in the active site, would require very long calculations for quantitative convergence. We also note that the gas-phase barrier to C5-protonation at the MP2/6-31+G* level (Figure 4) is ~18 kcal/mol and that some of the environmental stabilization of the protonated intermediate is likely to occur at the transition state for protonation. It could well be that the transition state in Figure 4 is stabilized by a significant fraction of the -5 to -10 kcal/mol stabilization that is experienced by the C5-protonated intermediate. The goal here is not to attain a quantitative ΔG^* , but to show that the enzyme environment stabilizes the C5-protonated intermediate over the reactant of an order of magnitude such that $\Delta G^* = \Delta E_{QM} + \Delta G_{FE}$ is significantly less than the ΔE_{QM} (Figure 4, ~10 kcal/mol), in contrast to what is suggested by the continuum calculations, Figure 5, where a general environmental effect destabilizes the protonated intermediate.

Table 1. Average MM Free Energy Differences^a (in kcal/mol) for the C5-Protonated Intermediate Relative to the Ground State Determined by QM-FE Calculations for Different Simulation Lengths and Nonbonded Cutoffs

	8 Å	10 Å	12 Å
10 ps	-18.3 ± 6.0	-10.2 ± 6.6	-15.2 ± 2.3
20 ps	-11.9 ± 7.4	-41.5 ± 17.5	-33.4 ± 2.6
40 ps	-41.4 ± 6.7	-32.9 ± 7.7	-41.2 ± 3.4
80 ps	-16.3 ± 1.9	-27.7 ± 17.8	-15.1 ± 4.9

^a Averages are from backward and forward simulations with the range given, e.g., -20 ± 10 means one simulation leads to a free energy of -30 kcal/mol, the other, -10 kcal/mol.

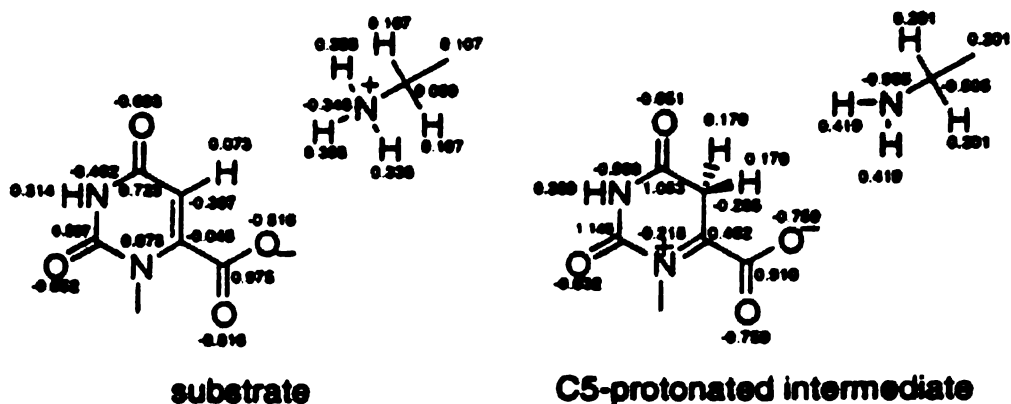


Figure 6. Partial charges for the QM atoms in the ground and C5-protonated intermediate states used in the QM-FE calculations.

Proton Affinities. The calculated gas-phase proton affinities (PA) of different sites of orotate and deprotonated uracil are shown in Figure 7. Since the addition of a proton to C6 will result in decarboxylation, the PA of C6 was calculated by constraining the C6-H distance at 1.09 Å and the C6-carboxyl carbon distance at 1.60 Å; these values are taken from the equilibrium distances of C5-protonated orotate. Lee and Houk compared the PAs of O2 and O4 positions of orotate and concluded that O4 is the preferred protonation site [4]. Surprisingly, the PA of C5 is higher than that of either O2 or O4, both in orotate and in deprotonated uracil. These results show that C5 is intrinsically more basic than O2, O4, or C6.

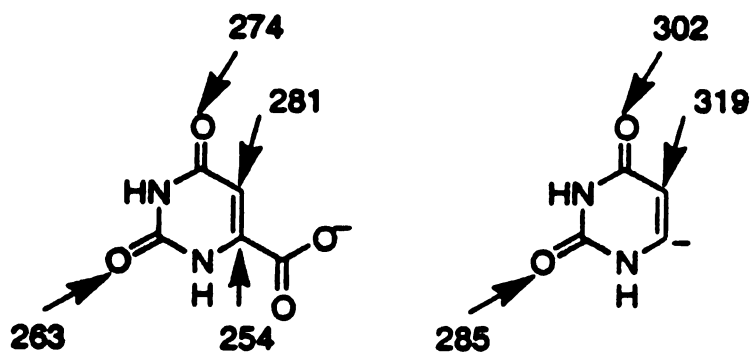


Figure 7. Calculated gas-phase proton affinity for different sites of orotate and deprotonated uracil in kilocalories per mole at the MP2/6-31+G**/HF/6-31+G* level.

PAs for some derivatives of orotate were also computed (Figure 8). The fact that the PA of C5 is nearly 100 kcal/mol lower in uracil than in orotate indicates that the carboxylate group plays a critical role in raising the PA of C5. The C5 PA of both the 2-thio and 4-thio substrate analogues is similar to that of the unsubstituted substrate, suggesting that these substitutions do not have a large effect on the electronic structure near the C5 position. The fact that the proton affinity increases upon replacing the N1-H by CH₂ suggests that there is little or no stabilization of protonation at C5 by resonance delocalization effects due to N1.

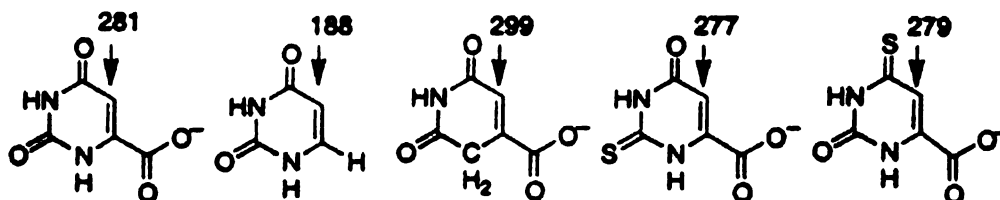


Figure 8. Calculated gas-phase proton affinity for different derivatives of orotate in kilocalories per mole at the MP2/6-31+G**/HF/6-31+G* level.

Energy Profile of Decarboxylation. As shown in Figure 9, the barrier to decarboxylation of orotate is 35–40 kcal/mol, in accord with experimental results [1] and other computational results [4, 13] for the nonenzymatic reaction. This barrier is greatly reduced to ~10 kcal/mol for the C5-protonated intermediate (Figure 9). At the MP2/cc-pVDZ level, the same barrier is reduced to ~5 kcal/mol. Incorporation of environmental effects using the PCM model increases this barrier to 10, 15, and 21 kcal/mol using dielectric constant values of 2, 4, and 80, respectively. Barriers to decarboxylation for the 2-thio and 4-thio substrate analogues (data not shown) are within 1 kcal/mol of that of the unsubstituted substrate.

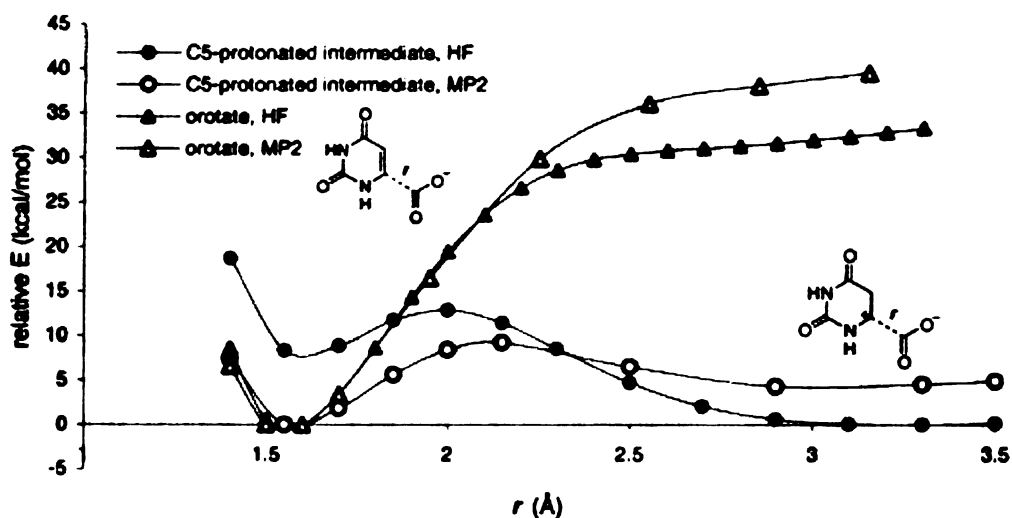


Figure 9. Energy profiles for the decarboxylation of orotate and of the C5-protonated intermediate at the HF/6-31+G**/HF/6-31+G* and MP2/6-31+G**/HF/6-31+G* levels.

Combining the results from the QM calculations of both protonation and decarboxylation and our less rigorously derived, but plausible, estimates for the free energy of environmental stabilization for the protonation step, we suggest a free energy profile for the reaction as shown in Figure 10. As noted above, our free energy

calculations lead to an unrealistically large stabilization of the protonated intermediate by -11 to -40 kcal/mol, but using ~10 kcal/mol for the QM barrier to protonation and an environmental stabilization of ~-5 kcal/mol leads to the illustrated barrier and relative energy for the C5-protonated intermediate. For the decarboxylation step, the QM calculations suggest a barrier of ~10 kcal/mol (MP2/6-31+G*) and ~5 kcal/mol (MP2/cc-pVDZ), with continuum solvation raising this value by 5-16 kcal/mol (ϵ values of 2-80). Free energy perturbation calculations on this step would be desirable but are likely to suffer from the same uncertainty in the electrostatic energy as observed in the QM-FE calculations on the protonation step. All we can claim is that the barrier to decarboxylation of the C5-protonated intermediate is much smaller than that of orotate and that our calculated values (5-21 kcal/mol) are in the range consistent with the experimental ΔG^\ddagger of orotate. The rate-limiting step is therefore the decarboxylation with a net ΔG^\ddagger of ~15 kcal/mol relative to reactants. The protonated intermediate is less stable than reactants by ~5 kcal/mol, but the barrier to achieve it is significantly less than 15 kcal/mol, such that the decarboxylation is clearly rate-limiting. This reaction profile is consistent with the finding by Ehrlich *et al.* [5] that the protonated intermediate deprotonates with greater probability than proceeding through decarboxylation.

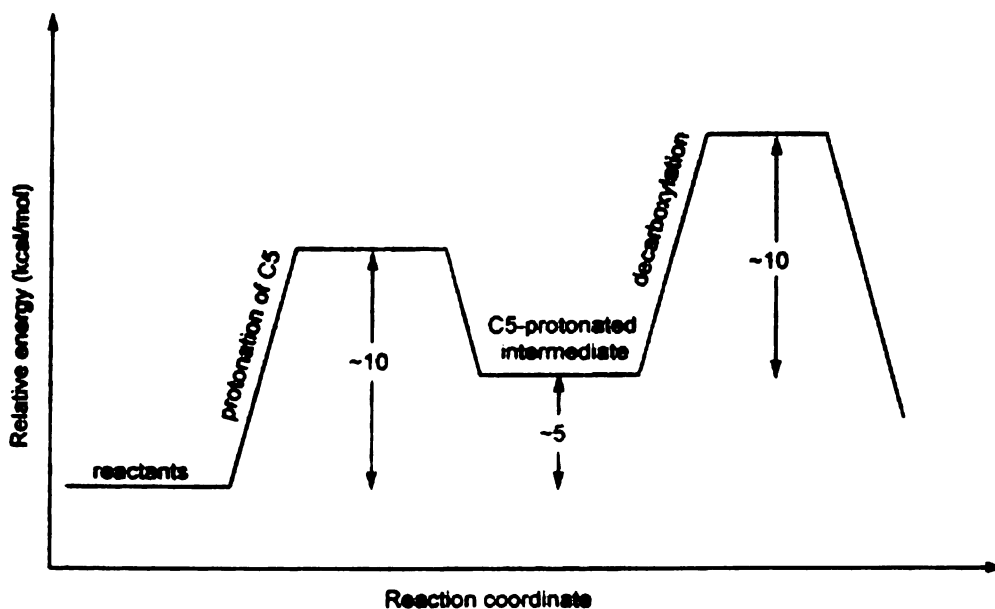


Figure 10. Qualitative energy profile illustrating the plausibility of the proposed mechanism of catalysis by ODCase. The forward barrier height of decarboxylation of the C5-protonated intermediate was computed by gas-phase *ab initio* QM calculations at the MP2/6-31+G**/HF/6-31+G* level (Figure 9). The energy of the C5-protonated intermediate relative to reactants ($\Delta G^* = \Delta E_{QM} + \Delta G_{FE}$) was estimated from QM-FE calculations, which yielded a QM energy difference (ΔE_{QM}) of ~10 kcal/mol (Figure 4) at the MP2/6-31+G**/HF/6-31+G* level for C5-protonation by methylammonium ion and a stabilizing effect of the enzyme environment (ΔG_{FE}) that is unrealistically favorable as well as not converged due to inaccurate treatment of long-range electrostatics (Table 1). Given that decarboxylation is the rate-limiting step and that the experimental ΔG^\ddagger is 15 kcal/mol [1], the stabilizing effect of the enzyme environment would be expected to be less than -5 kcal/mol, such that the relative energy of the intermediate to reactants is 5 kcal/mol as shown. A significant fraction of the -5 to -10 kcal/mol environmental stabilization of the protonated intermediate is likely to be experienced by the transition state of the protonation step such that the computed barrier of ~18 kcal/mol at the MP2/6-31+G**/HF/6-31+G* level is reduced to an estimate of ~10 kcal/mol.

MD:

Stable dynamics of all the enzyme-ligand complexes were obtained from the simulation protocol and force field parameters as indicated by low belly atom root-mean-square deviations (RMSD) from the average structure. For the enzyme-(6-aza-UMP)

complex, the RMSD of the belly atoms from the average structure ranged from 0.3 to 0.4 Å, while the RMSD from the minimized crystal structure remained steady at 0.6 Å over the entire 300 ps trajectory. As the other ligands were constructed from the 6-aza-UMP complex crystal structure, the first 150 ps were omitted from analysis to allow for additional equilibration. This equilibration period was sufficient for all cases except for the 2-thio-OMP complex trajectory, in which considerable RMSD fluctuations occurred in the 150-300 ps interval. This trajectory was therefore extended to 1 ns. The belly atom RMSDs from the average structures over the final 150 ps of the substrate, 4-thio-OMP, and extended 2-thio-OMP complex trajectories ranged from 0.3 to 0.6 Å.

To explore the feasibility of the C5-protonation mechanism, key interatomic distances were computed. Specifically, it was noted that neutral Asp70 forms a hydrogen bond with the substrate carboxylate throughout the complex trajectory. Furthermore, Lys72 lies out of the plane of the orotate ring with the distance between the N ζ atom of Lys72 and C5 in the substrate (OMP) fluctuating between 3.4 and 5.2 Å. The proximity of these two atoms at various points of the trajectory makes it possible for Lys72 to donate a proton to C5 of the substrate. Thus, the dynamics of the enzyme-substrate complex are consistent with the proposed C5-protonation mechanism.

The dynamics of the 4-thio-OMP and 2-thio-OMP complexes are also consistent with experimental kinetics data. Shostak and Jones observed no enzyme activity for the 2-thio-OMP ligand, while the reaction rate was only slightly diminished for the 4-thio-OMP ligand (50% reduction in k_{cat}) [8]. Since the experimental binding free energies of these thio-substituted ligands are similar to that of the natural substrate, subtle differences in the dynamics of the 2-thio-OMP complex are likely to be responsible for the dramatic

decrease in catalytic activity. To examine differences in the dynamics of the substrate, 4-thio-OMP, and 2-thio-OMP complexes, the latter two were superimposed onto the substrate through least-squares fits of the active site residues pictured in Figure 11A. As pictured in Figure 11B,C, the position of the orotate ring in 2-thio-OMP has shifted, and its phosphoribosyl portion has changed in conformation from that of the substrate while the position of 4-thio-OMP is similar to that of the substrate. As the 2-carbonyl in the substrate participates in a water-mediated hydrogen bond to Gln185, its replacement with a thiocarbonyl likely causes a disruption in the network of hydrogen bonds. This disruption somehow leads to a loss in catalytic activity.

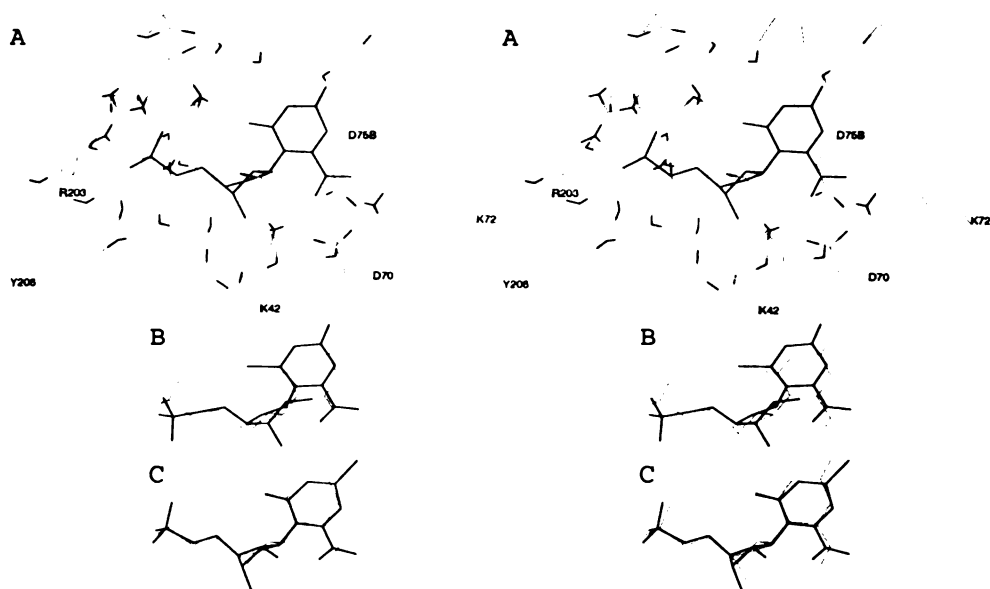


Figure 11. Stereoviews of final structures from MD trajectories. (A) Active site residues (gray) and OMP (black); (B) 4-thio-OMP (black) superimposed onto OMP (gray) using a least-squares fit of the residues hydrogen bonding to OMP; and (C) 2-thio-OMP (black) superimposed onto OMP (gray) using the same method. Produced with the MidasPlus graphics program [53].

Table 2. Statistical Data on Distances (in Å) between N Lys72 and C5 of the Ligand As Sampled by Various ODCase-Ligand Complex Trajectories^a

ligand	min	max	avg (SD)	% less than 4.6 Å
OMP	3.4	5.2	4.6 (0.3)	47.3
4-thio-OMP	3.7	5.4	4.6 (0.3)	46.7
2-thio-OMP	4.6	5.7	5.1 (0.2)	0.7

^a Structures were sampled every picosecond in the 150-300 ps interval for OMP and 4-thio-OMP, and in the 850-1000 ps interval for 2-thio-OMP.

Since protonation of C5 of the ligand by Lys72 is key in our proposed mechanism, it was predicted that Lys72 would be positioned close to C5 for only a small percentage of the time, if at all, during the 2-thio-OMP complex trajectory. Table 2 shows the statistical data on the distances between N ζ of Lys72 and C5 of the ligand as sampled by the substrate, 4-thio-OMP, and 2-thio-OMP complex trajectories. For both the substrate and 4-thio-OMP complexes, the average distance between the two atoms is 4.6 Å, with 47% of the 150 configurations having values less than this average distance. For the 2-thio-OMP complex, this percentage is only 0.2%, with an average distance of 5.1 Å. In the 150-300 ps interval of the 2-thio-OMP complex trajectory, the percentage is comparable with a value of 2.7%. Overall, these results suggest that the loss of catalytic activity for the 2-thio-OMP ligand is due to a greatly reduced occupancy of configurations that allow for efficient delivery of a proton by Lys72 to C5 relative to the cases of the substrate and 4-thio-OMP ligand.

Interestingly, similar results were obtained for C6-protonation. The average distance between N of Lys72 and C6 of the ligand was 4.3 Å for the substrate complex trajectory, with 48.7% of the substrate complex configurations, 33.3% of the 4-thio-OMP

complex configurations, and only 4% of the 2-thio-OMP configurations (5.3% for the 150-300 ps interval) with values less than this distance.

MM-PBSA:

Table 3. Average Energetic Contributions to Enzyme Complex Formations with the 6-aza-UMP Inhibitor^a (in kcal/mol) with Standard Errors of the Mean in Parentheses

	with charged Asp70		with neutral Asp70	
	$\epsilon_{\text{int}} = 1$	$\epsilon_{\text{int}} = 4$	$\epsilon_{\text{int}} = 1$	$\epsilon_{\text{int}} = 4$
$\langle \Delta E_{\text{es}} \rangle$	107.5 (4.3)	26.9 (1.1)	22.0 (2.3)	5.5 (0.6)
$\langle \Delta E_{\text{vdW}} \rangle$	-31.0 (0.9)	-31.0 (0.9)	-32.9 (1.0)	-32.9 (1.0)
$\langle \Delta G_{\text{PB}} \rangle$	-54.2 (4.4)	-17.3 (1.0)	38.8 (2.0)	3.4 (0.5)
$\langle \Delta G_{\text{SA}} \rangle$	-4.2 (0.1)	-4.2 (0.1)	-4.2 (0.1)	-4.2 (0.1)
$-T\Delta S^b$	20	20	20	20
$\langle \Delta G_{\text{bind}} \rangle$	38.2 (3.1)	-5.5 (0.8)	43.6 (2.1)	-8.2 (0.9)
$\Delta G_{\text{bind}}(\text{exp})$	-9.0			

^aFifteen configurations sampled every 10 ps from 150 to 300 ps in trajectory.

^bEstimated value based on published results for similarly sized ligands [46].

The combined molecular dynamics and continuum solvent model approach (MM-PBSA) of Srinivasan *et al.* [21] was performed to obtain the binding free energies for each enzyme-ligand complex in this study. Given the fact that use of a larger interior dielectric ($\epsilon_{\text{int}} = 4$) leads to a more realistic absolute ΔG_{bind} for highly charged species, [43] the electrostatic contributions to the binding free energy were computed using $\epsilon_{\text{int}} = 4$ as well as $\epsilon_{\text{int}} = 1$. Solute entropic contributions ($-T\Delta S$) to the binding free energy were assumed to be similar for all enzyme-ligand complexes and hence were not computed. An assumed entropic contribution of +20 kcal/mol was used, based on published results

for similarly sized ligands [45]. Computed binding free energies for the enzyme-(6-aza-UMP) complex, with both the neutral and anionic forms of Asp70, are shown in Table 3. Even though the absolute calculated free energies of binding are unrealistically positive for $\epsilon_{\text{int}} = 1$, they are reasonably close to experiment with $\epsilon_{\text{int}} = 4$, given the ~ 20 kcal/mol expected contribution from the $-T\Delta S$ term for ligand binding [45]. Furthermore, both the OMP substrate and the unimolecular decarboxylation transition-state model (Figure 4) are computed to have significantly more favorable binding free energies than the 6-aza-UMP inhibitor, independent of the assumed dielectric model or protonation state. This argues against “ground-state destabilization” [13] since, if anything, the substrate is found to interact more favorably with the enzyme than the inhibitor, which lacks the negatively charged carboxylate group.

Table 4. Average Energetic Contributions to GS and TS Complex Formations with Neutral Asp70^a (in kcal/mol) with Standard Errors of the Mean in Parentheses

	GS complex		TS complex	
	$\epsilon_{\text{int}} = 1$	$\epsilon_{\text{int}} = 4$	$\epsilon_{\text{int}} = 1$	$\epsilon_{\text{int}} = 4$
$\langle \Delta E_{\text{es}} \rangle$	47.6 (4.6)	11.9 (1.2)	57.9 (4.7)	14.5 (1.2)
$\langle \Delta E_{\text{vdw}} \rangle$	-29.7 (1.5)	-29.7 (1.5)	-33.4 (1.3)	-33.4 (1.3)
$\langle \Delta G_{\text{PB}} \rangle$	-21.8 (4.6)	-14.7 (1.0)	-34.2 (4.7)	-18.1 (1.0)
$\langle \Delta G_{\text{SA}} \rangle$	-4.5 (0.1)	-4.5 (0.1)	-4.6 (0.0)	-4.6 (0.0)
$-T\Delta S^{\text{b}}$	20	20	20	20
$\langle \Delta G_{\text{bind}} \rangle$	11.7 (2.5)	-16.9 (1.5)	5.6 (3.2)	-21.7 (1.0)
$\Delta G_{\text{bind}}(\text{exp})$	-8.3			

^aFifteen configurations sampled every 10 ps from 150 to 300 ps in trajectory.

^bEstimated value based on published results for similarly sized ligands [46].

Binding free energy contributions for the substrate and transition state to unimolecular decarboxylation are shown in Table 4 for the neutral form of Asp70 in the enzyme. The binding free energies of both the substrate and transition-state complexes are more favorable for the neutral form of Asp70 than the anionic form, regardless of the choice of interior dielectric constant (data not shown). Considering the results with the neutral Asp70, the binding free energy of the transition-state complex is 5-6 kcal/mol more favorable than that of the substrate complex with the neutral Asp70, comparable to the 2 kcal/mol stabilization noted by Wu *et al.* [13]. This lack of differential stabilization of the ground state and transition state is inconsistent with ground-state destabilization, which would require a 20 kcal/mol difference between the destabilization of the ground state (18 kcal/mol) and stabilization of the transition state (-2 kcal/mol) [13]. Binding free energies for the 2-thio and 4-thio substrate analogues in complex with the neutral form of Asp70 of the enzyme are shown in Table 5. Although the calculated ΔG 's are unusually favorable for the choice of dielectric constant $\epsilon_{\text{int}} = 1$, they are comparable to that of the unsubstituted substrate for $\epsilon_{\text{int}} = 4$, which is consistent with experimental data [8]. It is clear that there are difficulties in quantitatively applying MM-PBSA to highly charged ligands, but, nonetheless, our results provide evidence against ground-state destabilization of the substrate.

Table 5. Average Energetic Contributions to Enzyme Complex Formations with 2-thio-OMP^a and 4-thio-OMP^b with Neutral Asp70 (in kcal/mol) with Standard Errors of the Mean in Parentheses

	2-thio-OMP complex		4-thio-OMP complex	
	$\epsilon_{\text{int}} = 1$	$\epsilon_{\text{int}} = 4$	$\epsilon_{\text{int}} = 1$	$\epsilon_{\text{int}} = 4$
$\langle \Delta E_{\text{es}} \rangle$	59.9 (5.6)	15.0 (1.4)	45.0 (4.2)	11.3 (1.0)
$\langle \Delta E_{\text{vdW}} \rangle$	-26.8 (1.5)	-26.8 (1.5)	-29.4 (1.1)	-29.4 (1.1)
$\langle \Delta G_{\text{PB}} \rangle$	-51.7 (4.2)	-20.4 (0.9)	-32.9 (3.8)	-17.9 (0.9)
$\langle \Delta G_{\text{SA}} \rangle$	-4.5 (0.1)	-4.5 (0.1)	-4.7 (0.0)	-4.7 (0.0)
$-T\Delta S^{\text{c}}$	20	20	20	20
$\langle \Delta G_{\text{bind}} \rangle$	-3.0 (2.6)	-16.7 (1.1)	-1.9 (2.6)	-20.7 (1.1)
$\Delta G_{\text{bind}}(\text{exp})$	-6.5		-6.7	

^aFifteen configurations sampled every 10 ps from 850 to 1000 ps in trajectory.

^bFifteen configurations sampled every 10 ps from 150 to 300 ps in trajectory.

^cEstimated value based on published results for similarly sized ligands [46].

To this point, we have assumed Asp70 to be in its neutral state. Given the proximity of Asp70 (with an intrinsic pK_{a} of 4.0) to the orotate carboxylate group (intrinsic pK_{a} of 2.5), one would expect it to pick up a proton. Yet, several charged residues are also present in the active site (Lys72, Lys42, and Asp75B), which could have the effect of compensating for the strong repulsion expected between an anionic Asp70 and the orotate carboxylate group. To address this issue, we have employed a new methodology that uses the MM-GBSA approach to calculate the pK_{a} 's of ionizable groups (B. Kuhn, unpublished results). This method has led to an average error of 1.5 pK_{a} units for known pK_{a} 's in hen egg white lysozyme when the computations were performed on a single trajectory (B. Kuhn, unpublished results). Due to the significant differences in local environment for the anionic and neutral states of Asp70, its pK_{a} was determined by taking the average of the results obtained using the trajectory with the

anionic form of Asp70 and that with the neutral form of Asp70. The resulting pK_a of Asp70 is 7.7 ± 2.2 , which is raised from its intrinsic value by $\sim 3-4$ pK_a units.

Discussion and Conclusions:

QM calculations performed here and by others [4, 13] as well as experiments [1] find the activation barrier for the nonenzymatic decarboxylation of the OMP substrate to be in the range of 35-40 kcal/mol. For a simple unimolecular reaction where the C6-CO₂⁻ dissociates to leave an anionic base, it is hard to imagine what kind of noncovalent interactions could differentially stabilize the more charge-delocalized transition state relative to the reactant. Thus, several research groups [9, 10, 13] have invoked the concept of ground-state destabilization to explain the catalysis of decarboxylation by ODCase.

A ground-state destabilization of 18 kcal/mol has been reported by Wu *et al.* [13] on the basis of QM/MM calculations with a model where a severe repulsion exists between the anionic form of Asp70 and the substrate carboxylate. Using potential of mean force (PMF) free energy calculations, they determined the ΔG^\ddagger for the enzyme-catalyzed reaction to be 15 kcal/mol, in excellent agreement with that found experimentally. It should be noted that Wu *et al.* used a system containing only *N*-methyl orotate in their QM solution calculations and a QM/MM system with C1' linked by a classical link atom in their enzyme calculation. Since the solution PMF was not computed with the same QM/MM protocol, it is uncertain whether the inaccurate forces introduced by the link atom resulted in artifacts appearing in the PMF. In addition, Warshel *et al.* [16] have criticized the decision not to include Lys72 in the QM region of

the QM/MM system due to its role in donating a proton in the concerted ground-state destabilization mechanism.

Our MD simulations of the enzyme-substrate complex show that Lys72 is well-positioned to deliver a proton to C5 or C6. However, C5-protonation has a lower activation barrier than C6-protonation, as determined by QM reaction energy profiles and QM-FE calculations. Furthermore, MM-PBSA binding free energy calculations of the enzyme-substrate, enzyme-transition state, and enzyme-inhibitor complexes show that there is little or no ground-state destabilization by the enzyme. Instead, Asp70 is preferentially protonated, as indicated by a computed pK_a of 7.7 ± 2.2 . In principle, our pK_a calculations should be reasonably accurate, given the average error observed on similar systems. The large error of 2.2 pK_a units in our pK_a calculation for Asp70 is most likely due to the significant difference between the complex structure with the neutral Asp70 and that with the anionic Asp70, as observed in the MD simulations.

Furthermore, the MD simulation of the enzyme-substrate complex remains closer to the crystal structure of the inhibitor-bound enzyme with the neutral state of Asp70. This observation, combined with our pK_a calculation and the fact that the binding affinity of the enzyme-substrate complex is more favorable with a neutral Asp70 than with an anionic Asp70, supports the assertion that Asp70 is neutral. The preference of Asp70 to be in a neutral state has precedent in enzyme systems such as HIV protease, where it is likely that the catalytically active form has one neutral and one charged aspartic acid residue [46-48]. Based on the role of Asp70 in stabilizing the substrate carboxylate group in the correct geometry, its role in controlling the position of Lys72 with the help of Asp75B from the adjacent monomer, and its total conservation in all known forms of

ODCase, we would expect Asp70 to be important for catalysis. We would also expect that a D70N mutant would retain some catalytic activity. In solution, the mechanism of decarboxylation is different, since any excess protons would first protonate the substrate carboxylate group, blocking decarboxylation through the C5-protonation mechanism. The enzyme has thus created a microenvironment that protects the substrate carboxylate group from protonation through hydrogen bonding with Asp70, enabling facile decarboxylation through protonation of C5.

To relate our calculations to the observed ΔG^\ddagger of 15 kcal/mol [1], the QM-FE calculations for C5-protonation would have to lead to a significant lowering of the barrier to protonation and stabilization of the C5-protonated intermediate over those found in the gas phase. Our calculations do show this, although the quantitative stabilization is clearly too large. Postulating a more plausible barrier to protonation of ~ 10 kcal/mol that accounts for the observed stabilization of the enzyme environment and proton tunneling effects, we can construct a free energy profile for the reaction that explains the experimental data. As shown in Figure 10, the barrier to decarboxylation of the C5-protonated species (Figure 9) would lead to an effective barrier of ~ 15 kcal/mol for decarboxylation. We emphasize that we have not precisely calculated the free energy profile in Figure 10. Thus, all that we can claim at this point is the plausibility of this mechanism.

Our proposed mechanism has similarities to the mechanism proposed by Beak and Siegel [2], in which protonation of the substrate at O2 was the critical step to catalysis, and to that proposed by Lee and Houk,[4] in which O4-protonation was instrumental. In our view, supported by the greater intrinsic proton affinity of C5 than

O2, O4, or C6 of orotate, the smaller energy barrier of C5-protonation than C6-protonation, the modest calculated barrier for methylammonium ion (as a model of Lys72) to orotate proton transfer, and the subsequent smaller barrier for decarboxylation once C5 is protonated, the mechanism here is more consistent with the evidence from Ehrlich *et al.* [5] that suggests equilibrium pre-protonation followed by a rate-limiting decarboxylation.

Recent ^{15}N kinetic isotope effect experiments [49] have implied that no bond order changes take place at N1, thereby ruling out mechanisms such as the one proposed by Beak and Siegel [2] which involve formation of an ylide intermediate. To assess the magnitude of bond order change at N1, we computed the N1 ^{15}N equilibrium isotope effect on C5-protonation using both orotate and N-methyl orotate. The computed ^{15}N equilibrium isotope effects for orotate and N-methyl orotate are 0.994 and 0.995, respectively. Experimentally, the intrinsic ^{15}N isotope effect for the enzyme-catalyzed reaction was determined to be 1.007, while the observed isotope effects for model compounds N-methyl picolinic acid and picolinic acid were found to be 1.007 and 0.995, respectively (extrapolated to 298 K) [49]. The normal isotope effect for N-methyl picolinic acid is interpreted as arising solely from the loss of N-C-C and N-C-C-O vibrational modes in the decarboxylation step, while the inverse effect for picolinic acid is interpreted as being due to the product of a similar normal value for the decarboxylation step and a larger inverse equilibrium value arising from the protonation of N1 to give the quaternary ammonium ion prior to the decarboxylation step. The overall ^{15}N isotope effect in our mechanism will be the product of our calculated value of 0.995 for the equilibrium protonation at C5 and a normal secondary kinetic isotope effect

for the subsequent decarboxylation step due to the loss of the same vibrational modes as in the two model compounds. Although the identity of the ^{15}N isotope effects in the enzyme and *N*-methyl picolinic acid was used as an argument against the pre-equilibrium formation of the ylide intermediate with a much larger predicted inverse isotope effect of 0.97, the value of 1.007 observed in the enzyme would also be consistent with our much smaller inverse isotope effect (0.995) multiplied by a slightly larger normal effect for the decarboxylation step than the value observed for the *N*-methyl picolinic acid model. The fact that the ^{13}C isotope effect for the enzymatic reaction (1.049) is substantially larger than that observed for the model reaction (1.028) supports the idea that the ^{15}N effect for the decarboxylation step in the enzyme is larger than that for the model. Consistent with these results are the observed opposite changes in the N1-C6 and N1-C2 bond lengths upon C5-protonation (1.37 Å \rightarrow 1.28 Å and 1.35 Å \rightarrow 1.42 Å, respectively) as well as the fact that there is little or no stabilization of C5-protonation by resonance delocalization effects due to N1, as indicated by an increase in PA upon replacing N1-H of orotate by CH₂.

The possibility of a significant involvement of C5 in the mechanism has been explored by a number of kinetic isotope effect experiments. Even though the results of these experiments are not suggestive by themselves of an important role for C5 in the mechanism, our proposed mechanism is consistent with the experimental data.

Acheson *et al.* found no 5-D isotope effect on either k_{cat} (0.99 ± 0.06) or $k_{\text{cat}}/K_{\text{M}}$ (1.00 ± 0.06) [6]. Since the hybridization does not change at C5 during the loss of CO₂, no secondary isotope effect is expected for the decarboxylation step. However, due to the $\text{sp}^2 \rightarrow \text{sp}^3$ change upon protonation at C5 in the first step of our mechanism, one would

expect an inverse secondary isotope effect that would be observed as an overall isotope effect on k_{cat} and k_{cat}/K_M . We thus computed the isotope effect expected upon 5-D substitution from the vibrational frequencies of the isotopically labeled and unlabeled versions of orotate and of the corresponding C5-protonated intermediate. This approach (see *Methods*) has reproduced measured isotope effects in many studies on enzyme mechanisms, with the most recent by Fitzpatrick [50], Houk [51], and Schramm [52]. Our calculations found a deuterium equilibrium isotope effect (k_H/k_D) of 1.012 ($\phi_{\text{orotate}} = 0.950$ and $\phi_{\text{intermediate}} = 0.938$ relative to water). This value is within experimental error of the values found by Acheson *et al.*[6] The absence of a significant isotope effect may be due to the electrostatic attraction between the hydrogen on the C5 and the carboxylate group, which could constrain the C5-H bond in such a way that the vibrational frequency of its out-of-plane bending motion is much higher than usual, thereby compensating for any inverse isotope effect.

Shostak and Jones found that F-substitution at C5 increased k_{cat} by 30-fold but reduced K_M by roughly the same amount, leading to an overall k_{cat}/K_M comparable to that for unsubstituted OMP [8]. To examine the consistency of our mechanism with this result, we carried out calculations to assess the effect of 5-F substitution on both the C5-protonation and decarboxylation steps in our proposed mechanism. As shown in Figure 8, the C5 PA decreases by 6 kcal/mol. On the other hand, the barrier to decarboxylation was computed to be only 4 kcal/mol, in contrast to the 10 kcal/mol found with the unsubstituted OMP substrate. These two results—reduction of the proton affinity and increase in the facility of decarboxylation—make qualitative sense, given the electron-withdrawing nature of the fluorine. While these calculations do not consider the

environment of the enzyme, they suggest that, through raising the energy of the intermediate but diminishing the barrier to decarboxylation, the overall effective barrier (see Figure 10) remains approximately the same, leading to an observed k_{cat}/K_M that is similar to that of the unsubstituted OMP. In addition, an MD simulation of the enzyme in complex with the 5-F substrate analogue has revealed that the distance between $N\zeta$ of Lys72 and C5 of the ligand throughout the 150-300 ps interval averages only 3.6 Å, which is shorter than the average distance between the two atoms in the enzyme-OMP complex trajectory and could rationalize the greater k_{cat} for the 5-F analogue. Shostak and Jones also found that the 5-aza substrate analogue was a good substrate for ODCase, but that the 5-Cl and 5-Br analogues were inhibitors [8]. Our calculations for the unsubstituted substrate and the 5-F analogue suggest that the 5-Cl and 5-Br analogues would also have PAs comparable to that of the unsubstituted substrate, as well as a low barrier to decarboxylation, as found for the unsubstituted substrate. Thus, it is hard to explain why the 5-Cl and 5-Br analogues are not substrates, but perhaps their larger size and greater hydrophobicity cause them to move toward the hydrophobic region of the binding site (away from the four charged residues, thus placing the C5 farther from the $N\zeta$ of Lys72). Clearly, further calculations or experiments (*e.g.*, solving the crystal structure of the enzyme in complex with either its 5-Cl or 5-Br inhibitor) are necessary to sort this out.

Finally, our mechanism is consistent with the fact that replacement of O2 with sulfur has a large effect on k_{cat} , reducing it by more than 4-5 orders of magnitude, whereas replacement of O4 has little effect on k_{cat} [8]. We have shown that the energies for C5-protonation and subsequent decarboxylation are little affected by O4 → S4 or O2

→ S2 substitution, but the structure of the S2-substituted complex is significantly different from that of the substrate or the S4-substituted complex. Indeed, MD simulations of 2-thio-OMP show that its phosphoribosyl group occupies a significantly different position in the active site than the OMP substrate or 4-thio-OMP, leading to a much less favorable orientation of the orotate ring for protonation of C5 by Lys72. This observation ties in nicely with the results of Wolfenden *et al.* [17], which show a greater than 10^7 -fold reduction in the k_{cat}/K_M upon removal of phosphoryl contacts with Tyr206 and Arg203 in the yeast enzyme via mutations to alanine residues. In our proposed mechanism, the O2 → S2 substitution places the substrate in a position where it is much less favorable for the lysine residue to protonate C5, thus significantly increasing the barrier to enzyme catalysis.

Our mechanism makes a rather interesting prediction for the enzyme-catalyzed reaction: if it is carried out in D_2O , the product could have some amount of deuterium at C5 as well as complete deuteration at C6. Simple unimolecular cleavage suggested by Wu *et al.* [13], O4-protonation suggested by Lee and Houk [4], and C6-protonation mechanisms suggested by Appleby *et al.* [9] and Harris *et al.* [10] would not incorporate any D at C5. Thus, product analysis by NMR or mass spectroscopy could shed light on our proposed mechanism. On the other hand, if the proton delivered by the lysine residue is the one that is preferentially removed or transferred from C5, once the CO_2 departs from the substrate, one could see significantly less than 50% D at C5 after the reaction. There are many precedents for such regioselectivity in enzyme reactions, such as in triose phosphate isomerase, where Glu165 always abstracts the pro-R hydrogen at C1 and delivers the proton to the re face of C3. In ODCase, one face of the bound ring has the

hydrophilic groups such as Lys72 and Asp70 and the other, mainly hydrophobic side chains (*e.g.*, so the H(D) lost from C5 could predominantly be the D).

To summarize, an observation of D-incorporation at C5 upon running the enzymatic reaction in D₂O would strongly support our mechanism, but the absence of significant D at C5 would not rule it out. Once the CO₂ of the substrate leaves, there are two hydrogens at C5 and none at C6. To probe whether the substrate reaches its more stable state with one hydrogen at C5 and one at C6 by an intramolecular C5 → C6 hydride transfer, we calculated the energies for this process at the MP2/6-31+G**//HF/6-31+G* level and found the barrier for it to be 28 kcal/mol. Thus, it is very likely that the hydrogen on C5 transfers to Lys72, particularly given that the loss of CO₂ dramatically reduces the favorable energy for C5-protonation (Figure 8). Independently, the neutral Asp70 or a water molecule donates a proton to C6, given the instability of the carbanion at C6, thereby regenerating the system for another round of catalysis. Thus, once the C6-CO₂ bond breaks, both deprotonation of C5 and protonation of C6 by the environment would occur in an independent and rapid manner.

We have considered five reaction mechanisms for the decarboxylation of OMP catalyzed by ODCase (Figure 12). The first is simple decarboxylation, in which ground-state destabilization drives catalysis. Although supported by the QM/MM and free energy calculations by Wu *et al.* [13], it is nonintuitive for such a mechanism to produce such a large catalytic effect, since the ground state and transition state are similarly charged. Indeed, the free energy calculations by Wu *et al.* found that discharging the ground state and transition state gave similar $\Delta\Delta G^\ddagger$'s, although their calculated reaction profile reproduces a very large effect of the enzyme. Likewise, our MM-PBSA free

energy calculations found similar binding free energies for the enzyme-bound ground state and transition state. These two pieces of data argue against preferential stabilization of the transition state or destabilization of the ground state for decarboxylation.

Furthermore, our pK_a calculations have found that Asp70 prefers to be neutral in the presence of the substrate rather than creating electrostatic stress with the substrate in its anionic form.

The other four mechanisms for ODCase suggest that the enzyme mechanism is fundamentally different than in solution, with protonation of either O2, O4, C6, or C5 preceding decarboxylation. Given the location of Lys72 in the enzyme, the environment of O2 and O4, and the fact that O4 \rightarrow S4 has little effect on k_{cat} , the O2-protonation and O4-protonation mechanisms are unlikely. During our MD simulations, Lys72 tended to be approximately equidistant from C5 and C6, suggesting that this group was reasonably located to protonate either site.

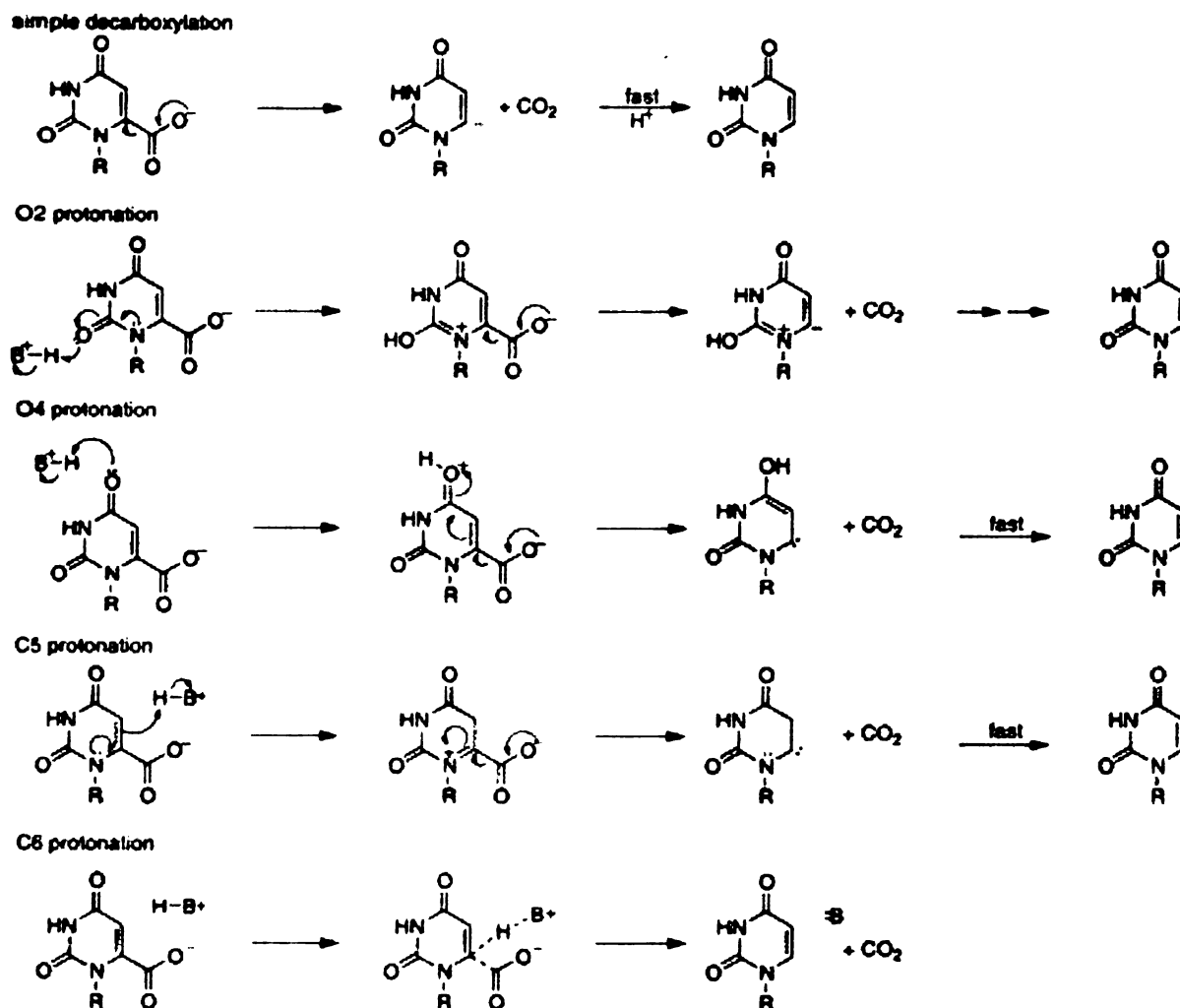


Figure 12. Reaction mechanisms considered in this study.

Our QM calculations found that C5-protonation is favored over C6-protonation, although our QM energy profiles do not consider any proton tunneling effects that would presumably lower the barriers to protonation. C6-protonation would immediately lead to decarboxylation, whereas C5-protonation would lead to a metastable intermediate, whose barrier to loss of CO₂ is only ~10 kcal/mol (Figure 9). Thus, our calculations do not fully rule out C6-protonation since we have not performed FE calculations on an intermediate

with C6-protonation. It is conceivable that environmental effects would also stabilize protonation at this position. Based on our lower calculated barrier for C5-protonation and the implication by Ehrlich *et al.* [5] of an equilibrium protonation step, our results would favor C5- over C6-protonation.

We should note, however, that accurate calculations on the enzyme environment effect were not performed along the entire reaction profile. Furthermore, Warshel *et al.* [16] obtained very large TS stabilization in a consistent calculation that involved both energy profile and binding calculations. This might be significant since the intermediate in the calculations by Warshel *et al.* (uridine⁻ + Lys72⁺ + CO₂ + Asp70⁻) is similar to ours (uridine⁻ + Lys72⁺ + CO₂ + Asp70).

In conclusion, we have presented results using a wide variety of computational methods-QM calculations, classical free energy calculations, molecular dynamics simulations, and MM-PBSA free energy calculations-to support the hypothesis that ODCase has a different mechanism of decarboxylation than that found in the reference reaction in solution. In the enzyme mechanism, we propose that C5-protonation precedes decarboxylation. There are a number of uncertainties in our results, most of which arise from not having the exact geometry of the enzyme-substrate complex and the large size of this enzyme system, which precludes the use of accurate long-range (*e.g.*, particle mesh Ewald) electrostatics. Thus, the mechanism of ODCase is far from resolved. Nonetheless, some of the structure/activity relationships (high rate of catalysis of the 4-thio substrate analogue, but low activity of the 2-thio analogue; and the high rate of catalysis of the 5F analogue, but not the 5Cl analogue) argue for a mechanism which

involves precise placement of groups in the enzyme (*e.g.*, proton transfer) rather than just generalized electrostatic stabilization/destabilization during unimolecular decomposition.

Acknowledgments:

This article is dedicated to Peter A. Kollman, a wonderful mentor and friend who succumbed to cancer during the lengthy review process, but whose enthusiasm to see this work published never waned. The authors thank B. Kuhn for his assistance and methodology for computing pKa values, and M. Saunders for providing the QUIVER program. We also thank S. Miller and I. Kuntz for helpful discussions. This work was supported in part by a National Institutes of Health (NIH) Grant (GM29072) to P.A.K., an NIH training grant to T.-S.L. (F32-GM19410-02), a National Science Foundation (NSF) Fellowship to L.T.C., and an NIH training grant (GM08284) to J.D.C. J.D.C. is a Howard Hughes Medical Institute Predoctoral Fellow. Graphics were provided by the Computer Graphics Laboratory, University of California, San Francisco (T. Ferrin, P.I., NIH P41 Grant RR-01081). NSF is acknowledged for providing computational resources at the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign.

References:

1. Radzicka, A. and R. Wolfenden, *A proficient enzyme*. Science, 1995. **267**(5194): p. 90-93.
2. Beak, P. and B. Siegel, *Mechanism of Decarboxylation of 1,3-Dimethylorotic Acid. A Possible Role for Orotate Decarboxylase*. J. Am. Chem. Soc., 1976. **98**: p. 3601-3606.
3. Silverman, R. and M. Groziak, J. Am. Chem. Soc., 1982. **104**: p. 6434-6439.
4. Lee, J.K. and K.N. Houk, *A proficient enzyme revisited: the predicted mechanism for orotidine monophosphate decarboxylase*. Science, 1997. **276**(5314): p. 942-5.
5. Ehrlich, J.I., et al., *Evidence for a stepwise mechanism of OMP decarboxylase*. Journal of the American Chemical Society, 1999. **121**(29): p. 6966-6967.
6. Acheson, S.A., et al., *Orotidine-5'-monophosphate decarboxylase catalysis: kinetic isotope effects and the state of hybridization of a bound transition-state analogue*. Biochemistry, 1990. **29**(13): p. 3198-202.
7. Smiley, J.A., et al., *Investigation of the enzymatic mechanism of yeast orotidine-5'-monophosphate decarboxylase using ¹³C kinetic isotope effects*. Biochemistry, 1991. **30**(25): p. 6216-23.
8. Shostak, K. and M.E. Jones, *Orotidylate decarboxylase: insights into the catalytic mechanism from substrate specificity studies*. Biochemistry, 1992. **31**(48): p. 12155-61.
9. Appleby, T.C., et al., *The crystal structure and mechanism of orotidine 5'-monophosphate decarboxylase*. Proc Natl Acad Sci U S A, 2000. **97**(5): p. 2005-10.
10. Harris, P., et al., *Structural basis for the catalytic mechanism of a proficient enzyme: orotidine 5'-monophosphate decarboxylase*. Biochemistry, 2000. **39**(15): p. 4217-24.
11. Feng, W.Y., et al., *The mechanism of orotidine 5'-monophosphate decarboxylase: catalysis by destabilization of the substrate*. Biochemistry, 2000. **39**(7): p. 1778-83.
12. Miller, B.G., et al., *Anatomy of a proficient enzyme: the structure of orotidine 5'-monophosphate decarboxylase in the presence and absence of a potential transition state analog*. Proc Natl Acad Sci U S A, 2000. **97**(5): p. 2011-6.

13. Wu, N., et al., *Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase*. Proc Natl Acad Sci U S A, 2000. **97**(5): p. 2017-22.
14. Smiley, J.A. and M.E. Jones, *A unique catalytic and inhibitor-binding role for Lys93 of yeast orotidylate decarboxylase*. Biochemistry, 1992. **31**(48): p. 12162-8.
15. Houk, K.N., et al., *Crystal Structures of Orotidine Monophosphate Decarboxylase: Does the Structure Reveal the Mechanism of Nature's Most Proficient Enzyme?* Chembiochem Europ J Chem Biol, 2001. **2**(2): p. 113-8.
16. Warshel, A., et al., *Remarkable rate enhancement of orotidine 5'-monophosphate decarboxylase is due to transition-state stabilization rather than to ground-state destabilization*. Biochemistry, 2000. **39**(48): p. 14728-38.
17. Miller, B.G., et al., *Contribution of enzyme-phosphoribosyl contacts to catalysis by orotidine 5'-phosphate decarboxylase*. Biochemistry, 2000. **39**(28): p. 8113-8.
18. Warshel, A., et al., *Circe Effect versus Enzyme Preorganization: What Can Be Learned from the Structure of the Most Proficient Enzyme?* Chembiochem Europ J Chem Biol, 2001. **2**(2): p. 109-11.
19. Stanton, R.V., et al., *Combined ab initio and free energy calculations to study reactions in enzymes and solution: Amide hydrolysis in trypsin and aqueous solution*. Journal of the American Chemical Society, 1998. **120**(14): p. 3448-3457.
20. Kuhn, B. and P.A. Kollman, *QM-FE and molecular dynamics calculations on catechol O-methyltransferase: Free energy of activation in the enzyme and in aqueous solution and regioselectivity of the enzyme-catalyzed reaction*. Journal of the American Chemical Society, 2000. **122**(11): p. 2586-2596.
21. Srinivasan, J., et al., *Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices*. Journal of the American Chemical Society, 1998. **120**(37): p. 9401-9409.
22. Frisch, M.J., et al., *Gaussian 98*. revision A.6. 1998, Pittsburgh, PA: Gaussian, Inc.
23. Bigeleisen, J. and M. Goepfert-Mayer, *J. Chem. Phys.*, 1947. **15**: p. 261-267.
24. Saunders, M., K.E. Laidig, and M. Wolfsberg, *Theoretical Calculation of Equilibrium Isotope Effects Using Abinitio Force Constants - Application to Nmr Isotopic Perturbation Studies*. Journal of the American Chemical Society, 1989. **111**(25): p. 8989-8994.

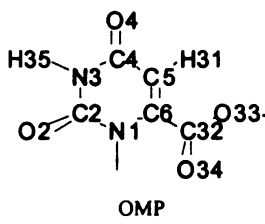
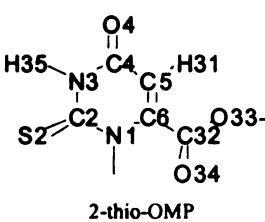
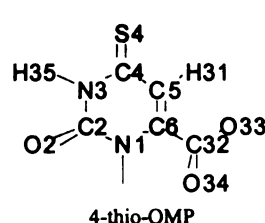
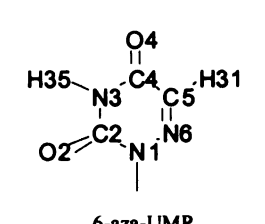
25. Chandrasekhar, J., S. Smith, and W. Jorgensen, *J. Am. Chem. Soc.*, 1985. **107**: p. 154-162.
26. Chandrasekhar, J. and W. Jorgensen, *J. Am. Chem. Soc.*, 1985. **107**: p. 2974-2975.
27. Severance, D. and W. Jorgensen, *ACS Symp. Ser.*, 1993. **568**: p. 243-259.
28. Bayly, C.I., et al., *A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model*. *Journal of Physical Chemistry*, 1993. **97**(40): p. 10269-10280.
29. Beveridge, D.L. and F.M. DiCapua, *Free energy via molecular simulation: applications to chemical and biomolecular systems*. *Annu Rev Biophys Biophys Chem*, 1989. **18**: p. 431-92.
30. Porter, D.J. and S.A. Short, *Yeast orotidine-5'-phosphate decarboxylase: steady-state and pre-steady-state analysis of the kinetic mechanism of substrate decarboxylation*. *Biochemistry*, 2000. **39**(38): p. 11788-800.
31. Case, D.A., et al., *Amber 5.0*. 1997, San Francisco, CA: University of California.
32. Jorgensen, W., et al., *J. Chem. Phys.*, 1983. **79**: p. 926-935.
33. Cornell, W.D., et al., *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules*. *Journal of the American Chemical Society*, 1995. **117**(19): p. 5179-5197.
34. Ryckaert, J., G. Ciccotti, and H. Berendsen, *J. Comput. Phys.*, 1977. **23**: p. 327-341.
35. Berendsen, H., et al., *J. Comput. Phys.*, 1984. **81**: p. 3684-3690.
36. Sharp, K.A. and B. Honig, *Electrostatic interactions in macromolecules: theory and applications*. *Annu Rev Biophys Biophys Chem*, 1990. **19**: p. 301-32.
37. Weiser, J., P.S. Shenkin, and W.C. Still, *Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO)*. *Journal of Computational Chemistry*, 1999. **20**(2): p. 217-230.
38. Jayaram, B., D. Sprous, and D.L. Beveridge, *Solvation free energy of biomacromolecules: Parameters for a modified generalized born model consistent with the AMBER force field*. *Journal of Physical Chemistry B*, 1998. **102**(47): p. 9571-9576.

39. Case, D.A., et al., *AMBER 6.0*. 1999, San Francisco, CA: University of California.
40. Warshel, A., *Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites*. J Biol Chem, 1998. **273**(42): p. 27035-8.
41. Warshel, A. and J. Florian, *Computer simulations of enzyme catalysis: finding out what has been optimized by evolution*. Proc Natl Acad Sci U S A, 1998. **95**(11): p. 5950-5.
42. Kollman, P.A., et al., *Elucidating the nature of enzyme catalysis utilizing a new twist on an old methodology: quantum mechanical-free energy calculations on chemical reactions in enzymes and in aqueous solution*. Acc Chem Res, 2001. **34**(1): p. 72-9.
43. Warshel, A., *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. 1991, New York: Wiley.
44. Donini, O., T. Darden, and P.A. Kollman, *QM-FE calculations of aliphatic hydrogen abstraction in citrate synthase and in solution: Reproduction of the effect of enzyme catalysis and demonstration that an enolate rather than an enol is formed*. Journal of the American Chemical Society, 2000. **122**(49): p. 12270-12280.
45. Chong, L.T., et al., *Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7*. Proc Natl Acad Sci U S A, 1999. **96**(25): p. 14330-5.
46. Wang, Y.X., et al., *Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272*. Biochemistry, 1996. **35**(31): p. 9945-50.
47. Luo, R., et al., *pK(a) shifts in small molecules and HIV protease: Electrostatics and conformation*. Journal of the American Chemical Society, 1998. **120**(24): p. 6138-6146.
48. Trylska, J., et al., *Thermodynamic linkage between the binding of protons and inhibitors to HIV-1 protease*. Protein Sci, 1999. **8**(1): p. 180-95.
49. Rishavy, M.A. and W.W. Cleland, *Determination of the mechanism of orotidine 5'-monophosphate decarboxylase by isotope effects*. Biochemistry, 2000. **39**(16): p. 4569-74.
50. Moran, G.R., et al., *On the catalytic mechanism of tryptophan hydroxylase*. Journal of the American Chemical Society, 2000. **122**(19): p. 4535-4541.

51. Tantillo, D.J. and K.N. Houk, *Fidelity in Hapten Design: How Analogous Are Phosphonate Haptens to the Transition States for Alkaline Hydrolyses of Aryl Esters?* J Org Chem, 1999. 64(9): p. 3066-3076.
52. Chen, X.Y., P.J. Berti, and V.L. Schramm, *Transition-state analysis for depurination of DNA by ricin A-chain.* Journal of the American Chemical Society, 2000. 122(28): p. 6527-6534.
53. Ferrin, T., et al., *The MIDAS display system.* J. Mol. Graphics, 1988. 6: p. 13-27.

Appendix:

Table A. Atom types and partial charges of ligands.

	name	type	charge
 <p>OMP</p>	N1	N*	0.07317
	C2	C	0.58651
	N3	NA	-0.46189
	C4	C	0.72321
	C5	CM	-0.30652
	C6	CM	-0.04823
	O2	O	-0.65236
	O4	O	-0.68349
	C32	C	0.97450
	O33	O2	-0.81585
	O34	O2	-0.81585
	H31	HA	0.11238
	H35	H	0.31442
	 <p>2-thio-OMP</p>	N1	N*
C2		C	-0.01057
N3		NA	-0.11316
C4		C	0.65243
C5		CM	-0.21965
C6		CM	-0.07615
S2		SD	-0.50995
O4		O	-0.65887
C32		C	0.94958
O33		O2	-0.80537
O34		O2	-0.80537
H31		HA	0.09905
H35		H	0.21142
 <p>4-thio-OMP</p>		N1	N*
	C2	C	0.54726
	N3	NA	-0.08368
	C4	C	0.00978
	C5	CM	0.10357
	C6	CM	-0.06857
	O2	O	-0.63302
	S4	SD	-0.51726
	C32	C	0.95085
	O33	O2	-0.80021
	O34	O2	-0.80021
	H31	HA	0.02022
	H35	H	0.21638
	 <p>6-aza-UMP</p>	N1	N*
C2		C	0.45733
N3		NA	-0.34275
C4		C	0.56115
C5		CM	0.35510
N6		NC	-0.46094
O2		O	-0.57740
O4		O	-0.57397
H31		HA	-0.01475
H35		H	0.32241

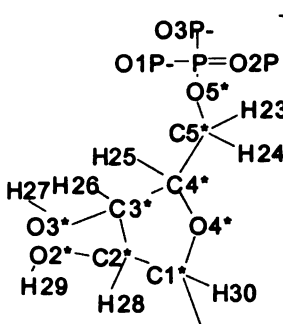
	label	type	charge
 <p>phosphoribosyl group</p>	C1*	CT	-0.14840
	C2*	CT	0.11558
	C3*	CT	0.43058
	C4*	CT	0.11307
	O2*	OH	-0.77247
	O3*	OH	-0.82734
	O4*	OS	-0.48084
	C5*	CT	0.03216
	O5*	OS	-0.57050
	P	P	1.59321
	O1P	O2	-0.99306
	O2P	O2	-0.99306
	O3P	O2	-0.99306
	H30	H2	0.15878
	H28	H1	0.08424
	H26	H1	-0.01628
	H25	H1	0.17510
	H29	HO	0.49639
H27	HO	0.49722	
H23	H1	0.04934	
H24	H1	0.04934	

TABLE B. Additional force field parameters used for ligands.

MASS	<i>m</i> (amu)				
ZC	12.01				
SD	32.06				
BOND	<i>K_r</i> (kcal mol ⁻¹ Å ⁻²)	<i>r_{eq}</i> (Å)			
C-SD	395.0	1.69	*		
N*-NC	448.0	1.365	=CM-N*		
CM-NC	461.0	1.354	=CB-NC		
CM-ZC	410.0	2.40	** <i>K_r</i> from CM-C		
ZC-O2	656.0	1.16	** <i>K_r</i> from C-O2		
ANGLE	<i>K_θ</i> (kcal mol ⁻¹ rad ⁻²)	<i>θ_{eq}</i> (deg)			
N*-C-SD	80.0	122.9	*		
NA-C-SD	80.0	122.9	*		
CM-C-SD	80.0	125.3	=CM-C-O		
N*-CM-C	70.0	121.2	=CM-CM-N*		
CM-C-O2	70.0	117.0	=CT-C-O2		
CT-N*-NC	70.0	121.2	=CM-N*-CT		
N*-NC-CM	70.0	121.2	=CM-CM-N*		
C-N*-NC	70.0	121.6	=C-N*-CM		
C-CM-NC	63.0	120.7	=C-CM-CM		
NC-CM-HA	50.0	119.7	=CM-CM-HA		
N*-CM-ZC	70.0	129.53	** <i>θ_{eq}</i> from CM-CM-N*		
CM-ZC-O2	70.0	97.48	** <i>θ_{eq}</i> from CT-C-O2		
O2-ZC-O2	80.0	159.49	** <i>θ_{eq}</i> from O2-C-O2		
CM-CM-ZC	63.0	113.68	** <i>θ_{eq}</i> from CM-CM-C		
DIHEDRAL	<i>idiv</i> [†]	<i>V_n/2</i> (kcal mol ⁻¹)	<i>γ</i> (deg)	<i>n</i>	
SD-C-N-X	4	11.2	180.0	2	*
X-NC-CM-X	4	26.6	180.0	2	=X-CM-CM-X
X-N*-NC-X	4	7.4	180.0	2	=X-N*-CM-X
N*-CM-ZC-O2	4	8.7	180.0	2	=X-C-CM-X
CM-CM-ZC-O2	4	8.7	180.0	2	=X-C-CM-X
IMPROPER		<i>V_n/2</i> (kcal mol ⁻¹)	<i>γ</i> (deg)	<i>n</i>	
N-N-C-SD		10.5	180.0	2	*
NONBONDED	<i>σ</i> (Å)	<i>ε</i> (kcal/mol)			
SD	2.00	0.25	*		
ZC	1.91	0.10	=C		

*From Miyamoto and Kollman, *PROTEINS: Structure, Function, and Genetics* 16:226-245 (1993).

***r_{eq}* computed from HF/6-31+G* geometry optimization.

[†]Factor by which torsional barrier is divided.

The research described in this dissertation has helped us understand the dynamics and energetic aspects of antibody and enzyme catalysis on an atomic level. This understanding has laid a foundation for the challenging design of biocatalysts with improved activities or novel functions. In particular, we have shown that state-of-the-art molecular dynamics, free energy, and quantum mechanical approaches are viable tools for dissecting mechanisms of catalysis. Such tools can guide the improvement of energy functions used in computational protein design approaches, which have given encouraging results in recent years [1].

The first validation of the MM-PBSA free energy approach [2] for protein-ligand interactions is presented in Chapter 2, which explores the energetic aspects of affinity maturation of esterase antibody 48G7. Using the MM-PBSA approach, the experimentally measured difference in binding free energies of the hapten complexes of the germ line and mature forms of 48G7 was reproduced within 1 kcal/mol with reasonable absolute binding free energies. Further analysis of the energetic components revealed that the dramatic increase in affinity of $>10^4$ -fold upon affinity maturation of 48G7 arises due to optimization of the electrostatic interactions between the residues in the binding site and the hapten. These interactions are fine-tuned by the nine mutated residues in the mature form that are distant from the binding pocket. Scaffold residues are likely to have significant effects on binding in other systems as well and should therefore be considered in efforts towards improving binding or catalytic activity.

In Chapter 2, we have also introduced an approach for the determination of energetic contributions by individual residues. Application of this approach revealed that Tyr H33 contributes more favorably to the mature complex formation than to the germ

line complex formation due to a smaller desolvation penalty upon binding. Optimization of binding by decreasing the desolvation penalty is a less geometrically demanding strategy than optimizing specific van der Waals or hydrogen-bonding interactions. Thus, incorporation of this strategy in molecular design efforts [3, 4] is likely to be effective.

We went on to explore the mechanisms of catalysis by two esterase antibodies, 43C9 and 17E8, which are elicited by haptens that are similar to that of 48G7. Our results showed that direct hydroxide attack on the substrate, as programmed by the hapten, is a plausible mechanism for both 43C9 and 17E8. Nucleophilic attack by an active-site histidine residue of 43C9 could not, however, be ruled out. The unprogrammed presence of the nucleophilic histidine residue has been suggested to be the key to the unique amidase activity of 43C9. Even so, our results have demonstrated that active-site features of a catalytic antibody are strongly linked to the structure of the hapten. Thus, a greater understanding of this link, as is being pursued through quantum mechanical calculations by Houk and co-workers [5], should aid in the design of haptens that lead to improved catalytic antibodies.

Finally, we extended our understanding of antibody catalysis to that of orotidine 5'-phosphate decarboxylase (ODCase), the most proficient enzyme known. The combined results of molecular dynamics simulations, quantum mechanical calculations, and free energy calculations support our proposed mechanism, which was previously ruled out by non-conclusive experimental approaches. Contrary to the ground-state destabilization mechanisms that were previously proposed [6, 7], our results show that there is little differential stabilization of the ground state and transition state, which are both stabilized through binding by the enzyme. Thus, the role of binding energy in

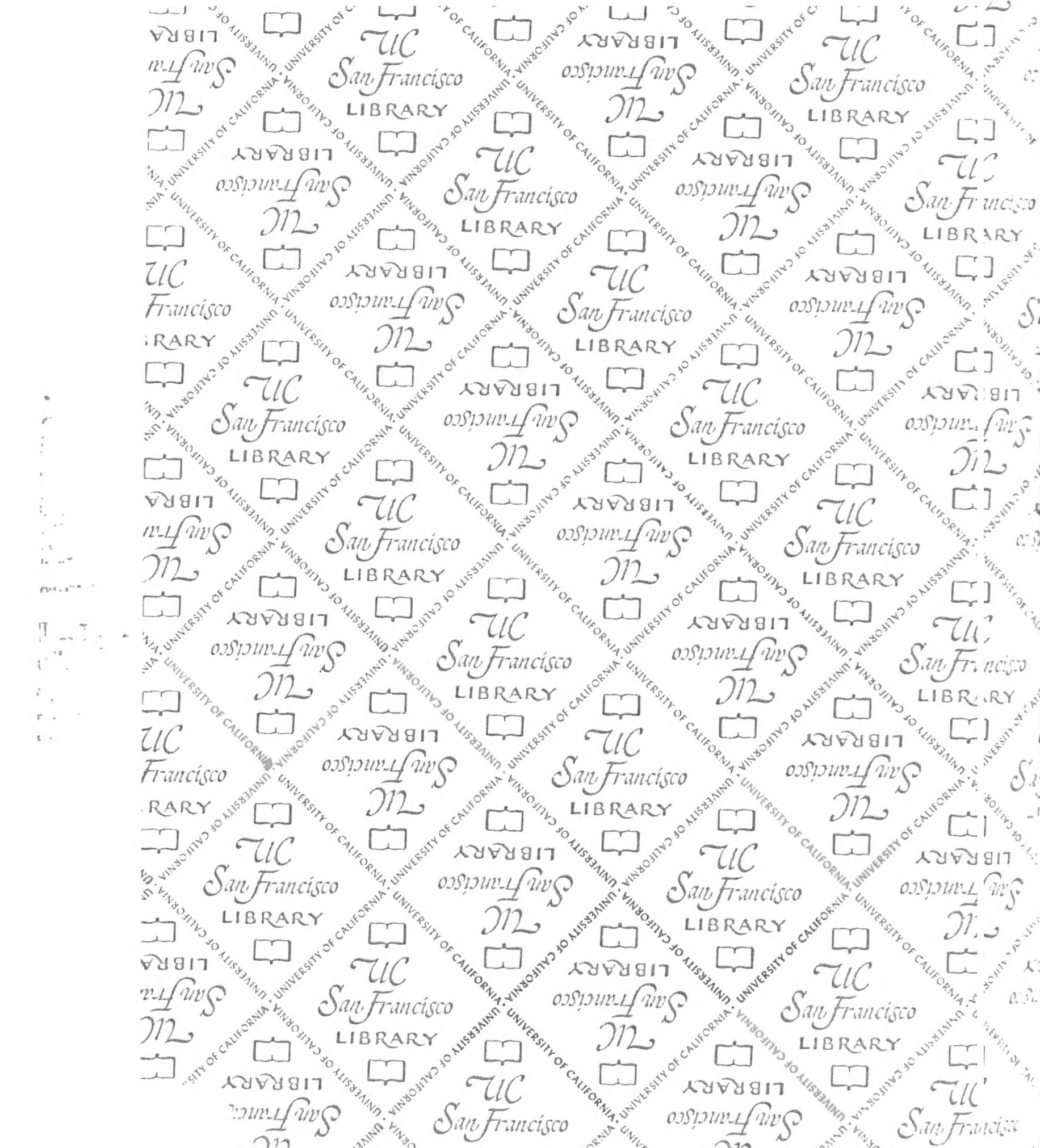
catalysis is not sufficient to explain the 10^{17} -fold rate enhancement by ODCase. Instead, key factors leading to the high catalytic proficiency include the donation of a catalytic proton by an active-site lysine residue to the C5 position of the orotate ring in the substrate and the unique microenvironment provided by the highly charged active site of the enzyme. Due to the size and large number of charged residues in this dimeric enzyme, our computational approaches afforded only semi-quantitative results. Such results, however, are still informative and contribute to the ongoing debate about the mechanism of catalysis by ODCase. This study also underscores the importance of understanding the interplay between the enzyme microenvironment, catalytic residues, and binding energy for enzyme design efforts.

To date, the use of transition-state analogs to elicit catalytic antibodies has been the most successful approach for generating novel protein catalysts. However, enzyme-like activity remains elusive for catalytic antibodies, which are impaired by factors such as product inhibition and the inefficient selection for catalytic side chains by the current technology. Now that the combinatorial processes of the immune system can be mimicked *in vitro* and that computational design approaches are capable of exploring an even larger sequence space than these experimental methods, there is no need to restrict design efforts to a single protein fold. In recent work, Bolon and Mayo [8] have demonstrated the successful computational design of esterase activity into the catalytically inert thioredoxin scaffold. The activity of this “protozyme” is inferior to that of the most efficient catalytic antibodies, but is nonetheless an encouraging step towards the *in silico* optimization of catalytic activity as well as the design of novel function. Substantially more computational efforts of this kind are underway, which should provide

insight into the relationship between catalytic activity and the protein scaffold. My doctoral work has laid a foundation for upcoming computational design efforts toward improved or novel protein catalysts. In addition, application of the approaches described in this dissertation to newly designed protozymes is likely to further our understanding of enzyme catalysis.

References:

1. Pokala, N. and T.M. Handel, *Review: Protein Design -- Where We Were, Where We Are, Where We're Going*. Journal of Structural Biology, 2001. **134**: p. 269-281.
2. Srinivasan, J., et al., *Continuum solvent studies of the stability of RNA hairpin loops and helices*. J Biomol Struct Dyn, 1998. **16**(3): p. 671-82.
3. Kangas, E. and B. Tidor, *Optimizing electrostatic affinity in ligand-receptor binding: Theory, computation, and ligand properties*. Journal of Chemical Physics, 1998. **109**(17): p. 7522-7545.
4. Lee, L.P. and B. Tidor, *Optimization of electrostatic binding free energy*. Journal of Chemical Physics, 1997. **106**(21): p. 8681-8690.
5. Tantillo, D.J. and K.N. Houk, *Fidelity of Hapten Design: How Analogous Are Phosphonate Haptens to the Transition States for Alkaline Hydrolyses of Aryl Esters?* J. Org. Chem., 1999. **64**: p. 3066-3076.
6. Appleby, T.C., et al., *The crystal structure and mechanism of orotidine 5'-monophosphate decarboxylase*. Proc Natl Acad Sci U S A, 2000. **97**(5): p. 2005-10.
7. Wu, N., et al., *Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase*. Proc Natl Acad Sci U S A, 2000. **97**(5): p. 2017-22.
8. Bolon, D.N. and S.L. Mayo, *Enzyme-like proteins by computational design*. Proc Natl Acad Sci U S A, 2001. **98**: p. 14274-14279.



For reference

Not to be taken
from the room.

