# Theoretical Calculations on Enzyme–Substrate Complexes: The Basis of Molecular Recognition and Catalysis

MATTHEW R. PINCUS

*Laboratory of Theoretical Biology, National Institutes of Health, Bethesda, Maryland 20205*

HAROLD A. SCHERAGA*

*Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853*

*Received April 16, 1981*

Of all processes involving catalysis, enzyme reactions are unique because of their remarkable specificity. Most enzyme-catalyzed reactions occur in two distinct steps: one is a binding step and the other is a catalytic one.[1] While progress has been made in understanding both of these processes, the fundamental questions of how enzymes recognize substrates and catalyze their reactions remain unanswered. Recently, theoretical techniques have been developed which are suited to the study of these problems. These may be divided into two categories: conformational energy calculations, used to compute the three-dimensional structures of enzyme–substrate complexes formed in the first step, and quantum mechanical calculations, used to compute the specific interactions between enzyme and substrate that result in rate enhancements in the second step.

It is now possible to combine both of these approaches so that quantum mechanical calculations may be applied to the structures of those noncovalent complexes (obtained by conformational energy calculations) to identify the ones that are likely to undergo catalysis or to constitute "productive" complexes. In addition, quantum mechanical calculations may be used to determine whether any given molecule, whose conformations have been obtained from ground-state calculations, will be a "good" or a "bad" substrate.

These objectives may be achieved by a quantum mechanical examination of the feasibility of proton transfers between enzyme and substrate and of bond distortions in the substrate for the lowest energy conformers calculated previously for the noncovalent complexes. The quantum mechanical calculations may be applied to the reactive portions of enzyme and substrate while the feasibility of attendant conformational changes in enzyme and substrate may be assessed by conformational energy calculations. We discuss each aspect of these calculations in turn.

Matthew R. Pincus was born in Brooklyn, NY, in 1948. He received his A.B. degree from Bowdoin College and his Ph.D. (in Biochemistry) and M.D. from S.U.N.Y. Downstate Medical Center. He was a Weizmann Fellow at the Weizmann Institute of Science and a postdoctoral fellow with Professor Scheraga at Cornell. He is currently in the Laboratory of Theoretical Biology at N.I.H.

Harold A. Scheraga was born in Brooklyn, NY, in 1921. He attended The City College of New York, where he received his B.S. degree, and went on to graduate work at Duke University, receiving the Ph.D. in 1946, and, in 1961, an Sc.D. (Hon). Following postdoctoral work at Harvard Medical School, he joined the faculty at Cornell University, where he is Todd Professor of Chemistry. His research interests are in the physical chemistry of proteins and other macromolecules, chemistry of blood clotting, and structure of water and dilute aqueous solutions.

## Conformational Energy Calculations of Structures of Enzyme–Substrate Complexes

All of these calculations are based on the assumption that any observed molecular conformation is the one of lowest free energy. If many structures contribute to the partition function for a given molecule, then any property of the system, $P$, will be a weighted normalized average over all such structures, i.e.,

$$\langle P \rangle = \int P e^{-U/kT}\, d\tau \qquad (1)$$

where $U$ is the energy of any conformation, $k$ is Boltzmann's constant, $T$ is temperature (in kelvin), and $d\tau$ is the volume element of conformation space. For large molecules, including enzyme–substrate complexes, only a few terms will contribute in eq 1, so that we are really concerned with finding the global minimum, i.e., the lowest of all possible energies satisfying the condition that

$$\frac{\partial U}{\partial q_i} = 0 \qquad (2)$$

for all $q_i$, where $q_i$ are the coordinates for the system in question. The methods used in calculating conformational energies have been reviewed extensively elsewhere.[2-6] To summarize these methods: the energy of any given conformation of a molecule is the sum of nonbonded, hydrogen-bonding, electrostatic, and torsional energy terms, i.e.,

$$U = \sum_{i \neq j} \epsilon_{ij} \left[ \left( \frac{r^0_{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{r^0_{ij}}{r_{ij}} \right)^{M} \right] + \sum_{i \neq j} \frac{q_i q_i}{D r_{ij}} + \sum_k \frac{A_k}{2} (1 \pm \cos n\, \theta_k) \qquad (3)$$

where $\epsilon_{ij}$ and $r^0_{ij}$ are the potential well depth and position of the minimum of the energy (for nonbonded energy, $M = 6$; for hydrogen-bonding energy, $M = 10$), $q$ is the partial atomic charge, $D$ is the dielectric constant, $r_{ij}$ is the distance between two interacting atoms, $A_k$ is the barrier height for rotation around the $k$th bond, $\theta_k$ is the dihedral angle, and $n$ is the $n$-fold degeneracy of the torsional potential.[2-6]

(1) M. Dixon and E. C. Webb, "Enzymes", 2nd ed., Academic Press, New York, 1964.
(2) C. Pottle, M. S. Pottle, R. W. Tuttle, R. J. Kinch, and H. A. Scheraga, *J. Comput. Chem.*, 1, 46–58 (1980).
(3) G. Némethy and H. A. Scheraga, *Q. Rev. Biophys.*, 10, 239–352 (1977).

0001-4842/81/0114-0299$01.25/0 

Table I
Conformational Parameters for the Acyl-Enzyme Complex between *N*-Acetyl-L-phenylalanine and α-Chymotrypsin[a]

| conformer | backbone conformation of *N*-acetyl-L-phenylalanine amide substrate | χ,[b] deg | rel energy,[c] kcal/mol | hydrogen bonds[d] |
|---|---|---|---|---|
| 1 | right-handed α helix | 111.8 | 0.0 | Ser-195 NH to Phe C=O<br>Phe NH to Ser-195 $O^{\gamma}$<br>Phe NH to His-57 $N^{\epsilon_2}$<br>Asp-194 NH to Phe C=O<br>Gly-193 NH to Phe C=O |
| 2 | extended conformation | 131.7 | 2.9 | |
| 3 | extended conformation | −155.3 | 8.4 | Phe NH to Ser-214 C=O |

[a] This table is adapted from ref 19. [b] χ is the dihedral angle involving the ester bond between enzyme and substrate. It is defined in terms of $C^{\beta}$ of Ser-195, $O^{\gamma}$ of Ser-195, C′ of the Phe substrate, and $C^{\alpha}$ of the Phe substrate. [c] All energies expressed relative to that of conformer 1. [d] Only hydrogen bonds *between* enzyme and substrate are reported here.

These calculations have been applied to a wide variety of molecules ranging from terminally blocked amino acids[7] to large peptides including gramicidin S[8] and synthetic models of collagen,[9-12] all with excellent agreement between predicted structures and experimentally determined ones. They are readily extended to the problem of predicting the structures of enzyme–substrate complexes.

The conformational energy ($E_{TOT}$) of an enzyme–substrate complex may be expressed as

$$E_{TOT} = E_{SUB} + E_{ENZ} + E_{INT} \quad (4)$$

where $E_{SUB}$, $E_{ENZ}$, and $E_{INT}$ are the conformational energy of the substrate, enzyme, and interaction between enzyme and substrate, respectively. This energy is, therefore, a function of the dihedral angles of the substrate and of those of the enzyme, and of the six external degrees of freedom (three translational and three rotational) of the substrate. Because of the large number of degrees of freedom and the many interactions present in this type of calculation, there are many solutions to eq 2, i.e., many minima exist. To find the global minimum, it is necessary to search conformational space in a systematic way.

In these calculations, the effects of solvent on the conformational energies are omitted. These effects would presumably influence such thermodynamic quantities as the enthalpy of binding in solution. Recently, an approach for estimating the effects of solvation on the thermodynamics of binding (given a specific structure for the enzyme–substrate complex) has been proposed.[13] The objective of the calculations discussed here, however, is to predict *structures* as they would occur in crystals where direct interactions between enzyme and substrate predominate. Most water molecules in crystal structures of enzyme–inhibitor complexes, wherever their positions can be inferred accurately, occur around the surface of the enzyme and only sparsely within the active-site cleft. Thus, the effects of solvent are not expected to alter the conformations of the complexes in any significant way.

We now show how the methods described above have been applied to the prediction of the structures of enzyme–substrate complexes of chymotrypsin and lysozyme.

## α-Chymotrypsin

This enzyme hydrolyzes peptide bonds in proteins, with a marked propensity to hydrolyze the bonds at the carboxyl ends of aromatic amino acid residues. The mechanism of catalysis involves nucleophilic attack of a highly reactive serine, Ser-195, with the aid of the neighboring His-57, on the peptide carbonyl group to yield an acyl-enzyme intermediate,[14,15] which is hydrolyzed by $H_2O$ in a subsequent step. The three-dimensional structure of this enzyme, determined by X-ray crystallographic methods,[16] shows that the active site lies in a cleft lined by nonpolar residues, in particular Phe-41 and Trp-215.[16]

Using the conformational energy approach, Platzer et al. mapped the potential energy surface for this enzyme with the simple substrate, *N*-acetyl-L-phenylalaninamide (NAPA).[17-20] Four regions of good binding were found near the active site. More detailed exploration of these sites revealed that, in many binding conformations, the carbonyl group of the scissile peptide bond was situated close to $O^{\gamma}$ of Ser-195. In all binding conformations, the phenyl group of the substrate made good contacts with either of the two nonpolar residues in the cleft, Phe-41 and Trp-215,[18,19] indicating the im-

(4) F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361–2381 (1975).
(5) F. A. Momany, L. M. Carruthers, R. F. McGuire, and H. A. Scheraga, *J. Phys. Chem.*, **78**, 1595–1620 (1974).
(6) F. A. Momany, L. M. Carruthers, and H. A. Scheraga, *J. Phys. Chem.*, **78**, 1621–1630 (1974).
(7) S. S. Zimmerman, G. Némethy, M. S. Pottle, and H. A. Scheraga, *Macromolecules*, **10**, 1–9 (1977).
(8) M. Dygert, N. Go, and H. A. Scheraga, *Macromolecules*, **8**, 750–761 (1975).
(9) M. H. Miller and H. A. Scheraga, *J. Polym. Sci., Polym. Symp.*, **54**, 171–200 (1976).
(10) M. H. Miller, G. Némethy, and H. A. Scheraga, *Macromolecules*, **13**, 470–478 (1980).
(11) M. H. Miller, G. Némethy, and H. A. Scheraga, *Macromolecules*, **13**, 910–913 (1980).
(12) G. Némethy, M. H. Miller, and H. A. Scheraga, *Macromolecules*, **13**, 914–919 (1980).
(13) J. Moult, D. M. Barnett, and A. T. Hagler in "Molecular Mechanisms of Biological Recognition, Proceedings of the Sixth Aharon Katzir-Katchalsky Conference", M. Balaban, Ed., Elsevier, Amsterdam, 1979, pp 131–143.
(14) M. L. Bender, F. J. Kézdy, and C. R. Gunter, *J. Am. Chem. Soc.*, **86**, 3714–3721 (1964).
(15) M. L. Bender and F. J. Kézdy, *Annu. Rev. Biochem.*, **34**, 49–76 (1965).
(16) J. J. Birktoft and D. M. Blow, *J. Mol. Biol.*, **68**, 187–240 (1972).
(17) K. E. B. Platzer, F. A. Momany, and H. A. Scheraga, *Int. J. Peptide Protein Res.*, **4**, 187–200 (1972).
(18) K. E. B. Platzer, F. A. Momany, and H. A. Scheraga, *Int. J. Peptide Protein Res.*, **4**, 201–219 (1972).
(19) K. E. B. Platzer, Ph.D. Dissertation, Cornell University (1973); University Microfilms, Ann Arbor, MI, Order No. 73-22,521 [*Diss. Abstr. Int. B*, **34**, 1020–1021 (1973)].
(20) H. A. Scheraga in "Current Topics in Biochemistry—1973", C. B. Anfinsen and A. N. Schechter, Eds., Academic Press, New York, 1974, pp 1–42.

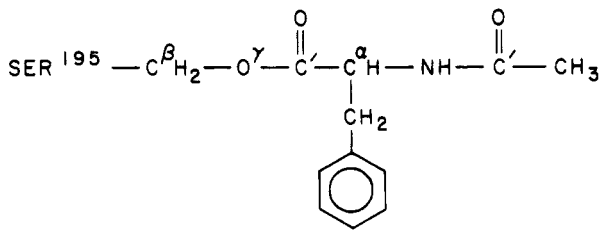


**Figure 1.** Acyl-enzyme intermediate of *N*-acetyl-L-phenylalanine.

portance of such contacts in the recognition process. Using the results for the binding of NAPA, these authors computed the binding conformations of three other simple peptides, two containing tyrosine and one containing tryptophan. In the case of the tryptophan-containing peptide, *N*-acetyl-L-tryptophanamide, the calculated coordinates for the indole ring in the active site were compared with those determined for a similar substrate by X-ray crystallography (subjected to energy refinement),[18,19,21] with close agreement between theory and experiment.[18,19]

Further, the calculated relative binding energies for each of the four simple substrates correlated well with the relative affinities determined in solution studies,[18,19] *N*-acetyl-L-tryptophanamide having the highest and *N*-acetyl-L-phenylalaninamide the lowest affinity. The binding of larger substrates was then studied with a flexible enzyme whose active-site side chains were allowed to move. A total of four binding subsites in addition to the main hydrophobic site have been identified.[19,20]

These calculations were extended to the study of the structure of the acyl-enzyme intermediate of *N*-acetyl-L-phenylalanine, shown in Figure 1. An extensive search for all allowed conformations for this acyl-enzyme intermediate resulted in only three low-energy minima (listed in Table I), indicating that this good deacylating substrate may be "frozen" in its catalytically active conformation.[14] The lowest energy conformer, no. 1 in Table I and shown in Figure 2, is involved in a network of strong hydrogen bonds with the enzyme, which include (among others) the carbonyl oxygen of the substrate and the backbone NH's of Gly-193 and Ser-195. A similar hydrogen-bonding arrangement has been postulated to facilitate nucleophilic attack by a water molecule on the carbonyl group by stabilizing incipient charge on the oxyanion of the forming tetrahedral intermediate.[22] A recent calculation[23] of the interactions of the tetrahedral intermediate between D and L isomers of *N*-acetyl-L-tryptophan derivatives and the enzyme has confirmed the importance of these hydrogen bonds.

None of the conformers in Table I have low-energy ester conformations; i.e., the dihedral angle $\chi$ (defined in footnote *b*) departs significantly from its most stable trans value of 180°, implying significant disruption of p-orbital overlap between $O^{\gamma}$ and the carbonyl group.[19,24] Such a disruption of the $\pi$-electron system has been postulated to occur on the basis of spectral shifts observed in acyl derivatives of the enzymes.[24] The resulting polarization of the C=O group would increase its electrophilicity toward the incoming nucleophile, $H_2O$.

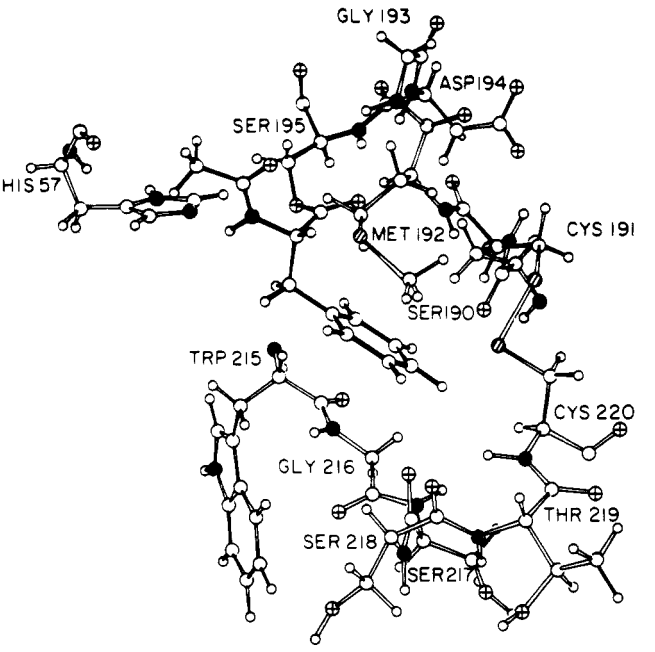


**Figure 2.** Lowest energy conformer for the acyl-enzyme intermediate[19,20] (conformer 1 in Table I) for *N*-acetyl-L-phenylalanine and $\alpha$-chymotrypsin. The symbols O, ●, ⊕, ◉, and ○ represent C, N, O, S, and H atoms, respectively.

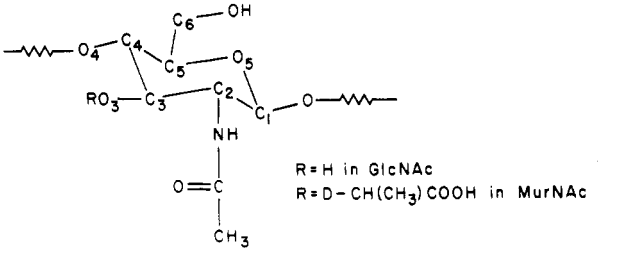


**Figure 3.** Structures of GlcNAc and MurNAc.

The calculations on enzyme–substrate complexes of chymotrypsin have thus predicted not only experimentally observed structures[18,19] but also observed binding trends and have provided insights into possible mechanisms of catalysis.

## Lysozyme

This enzyme hydrolyzes cell wall polysaccharides containing alternating copolymers (in $\beta$-1,4 linkage) of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc)[25] whose structures are shown in Figure 3.

Solution studies indicate that there are six subsites labeled A through F (see Figures 4 and 5) for the binding of polysaccharide substrates.[25] Maximal rates of hydrolysis and the highest affinities are achieved when six saccharide units are bound to the enzyme.[25] Hydrolysis occurs between sites D and E[25] in a cleft lined by the two catalytic residues Glu-35 and Asp-52. The mechanism is thought to involve a carbonium ion at $C_1$ of the D-site residue, stabilized by the carboxylate anion of Asp-52, with Glu-35 serving as a general acid that protonates the oxygen of the departing saccharide.[25] The X-ray crystal structure of the oligomer

(21) T. A. Steitz, R. Henderson, and D. M. Blow, *J. Mol. Biol.*, **46**, 337–348 (1969).
(22) R. Henderson, *J. Mol. Biol. 54*, 341–354 (1970).
(23) D. F. DeTar, *J. Am. Chem. Soc.*, **103**, 107–110 (1981).
(24) S. A. Bernhard and S. J. Lau, *Cold Spring Harbor Symp. Quant. Biol.*, **36**, 75–83 (1971).
(25) T. Imoto, L. N. Johnson, A. C. T. North, D. C. Phillips, and J. A. Rupley in "The Enzymes", Vol. VII, P. D. Boyer, Ed., Academic Press, New York and London, 1972, pp 665–868.

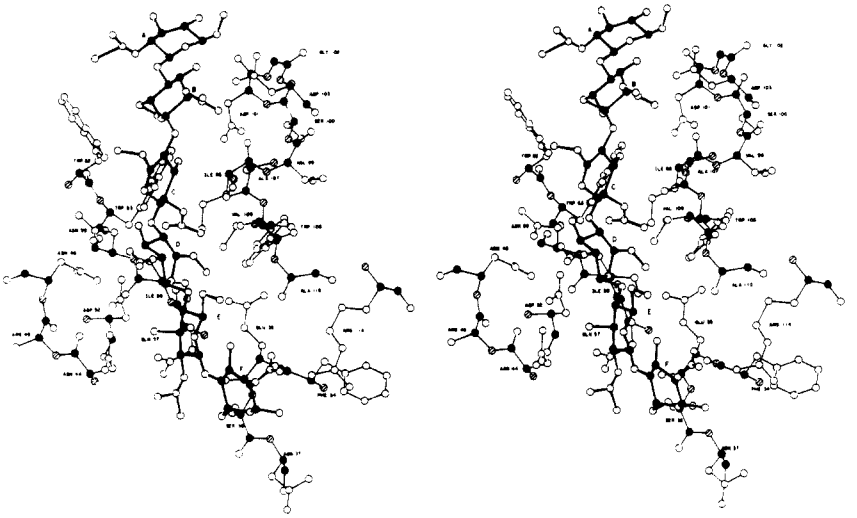

**Figure 4.** "Right-sided" binding mode for $(GlcNAc)_6$ bound to lysozyme.[30]

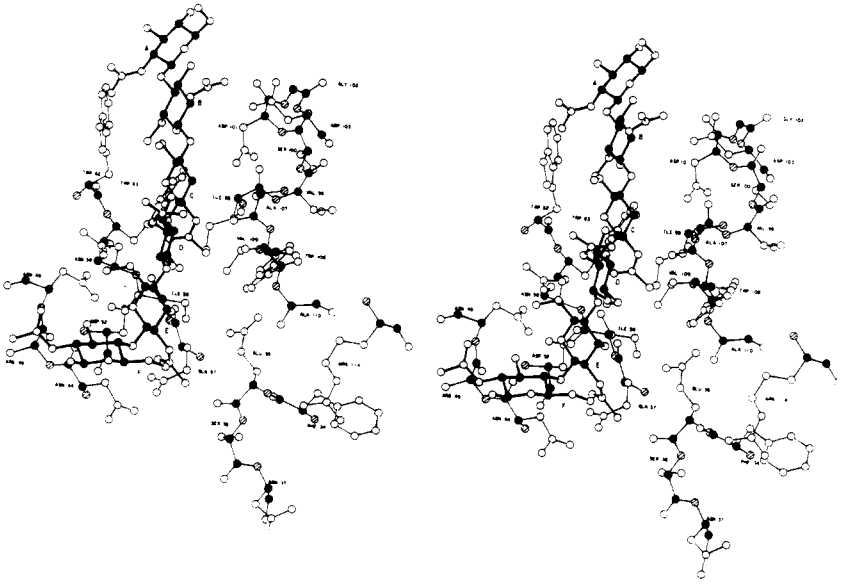

**Figure 5.** "Left-sided" binding mode for $(GlcNAc)_6$ bound to lysozyme.[30]

$(GlcNAc)_3$ bound to sites A–C of lysozyme was determined.[25] Three more residues were added in sites D–F by model building[25] (Figure 4). Because the residue in the D site was found to make bad contacts with the enzyme, it was deemed necessary to distort the D ring of this saccharide from the normally most stable chair form to a half-chair form, which was thought to be favorable for catalysis because it resembled the transition state (a carbonium ion would force the sugar ring into a half-chair form).[25] Whether distortion of the ring is necessary is a question easily answered by conformational energy calculations.

The conformational energy of the model-built structure was subjected to an energy minimization[26] in which the bond lengths, bond angles, and dihedral angles of the substrate were all allowed to vary. The most striking finding was that the distorted ring in site D relaxed to a full chair form while the D-site saccharide had a high affinity for this site. If the D-site residue were allowed to assume a series of conformations between chair and half-chair forms, the nonbonded energy was found to remain constant while only the strain energy increased with increasing ring distortion.[26] These results strongly suggested[26] that the proposed mechanism for lysozyme action involving ground-state distortion had to be modified and that a more systematic search for substrate binding conformations was necessary.

A systematic procedure for the global mapping of the favorable binding regions of $(GlcNAc)_2$ at the active site was devised by Pincus et al.[27,28] All of the low-energy minima for this disaccharide at the active site were used as starting points for extension of the saccharide to higher oligomers.[28-30]

A total of six GlcNAc residues were found to bind to the active site, in agreement with experiment.[25] The calculated structure of lowest energy for $(GlcNAc)_3$ bound to sites A–C was quite similar to that of the first

(26) M. Levitt in "Peptides, Polypeptides and Proteins", E. R. Blout, F. A. Bovey, M. Goodman, and N. Lotan, Eds., Wiley, New York, 1974, pp 99–113.

(27) M. R. Pincus, A. W. Burgess, and H. A. Scheraga, *Biopolymers*, 15, 2485–2521 (1976).

(28) M. R. Pincus, S. S. Zimmerman, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 4261–4265 (1976).

(29) M. R. Pincus, S. S. Zimmerman, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 2629–2633 (1977).

(30) M. R. Pincus and H. A. Scheraga, *Macromolecules*, 12, 633–644 (1979).

three residues of the energy-minimized X-ray crystal structure of $(GlcNAc)_4$ δ-lactone bound to sites A–D.[25,28-31] The calculated lowest energy tetramer, however, binds with its D residue in the chair form and in a position *removed* from the active-site cleft between the two acid residues Glu-35 and Asp-52 (unlike the D residue of the model-built structure). This calculated position for the binding of a GlcNAc residue in site D received strong experimental support from solution studies[32] and was further confirmed by a recent X-ray crystallographic structure of the trimer, MurNAc-GlcNAc-MurNAc in sites B–D.[33] Two lowest energy structures were computed for $(GlcNAc)_6$ bound to the enzyme,[30] one with its F site residue binding to the "right" side of the active site as in Figure 4 and one with this residue binding to the "left" side of the active site as in Figure 5. The latter is more stable and is the only one that can exist with the D residue outside of the cleft.[30] (Space-filling stereoviews of these structures, in color, are shown in ref 30.)

The prediction of two *stable* hexamer binding modes, in which all six residues contact the enzyme with no distortion of the D-site sugar residue, is in agreement with the results of temperature-jump experiments.[34]

These calculations have been extended to the binding of copolymers of GlcNAc and MurNAc to the active site.[35] It has been found[35] that MurNAc residues can bind only to sites B, D, and (right side) F, in good agreement with experiment.[25] The highest affinities for the dimer GlcNAc-MurNAc were calculated to be for sites C, D and E, F, a result corroborated by X-ray crystallographic data.[36] Three lowest energy conformations were computed for the trimer MurNAc-GlcNAc-MurNAc bound to sites B–D; the lowest energy one was close in conformation to that of the X-ray crystal structure for the same trimer bound to sites B–D.[35] Two lowest energy binding modes were calculated for copolymer hexamers containing GlcNAc and MurNAc, as found for $(GlcNAc)_6$,[35] except that the right-side binding mode is favored for copolymers with a MurNAc residue in site F.[35]

Approximately one-third of the binding energy for all complexes is attributable to the interactions of the *N*-acetyl groups of the substrate with the enzyme (mainly with the three active-site tryptophans 62, 63, and 108 and through hydrogen-bonding interactions in site C with Asn-59 and Ala-107), in agreement with experimental results.[25,30,35] Hydrogen-bonding interactions account for about one-fourth of the total favorable interaction energy. Polymers of saccharides without the *N*-acetyl group (e.g., glucose) are therefore predicted to bind with low affinities, in agreement with experiment.[25,30] The hydrogen-bonding scheme calculated for residues in sites A–C is identical with that inferred from X-ray crystallographic results.[25,30,33] From this kind of information, we can compute putative substrates that would maximize such contacts and would bind with the highest affinities.

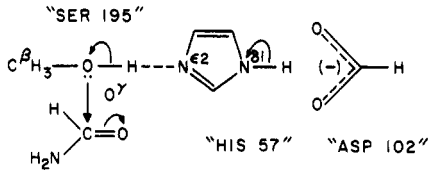


**Figure 6.** Model system for chymotrypsin- (or trypsin-) catalyzed hydrolysis of the model substrate, formamide. The arrows show the postulated flow of electrons in the formation of a tetrahedral intermediate.

## Quantum Mechanical Calculations on the Reaction Steps of Enzyme-Catalyzed Reactions

Having discussed the binding process, we next consider the catalytic step. Until now, we have been concerned with the recognition of structure and the manner in which enzymes interact with substrates in the *ground state* to produce stable complexes. We know that, despite the stability of these complexes, many of them are highly reactive. Because these reactions involve bond-making and bond-breaking processes, electronic transitions (which do not occur in the noncovalent interactions in ground-state complexes) now appear. These reactions can be accounted for satisfactorily only with a quantum mechanical treatment.

The quantum mechanical approach allows for systematic investigation of interactions that contribute toward stabilization of transition states. The theoretical basis for the quantum mechanical methods is presented extensively elsewhere.[37] While the details of the calculations vary widely from case to case, a clear pattern in the results emerges: to account for stabilization of transition states, the effects of residues in enzymes that are *not* directly in the active site are critical. These latter residues provide mainly electrostatic and hydrogen-bonding stabilization for charge separation involved during or prior to the rate-determining step in the reaction. We now consider several enzyme systems.

## Proteolytic Enzymes (Trypsin, Chymotrypsin, and Papain)

A putative reaction path for both acylation and deacylation has been calculated for a model "serine protease" by Scheiner et al.[38,39] The system is shown in Figure 6, where imidazole represents His-57, $CH_3OH$ represents Ser-195, formate represents Asp-102, and formamide is a prototypical substrate. Using a semiempirical method (partial retention of diatomic differential overlap, PRDDO),[40] these investigators calculated that the nucleophilicity of Ser-195 was due mainly to its dissociation of a proton to Asp-102 using His-57 in a "charge relay" system as depicted by the arrows in Figure 6. Attack of the resulting serine anion on the C=O of the substrate results in the formation of a relatively high energy tetrahedral intermediate which then breaks down to form the acyl-enzyme. The rate-determining step in acylation was computed to be the breakdown of the tetrahedral intermediate, in

(31) L. O. Ford, L. N. Johnson, P. A. Mackin, D. C. Phillips, and R. Tjian, *J. Mol. Biol.*, 88, 349–371 (1974).
(32) M. Schindler, Y. Assaf, N. Sharon, and D. M. Chipman, *Biochemistry*, 16, 423–431 (1977).
(33) J. A. Kelly, A. R. Sielecki, B. D. Sykes, M. N. G. James, and D. C. Phillips, *Nature (London)* 282, 875–878 (1979).
(34) E. Holler, J. A. Rupley, and G. P. Hess, *Biochemistry*, 14, 2377–2385 (1975).
(35) M. R. Pincus and H. A. Scheraga, *Biochemistry*, 20, 3960–3965 (1981).
(36) R. Sarma and R. Bott, *J. Mol. Biol.*, 113, 555–565 (1977).
(37) J. A. Pople and D. L. Beveridge, "Approximate Molecular Orbital Theory", McGraw-Hill, New York, 1970.
(38) S. Scheiner, D. A. Kleier, and W. N. Lipscomb, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 2606–2610 (1975).
(39) S. Scheiner and W. N. Lipscomb, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 432–436 (1976).
(40) T. A. Halgren and W. N. Lipscomb, *J. Chem. Phys.*, 58, 1569–1591 (1973).

Table II
Proton Affinities[a] of Active-Site Residues Involved in Proton Transfers in Proteolytic Enzymes

| enzyme | residue | proton affinity kcal/mol | method |
|---|---|---|---|
| trypsin or chymotrypsin | 1. His-57 | −260.0 | PRDDO[b] |
| | Asp-102 | −473.0 | |
| | 2. His-57 | −246.1 | 4-31G ab initio[c] |
| | Asp-102 | −350.9 | |
| | 3. His-57 | −307.7 | 4-31G ab initio[d] |
| | Asp-102 | −281.8 | |
| papain | 4. His-159 | −277.0 | STO-3G ab initio[e] |
| | CysH-25 | −498.0 | |

[a] Defined as the negative of the energy to dissociate a proton completely in the gas phase from the acid form. [b] Partial retention of diatomic differential overlap, as in ref 38–40. The proton affinities are for the *isolated* species with no long-range or hydrogen-bonding interactions considered. [c] Computed in ref 42, in a manner similar to that in *b*, but using an extended basis set. [d] Here, the two residues interact, and hydrogen bonds between Asp-102 and surrounding residues were taken into account, together with the electrostatic potentials of all enzyme residues.[42] Charge delocalization effects were also included. [e] As reported in Table II of ref 44. These values are calculated for the *isolated* components.

agreement with experiment.[41] Deacylation was computed to follow a qualitatively similar path, with $H_2O$ replacing Ser-195 as the nucleophile.[39]

The essential feature of these computations is the calculated relative proton affinities of Asp-102 and His-57 (shown as the first entry in Table II) which involves extensive charge delocalization due to the higher relative proton affinity of Asp-102. Allen and co-workers[42] have performed extended basis set (4-31G) ab initio calculations on a more extensive model system and have included the effects of (i) hydrogen bonding of Ser-214 to the $COO^-$ group of Asp-102 and (ii) point charges of all of the atoms of the protein. With these effects included, there was a reversal in proton affinity between His-57 and Asp-102, the former residue now having the higher proton affinity (entry 3 of Table II). If these latter effects are omitted, the proton affinity of Asp-102 is higher (entry 2 of Table II). This result casts doubt on the charge relay system, and this doubt is supported by recent experimental data.[43] The reaction path calculated by Allen and co-workers (discussed in ref 42) involved concerted nucleophilic attack by Ser-195 on the C=O of the substrate and proton transfer of the $O^\gamma$–H of Ser-195 (Figure 6) to the imidazole ring of His-57 which, in turn, is anchored in a favorable position to accept this proton by its *hydrogen bonding* (not charge relay) to Asp-102.[42] These latter calculations clearly demonstrate the effect of residues around the active site on the mechanism.

Similar conclusions were reached by two different groups[44,45] in calculations on the active site of papain, which differs from the "serine proteases" like trypsin and chymotrypsin in that the SH of a cysteine residue (CysH-25) is the nucleophile rather than the OH of serine. Using extended basis set ab initio calculations, van Duijnen et al.[44] have computed the electrostatic effect of the N-terminal $\alpha$ helix on the dissociation (transfer) of a proton from CysH-25 to His-159 (which has an analogous function to that of His-57 of the serine proteases). The proton was found to have two stable isoenergetic positions, one on the S of CysH-25 and the other on the ring $N^{\epsilon 2}$ of the histidine, the latter species being one with charge separation, i.e., Cys $S^-$ ··· $HIm^+$ (Im = imidazole), and stabilized almost solely by the electrostatic field from the N-terminal $\alpha$ helix. Without this latter effect, the proton affinity of the CysH residue is much higher (entry 4 of Table II). Similar conclusions were reached by Bolis et al.[45] who also found that a hydrogen-bonding network with residues around the active site stabilized this ion pair. It is clear that the enhanced nucleophilicity of CysH-25 is due to dissociation (transfer) of its proton to His-159 in a system stabilized by electrostatic fields and hydrogen-bonding networks from residues near but not actually in the active site.

## Lysozyme

As discussed above, it is not likely that ground-state strain is involved in the catalytic action of this enzyme. Warshel and Levitt[46] have investigated the electrostatic effects of all atoms of the enzyme on the formation of the carbonium ion intermediate in the hydrolysis of $(GlcNAc)_6$. Their approach was to divide the complex into three regions: a quantum mechanical region (treated with a semiempirical method[46]), including atoms of the reactive portion of the substrate and of the enzyme, a classical region (on which conformational energy calculations were performed) involving the rest of the atoms of the complex, and a surrounding water region represented by water dipoles. The interaction energy between the classical and quantum regions was a sum of electrostatic and nonbonded energies plus charge-induced dipole terms. The interaction energies between the water dipoles and the atoms of the complex were evaluated as charge-dipole terms. The major energy of stabilization for separation of charge in the formation of the carbonium ion was calculated to result from the permanent dipoles of the enzyme which could align themselves with the fields created by the charges. The carboxylate of Asp-52 (treated classically) was found to contribute relatively little to the stabilization of the carbonium ion, a finding which has experimental support in model systems.[47]

The above approach has been extended to compare electrostatic stabilization effects, in charge separation, of the active site of lysozyme with those of $H_2O$ in the nonenzymatic reaction.[48-50] For solvation of ions in the nonenzymatic reaction, a surface-constrained soft-sphere dipole (SCSSD)[49] model of solvation was used. Each ion in this model is surrounded by concentric spherical layers of tightly bound $H_2O$ (treated as dipoles with van der Waals radii), a layer of fixed water molecules, and a continuum. Charge–charge interactions were evaluated either from experimental data[48,49] or

(41) M. Caplow, *J. Am. Chem. Soc.*, **91**, 3639–3645 (1969).
(42) L. C. Allen, *Ann. N.Y. Acad. Sci.*, **367**, 383–406 (1981).
(43) W. W. Bachovchin and J. D. Roberts, *J. Am. Chem. Soc.*, **100**, 8041–8047 (1978).
(44) P. T. Van Duijnen, B. T. Thole, R. Broer, and W. C. Nieuwpoort, *Int. J. Quant. Chem.*, **17**, 651–671 (1980).
(45) G. Bolis, E. Clementi, M. Ragazzi, D. Salvaderi, and D. R. Ferro, *Int. J. Quant. Chem.*, **14**, 815–838 (1978).
(46) A. Warshel and M. Levitt, *J. Mol. Biol.*, **103**, 227–249 (1976).
(47) B. M. Dunn and T. C. Bruice, *Adv. Enzymol.*, **37**, 1–60 (1973).
(48) A. Warshel, *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 5250–5254 (1978).
(49) A. Warshel, *J. Phys. Chem.*, **83**, 1640–1652 (1979).
(50) A. Warshel and R. M. Weiss, *J. Am. Chem. Soc.*, **102**, 6218–6226 (1980).

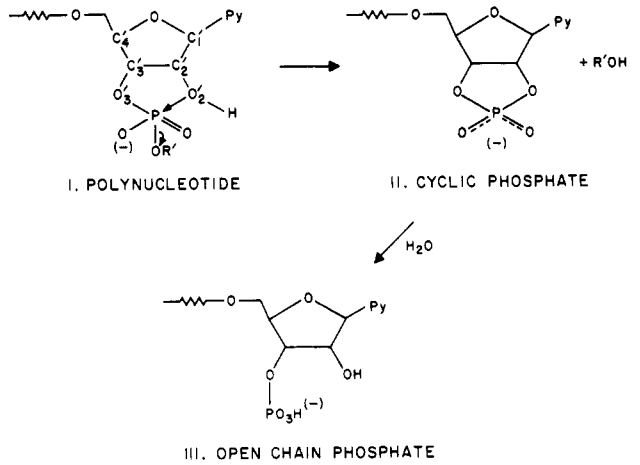


**Figure 7.** Steps in ribonuclease A catalyzed hydrolysis of RNA. Py = pyrimidine, R′ = a pyrimidine or purine nucleoside. The first step is a transesterification, the second a hydrolysis.

from quantum mechanical (valence bond) calculations.[50] The charge–water dipole interactions were evaluated as above. The potential due to the continuum was also taken into account.[48] This model does not consider hydrogen bonding of water or the directional interactions that result from the lone pairs of electrons on the oxygens of the water molecules.

The enthalpies of charge separation were then evaluated by minimizing the energy of the system for a number of random configurations of the "clusters" at different distances of separation of the ions in their hydrated layers. Though no attempt was made to search conformation space for all (or a representative sampling) of the minimum energy conformations or to calculate entropies of solvation, it was assumed that all the low-energy minima would be approximately isoenergetic and that entropy effects would be minimal.

When the two processes, solvation in the nonenzymatic and enzymatic reactions, respectively, were compared by this theoretical approach, it was found that the enzyme stabilizes charge separation by ~7 kcal/mol over the nonenzymatic (solution) process. This 7 kcal/mol would account for the large rate enhancements seen in catalysis.[48] The difference was attributed to the ability of the dipoles of the enzyme to orient themselves with the fields created by the charges whereas the $H_2O$ dipoles were not able to align themselves freely with the fields because of competition from the bulk $H_2O$ molecules.[48-50] The results of these calculations are strikingly similar (in the involvement of surrounding groups) to those based on quantum calculations on the proteolytic enzymes described above. It is clear that electrostatic fields around the active site play a major role in rate enhancements.

## Ribonuclease A

This enzyme has 124 amino acid residues and hydrolyzes RNA specifically at pyrimidine sites in a two-step process as shown in Figure 7.[51] In the first step, transphosphorylation, the 2′-OH group of the ribofuranosyl ring attacks the phosphorus atom in a direct, in-line, displacement of the outgoing 5′ oxygen to form a 2′,3′-cyclic phosphate. The reaction proceeds through a pentacovalent intermediate.[51-53] In the second step, the cyclic phosphate is hydrolyzed to yield exclusively the 3′-monophosphate. Three residues are involved in the reaction: His-12, which is thought to serve as a general base in the first step to aid nucleophilic attack of the 2′-O on the phosphorus atom; His-119, thought to serve as a general acid in the first step which protonates the leaving 5′-O; and Lys-41, thought to stabilize the dianionic pentacovalent intermediate.[51] A recent quantum mechanical treatment confirms these features of the mechanism.[42,54]

A molecular mechanics calculation[55] was performed on the (refined) X-ray crystal structure of the complex of uridyl 3′,5′-adenylate (UpA) with ribonuclease A (the crystal structure[56] was determined for a complex with UpcA where a $CH_2$ group replaced the 5′-oxygen). The three active-site residues were included. No electrostatic energies were computed. It was found that the (2′-O)–P interatomic distance of the uridyl residue was shorter than in the crystal structure of *isolated* UpA and that the 2′-O was in an advantageous position for nucleophilic attack. A reaction path was then calculated for the transphosphorylation of UpA. The activation energy was computed as only 0.5 kcal/mol (the free energy of activation for transphosphorylation is 13.3 kcal/mol, and the enthalpy of activation is 2.1 kcal/mol[51]). It was concluded that ground-state distortion may be important in forcing the substrate into a catalytically active conformation. The treatment omits a number of components in the energy that were included in the quantum mechanical treatment,[42,54] particularly electrostatics. It would be of interest, therefore, to determine quantum mechanically how much ground-state strain may actually contribute to observed rate enhancements.

## Carbonic Anhydrase

This enzyme catalyzes the hydration of $CO_2$ to produce $H_2CO_3$ with the fastest turnover of any enzymic system. It contains a $Zn^{2+}$ ion coordinated to three histidine residues, one $H_2O$, and one $CO_2$ molecule.[42] Jönsson et al.,[57] using ab initio methods, have calculated that the observed barrier to attack by $OH^-$ on $CO_2$ must be due to desolvation of the attacking $OH^-$ ion. Any system that would desolvate the nucleophile would presumably increase the reaction rate. Pullman and Demoulin[58] have calculated that the proton affinity of the complex $[Zn(NH_3)_3 \cdot H_2O]^{2+}$ is greatly reduced, i.e., that $H_2O$ coordinated to the complex readily dissociates a proton, leaving a nucleophilic $OH^-$ coordinated to Zn.

The active site of carbonic anhydrase combines both of these features in that (i) catalysis occurs in a relatively nonpolar environment and (ii) it contains a coordinated $Zn^{2+}$ similar to that in the model system described above. It has been further postulated[42] that

(51) F. M. Richards and H. W. Wyckoff in "The Enzymes", P. D. Boyer, Ed., Academic Press, New York, 1971, pp 647–806.

(52) D. A. Usher, D. I. Richardson, Jr., and F. Eckstein, *Nature (London)*, **228**, 663–665 (1970).

(53) D. A. Usher, E. S. Erenrich, and F. Eckstein, *Proc. Natl. Acad. Sci. U.S.A.*, **69**, 115–118 (1972).

(54) C. A. Deakyne and L. C. Allen, *J. Am. Chem. Soc.*, **101**, 3951–3959 (1979).

(55) R. R. Holmes, J. A. Deiters, and J. C. Gallucci, *J. Am. Chem. Soc.*, **100**, 7393–7402 (1978).

(56) F. M. Richards and H. W. Wyckoff, "Atlas of Molecular Structure in Biology", Vol. I, D. C. Phillips and F. M. Richards, Eds., Clarendon Press, Oxford, 1973.

(57) B. Jönsson, G. Karlstrom, and H. Wennerstrom, *J. Am. Chem. Soc.*, **100**, 1658–1661 (1978).

(58) A. Pullman and D. Demoulin, *Int. J. Quant. Chem.*, **16**, 641–653 (1979).

the stability of the $OH^-$ at the active site is increased even more by long-range hydrogen-bonding interactions between the active (nucleophilic) $H_2O$ and components of what has been termed the "water chain" in the active site, indicating again the importance of residues around but not in the active site.

Quantum and molecular mechanical calculations have thus provided insights into rate-enhancing factors in enzyme catalysis. Long-range electrostatic and hydrogen-bonding effects are crucial. It will be of interest to investigate these in a systematic manner to discover where the major portion of electrostatic stabilization energy in charge separation processes is centered. It will be further desirable to develop methods to calculate reaction paths so that reliable energies of activation can be computed to allow accurate estimates of rate enhancements of enzyme-catalyzed reactions over the corresponding nonenzymatic ones. Finally, it will be of value to combine the conformational energy and quantum mechanical approaches so that, of the favorable binding modes calculated in the ground state, we can predict which of these modes will be catalytic or productive ones.

*The work described here that was carried out at Cornell University was supported by research grants from the National Science Foundation (PCM79-20279) and the National Institute of General Medical Sciences of the National Institutes of Health, U.S. Public Health Service (GM-14312).*

# Direct Observation of Simple Tetrahedral Intermediates

BRIAN CAPON,* ASHIM K. GHOSH, and DUNCAN MCL. A. GRIEVE

*Department of Chemistry, Glasgow University, Glasgow G12 8QQ, Scotland, United Kingdom*

*Received April 17, 1981*

Many reactions are thought to pass through transient intermediates which have not been detected but whose incursion is inferred on the basis of indirect evidence. The potential energy–reaction coordinate diagram for such a reaction may be written as shown in Figure 1. For reactions of polyatomic molecules, this is a cross section of a potential energy surface in a multidimensional hyperspace. The complete and rigorous calculation of such a surface is virtually impossible, and it is necessary to employ approximate methods in order to make the calculations tractable.[1] The experimental approach to the study of the reactions of the intermediate is apparently limited by the fact that the potential energy differences $E_r$ and $E_p$ and the corresponding free energy differences are so large that the intermediate is impossible to detect. Nevertheless, in some reactions the value of $E_a$, the potential energy of activation for the breakdown of the intermediate, and the corresponding free energy of activation may not be so small that this reaction could not be studied. Therefore, it is a general principle that if one wants to study directly the reactions of such an intermediate, it is useless to approach it directly along the reaction coordinate shown in Figure 1. Instead it must be approached from some other direction on the potential energy surface from a higher energy state. The work described in this Account and also our work on vinyl alcohol[2] are based on this principle.

One of the most widely studied class of reactions for which a potential energy–reaction coordinate diagram of the type shown in Figure 1 may be written is acyltransfer reactions which are generally thought to proceed through tetrahedral intermediates.[3] Unlike the reactions of aldehydes, ketones, and their derivatives in which the tetrahedral intermediate is frequently detectable,[4] in the reactions of derivatives of carboxylic acids it usually is not. At this oxidation level, the evidence for the incursion of a tetrahedral intermediate is therefore normally indirect, based on, for example, experiments which demonstrate the exchange of the oxygen of the carbonyl group with that of water when this is the solvent[5] and on experiments which demonstrate a change in the rate-determining step, and hence the incursion of an intermediate.[6]

The tetrahedral intermediate which results from the attack of the nucleophile $Y^-$ on the carboxylic acid derivative **1** is, after protonation, **2**. The intermediate from the corresponding thione derivative would have an SH group instead of an OH group. The groups X

Brian Capon obtained his Ph.D. in physical organic chemistry at the University of Southampton. He subsequently worked in the Semiconductor Materials Group at the GEC Research Laboratories, Wembley, England, at Birkbeck College, University of London, as a Lecturer in physical chemistry, and at the University of Leicester as a Lecturer in organic chemistry. In 1968 he moved to the University of Glasgow where he is now Titular Professor in organic chemistry.

Ashim Kumar Ghosh obtained his B.Sc., M.Sc., and Ph.D. at the University of Calcutta. He subsequently worked as a postdoctoral research fellow at the University of Glasgow and now holds a similar position at the University of Western Ontario.

Duncan McL. A. Grieve obtained his B.A. at the University of Stirling and Ph.D. at the University of Glasgow. He subsequently worked as a postdoctoral research fellow at ETH, Zürich, and again at the University of Glasgow. He is now a Senior Research Scientist at Minnesota 3M Research Ltd., Harlow, Essex, England.

(1) Cf. K. Müller, *Angew. Chem., Int. Ed. Engl.*, **19**, 1 (1980).
(2) B. Capon, D. S. Rycroft, and T. W. Watson, *J. Chem. Soc., Chem. Commun.*, 724 (1979); B. Capon, D. S. Rycroft, T. W. Watson, and C. Zucco, *J. Am. Chem. Soc.*, **103**, 1761 (1981).
(3) Cf. M. L. Bender, *Chem. Rev.*, **60**, 53 (1960); "Mechanisms of Homogeneous Catalysis from Protons to Proteins", Wiley-Interscience, New York, 1971, p 108; W. P. Jencks, "Catalysis in Chemistry and Enzymology", McGraw-Hill, New York, 1968, pp 465, 526; T. C. Bruice and S. J. Benkovic, "Bio-Organic Mechanisms", Benjamin, New York, 1966, Vol. 1; S. L. Johnson, *Adv. Phys. Org. Chem.*, **5**, 237 (1967); A. J. Kirby, *Compr. Chem. Kinet.*, *10*, 104 (1972).
(4) Cf. W. P. Jencks, *J. Am. Chem. Soc.*, **81**, 475 (1959); *Prog. Phys. Org. Chem.*, **2**, 63 (1964); M. Cocivera, C. A. Fyfe, A. Effio, S. P. Vaish, and H. E. Chen, *J. Am. Chem. Soc.*, **98**, 1573 (1976); M. Cocivera and A. Effio, *J. Chem. Soc., Chem. Commun.*, 393 (1976); M. Cocivera, A. Effio, H. E. Chen, and S. Vaish, *J. Am. Chem. Soc.*, **98**, 7362 (1976).
(5) M. L. Bender, *J. Am. Chem. Soc.*, **73**, 1626 (1951).
(6) W. P. Jencks and M. Gilchrist, *J. Am. Chem. Soc.*, **86**, 5616 (1964).

0001-4842/81/0114-0306$01.25/0 