Published in final edited form as:
J Am Chem Soc. 2008 November 19; 130(46): 15361–15373. doi:10.1021/ja803213p.

Structural Reorganization and Preorganization in Enzyme Active Sites: Comparisons of Experimental and Theoretically Ideal Active Site Geometries in the Multi-step Serine Esterase Reaction Cycle

Adam J. T. Smith[1], **Roger Müller**[2], **Miguel D. Toscano**[2], **Peter Kast**[2], **Homme W. Hellinga**[3,*], **Donald Hilvert**[2,*], and **K. N. Houk**[1,*]

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095. [2]Laboratory of Organic Chemistry, ETH Zurich, Hönggerberg HCI F 339, CH-8093 Zurich, Switzerland. [3]Department of Biochemistry, Duke University Medical Center, Durham, NC 27710.

Abstract

Many enzymes catalyze reactions with multiple chemical steps, requiring the stabilization of multiple transition states during catalysis. Such enzymes must strike a balance between the conformational reorganization required to stabilize multiple transition states of a reaction and the confines of a preorganized active site polypeptide tertiary structure. Here we investigate the compromise between structural reorganization during the catalytic process, and preorganization of the active site for a multi-step enzyme-catalyzed reaction, the hydrolysis of esters by the Ser-His-Asp/Glu catalytic triad. Quantum mechanical transition states were used to generate ensembles of geometries that can catalyze each individual step in the mechanism. These geometries are compared to each other by superpositions of catalytic atoms to find "consensus" geometries that can catalyze all steps with minimal rearrangement. These consensus geometries are found to be excellent matches for the natural active site. Preorganization is therefore found to be the major defining characteristic of the active site, and reorganizational motions often proposed to promote catalysis have been minimized. The variability of enzyme active sites observed by X-ray crystallography was also investigated empirically. A catalog of geometrical parameters relating active site residues to each other and to bound inhibitors was collected from a set of crystal structures. The crystal-structure-derived values were then compared to the ranges found in quantum mechanically optimized structures along the entire reaction coordinate. The empirical ranges are found to encompass the theoretical ranges when thermal fluctuations are taken into account. Therefore, the active sites are preorganized to a geometry that can be objectively and quantitatively defined as minimizing conformational reorganization while maintaining optimal transition state stabilization for every step during catalysis. The results provide a useful guiding principle for *de novo* design of enzymes with multi-step mechanisms.

Introduction

Enzymes have been recognized as masters of catalysis, capable of catalyzing reactions at diffusion-limited rates and able to achieve rate accelerations of as much as 10^{21} over the uncatalyzed reaction.[1,2] A primary role-player in this rate enhancement is the preorganization of the active site to a geometry that electrostatically stabilizes the transition state.[3,4] In solution, solvent organization around a reacting molecule can be slow compared to the reaction itself and limit the rate of reaction,[5] and solvent dipoles are free to reorient around different charge

* To whom correspondance should be addressed. E-mail: houk@chem.ucla.edu; hilvert@org.chem.ethz.ch; hwh@biochem.duke.edu.

distributions such as those in a reactant as compared to a transition state.[3] A preorganized enzyme active site is not subject to such limitations and can employ relatively fixed dipoles oriented towards transition state stabilization.[3,4,6–9]

Concordant with the relative rigidity implied by preorganization, catalytic residues are often observed to be less mobile than non-catalytic residues. This has been demonstrated by comparisons of unliganded and liganded crystal structures,[10] temperature factors,[11] vibrational modes from Gaussian Network Modeling (GNM),[12] and calculated motions in dynamics simulations.[13] A decrease in flexibility has also been shown to be essential to the evolutionary drive towards a selective and proficient catalyst.[14] Promiscuous enzymes have been proposed to represent an early point in protein evolution. These promiscuous enzymes display more conformational flexibility than their more evolved, more selective, and more proficient counterparts.[14–16] Additionally, an engineered molten globular chorismate mutase enzyme shows a large reduction in dynamics upon binding of a transition state analogue, suggesting that a decrease in flexibility is associated with catalysis in this case.[17,18]

Preorganization does not, however, imply a completely rigid geometry. A recent study of enzyme mechanisms has shown that they have an average of 4.3 steps, and 2.7 intermediates per reaction.[19] Enzymes must stabilize each transition state in multi-step reactions. These necessitate at least some degree of active site reorganization, as each of the transition states will have slightly different geometries and electrostatic profiles. Indeed, the prevailing dogma about enzyme catalysis has changed over the last century from a static model to a highly dynamic model. The original “lock-and-key” postulate[20] gave way to the more fluid “induced fit” model,[21] and currently to the idea that enzyme dynamic motions actually promote catalysis.[22–24]

Motions that aid catalysis may be broadly divided into two categories: 1) protein domain hinge and shear motions that facilitate binding and release of substrate and product,[25] and 2) protein vibrational modes that are coupled to the reaction coordinate as it proceeds from reactant to product.[26,27] While all argue that coupled motion exists, Warshel and coworkers have demonstrated that the catalytic effects of protein motions are relatively minor by comparing simulations of reactions taking place in enzymes and in solution.[9]

Motions in the first category do not directly influence the chemical barriers to the reaction but are essential for the substrate to bind or product to be released; these domain motions can be rate limiting. They are pre- and post-catalytic motions, and therefore do not play a role in structural reorganization of the active site along the chemical reaction coordinate. Residues that interact with the substrate are moved into place to stabilize and desolvate it in the binding site. These motions take place with lifetimes on the order of msec-μsec. Enzymes that exemplify these motions include triose phosphate isomerase,[28] RNase A,[29] HIV protease,[30] chorismate mutase,[31] and adenylate kinase among others.[32–34]

The second category are co-catalytic motions that guide the reaction from one stationary point to the next as it proceeds from reactant complex to product complex. These motions involve the catalytic residues directly, and thus can play an active role in structural reorganization of the active site. A subtype of motions proposed to be coupled to catalysis are very fast, low amplitude atomic vibrations within the residues in and around the active site.[26,35,36] Their timescales reflect the lifetimes of atomic vibrations, on the order of psec-100 fsec, and as such they are too fast and small in amplitude to effect significant active site reorganization and are not considered here.

Those motions that do reorganize the active site may be divided into intra-step and inter-step reorganizational motions (Figure S1). Both reorient the enzyme active site catalytic groups along the reaction coordinate, but the distinction is made based on the stationary points they

connect. Intra-step reorganizations connect the ground state immediately preceding a chemical transformation to the transition state of the step. These are reorientations of the active site functionality involved in bond formation and breaking, and of active site dipoles to enhance transition state stabilization.[37] Inter-step reorganizations connect different chemical steps that must have different transition states. These involve reorienting the catalytic units themselves to catalyze a chemical reaction at a different position on the substrate, but no bonds are formed or broken during the reorganization.

Intra-step reorganizational motions include, for example, a compressing motion on a hydride donor-acceptor pair that brings them from the Michaelis complex to the transition state for hydride transfer.[26] The timescales of these motions are coincident with the rates of the catalyzed reactions, with lifetimes on the order of 10 msec-100 μsec. They have been found in liver alcohol dehydrogenase,[38] cyclophilin A,[39] and dihydrofolate reductase.[40] Previous work has investigated the intra-step reorganizational energies of wild-type enzymes compared to mutants or to the uncatalyzed reaction in aqueous solution. In these studies it was shown that reactions in water or in mutant enzymes each have higher intra-step reorganization energy barriers in going from the ground state to the transition state than the wild-type enzymes, emphasizing that enzymes are well preorganized with respect to minimizing intra-step motions.[7,8,41,42]

Inter-step reorganization is less studied, but it is necessary in order for an enzyme to stabilize multiple transition states.[43,44] It is conceivable that such enzymes might extensively reorganize their active site geometries between chemical steps to afford maximal stabilization of each transition state. On the other hand, the exquisite placement of catalytic residues in a geometry that seems ideal for catalyzing multiple chemical steps has been noted previously for triose phosphate isomerase on qualitative grounds,[23,45] but quantitative studies have yet to address this in comparison to inter-step reorganization.

Enzymes unquestionably make use of both preorganization and reorganization. Any protein with a defined tertiary structure may be said to be “preorganized,” but the requirement that an enzyme must stabilize the transition state and also bind the substrate and product necessitates some reorganization along the reaction coordinate. A key issue remaining, however, is whether enzymes have evolved active sites that may be objectively defined as minimizing active site reorganization,[3,6] or if coupled catalytic motions have evolved such that inter-step reorganizational barriers are low enough to allow significant changes in the shape of the active site to stabilize each chemical transition state.

The two questions addressed in this work are the following. 1) What are the extents of active site inter-step reorganizations during a complex multi-step reaction? 2) Does evolution give the theoretically optimal arrangement for all steps of catalysis, or do the protein backbone and surrounding sidechains limit catalysis and prevent the optimum catalytic geometry from being observed in one or more steps?

This study emphasizes the geometry of the active site, and why one particular transition-state-stabilizing geometry has evolved in preference to others. This is investigated by comparing the naturally evolved active site structure found in serine hydrolases to other plausible structures that have not evolved. The active site has been modeled using just the functional groups involved in catalysis, free of their attachment to the protein and in the absence of solvent. This neglect of the surrounding environment may affect the calculated energies and mechanism, but the advantage of this system is that it allows for the generation of alternative active sites not found in nature. If the full protein were modeled, even minor changes in active site architecture would require large scale backbone redesign to accommodate them, and the viability of such structures or energies calculated in these environments would be questionable.

For each chemical step in the hydrolysis of an ester by a serine esterase, quantum mechanical modeling has been used to calculate the optimum arrangement of catalytic side chains. Such QM structures, with catalytic residue positions fully optimized, have been termed "theozymes," short for theoretical enzymes.[46] The calculations lead to fully optimized geometries for each step considered, independent of the constraints of the polypeptide backbone or the noncatalytic sidechains. Molecular mechanics methods were then employed to generate an ensemble of structures that maintain the geometrical constraints necessary for catalysis and could, therefore, each catalyze the chemical step for which the ensemble was generated. The members of each ensemble were then compared to those in ensembles for other steps to determine consensus structures that can catalyze all the required steps of catalysis within a given root-mean-square tolerance. These consensus structures were then compared to the natural active site geometry found in the crystal structure. Figure S2 depicts the entire protocol from QM optimization to comparison of consensus structures to the natural geometry. Reorganization energies cannot be evaluated directly in the theoretical models of the active site, so the amount of motion required by a given active site geometry is used as a metric.

We also measured distance, angle, and dihedral parameters of the active sites of thirty-three high-resolution X-ray crystal structures of serine esterase enzymes. The geometrical variations between serine esterase active sites from different species and with different substrate specificities were measured and compared to the range of parameters measured in quantum mechanically optimized stationary points of the active site model system along the entire reaction coordinate.

The serine esterase system was chosen for its characteristic multi-step mechanism[47] and the availability of many crystal structures from different fold families and with different bound ligands. Additionally, serine esterases are very efficient catalysts, with acetylcholinesterase cited as an example of a perfectly evolved enzyme that is limited only by the diffusion of substrate to the active site.[48] The understanding of how nature has arrived at the architecture of these potent catalysts, and what restrictions are placed on preorganization and reorganization, are important for the field of *de novo* enzyme design, in which the goal is to be able to design enzymes to catalyze any desired reaction.[49–51]

Results and Discussion

Serine Esterase Reaction and QM Location of Stationary Points

The specific catalytic amino acid functional groups modeled were taken from the serine hydrolase enzyme butyrylcholinesterase (BChE). This enzyme uses the multi-step mechanism shown in Figure 1.[47] In this figure, ground states other than the unliganded enzyme are named according to the geometry of the substrate ester carbon, either tetrahedral or trigonal, and where in the cycle they occur. Transition states are numbered according to their temporal order. The hydrolysis mechanism involves the catalytic triad of Ser-His-Glu that provides the essential nucleophile. Starting from the unliganded enzyme, there is a complexation barrier that may involve a pre-catalytic gating motion (TS1) before formation of the enzyme-substrate complex (trig1). The catalytic serine is then deprotonated by the His-Glu general base and attacks the substrate (TS2), forming a covalent intermediate where the histidine is hydrogen bonding to the serine γ-oxygen (tet1). In this step, the oxyanion hole stabilizes the negative charge that develops on the substrate ester carbonyl oxygen that becomes an alkoxide in the tetrahedral intermediate. The enzyme must then at least slightly reorganize the active site (TS3) to a covalent intermediate in which the histidine is hydrogen bonding to the substrate leaving group oxygen (tet2).[44] This is an inter-step structural reorganization required by this multi-step mechanism. It is given as a formally separate step here and has been calculated previously in the serine protease elastase enzyme,[52] although it may be concerted with the next step, loss of the alcohol portion of the ester, assisted by general acid catalysis involving the protonated His-

Glu dyad (TS4). This generates the acylenzyme (trig2). The alcohol product clears the active site (TS5), making room for the attacking water (trig3). The acylenzyme then reacts with this water using His-Glu as a general base (TS6) and forming another tetrahedral intermediate with the His hydrogen bonding to the hydroxyl oxygen (tet3). The alkoxide is stabilized again by the oxyanion hole. Another inter-step conformational rearrangement then occurs analogous to TS3 to reposition the His-Glu dyad in coordination with the serine γ-oxygen (TS7, tet4). Finally, the acid is formed and released from the protein by the breaking of the serine-substrate CO bond, with protonated His-Glu serving now as general acid (TS8), and the catalyst is reformed (trig4). TS9 releases the acid from the active site, regenerating the unliganded enzyme.

Quantum mechanical geometries were optimized for as many structures in Figure 1 as possible. The unliganded geometry was not computed, as the absence of the ligand would lead to collapse of the active site to fill the vacuum in the gas phase. The potential involvement of non-catalytic residues in TS1, TS5, and TS9 prevented their calculation in the model system used here. All 14 other ground and transition states were optimized. Figure 2 shows examples of the optimized structures for the Michaelis complex (trig1, Figure 2a), TS2 (Figure 2b), and TS4 (Figure 2c). The TS2 and TS4 geometries demonstrate the chemistry that must take place at two different points in space—the serine Oγ and the leaving group/water oxygen, respectively (Figure 2b, c). All fourteen theozyme stationary points are shown in Figure 2d, overlaid over two atoms of each catalytic triad residue model (O, Cβ in Ser; N, N in His; O, Cδ in Glu), with TS3 as the reference. TS3 was chosen because it represents an average geometry of all the stationary points. RMSDs ranged from 0.18 Å, for TS7, to 1.22 Å, for trig2. Without trig2 and trig3 included, all RMSDs were below 0.80 Å. The trig2 and trig3 structures (magenta carbons, Figure 2d) contain an acylenzyme model system with a free molecule coordinated to the imidazole, either methanol in trig2 or water in trig3. In the gas phase, the partial negative charge of the oxygen of the methanol or water is drawn to a partial positive hydrogen of the peptide backbone model representing a part of the oxyanion hole. This attraction pulls the whole methanol/water, imidazole, and carboxylate unit along and results in the higher RMSDs. In solution, bulk water should serve as a hydrogen bond donor, and so it is possible that the higher RMSDs in these geometries are an artifact of theozyme modeling.

The mechanism of the theozyme-catalyzed hydrolysis of methyl acetate differs slightly in protonation states from the experimentally supported mechanism in the enzyme[47] and the mechanism calculated by empirical valence bond studies of esterases and proteases.[42,53] This is a consequence of unrestrained modeling in the gas phase, but has no effect on the structural ensembles described further on in this work. The trig1 (Figure 2a) and trig4 geometries contain an imidazolate and acetic acid, rather than the imidazole and acetate that would be expected from Figure 1. When a proton is transferred from the acetic acid to the imidazolate the partial negative charges on the acetate become stronger, and cause a reorientation to interact with a partial positive hydrogen on the 2-(formylamino)-N-methylacetamide. This reorientation would not be tolerated in the natural enzyme. A trig1 geometry with the imidazole/acetate pair was near converged with an energy 1.2 kcal/mol below the imidazolate stationary point, indicating the energy difference is not too significant. The TS2 geometry (Figure 2b) shows a stepwise rather than concerted mechanism, with initial proton transfer from the N-methyl-3-hydroxy-propanamide to the imidazolate. This is consistent with previous theozyme models of this reaction,[52] and in other work on a similar system the stepwise and concerted mechanisms were calculated to be isoenergetic.[54] Finally, the imidazolium/acetate ion pair in the tetrahedral intermediates in Figure 1 neutralizes in the gas phase, placing the proton on acetic acid in TS4 (Figure 2c). This is similar to the double proton transfer mechanism that has been found in previous computational work on butyrylcholinesterase, modeled by both theozymes[55] and combined quantum mechanics/molecular mechanics (QM/MM) calculations in the enzyme, [56] and that was used to calculate the effect of mutations in free energy perturbation studies.[57]

Comparison of Calculated Free Energy Barrier to Experimental k_{cat}

The free energies of the full system across the entire reaction path were evaluated in order to identify the effective barrier calculated for the reaction (Figure 3). This barrier is the greatest energy difference between any ground state and any transition state along the forward reaction from reactant to product. The calculated barrier may be compared to the experimental k_{cat}. The k_{cat} for butyrylcholinesterase-catalyzed hydrolysis of butyrylcholine is 2,850 s^{-1}, corresponding to a free energy barrier of 12.8 kcal/mol.[58]

The greatest barrier among the calculated structures is 12.1 kcal/mol, between trig3 and TS6, in good agreement with the experimental value. This comes on the deacylation side of the reaction, which has been demonstrated experimentally to be rate determining for BChE hydrolysis of a benzoylester substrate.[59] However, this is not always the case in serine esterases, and both acylation and deacylation have been shown to influence the rate of acetylcholinesterase-catalyzed hydrolysis of acetylcholine.[60] The largest barrier calculated on the acylation side is 11.2 kcal/mol, less than 1.0 kcal/mol below the deacylation barrier. It is possible that with a different substrate leaving group the acylation of the enzyme could contribute significantly to the rate as well. For example, acylation has been shown to be rate determining for BChE hydrolysis of a thioester substrate.[59]

As discussed above, not all stationary points in Figure 1 have been calculated. The missing stationary points correspond to the free, uncomplexed enzyme and the transition states to binding the substrate and releasing the products. Energies of these states were estimated based on the magnitude of k_{on} and k_{off} rates measured for enzymatic reactions. The rate constants for association of acetyl-L-tryptophan *p*-nitrophenyl ester with chymotrypsin, also a catalytic triad-containing hydrolase, is 6×10^7 s^{-1} M^{-1}, while the dissociation rate constant is 6×10^4 s^{-1}.[61] These rate constants have been used to estimate barriers to complexation and dissociation for the BChE reaction, completing every stationary point along the catalytic cycle. In this case, the greatest energy span is 12.6 kcal/mol, between trig1 and TS5, still in agreement with experiment.

Generation of Transition State Ensembles and Identification of Consensus Structures

The partial bond distances and angles are rather rigid in transition states,[62] but catalytic groups may readily rotate around single bonds and exist in other local minima while still preserving the saddle point nature of the transition state. That is, other side chain rotamers may give equally effective transition states for catalysis. Large ensembles of structures that may theoretically catalyze each single step were generated by restrained molecular mechanical Monte Carlo conformational searches on each QM structure for chemical transition states,[63] constraining the bonds and angles of the reaction coordinate but allowing all intramolecular torsional parameters to vary, and allowing for the individual molecules in the model system to rotate and translate relative to one another (Methods). The bonds and angles that were constrained are shown as dotted lines and arcs in the theozymes for the esterase chemical steps on the left in Figure 4a–d, and atoms that were frozen in space are shown as spheres. After a 5,000-step search, the 500 structures with the lowest molecular mechanical energies were saved for each chemical TS. These are shown for each step in Figure 4 to the right of the QM optimized structures. These 2,000 total structures represent the wide array of geometries which nature—or the rational enzyme designer of today—might have chosen to catalyze a single step of this reaction.

Structures from the generated ensembles were compared to each other by geometrical similarity. Similarities for each unique pair of structures within the full set of 2,000 were computed by superpositions of catalytically important heavy atoms involved in bond formation, bond breaking, or in transition state stabilization in the case of the oxyanion hole

(shown as spheres in Figure 2a–c). RMSDs of these atoms were measured to provide the metric of reorganizational motion required by one structure to achieve catalysis in the shape of the other structure. The next step was to use these RMSD data in order to identify a "consensus" geometry—a geometry from one step that structurally resembles a geometry from all other steps. How tight such resemblance should be was controlled by a variable model-to-model threshold RMSD value, *r*. If a given structure from one ensemble is similar to at least one member of every other ensemble by an RMSD less than *r*, it is considered a consensus geometry at that threshold. This adds a layer of complexity to traditional clustering, and is illustrated graph theoretically in Figure S3.

At lower threshold RMSDs the structures have partners from the other ensembles that are more and more similar to them, thus the less motion is required to interconvert between steps, and presumably the lower the inter-step reorganizational barriers are in TS3 and TS7 of Figure 1. Different threshold values give consensus ensembles of varying sizes. For example, the entire set of 2,000 structures is shown in Figure 5 alongside consensus geometry ensembles at thresholds of 1.2 Å and 0.95 Å. These structures have been superimposed over the serine γ-oxygen and substrate carbonyl carbon and oxygen. With a threshold of 1.2 Å, more rearrangement is tolerated, and there are a wide variety of "consensus" active site geometries at this level (Figure 5b). With a more strict threshold of $r = 0.95$ Å the geometries are more tightly clustered towards a region where the imidazole/acetate models of the His/Glu dyad are in positions that appear most able to perform the required general acid/base chemistry with both the serine γ-oxygen and substrate leaving group oxygen (Figure 5c). The active site from the crystal structure of butyrylcholinesterase (1XLV)[64] is shown in all black in Figures 5b,c, superimposed over the same three atoms. As the threshold goes down the structures cluster more tightly to the experimentally observed geometry.

Catalytic Atom RMSD Comparison of Consensus Geometries to Crystal Structure of Butyrylcholinesterase

The consensus structures obtained for different threshold *r* values were superimposed onto the crystal structure for butyrylcholinesterase (1XLV)[64] over the same set of catalytic atoms used to define the consensus threshold (the heavy atoms shown as spheres in Figure 2a–c). For each threshold value of *r*, the catalytic atoms of the consensus structures were superimposed onto the corresponding atoms of the crystal structure. Figure 6 plots the average and maximum RMSDs for the superpositions of each consensus model on the crystal structure in this way. At a high threshold of $r = 2.2$ Å, all 2,000 structures qualify as "consensus" geometries. The average and maximum RMSDs of these 2,000 structures from the crystal structure are 2.12 Å and 3.39 Å, respectively. At an intermediate threshold, $r = 1.25$ Å, 597 structures pass the cutoff. These give average and maximum RMSDs of 1.60 Å and 2.94 Å from the crystal. At the lowest threshold, $r = 0.85$ Å, only 4 structures remain, but both the average model-to-crystal and maximum model-to-crystal RMSDs are the lowest of any threshold, 1.25 Å and 1.28 Å. The red curve shows the greatest RMSD of a consensus structure from the crystal structure at a given threshold. It remains quite high until a threshold of around $r = 1.0$ Å, below which all consensus structures are relatively close matches to the crystal structure. This provides a quantitative complement to the qualitative picture given by the ensembles with $r = 1.2$ and 0.95 Å in Figure 5b,c. The tapering of both traces in Figure 6 down and to the left shows that at lower threshold RMSDs the resulting consensus structures are progressively better matches to the geometry that has been perfected by nature. This means that the natural active site has evolved to an architecture that may be objectively defined as requiring the least motion that is theoretically possible for catalysis of its reaction.

Figure 7a shows the structure from the catalytic ensembles with the lowest consensus threshold, $r = 0.81$ Å. This structure is superimposed onto the crystal structure for butyrylcholinesterase

in Figure 7b. The RMSD over the 10 catalytic atoms is 1.26 Å, and all of the functional groups in the theoretical consensus structure show good alignment with the experimental structure. The catalytically relevant heteroatoms on the acetate and imidazole molecules in the consensus geometry match up almost identically to those in the natural amino acids glutamate and histidine, respectively, and the oxyanion hole model 2-(formylamino)-N-methylacetamide is placed directly on top of the natural diglycine moiety. The N-methyl-3-hydroxy-propanamide is in a slightly different orientation to the natural serine, but the catalytic atoms are still very close, allowing the consensus serine model to provide hydrogen bond stabilization as part of the oxyanion hole just as the natural serine does. The structure that requires the least inter-step structural reorganization of all is thereby shown to be an excellent match to the geometry perfected by nature.

These results suggest that the active site of esterase enzymes minimizes necessary inter-step reorganization during the course of their multi-step catalytic mechanism, answering the first question given in the Introduction. In the context of the compromise between the preorganization dictated by the protein tertiary structure and the reorganization required to catalyze multiple steps of a reaction, these results show that active site preorganization plays a dominating role over reorganizational motions. While it has been proposed that enzyme motions facilitate catalysis,[23,24] the active site of the serine esterase enzymes may be objectively defined as minimizing reorganizational motion during the reaction, in that the less motion a theoretical active site geometry requires the more natural it becomes (Figure 6). The consensus geometry shown in Figure 7 was selected from the set of 2,000 shown in Figure 5a solely through the investigation of which theoretical structures require the least motion during catalysis, and without any reference to the natural active site. This does not mean the less motion the better. For example in cases where product release is rate-limiting, an increase in flexibility along a release channel can prove beneficial.[65] When conformational rearrangement is required in the process of catalyzing a multi-step reaction, however, it has been minimized as much as possible.

Preorganization of enzyme active sites is known to be an important contributor to the rate enhancements produced by enzymes,[37] and previous work has shown that enzymes have lower intra-step reorganization energy barriers than the same reaction in solution.[7,8,41] It has been proposed that enzymes may minimize conformational motion required for catalysis,[6,41] but here this proposal has been confirmed quantitatively by a unique combination of quantum and molecular mechanics. This finding represents an important advance in our understanding of enzyme catalysis, as it has been shown that there are a multitude of geometries to which an active site could preorganize if electrostatic stabilization of transition states were the only governing principle (Figure 5). Enzymes preorganize to an active site geometry that electrostatically stabilizes the transition states of the reaction and minimizes conformational reorganization.

Comparisons Between X-Ray Crystal Structures of Esterase Active Sites and Quantum Mechanical Theozyme Geometries

Does the minimization of reorganizational motion limit catalysis to the extent that theoretically ideal geometries are no longer possible for each of the different steps? To test this, internal coordinate parameters were measured for a set of thirty-three crystal structures of ester hydrolases (Methods). These parameters were compared to the parameters calculated in the unconstrained theozyme models to assess if ideal geometries are still attainable within the constraints of a preorganized active site that limits motion as much as possible.

The PDB structures come from three separate fold families—the carboxylic ester hydrolase, acetylcholinesterase, and bacterial lipase families. As a measure of similarity, the residues composing the catalytic triad were overlaid using a least-squares fitting method (Figure 8).

These overlays were focused on the side-chain heteroatoms of the triad residues, since their relative positions are the most important for catalysis. The positions of the backbone atoms are not important for catalysis, and the vectorial origin of the triad residues is mostly fold-dependent. The positions of the side chains that compose the catalytic triad are remarkably conserved (RMSD ≤ 0.8 Å for the 6 atoms overlaid) considering that they originate from different scaffolds, and most of the deviations arise from the C atom of the carboxylate residue, which may be either Asp or Glu. Within the same fold, the RMSD's are below 0.45 Å. The geometric similarity of catalytic triads from different folds has been noted previously by the Thornton group and used to help define catalytic unit templates, which in turn were used to identify the functions of proteins of known three-dimensional structure but unknown function.[66–68]

Internal geometric relationships were measured for each individual active site. The parameters of interest are listed in Figure 9. For each inter-residue or residue-substrate interaction, there are six parameters: one distance, two angles, and three dihedrals, the degrees of freedom needed to relate one rigid body to another. The ranges found in the crystal structures are tabulated in Table 1 in four columns, each composed of a different set of structures. These sets are 1) covalently inhibited structures, in which a tetrahedral intermediate analogous to tet1-4 in Figure 1 is mimicked, 2) noncovalently inhibited structures, in which the inhibitor interacts mainly with substrate-binding residues and less so with the catalytic triad, 3) unliganded proteins, and 4) the complete set of all enzymes from the prior three sets. The data for all 18 internal coordinate parameters are presented, although seven are unique to the covalently inhibited structures and cannot be compared across categories.

The magnitudes of the ranges in Table 1 are largely similar across the three subsets of structures. For example, the ω1 parameter of the His/carboxylate interactions, corresponding to the His-Cε1/His-Nδ1/carboxylate-Oδ1 angle, has a range of 15° in the covalently inhibited structures, 17° in the noncovalently inhibited structures, and 24° in unliganded structures. It might have been thought that covalently inhibited structures would have tighter ranges since they are potentially more restricted to a catalytically competent geometry. This is the case for the dihedral angles of the His/carboxylate interactions, but this expectation cannot be generalized. In fact, the ranges are slightly broader among the Ser/His angles and dihedrals. This may reflect the fact that in the covalent intermediate there is a second hydrogen bond acceptor present, O2 in Figure 9, to compete with the serine Oγ for histidine's hydrogen bond donor, although the differences are slight. The fact that each range is similar across the three groups indicates that systematic changes in conformation do not occur upon ligand binding, neither covalent nor noncovalent.

The comparisons of parameters obtained from experimental crystal structures to those calculated by quantum mechanics are shown in Table 2. The quantum mechanical parameters are from the fully optimized 14 structures along the reaction coordinate, shown in Figure 2d. The first column of values shows the ranges found in the full set of crystal structures, and the columns to the right summarize the QM parameters, divided into categories according to the type of stationary point: all ground states, trigonal ground states, all transition states, tetrahedral ground states, and transition states for histidine changing hydrogen bond coordination between the serine and substrate (TS3 and TS7 in Figure 1).

Since proteins are dynamic structures, additional tolerances have been added to the empirically observed values to account for thermal motion. The energy accessible by thermal fluctuation is on the order of RT, ~0.6 kcal/mol at 310K. Based on this energy and knowledge-based potentials for hydrogen bonds in proteins, the tolerances chosen were ± 0.1 Å for distances, ± 15° for angles, and ± 30° for dihedrals.[69] These potentials have been shown to be in good agreement with QM calculations.[70] The bottom rows of Table 2 indicate the number of internal

coordinate ranges measured for the QM structures that are outside of the ranges measured in the crystal structures, with and without thermal tolerances.

There are fourteen parameters from the full set of QM theozymes that are outside the empirically observed ranges, but this number drops to 3 with thermal tolerances. Most of the out-of-bounds parameters are associated with trigonal ground states. For instance, with thermal tolerances the transition states only have 2 values outside this range, while the tetrahedral ground states and the transition states for changing hydrogen bond coordination have none. Upon further databank mining, the trigonal geometries are the only ones with parameters outside experimental ranges (Supporting Information). The TS3 and TS7 geometries are the closest matches to the crystal structures, with 14 out of 18 parameters within the measured ranges (Table 2). These geometries may be seen as coming from a "consensus" step in the mechanism, between the chemical transition states involving the serine and leaving group/water oxygens.

In answer to the second question asked in the Introduction, nature has indeed evolved to produce the theoretically optimal arrangement for all steps of catalysis. Combining the first part of this study with the second, it has been shown that the esterase active site is preorganized to a geometry that minimizes inter-step conformational reorganization while still allowing for optimal transition state stabilization of each step.

Conclusions

The theoretical results obtained in this work show that serine esterases can deploy the catalytic triad and oxyanion hole in a three-dimensional array that optimally stabilizes every step in the catalytic cycle—what we refer to as the consensus catalytic geometry. Comparisons of this theoretical consensus catalytic geometry to that in natural enzymes show that active sites of serine esterases have all evolved to achieve this consensus geometry. Consequently, these serine esterases perform multiple steps of a reaction with a minimum of conformational rearrangement (Figure 5, 6, 7). The preorganization of enzyme active sites has been recognized as an important contributor to the acceleration of the catalyzed reaction over the uncatalyzed,[4,71] but here we show objectively that nature has chosen a reaction pathway involving the least motion possible during catalysis. These results are in accord with a recent NMR study of dihydrofolate reductase, which investigated loop motions required between successive bound states of combinations of ligand, cofactor, and product. The study concludes that "the number and heights of the energetic barriers between consecutive conformations have been minimized."[72]

The principle of minimal active site reorganization during catalysis mirrors the principle of minimal frustration during protein folding. The principle of minimal frustration states that proteins proceed from the unfolded to the folded state along an energy landscape that has minimized conformational barriers by avoiding thermodynamic traps.[73,74] Similarly, enzymes have minimized conformational barriers to active site reorganization during catalysis, and can proceed along the reaction coordinate by a "minimally frustrated" path.

Minimal active site reorganization may dictate not only the geometry of the active site, but also the mechanisms and catalytic units that are available for enzyme catalysis of a given reaction. For example, esterase and protease reactions involve attack of trigonal substrates, with intermediates and transition states in more tetrahedral geometries. Here the atoms participating in acid/base and nucleophilic catalysis are positioned along tetrahedral axes. These are close enough that only one general acid/base unit may be used for nucleophilic additions of both serine and water with minimal reorganization. In contrast, hydrolysis of phosphodiesters involves attack of a tetrahedral center, with intermediates and transition states

in trigonal bipyramidal geometries. When this reaction is catalyzed in nature two general acid/base units are employed, one for each nucleophilic addition occurring 180° apart.[75] It may be hypothesized that a single catalytic unit such as a catalytic triad would require too much reorganization to catalyze every step of the reaction, and the conformational barrier prohibits that mechanism.

The observation that a single consensus catalytic geometry describes the active site of an enzyme provides a useful paradigm for the field of enzyme design.[49–51,76,77] Just as a polypeptide sequence's native structure may be predicted based on its global minimum energy conformation,[78] the native architecture of an active site may be predicted from the mechanism based on the minimally reorganizing conformations. The recent *de novo* design of retro-aldolase enzymes incorporated the need to stabilize multiple transition states by superimposing several structures along the reaction coordinate to create a composite structure that also shows minimal rearrangement.[50]

The quantum and molecular mechanical modeling techniques used here have been demonstrated to be useful tools for identifying productive geometries,[79] and in the future they may be carried over to truly non-natural reactions without natural precedent. The structure with the lowest RMSD threshold for catalytic atom variation, shown in Figure 7 for the serine esterase mechanism, could be used directly to provide distance and angle values for algorithms such as DEZYMER or RosettaMatch.[80,81] Alternatively, the quantum mechanical structure of TS3 could be used, which is the single step that is best seen as a consensus of them all and the best match to the data set of thirty-three esterse crystal structures used here (Table 2). This step places the imidazole/carboxylate unit between the serine oxygen and leaving group oxygen, in a position in which it can interact with both. The resulting designs would have the best chance at providing optimal transition state stabilization of all steps with minimal motion, as has been shown here occurs in nature in the active sites of esterase enzymes.

Methods

Model System and Quantum Mechanical Calculation of Stationary Points Along the BChE Reaction Coordinate

The active site model system was taken from a crystal structure of a covalently inhibited butyrylcholinesterase (1XLV).[64] The nucleophilic serine (Ser198) is bonded to an ethyl dihydrogen phosphate, which takes on a tetrahedral geometry in the active site. Ser198 was modeled by an N-methyl-3-hydroxy-propanamide, His438 by an imidazole ring, Glu325 by an acetate, and the Gly115-Gly117 oxyanion hole by 2-(formylamino)-N-methylacetamide. Methyl acetate was the model for a substrate ester. Geometry optimizations were carried out by hybrid density functional theory at the B3LYP/6-31G(d) level[82,83] using Gaussian03 in the gas phase, with default convergence criteria.[84] The procedure for quantum mechanical theozyme geometry optimization is depicted in Figure 10. Frequency calculations were performed to confirm the nature of each stationary point and to evaluate free energies.

The goal was to compute stationary points in conformations that closely resemble the natural enzyme active site (Figure 1). While there are other quantum mechanical geometries that may be possible, this was done to enable the calculation of the effective barrier that could be compared to the experimentally observed barrier given by the k_{cat} of the reaction. Geometric diversity was obtained by the molecular mechanics-based procedure using the QM transition states as described below.

For the transition structures, geometries were taken from the 1XLV structure, and partial bonds of the reaction coordinate were manually adjusted to a near-TS guess value. These partial bonds were frozen during an initial QM optimization to allow all non-reaction internal coordinates

to optimize to energy minima. The entire system was then unfrozen during a full TS optimization.

The ground states were obtained by Intrinsic Reaction Coordinate (IRC) calculations.[85,86] In IRC calculations the geometry of the transition state is incrementally stepped along the reaction coordinate towards the reactants and products that the TS connects. The reaction coordinate is determined by the negative force constant that gave rise to the single imaginary frequency characteristic of transition states. In one case, trig3, the IRC calculation led to a geometry in which the 2-(formylamino)-N-methylacetamide backbone model had rotated out of position to form a hydrogen bond between one of its amides and the catalytic water molecule. In the native enzyme the backbone is constrained by its peptide bonds to remain relatively fixed, with the hydrogen bond donors pointing towards the acylenzyme oxygen. Previous theozyme calculations on butyrylcholinesterase have used explicit restraints to account for this,[87] but here the solution was to replace the methanol in trig2—which was optimized by IRC from TS4—with a water molecule and optimize to the nearest ground state.

Generation of Active Site Geometry Ensembles for Chemical Steps

The ensembles for the four chemical steps involving bond breaking and bond formation were generated by a Monte Carlo Multiple Minimum (MCMM)[63] search in MacroModel.[88] QM structures were assigned OPLS-AA atom types. The OPLS-AA atom types themselves would be expected to be poor at reproducing transition state geometries, so restraints were applied to maintain the partial bonds and angles characteristic of each individual transition state during ensemble generation. In this manner, the transition state internal coordinate values obtained quantum mechanically are maintained in the molecular mechanics ensembles. Carbons in between tetrahedral and trigonal hybridizations were frozen, as were their four connected atoms and heavy atoms making or breaking a bond. Distances between hydrogen-bonding heavy atoms were constrained with a force constant of 1,000.0 kJ/mol•Å^2. Angles involving partial bonds, hydrogen bonding heavy atom/hydrogen/heavy atom angles, and hydrogen bonding hydrogen/heavy atom/heavy atom angles where the first distance is nonbonded were restrained with a force constant of 1,000.0 kJ/mol•rad^2. The variable torsional and molecular rotational and translational parameters were set automatically using the “Perform Automatic Setup” feature of MacroModel. Each ensemble is made up of the 500 lowest energy fully converged minima by OPLS-AA under these constraints, following a 5,000 step conformational search.

Catalytic Atom Superpositions and Identification of Consensus Geometries

Heavy atoms that participate in bond forming, bond breaking, proton transfer, or transition state stabilization will be referred to as catalytic atoms. These ten atoms are: acetate H-bonding oxygen, two imidazole nitrogens, N-methyl-3-hydroxypropanamide nucleophilic oxygen and H-bonding nitrogen, two N-formyl-glycine-N-methylamide nitrogens, substrate carbonyl carbon and oxygen, and either the leaving group or water oxygen according to the stationary point. These form the set over which geometries are superimposed and RMSDs computed to compare structural similarities. The 500 structures constituting each theozyme ensemble were combined into one 2,000 structure Maestro formatted file. This file was processed by the program XCluster,[89] superimposing each structure on the other 1,999 and calculating an RMSD.

XCluster produces an output distance file, containing all 1,999,000 RMSDs in space-delimited form. The sequential order in which the data are written can be related to the order in which the structures were given in the Maestro file by the following formulae for the entry number i containing the RMSD between structures x and y out of a total of n structures: if $x < y$, $i = (x - 1) \times n - 0.5(x(x + 1)) + y$; if $x > y$, $i = (y - 1) \times n - 0.5(y(y + 1)) + x$. A data mining script was written to read these data into an array, and the entries comparing each structure from each

transition state ensemble to structures from each other ensemble were identified using these formulae. If the structure in question had a partner in all three other ensembles with an RMSD less than a specified threshold *r* value it was added to a list to be written to a separate Maestro file after all structures had been checked. These Maestro files were then used in turn to compare the consensus structures at each threshold to the crystal structure over the same ten catalytic atom set.

Geometrical Parameters Derived From Crystal Structures

Thirty-three structures of carboxylic ester hydrolases containing an active site catalytic triad (17 liganded and 16 unliganded) were chosen from the PDB for detailed analysis based on resolution (≤ 2.2 Å), R-value (≤ 0.21), and wild-type sequence (no mutants were considered). These structures are listed in Table S1. These include structures of 20 distinct enzymes from three different fold families (carboxylic ester hydrolase like, acetylcholinesterase like, and bacterial lipase like) that are either unliganded, bound noncovalently to inhibitors or covalently bound to tetrahedral inhibitors. All active sites with non-crystallographic symmetry were analyzed separately, expanding the structure data set to 79 individual structures.

The geometric definition for each residue was based on a distance, two angles and three dihedral angles between 6 atoms of the target residue, a reference residue and/or the bound ligand as described by Hellinga *et al*.[80] For example, the position of the serine residue relative to the covalently bound ligand is defined by one distance, two angles, and three dihedral angles using the Cα, Cβ, and Oγ atoms of serine and C1, O1, and O2 atoms of the ligand. The histidine residue is represented relative to the serine and to the bound ligand, and the geometric definition includes atoms from all three components. Finally, the carboxylate (Asp or Glu) is represented solely relative to the histidine. Including all different crystal structures from the data set, an experimental range was measured for each parameter in the geometric definition (which gives also an estimate for the conformational flexibility of the residue within the triad). Such ranges account for the different productive orientations that a residue participating in the catalytic triad can adopt in natural enzymes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

A. J. T. S. and K. N. H. are grateful for the support of the National Institutes of Health (Chemistry-Biology Interface Grant T32 GM008496, and the Medical Scientist Training Program Grant T32 GM08042 and research grant GM-36700. All authors are grateful for the support and computational time from DARPA as part of the PDP project.

References

1. Wolfenden R, Snider MJ. Acc. Chem. Res 2001;34:938–945. [PubMed: 11747411]
2. Lad C, Williams NH, Wolfenden R. Proc. Natl. Acad. Sci. U. S. A 2003;100:5607–5610. [PubMed: 12721374]
3. Warshel A. Proc. Natl. Acad. Sci. U. S. A 1978;75:5250–5254. [PubMed: 281676]
4. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MH. Chem. Rev 2006;106:3210–3235. [PubMed: 16895325]
5. Gertner BJ, Wilson KR, Hynes JT. J. Chem. Phys 1989;90:3537–3558.
6. Cannon WR, Singleton SF, Benkovic SJ. Nat. Struct. Biol 1996;3:821–833. [PubMed: 8836096]
7. Roca M, Moliner V, Tunon I, Hynes JT. J. Am. Chem. Soc 2006;128:6186–6193. [PubMed: 16669689]
8. Yadav A, Jackson RM, Holbrook JJ, Warshel A. J. Am. Chem. Soc 1991;113:4800–4805.
9. Olsson MH, Parson WW, Warshel A. Chem. Rev 2006;106:1737–1756. [PubMed: 16683752]

10. Gutteridge A, Thornton J. J. Mol. Biol 2005;346:21–28. [PubMed: 15663924]
11. Yuan Z, Zhao J, Wang ZX. Protein Eng 2003;16:109–114. [PubMed: 12676979]
12. Yang LW, Liu X, Jursa CJ, Holliman M, Rader AJ, Karimi HA, Bahar I. Bioinformatics 2005;21:2978–2987. [PubMed: 15860562]
13. Sacquin-Mora S, Lavery R. Biophys. J 2006;90:2706–2717. [PubMed: 16428284]
14. Hou L, Honaker MT, Shireman LM, Balogh LM, Roberts AG, Ng KC, Nath A, Atkins WM. J. Biol. Chem 2007;282:23264–23274. [PubMed: 17561509]
15. Demetrius L. J. Theor. Biol 1998;194:175–194. [PubMed: 9778432]
16. James LC, Tawfik DS. Trends Biochem. Sci 2003;28:361–368. [PubMed: 12878003]
17. Pervushin K, Vamvaca K, Vogeli B, Hilvert D. Nat. Struct. Mol. Biol 2007;14:1202–1206. [PubMed: 17994104]
18. Vamvaca K, Vogeli B, Kast P, Pervushin K, Hilvert D. Proc. Natl. Acad. Sci. U. S. A 2004;101:12860–12864. [PubMed: 15322276]
19. Holliday GL, Almonacid DE, Mitchell JB, Thornton JM. J. Mol. Biol 2007;372:1261–1277. [PubMed: 17727879]
20. Fischer E. Ber. Dtsch. Chem. Ges 1894;27:3189–3232.
21. Koshland DE. Proc. Natl. Acad. Sci. U. S. A 1958;44:98–104. [PubMed: 16590179]
22. Radkiewicz JL, Brooks CL 3rd. J. Am. Chem. Soc 2000;122:225–231.
23. Benkovic SJ, Hammes-Schiffer S. Science 2003;301:1196–1202. [PubMed: 12947189]
24. Tousignant A, Pelletier JN. Chem. Biol 2004;11:1037–1042. [PubMed: 15324804]
25. Gerstein M, Lesk AM, Chothia C. Biochemistry 1994;33:6739–6749. [PubMed: 8204609]
26. Hammes-Schiffer S. Biochemistry 2002;41:13335–13343. [PubMed: 12416977]
27. Hammes-Schiffer S, Benkovic SJ. Annu. Rev. Biochem 2006;75:519–541. [PubMed: 16756501]
28. Rozovsky S, McDermott AE. J. Mol. Biol 2001;310:259–270. [PubMed: 11419951]
29. Watt ED, Shimada H, Kovrigin EL, Loria JP. Proc. Natl. Acad. Sci. U. S. A 2007;104:11981–11986. [PubMed: 17615241]
30. Nicholson LK, Yamazaki T, Torchia DA, Grzesiek S, Bax A, Stahl SJ, Kaufman JD, Wingfield PT, Lam PY, Jadhav PK, et al. Nat. Struct. Biol 1995;2:274–280. [PubMed: 7796263]
31. Eletsky A, Kienhofer A, Hilvert D, Pervushin K. Biochemistry 2005;44:6788–6799. [PubMed: 15865424]
32. Arora K, Brooks CL 3rd. Proc. Natl. Acad. Sci. U. S. A 2007;104:18496–18501. [PubMed: 18000050]
33. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. Nature 2007;450:913–916. [PubMed: 18026087]
34. Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson MA, Petsko GA, Karplus M, Hubner CG, Kern D. Nature 2007;450:838–844. [PubMed: 18026086]
35. Nunez S, Antoniou D, Schramm VL, Schwartz SD. J. Am. Chem. Soc 2004;126:15720–15729. [PubMed: 15571394]
36. Quaytman SL, Schwartz SD. Proc. Natl. Acad. Sci. U. S. A 2007;104:12253–12258. [PubMed: 17640885]
37. Villa J, Warshel A. J. Phys. Chem. B 2001;105:7887–7907.
38. Billeter SR, Webb SP, Agarwal PK, Iordanov T, Hammes-Schiffer S. J. Am. Chem. Soc 2001;123:11262–11272. [PubMed: 11697969]
39. Eisenmesser EZ, Bosco DA, Akke M, Kern D. Science 2002;295:1520–1523. [PubMed: 11859194]
40. Agarwal PK, Billeter SR, Rajagopalan PT, Benkovic SJ, Hammes-Schiffer S. Proc. Natl. Acad. Sci. U. S. A 2002;99:2794–2799. [PubMed: 11867722]
41. Liu H, Warshel A. Biochemistry 2007;46:6011–6025. [PubMed: 17469852]
42. Fuxreiter M, Warshel A. J. Am. Chem. Soc 1998;120:183–194.
43. Guallar V, Jacobson M, McDermott A, Friesner RA. J. Mol. Biol 2004;337:227–239. [PubMed: 15001364]
44. Hedstrom L. Chem. Rev 2002;102:4501–4524. [PubMed: 12475199]
45. Knowles JR. Nature 1991;350:121–124. [PubMed: 2005961]

46. Tantillo DJ, Chen J, Houk KN. Curr. Opin. Chem. Biol 1998;2:743–750. [PubMed: 9914196]
47. Quinn DM. Chem. Rev 1987;87:955–979.
48. Bazelyansky M, Robey E, Kirsch JF. Biochemistry 1986;25:125–130. [PubMed: 3954986]
49. Bolon DN, Voigt CA, Mayo SL. Curr. Opin. Chem. Biol 2002;6:125–129. [PubMed: 12038994]
50. Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF 3rd, Hilvert D, Houk KN, Stoddard BL, Baker D. Science 2008;319:1387–1391. [PubMed: 18323453]
51. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik D, Baker D. Nature 2008;453:190–195. [PubMed: 18354394]
52. Topf M, Richards WG. J. Am. Chem. Soc 2004;126:14631–14641. [PubMed: 15521783]
53. Warshel A, Naray-Szabo G, Sussman F, Hwang J-K. Biochemistry 1989;28:3629–3637. [PubMed: 2665806]
54. Vagedes P, Rabenstein B, Åqvist J, Marelius J, Knapp E-W. J. Am. Chem. Soc 2000;122:12254–12262.
55. Zhan CG, Zheng F, Landry DW. J. Am. Chem. Soc 2003;125:2462–2474. [PubMed: 12603134]
56. Zhan CG, Gao D. Biophys. J 2005;89:3863–3872. [PubMed: 16319079]
57. Pan Y, Gao D, Yang W, Cho H, Zhan CG. J. Am. Chem. Soc 2007;129:13537–13543. [PubMed: 17927177]
58. Main AR, Tarkan E, Aull JL, Soucie WG. J. Biol. Chem 1972;247:566–571. [PubMed: 5009701]
59. Masson P, Bec N, Froment MT, Nachon F, Balny C, Lockridge O, Schopfer LM. Eur. J. Biochem 2004;271:1980–1990. [PubMed: 15128307]
60. Froede HC, Wilson IB. J. Biol. Chem 1984;259:11010–11013. [PubMed: 6469995]
61. Renard M, Fersht AR. Biochemistry 1973;12:4713–4718. [PubMed: 4773852]
62. Houk KN, Li Y, Evanseck JD. Angew. Chem. Int. Ed. Engl 1992;31:682–708.
63. Chang G, Guida WC, Still WC. J. Am. Chem. Soc 1989;111:4379–4386.
64. Nachon F, Asojo OA, Borgstahl GE, Masson P, Lockridge O. Biochemistry 2005;44:1154–1162. [PubMed: 15667209]
65. Codreanu SG, Ladner JE, Xiao G, Stourman NV, Hachey DL, Gilliland GL, Armstrong RN. Biochemistry 2002;41:15161–15172. [PubMed: 12484753]
66. Laskowski RA, Watson JD, Thornton JM. Nucleic Acids Res 2005;33:W89–W93. [PubMed: 15980588]
67. Torrance JW, Bartlett GJ, Porter CT, Thornton JM. J. Mol. Biol 2005;347:565–581. [PubMed: 15755451]
68. Wallace AC, Laskowski RA, Thornton JM. Protein Sci 1996;5:1001–1013. [PubMed: 8762132]
69. Kortemme T, Morozov AV, Baker D. J. Mol. Biol 2003;326:1239–1259. [PubMed: 12589766]
70. Morozov AV, Kortemme T, Tsemekhman K, Baker D. Proc. Natl. Acad. Sci. U. S. A 2004;101:6946–6951. [PubMed: 15118103]
71. Warshel A, Florian J. Proc. Natl. Acad. Sci. U. S. A 1998;95:5950–5955. [PubMed: 9600897]
72. Boehr DD, McElheny D, Dyson HJ, Wright PE. Science 2006;313:1638–1642. [PubMed: 16973882]
73. Bryngelson JD, Wolynes PG. Proc. Natl. Acad. Sci. U. S. A 1987;84:7524–7528. [PubMed: 3478708]
74. Onuchic JN, Wolynes PG. Curr. Opin. Struct. Biol 2004;14:70–75. [PubMed: 15102452]
75. Davies DR, Interthal H, Champoux JJ, Hol WG. J. Chem. Biol 2003;10:139–147.
76. Bolon DN, Mayo SL. Proc. Natl. Acad. Sci. U. S. A 2001;98:14274–14279. [PubMed: 11724958]
77. Lippow SM, Tidor B. Curr. Opin. Biotechnol. 2007
78. Bradley P, Misura KM, Baker D. Science 2005;309:1868–1871. [PubMed: 16166519]
79. DeChancie J, Clemente FR, Smith AJT, Gunaydin H, Zhao YL, Zhang X, Houk KN. Protein Sci 2007;16:1851–1866. [PubMed: 17766382]
80. Hellinga HW, Richards FM. J. Mol. Biol 1991;222:763–785. [PubMed: 1749000]
81. Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, Althoff EA, Röthlisberger D, Baker D. Protein Sci 2006;15:2785–2794. [PubMed: 17132862]

82. Becke AD. J. Chem. Phys 1993;98:5648–5652.
83. Lee C, Yang W, Parr RG. Phys. Rev. B Condens. Matter 1988;37:785–789. [PubMed: 9944570]
84. Frisch, MJ., et al. Wallingford, CT: Gaussian, Inc.; 2004.
85. Fukui K. Acc. Chem. Res 1981;14:363–368.
86. Gonzalez C, Schlegel HB. J. Phys. Chem 1990;94:5523–5527.
87. Suárez D, Díaz N, Fontecilla-Camps J, Field MJ. Biochemistry 2006;45:7529–7543. [PubMed: 16768449]
88. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrickson T, Still WC. J. Comput. Chem 1990;11:440–467.
89. Shenkin PS, McDonald DQ. J. Comput. Chem 1994;15:899–916.

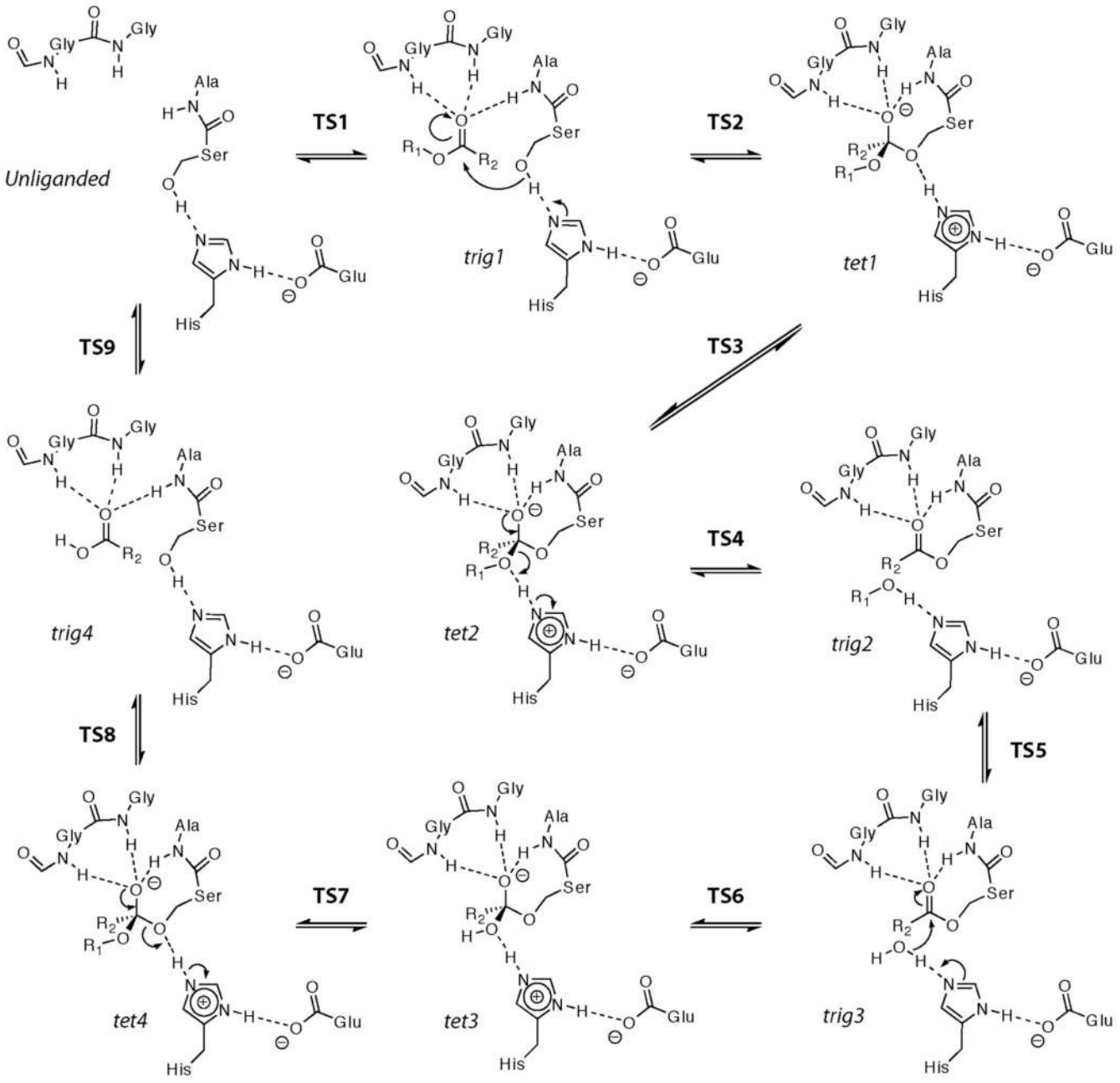

Figure 1.

Mechanism of catalysis of ester hydrolysis by butyrylcholinesterase. $R_1 = Me_3N^+CH_2CH_2$, R_2 = propyl for the natural substrate, butyrylcholine. Liganded ground states are named in italics according to the substrate geometry and their sequential order in the cycle, starting from the unliganded state and proceeding clockwise. Transition states are named in bold for their sequential order. TS1, TS3, TS5, TS7, and TS9 are conformational transition states, in which no bonds are made and broken, while TS2, TS4, TS6, and TS8 are chemical transition states, in which bonds are made and broken. TS1, TS5, and TS9 permit the binding of substrate and release of products from the active site. These may involve motions distal to the active site and have not been modeled explicitly. TS2 and TS8 are for the forming and breaking of the serine-

substrate bond, respectively. TS4 and TS6 are for the breaking of the substrate-leaving group bond and forming of the substrate-hydroxide bond, respectively. TS3 and TS7 are conformational transition states, in which the hydrogen bond coordination of the histidine changes from serine Oγ to substrate leaving group oxygen and from the nucleophilic water molecule oxygen to the serine Oγ, respectively.

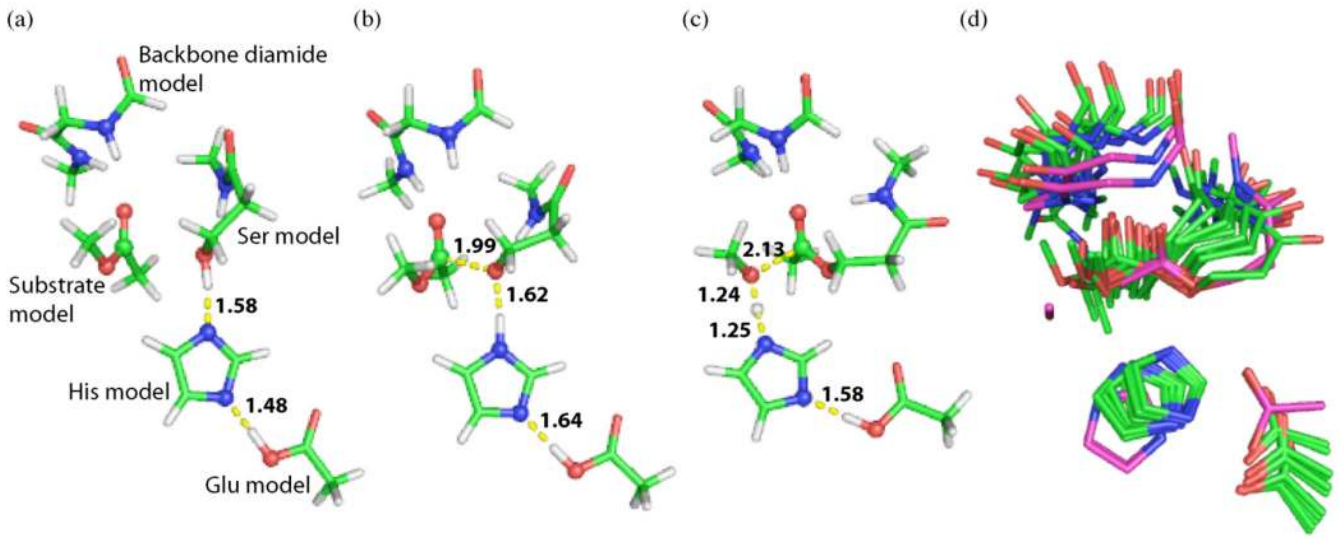

Figure 2.

Quantum mechanically optimized stationary point geometries, distances labeled in Ångstroms. a) The Michaelis complex, or trig1, as labeled in Figure 1. b) TS2, for serine nucleophilic addition to the substrate. c) TS4, for breaking the bond from the tetrahedral intermediate to the leaving group. Residue models are labeled in a). The heavy atoms shown as spheres represent the "catalytic atoms" over which theozyme models are superimposed in order to compare geometric similarity and identify consensus structures. d) All fourteen stationary points calculated for the esterase reaction, superimposed over six atoms in the catalytic triad (O, Cβ in Ser; N, N in His; O, Cδ in Glu). QM theozymes were superimposed with TS3 as a reference (trig2 and trig3 in magenta carbons, all others in green carbons).

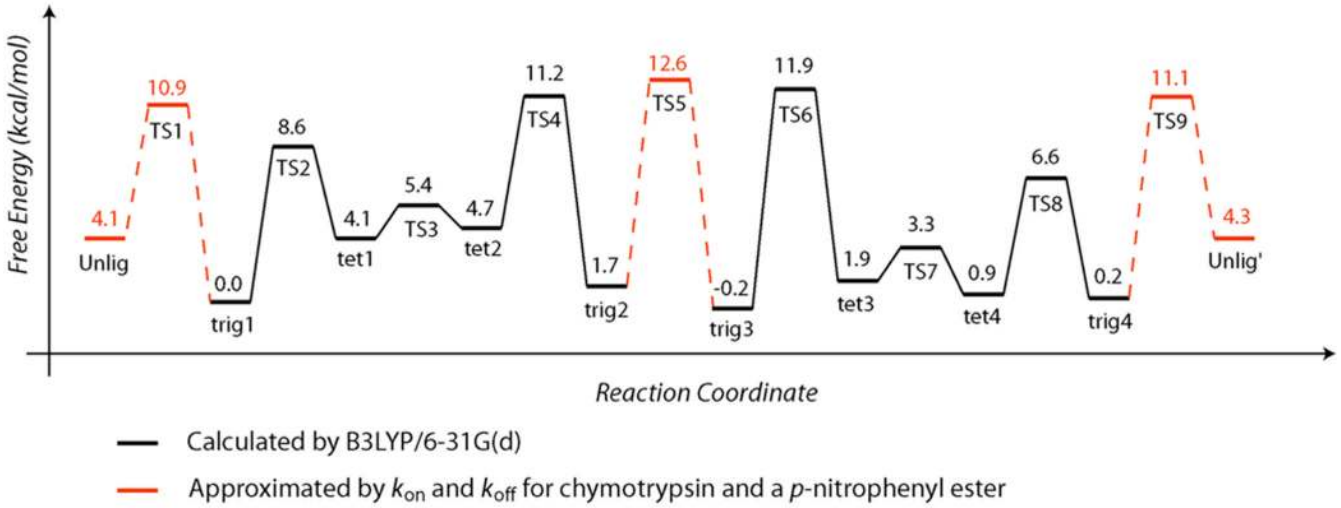

Figure 3.

Free energy diagram along the BChE reaction coordinate. Black lines correspond to stationary points whose energies were calculated quantum mechanically, while red energies were estimated by analogy to the known association and dissociation rates of chymotrypsin and a *p*-nitrophenyl ester. The energies of the unliganded geometries were estimated from bimolecular rate constants assuming standard state concentrations.

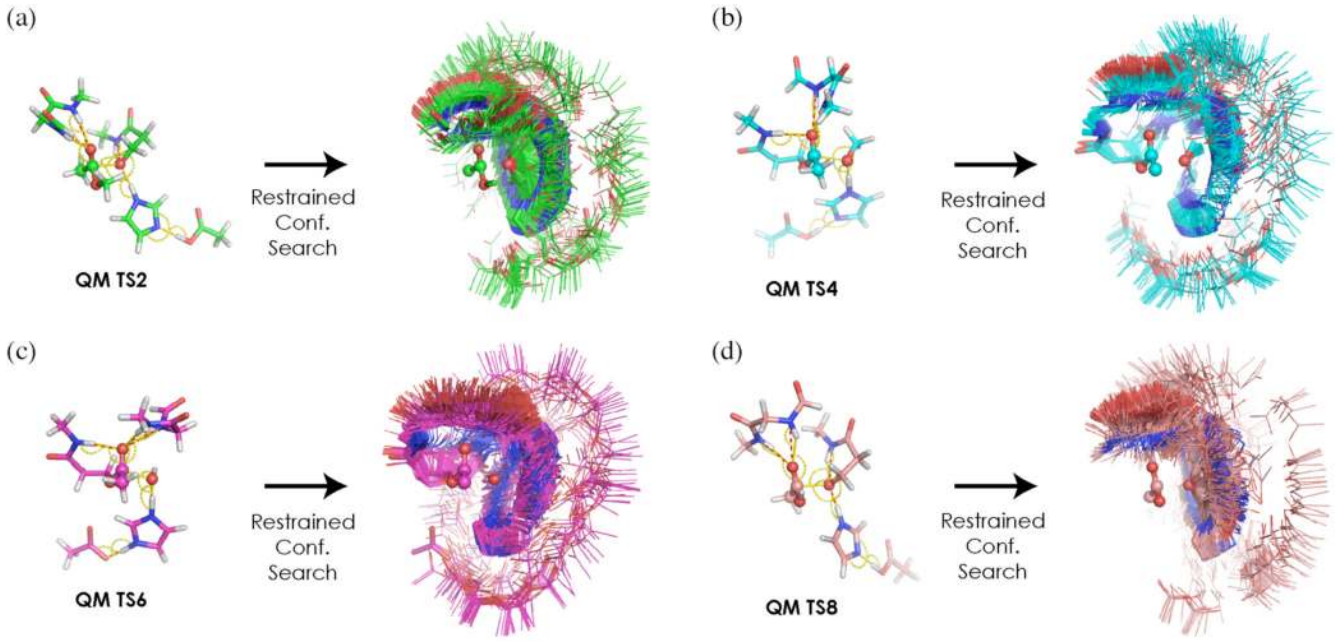

Figure 4.

Generation of ensembles of structures for each chemical step. Quantum mechanical transition states were calculated for each chemical step, a) TS2 in green carbons, b) TS4 in aqua carbons, c) TS6 in magenta carbons, and d) TS8 in salmon carbons. These were used as a template for restrained conformational searches to generate structures that could potentially catalyze that single step. Dashed lines and arcs represent the restrained distances and angles; atoms represented as spheres were frozen in space. After a 5,000 step conformational search the 500 structures with the lowest molecular mechanics energies were saved. These ensembles are shown on the right of panels a–d.

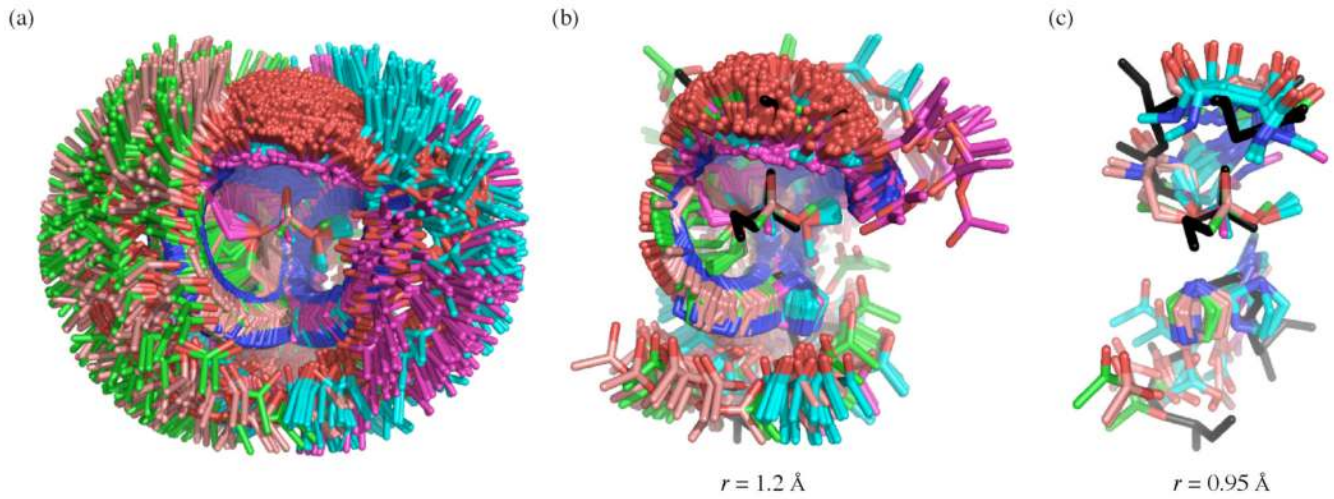

Figure 5.

Whole set of ensemble structures and example consensus geometries at different threshold *r* values. a) All 2,000 ensemble structures (those shown on the right side of Figure 4a–d) have been superimposed over the substrate carbonyl carbon and oxygen, and serine model Oγ. The TS2 ensemble is in green carbons, TS4 in aqua carbons, TS6 in magenta carbons, and TS8 in salmon carbons. Consensus geometries were identified using a threshold value *r* for RMSD similarity over the catalytic atoms shown in spheres in Figure 2a–c as described in the text and in Methods. Consensus geometries for *r* = 1.2 Å are shown in b) and for *r* = 0.95 Å in c). In panels b) and c) the active site of a crystal structure of butyrylcholinesterase (1XLV)[64] is shown in all black atoms. In panel b) the consensus structures show wide variability and no preference towards the crystal structure, while in panel c) the consensus structures at this lower *r* value cluster tightly around the natural active site geometry.

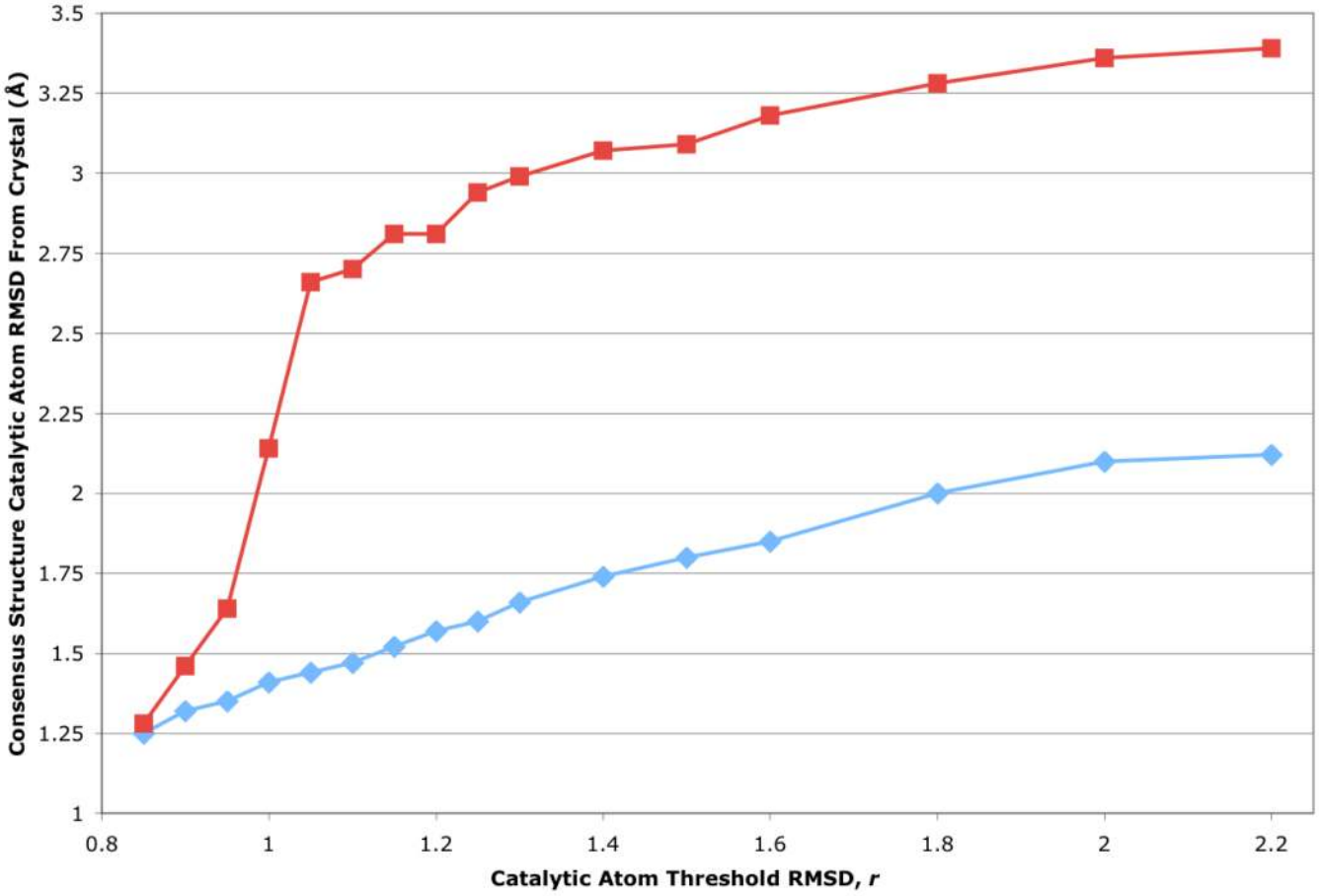

Figure 6.

Plot of average and maximum catalytic atom RMSDs of consensus structures from the crystal structure at different threshold RMSDs, *r*, for defining consensus structures. Blue diamonds give the average catalytic atom RMSD from the crystal structure over all consensus geometries at each threshold *r* value. Red squares give the largest catalytic atom RMSD from the crystal found for a single theoretical structure at each threshold *r* value.

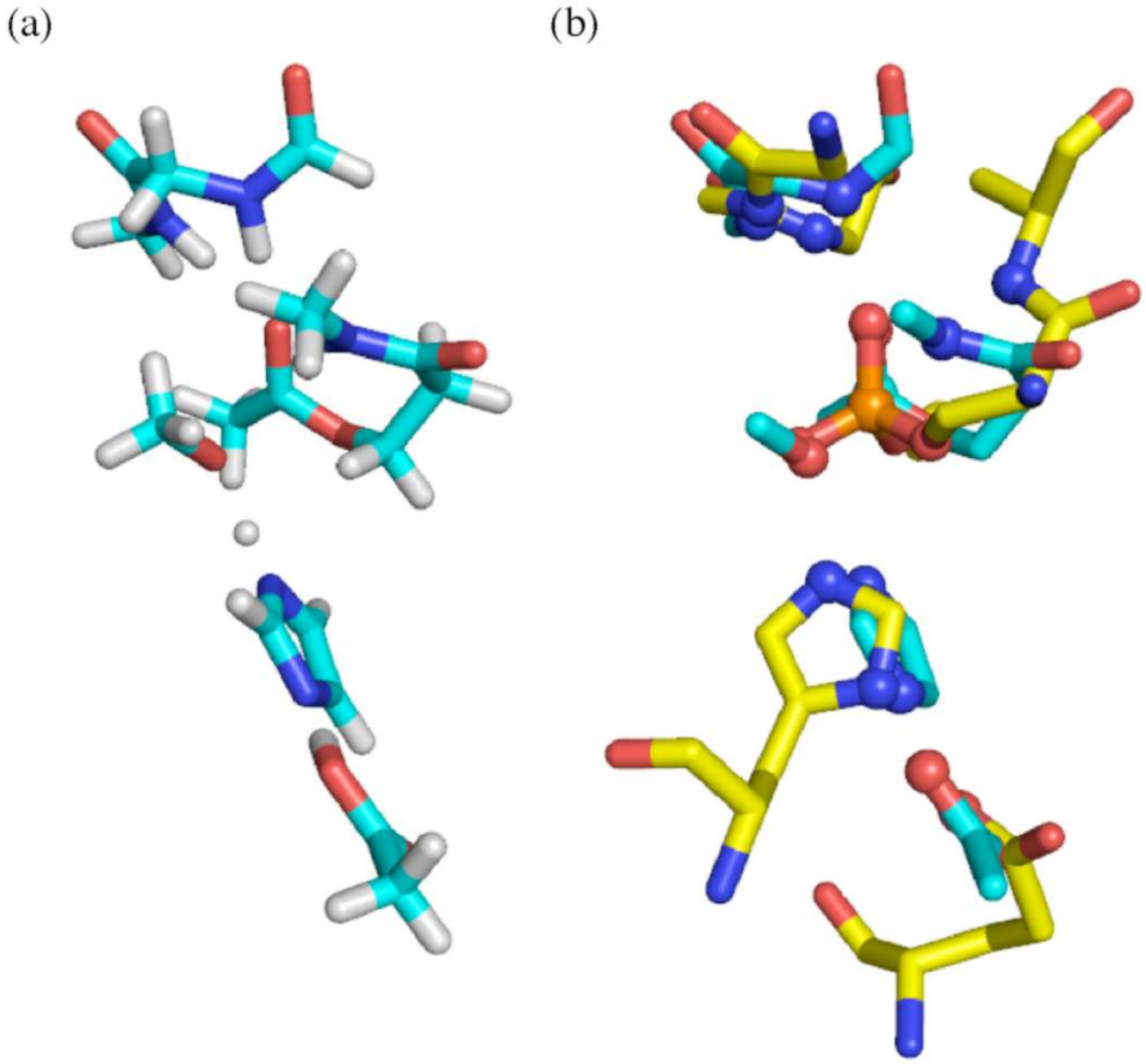

Figure 7.
Consensus structure with the lowest threshold *r* value and comparison to crystal structure. The structure with the lowest *r* value out of the 2,000 computed and shown in Figure 5a comes from the TS4 ensemble, in aqua carbons in panel a). This structure was compared to the crystal structure in yellow carbons by superimposition over ten catalytic atoms shown in spheres in panel b).

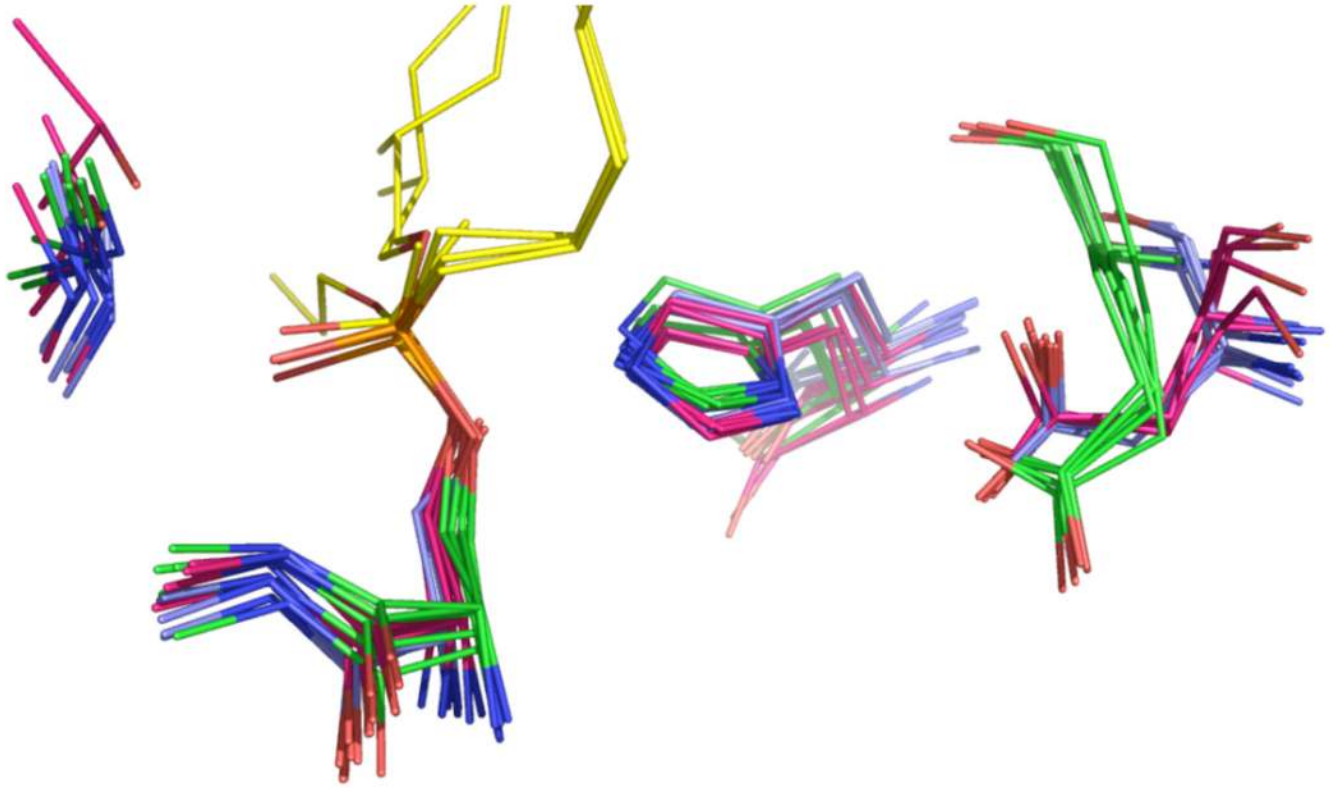

Figure 8.

Superpositions of esterase crystal structures from three different fold families over 6 heavy atoms: serine Oγ and Cβ, histidine nitrogens, and carboxylate residue's carboxylate carbon and His-coordinating oxygen. The 1TQH structure was used as a reference. Structures from the carboxylesterase-like fold colored in pink, acetylcholinesterase-like in green, and bacterial lipase-like in blue. Shown are the catalytic triad residues, Ser-His-Asp/Glu, the oxyanion hole and covalently bound inhibitors (yellow).

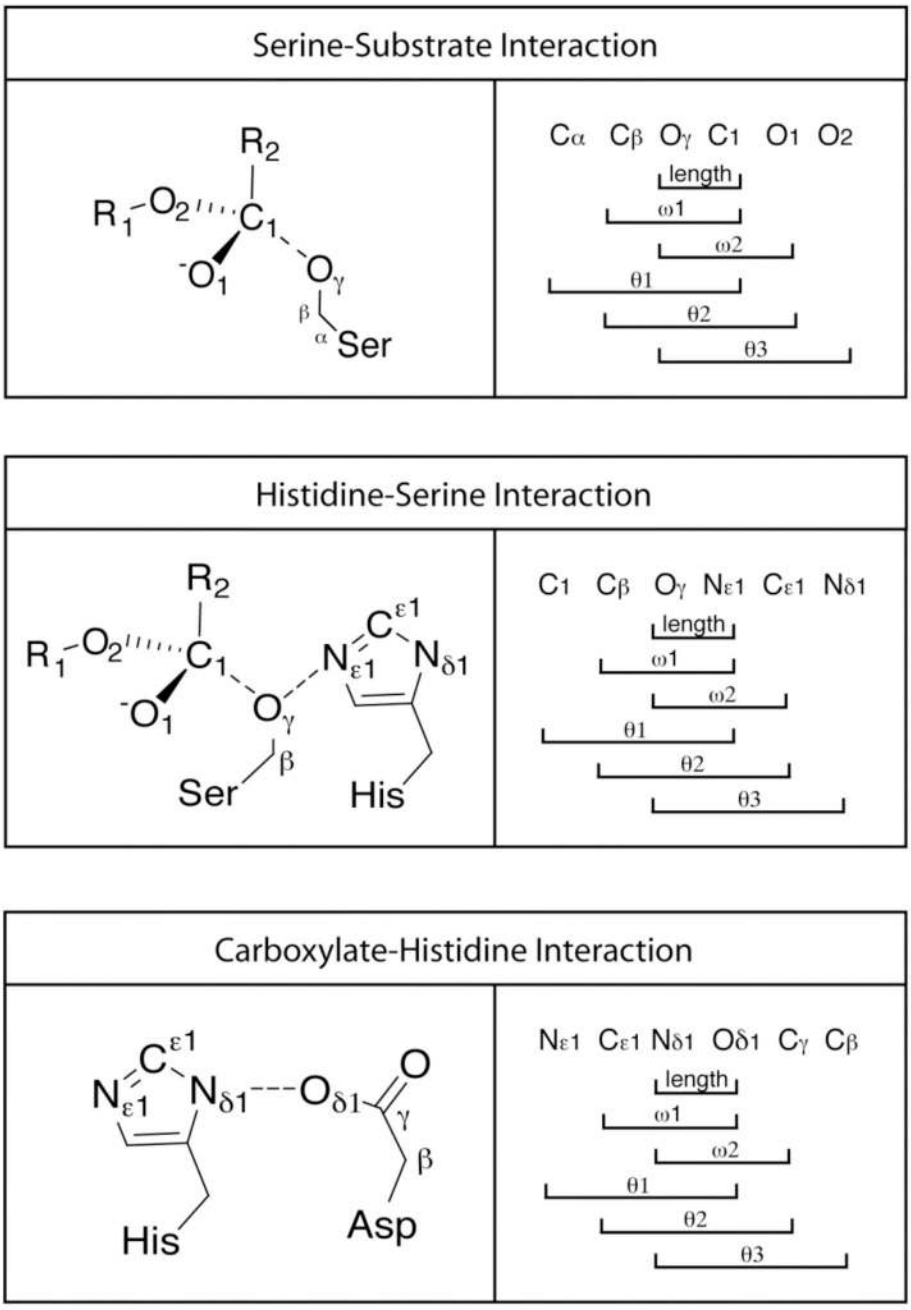

Figure 9.
Internal coordinate parameters extracted from the crystal structures and from theozyme geometries. The examples show how an interaction can be fully described by a distance, 2 angles (ω1, ω2), and three dihedrals (θ1, θ2, θ3). Only heavy atoms are shown.

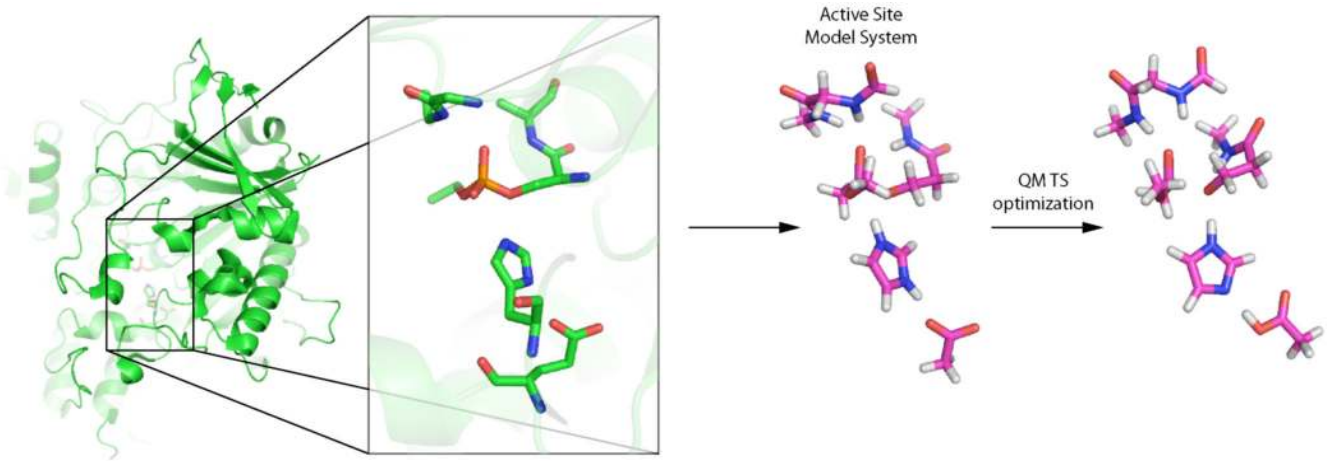

Figure 10.

Theozyme modeling of the butyrylcholinesterase active site. The entire butyrylcholinesterase enzyme (1XLV)[64] is shown on the left. The active site is enlarged to show the catalytic residues and the phosphonate inhibitor. These residues are truncated to the essential functional groups in the theozyme model system. The active site atoms are then fully optimized quantum mechanically to structures such as those shown on the right.

Table 1

Ranges of internal coordinate parameters derived from crystal structures, categorized as covalently inhibited, noncovalently inhibited, unliganded, and the full data set.

Interacting Pair	Parameter	Covalent	Noncovalent	Unliganded	All
Serine-Substrate	Length (Å)	1.43:1.73	n.d.	n.d.	1.43:1.73
	ω1 (deg.)	111:128	n.d.	n.d.	111:128
	ω2 (deg.)	107:113	n.d.	n.d.	107:113
	θ1 (deg.)	−145:−116	n.d.	n.d.	−145:−116
	θ2 (deg.)	3:76	n.d.	n.d.	3:76
	θ3 (deg.)	−125:−108	n.d.	n.d.	−125:−108
Serine-Histidine	Length (Å)	2.51:3.22	2.45:3.00	2.40:3.13	2.40:3.22
	ω1 (deg.)	81:138	94:137	79:119	79:138
	ω2 (deg.)	80:122	95:118	74:111	74:122
	θ1 (deg.)	−180:−90,165:180	n.d.	n.d.	−180:−90,165:180
	θ2 (deg.)	−95:−36	−116:−58	−109:−66	−116:−36
	θ3 (deg.)	−180:−152,164:180	−180:−174,159:180	−180:−164,168:180	−180:−152,159:180
Histidine-Carboxylate	Length (Å)	2.54:2.87	2.54:2.84	2.40:3.08	2.40:3.08
	ω1 (deg.)	128:143	123:140	123:147	123:147
	ω2 (deg.)	101:166	106:162	97:166	97:166
	θ1 (deg.)	−180:−175,163:180	−180:−172,167:180	−180:−158,139:180	−180:−158,139:180
	θ2 (deg.)	−145:−94	−81:−1	−139:−62,−14:−5	−145:−1
	θ3 (deg.)	−169:−131	−179:−89	−172:−136,−107:−90	−179:−89

Table 2

Ranges of internal coordinate parameters derived from crystal structures and QM stationary points. Bold numbers indicate where theozymes are outside of experimentally measured ranges. Bold and italicized numbers are theozyme ranges that are outside of the crystal ranges ± thermal tolerances of 0.1 Å for distances, 15° for angles, and 30° for dihedrals.

Interacting Pair	Parameter	Crystal Structures	QM: All	All Ground States	Trigonal Ground States
Serine-Substrate	Length (Å)	1.43:1.73	***1.32:2.92***	***1.32:2.92***	***1.32:2.92***
	ω1 (deg.)	111:128	**105:121**	**105:121**	**105:121**
	ω2 (deg.)	107:113	**95:124**	**95:124**	**95:124**
	θ1 (deg.)	−145:−116	**−137:−113**	**−131:−113**	**−121:−113**
	θ2 (deg.)	3:76	**6:77**	**6:77**	**6:77**
	θ3 (deg.)	−125:−108	**−134:−92**	**−134:−92**	**−134:−92**
Serine-Histidine	Length (Å)	2.40:3.22	***2.62:4.93***	***2.62:4.93***	***2.62:4.93***
	ω1 (deg.)	79:138	95:125	102:125	108:123
	ω2 (deg.)	74:122	***50:135***	***50:135***	***50:135***
	θ1 (deg.)	−180:−90,165:180	**−134:−87**	**−134:−87**	**−110:−87**
	θ2 (deg.)	−116:−36	**−118:−32**	**−118:−32**	**−118:−32**
	θ3 (deg.)	−180:−152,159:180	**−180:−150,170:180**	**−180:−150,174:180**	**−180:−150,178:180**
Histidine-Carboxylate	Length (Å)	2.40:3.08	2.53:2.70	2.57:2.70	2.57:2.69
	ω1 (deg.)	123:147	**117:123**	**118:123**	**121:123**
	ω2 (deg.)	97:166	106:113	106:113	106:113
	θ1 (deg.)	−180:−158,139:180	−180:−172,179:180	−180:−179,179:180	−180:−179,179:180
	θ2 (deg.)	−145:−1	**−1:0,0:7**	**−1:0,0:3**	**−1:0,0:3**
	θ3 (deg.)	−179:−89	**−180:−176,178:180**	**−180:−176**	**−180:−176**
QM ranges outside of crystal ranges		NA	14	14	14
QM ranges outside of crystal ranges with thermal tolerances		NA	3	3	3

All Transition States	Tetrahedral Ground States	TS3 and TS7
1.36:2.12	1.45:1.52	1.47:1.48
108:119	115:119	117:119
104:121	**111:114**	**113:114**
−137:−120	−133:−119	−129:−120
21:56	20:47	21:38
−120:−107	−121:−119	−120:−119

All Transition States	Tetrahedral Ground States	TS3 and TS7
2.64:3.42	**2.83:3.26**	3.08:3.09
95:121	102:125	104:121
70:133	**84:132**	100:102
−121:−99	−134:−97	−118:−104
−94:−62	−86:−50	−78:−62
−180:−177,170:180	−179:−180,174:180	−180:−177,179:180
2.53:2.70	2.69:2.70	2.69:2.70
117:119	**118:119**	**119:119**
108:111	109:110	109:109
−180:−172	−180:−179	−179:−179
−1:0,0:7	**0:3**	**1:2**
177:180	**−180:−180**	**180:180**
9	6	4
2	0	0