THREE HYDROLASES AND A TRANSFERASE: COMPARATIVE ANALYSIS OF ACTIVE-SITE DYNAMICS VIA THE BIOSIMGRID DATABASE†

*Kaihsu Tai[a], Marc Baaden[b], Stuart Murdock[c,‡], Bing Wu[a], Muan Hong Ng[d], Steven Johnston[c,d], Richard Boardman[c,d], Hans Fangohr[d], Katherine Cox[a], Jonathan W. Essex[c], Mark S. P. Sansom[a,]**

[a]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom

[b]Laboratoire de Biochimie Théorique, CNRS UPR 9080, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, F-75005 Paris, France

[c]School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

[d]School of Engineering Science, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

[‡]*Current address:*
Schrödinger Inc., 101 SW Main Street, Suite 1300, Portland, OR 97204, USA

*To whom correspondence should be addressed at:
Fax: +44 1865 275273; email: mark.sansom@bioch.ox.ac.uk

for submission to *Journal of Molecular Graphics and Modelling*

Abstract

Comparative molecular dynamics (MD) simulations enable us to explore the conformational dynamics of the active sites of distantly related enzymes. We have used the BioSimGrid (http://www.biosimgrid.org) database to facilitate such a comparison. Simulations of four enzymes were analyzed. These included three hydrolases and a transferase, namely acetylcholinesterase, outer-membrane phospholipase A, outer-membrane protease T, and PagP (an outer membrane enzyme which transfers a palmitate chain from a phospholipid to lipid A). A set of 17 simulations were analyzed corresponding to a total of ~0.1 μs simulation time. A simple metric for active site integrity was used to demonstrate the existence of clusters of dynamic conformational behaviour of the active sites. Small (i.e. within a cluster) fluctuations appear to be related to the function of an enzymatically active site. Larger fluctuations (i.e. between clusters) correlate with transitions between catalytically active and inactive states. Overall, these results demonstrate the potential of a comparative MD approach to analysis of enzyme function. This approach could be extended to a wider range of enzymes using current high throughput MD simulation and database methods.

Keywords: biomolecular simulation, database, data mining, molecular dynamics, catalytic triad, conformational change

Introduction

Molecular dynamics (MD) simulations of proteins allow us to address molecular motions of importance to biological function (*1, 2*). Such motions may be on a relatively small scale (e.g. sidechain and water mobility at a ligand binding site), or on a larger time and length scale (e.g. inter-domain motions in complex proteins). For example it has been shown that protein dynamics is important in the context of enzyme efficiency (*3*).

One aspect of such simulations that has received some attention (*4-6*) is that of incomplete sampling of conformational space over the finite duration of an MD simulation. Another is that of the more general biological relevance of any motion observed. Both of these aspects may be addressed via analysis of *multiple* simulations (*7*). In particular, comparative analysis of multiple simulations of several proteins from a given protein family may be used to exploit evolution to help overcome the sampling problem for larger (and slower) conformational changes (*8*). In the longer term, comparative analysis of a wide range of simulations may enable the development of heuristics making it possible to relate protein folds to patterns of protein dynamics.

It is also interesting to explore possible integration of biomolecular simulations and structural genomics. This requires development of a number of components, including tools for storage and analysis of simulation data, comparable to those provided for static protein structures by the protein data bank (*9*). We are developing a distributed database for storage and analysis of biomolecular simulations, BioSimGrid (*10*) (http://www.biosimgrid.org/), the use of which is illustrated in this paper. Comparable endeavours include SimDB (http://simdb.bch.msu.edu/) (*11*), GEMS (http://gipse.cse.nd.edu/GEMS/) (*12*), and the Ascona B-DNA Consortium (http://humphry.chem.wesleyan.edu:8080/MDDNA/) (*13*).

One aspect of comparative MD is that of analysis of the rigidity/flexibility of active sites in a number of (distantly) related enzymes. This is an attractive application as it adds a temporal dimension to comparisons of (static) active site geometries for a family of related enzymes. The latter (structural bioinformatics) approach has been used for e.g. comparative analysis of active site geometries of e.g. serine esterases (*14, 15*), and is embodied in e.g. the Catalytic Sites Atlas

(http://www.ebi.ac.uk/thornton-srv/databases/CSA/).

Using BioSimGrid, we compare MD simulations of three hydrolases and a transferase: (i) acetylcholinesterase (AChE; Enzyme Classification 3.1.1.7; Protein Data Bank code: 1MAH, 1KU6), a key enzyme of the nervous system which degrades the neurotransmitter acetylcholine (*16*); (ii) outer-membrane phospholipase A (OMPLA; EC 3.1.1.32; PDB codes: 1QD5, 1QD6), a bacterial outer membrane enzyme which degrades phospholipid (*17*); (iii) outer-membrane protease T (OmpT; EC 3.4.21.87; PDB code: 1I78) (*18*); and PagP, an outer membrane enzyme which transfers a palmitate chain from a phospholipid to lipid A (*19, 20*). (PDB code 1THQ). The first two enzymes are carboxylic ester hydrolases (EC 3.1.1). OmpT, a protease, belongs to the category of peptide hydrolases. PagP, is – strictly speaking – an acyl transferase.

Crystallographic structures show that these enzymes do not have related folds, but have similar active sites, a catalytic triad of amino acid residues (Figures 1 and 2). In mouse (*Mus musculus*) AChE, the active site (Ser203, His447, Glu334) has the favorable arrangement conducive to proton transfer where the imidazole ring of His447 is placed between the carboxyl group of Glu334 and the hydroxyl group of Ser203 (*16, 21, 22*). In *Escherichia coli* OMPLA, the active site consists of the analogous Ser144, His142, and Asn156. Earlier MD simulations suggested that the presence of a calcium ion induces hydrogen-bond networks near the active site (*17*). OmpT has the central His212, with a water molecule held by Asp83 (in turn held by Asp85) probably serving as the attack group equivalent to the role serine plays in the earlier two enzymes; finally Asp210 is the third residue in the triad (*18*). Though a transferase, the putative triad in PagP, consisting of His33, Asp76, and Ser77, fits nicely into the hydrolase triad paradigm. Similar to OmpT, the active site is extracellular (*23*); and similar to OMPLA, it has been suggested to act as a homodimer (*24*).

In this study we employ comparative analysis of multiple MD simulations of these enzymes to explore possible functionally significant patterns of catalytic side-chain dynamics in these four superficially unrelated enzymes, to investigate the relationship between side-chain mobility and catalytic mechanism. We show that comparative MD simulations are able to distinguish between "functional" and "non-functional" catalytic triads.

Methodology

In total, 17 trajectories of the four enzymes were deposited in the BioSimGrid database (Table 1). All except the PagP simulations have already been published (*16-18, 21, 22*). The PagP simulations were initiated from the X-ray structure of the PagP monomer (*20*), embedded in a DMPC bilayer (Cox & Sansom, ms. in preparation). The OMPLA, OmpT and PagP simulations were performed using Gromacs (http://www.gromacs.org/) and a modified Gromos96 forcefield (*25*). The AChE simulations were performed using the packages NWChem (http://www.emsl.pnl.gov/docs/nwchem/nwchem.html) (*26*) and TINKER (http://dasher.wustl.edu/tinker/), with the AMBER force field (*27*). All told, the trajectories amount to more than 100 ns of simulation time.

The three residues in the catalytic triads are Ser (in the case of OmpT, Asp83), His, and Asx|Glx (Asn156 in OMPLA; Glu334 in AChE; Asp210 in OmpT; Asp76 in PagP) (see Fig. 1). To provide metrics for the intactness of the active sites, we analysed a number of measurements, using methods of multivariate analysis and visualization before settling on a metric illustrated in Fig. 2. Selecting atoms Cγ (CG) of His, the sidechain carboxyl/amide atom in Asx|Glx, and atom Oγ (OG) in residue Ser (in the case of OmpT, Cγ of Asp83), we measured the pairwise distance between these three atoms. The retrieval of these distance time series was performed using the Python-language interface to BioSimGrid (*10*) to produce a Python program (see supporting information, triad_metric.py) which generates the relevant scatter plots.

Structural bioinformatics analyses have focused on "template" positions of catalytic triads derived from crystal structures of serine proteases and lipases (*14, 15*). To explore the application of this analysis to MD trajectories we superimposed the active sites studied so as to identify such recurring clusters in the trajectories. Again using a Python program in the BioSimGrid interface (see supporting information , triad_super.py), we extracted 9 atoms from each of the active sites: the imidazole ring (5 atoms) and the β-carbon atom in His, the post-branch atoms (δ- or ε-, oxygen or nitrogen; 2 atoms) in Asx|Glx, and the γ-oxygen atom in Ser. The His-fragment was superimposed

by minimizing the root mean square deviation (RMSD) of the atoms.

Results

Our underlying hypothesis is that, given the structural bioinformatics studies of Thornton and coworkers (*15*), one might expect a degree of conservation of catalytic triad dynamics between the active sites of the different enzymes, even though their overall structures are rather different. The simulations analysed, summarised in Table 1, were selected to provide a wide range of simulation configurations and conditions. AChE is a "classical" example of a Ser esterase. Simulations with/without a bound inhibitor (fasciculin) and with two different protonation states of the active site histidine were compared. OMPLA is a rather complex enzyme, a lipase which shows differences in catalytic activity between its monomeric and dimeric states. Simulations with/without covalently bound (to the active-site Ser) inhibitor and of the monomeric versus dimeric enzyme were compared. We also analysed two sets of simulations, with different protocols for the treatment of long-range electrostatic interactions in the simulation (*28, 29*) to explore whether such differences had any profound effect on the active site dynamics. OmpT is a protease, with an active site somewhat more distantly related to that of the classical Ser-His-Asn triad. Again, multiple simulations were compared with respect to the possible influence of the long-range electrostatic interaction protocols used in the simulations. Finally, PagP is a much more distantly related enzyme. It is a transferase, which in both the nuclear magnetic resonance (NMR) and the X-ray structures appears to have an incompletely formed active site. We included PagP to enable comparison of its active site conformational dynamics with those of some better understood enzymes, and also to explore the influence of His33 protonation on such dynamics.

Two metrics were selected (see above) to measure the integrity of each active site as a function of time in the simulations (Fig. 2): d_1 measures the Ser-His (or equivalent) distance, and d_2 the His-Asx|Glx distance. (The third possibility d_3 which measures the Ser-Asx|Glx distance was shown not to result in as significant a spread as d_1 and d_2 (not shown) and so was excluded from further analysis). The most notable feature of the data is that (d_1,d_2) shows significant clustering by

enzyme. This is especially evident if we reduce the "raw" metric data (sampled every 1 or 10 ps; Fig. 3A) to the corresponding metrics sampled every 1 ns (Fig. 3B). This reveals that to a large extent the (d_1,d_2) clusters for the various enzymes do not overlap. The exception is AChE which shows three distinct clusters, one of which (for simulation AChEδ) overlaps with the OMPLA cluster, revealing a sensitivity to the protonation state of the His sidechain (see below).

Significantly, the enzymatically active His catalytic configurations (with a singly δ-protonated His) of OMPLA and AChE seem to form the "tightest" active-site conformation, centred around (d_1,d_2) = (0.5 nm, 0.5 nm). As expected, the OmpT active site showed a similar d_1 (i.e. His-Asx|Glx) distance to that of OMPLA (d_1 ~ 0.4 nm) but the OmpT d_2 (His:Cγ–Asp83:C) distances are about ~0.5 nm more distant, at d_2 ~ 0.9 nm. This is consistent with the suggestion (*18*) that an intervening water molecule may form the nucleophilic group instead of a serine in the productive active site of this enzyme.

In the OMPLA dimer simulations, especially the longer ones (i.e. OMPLA2 and OMPLA3) the two dimers in the same simulation do not exhibit the same (d_1,d_2) distribution (data not shown separately for the constituent monomers). This may reflect an intrinsic asymmetry in the dynamics of the OMPLA dimmer, such as would be anticipated for an alternating active sites mechanism of catalytic activity. Alternatively, it may be a manifestation of each monomer's "memory" of its starting conformation, i.e. a reflection of incomplete sampling. However, we do see an overall reduction of active-site flexibility upon ligand binding, taking all samples (monomers and dimers) into account (Fig. 3A). Curiously, the corresponding monomer simulation (OMPLA1) does not have as wide a spread as in each dimer simulation. Using methods discussed in (*6*), we show in Fig. 3B the average measurements of 1 ns trajectory segments, which provides us with some indication of the sampling adequacy in the active sites. Within the coverage of these two active-site metrics, the sampling may be seen to be by and large sufficient, as evidenced by the revisiting of previous states. The major exceptions are the PagP trajectories whose spreads are wide (as discussed further below), and whose active sites do not seem to have revisited earlier conformational states.

In the AChE simulations, two protonation states of His447 (ε vs. δ) and the

presence/absence of bound fasciculin have been explored. These form three distinct clusters in (d_1,d_2). The unproductively configured AChE active sites (with ε-protonated His447) have longer and more dispersed (d_1,d_2) values than the productive (δ-protonated) active-site cluster, which overlaps with the OMPLA clusters. This suggests that the active site of AChE is dynamically "tighter" under enzymatically active conditions and that this property is conserved between two serine esterases.

As mentioned above, PagP is a more distantly related enzyme, a transferase whose active site is incompletely formed in the X-ray structure used as the starting point of the simulations. For both of the PagP simulations, the (d_1,d_2) clusters are highly disperse and far removed from those for the other enzymes. In the "PagP (charged)" simulation, the His33 is doubly protonated (and hence positively charged) whilst the Asp76 is also ionised. Despite the relative shortness of these simulations, the charged simulation results in a (d_1,d_2) cluster which drifts with respect to time such that by the end of the simulation it overlaps with that of AChE. This suggests that the active site of PagP may be highly flexible, forming only once the His33 is protonated (and perhaps after a substrate molecule binds), and that electrostatic interactions may drive the association of the active site loops.

Finally, in simulations of OMPLA and OmpT we have explored the effects of the treatment of long-range electrostatics on the conformational dynamics of their respective active sites. Interestingly, given the continued discussion of how best to treat such long-range interactions (*28, 29*) there do not seem to be any significant differences in (d_1,d_2) behavior between simulations employing the Ewald and cutoff methods. This suggests a degree of robustness of these simulation results to changes in the simulation protocol.

Following Thornton et al. (*14, 15*) we extracted active-site "template" clusters, sampling frames at each nanosecond from the trajectories OMPLA3, Fas-AChE, and PagP (charged) (Fig. 4), and superimposing the His residues. Examination of these revealed that: (i) the OMPLA3 frames formed a tight cluster of the Ser and Asn residues around the His; (ii) the "His-unproductive" Fas-AChE simulation frames had their Glu more distant from the His than was the Asn of OMPLA3;

and (iii) the Asp and Ser in PagP (charged) were even further from the His, and failed to form a defined cluster. Inspecting the average structure of this "template" set of atoms for all the trajectories (data not shown) yielded similar conclusions to those derived from Fig. 3. In the average structures, the Asx and Ser from OMPLA, OmpT, and the "His-productive" AChE trajectories formed defined clusters on either side of the His. For the PagP trajectory and the "His-unproductive" AChE trajectories, the Ser and Asx|Glx did not fall closely into the above clusters.

Discussion and conclusions

Comparative MD simulations have allowed us to explore both relatively small (<0.1 nm) and and more extensive dynamic fluctuations in active site geometry for a number of (distantly) related enzymes. The dynamic differences between the active sites are quite striking. Small (i.e. within a cluster) fluctuations are likely to be related to the function of an enzymatically active site. Indeed, evidence for changes in conformation of this magnitude during the catalytic process has been presented in e.g. a recent high resolution X-ray study of trypsin reaction intermediates (*30*). Larger fluctuations (i.e. between clusters) in active site metrics are also seen. These appear to correlate with transitions between catalytically active and inactive states. Probably the clearest example of this is for PagP, although the mechanism remains somewhat uncertain, in which 'nucleation' of the active site via changes in extracellular loop conformation is seen during the simulations.

The uneven durations of the simulations do invoke a caveat. The shorter simulations, such as OMPLA4–6 and each of OmpT (cutoff) simulations, may not have as effectively sampled active site conformational space as have the longer simulations, and therefore may not be as representative. In contrast, the wide spread of PagP conformations, despite the relatively short simulations, is indicative of the flexibility of the active site.

In MD studies in general is important that we consider the methodological limitations of individual simulations, especially those relating to simulation forcefields, to protocols (e.g. the use of cutoff vs. PME methods for long range electrostatic interactions), and to the duration of the simulations. Such considerations are of particular relevance in comparative studies, where potential

heterogeneity of simulation methodologies may complicate comparisons. Indeed, it is for this reason that we have made provision for simulation metadata and annotation in within the BioSimGrid database (*10, 31*).

It is valuable to consider also the more general implications of this study. We have analysed 17 MD simulations for 4 enzymes; even this would be difficult 'by hand'. Since the 1970s when the first protein MD appeared (*32*) there has been a near exponential increase in such *in silico* data (*31*). This will be exacerbated by a number of ongoing large scale simulation studies to explore protein stability and folding (e.g. the Dynameomics project http://www.dynameomics.org and the Molecular Dynamics Extended Library project http://mmb.pcb.ub.es/MODEL/). To extract a complete range of information and knowledge from these data, automatic methods of data mining may be needed (*33*). This will be likely to involve some form of model-free machine-learning. However, in the meantime, BioSimGrid provides tools for user-driven comparative analysis of such simulations, and is sufficiently flexible to be scalable to queries across much larger numbers of trajectories than in the current study.

A number of future directions are suggested by the current results. From an enzymological perspective nascent automatic methods for classification of enzyme family and for recognition of active sites (*34*) have emerged. One future extension of our current study would be to apply such methods to search for catalytic triads in the PDB and to use comparative MD simulations to search a wider family of related enzymes. A more specific direction of further investigation would be to our extend coverage of membrane-bound enzymes. Of the ~70 distinct membrane proteins of known structure (*35*), ~16 of those are enzymes. In particular, it would be interesting to modify our metrics for application to fatty acid amide hydrolase (*36*), which has a catalytic dyad with a lysine and a serine, but no histidine in the middle. In general, our results demonstrate the potential of a comparative MD approach to analysis of enzyme function. This is an approach which could be readily extended to a wider range of enzymes, and of other proteins (*8*), and which will benefit from ongoing developments in high throughput MD simulation (e.g. NAMD-G http://www.ncsa.uiuc.edu/UserInfo/SAP/project4.html (*37*); and Dynome http://www.dynome.org).

Acknowledgments

This work was funded by the Biotechnology and Biological Sciences Research Council, United Kingdom. We thank our collaborators in the BioSimGrid Consortium: Leo Caves, Simon Cox, Paul Jeffreys, Syma Khalid, Charles Laughton, David Moss, Adrian Mulholland, and Oliver Smart. We thank Jeremy Kua, J. Andrew McCammon, Christoph Meier, and Yingkai Zhang for trajectories, and Oliver Beckstein, Matthew Dovey, and Kim Henrick for valuable advice and discussions.

References

(1) Karplus, M. J., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules. *Nature Struct. Biol. 9*, 646-652.
(2) Adcock, S. A. M., and McCammon, J. A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev. (in press; DOI: 10.1021/cr040426m)*.
(3) Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., and Kern, D. (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature 438*, 117-121.
(4) Hess, B. (2000) Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E 62*, 8438-8448.
(5) Hess, B. (2002) Convergence of sampling in protein simulations. *Phys. Rev. E 65*, art. no. 031910.
(6) Faraldo-Gómez, J. D., Forrest, L. R., Baaden, M., Bond, P. J., Domene, C., Patargias, G., Cuthbertson, J., and Sansom, M. S. P. (2004) Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations. *Proteins: Struct. Func. Bioinf. 57*, 783-791.
(7) Caves, L. S. D., Evanseck, J. D., and Karplus, M. (1998) Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Sci. 7*, 649-666.
(8) Pang, A., Arinaminpathy, Y., Sansom, M. S. P., and Biggin, P. C. (2005) Comparative molecular dynamics - similar folds and similar motions? *Proteins: Struct. Funct. Bioinf. 61*, 809-822.
(9) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucl. Acids Res. 28*, 235-242.
(10) Tai, K., Murdock, S., Wu, B., Ng, M. H., Johnston, S., Fangohr, H., Cox, S. J., Jeffreys, P., Essex, J. W., and Sansom, M. S. P. (2004) BioSimGrid: towards a worldwide repository for biomolecular simulations. *Org. Biomol. Chem. 2*, 3219-3221.
(11) Feig, M., Abdullah, M., Johnsson, L., and Pettitt, B. M. (1999) Large scale distributed data repository: design of a molecular dynamics trajectory database. *Future Generat. Comp. Sys. 16*, 101-110.
(12) Wozniak, J. M., Brenner, P., Thain, D., Striegel, A., and Izaguirre, J. A. (2005) in *Proc. High Performance Distributed Computing*.
(13) Dixit, S. B., Beveridge, D. L., Case, D. A., Cheatham, T. E., Guidice, E., Lankas, F., Lavery, R., Maddocks, J. H., Osman, R., Sklenar, H., Thayer, K., and P., V. (2005) Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II. Sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys. J. 89*, 3721-3740.
(14) Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1996) Derivation of 3D coordinate templates for searching structural databases: Application to Ser-His-Asp catalytic triads in the serine proteinases and lipases *Protein Sci. 5*, 1001-1013.
(15) Bartlett, G. J., Porter, C. T., Borkakoti, N., and Thornton, J. M. (2002) Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol. 324*, 105-121.
(16) Tai, K., Shen, T., Henchman, R. H., Bourne, Y., Marchot, P., and McCammon, J. A. (2002)

Mechanism of acetylcholinesterase inhibition by fasciculin: a 5-ns molecular dynamics simulation. *J. Am. Chem. Soc. 124*, 6153-6161.

(17) Baaden, M., Meier, C., and Sansom, M. S. P. (2003) A molecular dynamics investigation of mono- and dimeric states of the outer membrane enzyme OMPLA. *J. Mol. Biol. 331*, 177-189.

(18) Baaden, M., and Sansom, M. S. P. (2004) OmpT: molecular dynamics simulations of an outer membrane enzyme. *Biophys. J. 87*, 2942-2953.

(19) Hwang, P. M., Choy, W. Y., Lo, E. I., Chen, L., Forman-Kay, J. D., Raetz, C. R. H., Privé, G. G., Bishop, R. E., and Kay, L. E. (2002) Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc. Nat. Acad. Sci. USA 99*, 13560-13565.

(20) Ahn, V. E., Lo, E. I., Engel, C. K., Chen, L., Hwang, P. M., Kay, L. E., Bishop, R. E., and Privé, G. G. (2004) A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. *EMBO J. 23*, 2931-2941.

(21) Tai, K., Shen, T., Börjesson, U., Philippopoulos, M., and McCammon, J. A. (2001) Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. *Biophys. J. 81*, 715-724.

(22) Kua, J., and Zhang, Y. (2002) Studying enzyme binding specificity in acetylcholinesterase using a combined molecular dynamics and multiple docking approach. *J. Amer. Chem. Soc. 124*, 8260–8267.

(23) Hwang, P. M., Bishop, R. E., and Kay, L. E. (2004) The integral membrane enzyme PagP alternates between two dynamically distinct states. *Proc. Natl. Acad. Sci. USA 101*, 9618-9623.

(24) Bishop, R. E., Gibbons, H. S., Guina, T., Trent, M. S., Miller, S. I., and Raetz, C. R. (2000) Transfer of palmitate from phospholipids to lipid A in outer membranes of gram-negative bacteria. EMBO J. 2000 Oct 2;19(19):5071-80. *EMBO J. 19*, 5071-5080.

(25) van Gunsteren, W. F., Kruger, P., Billeter, S. R., Mark, A. E., Eising, A. A., Scott, W. R. P., Huneberger, P. H., and Tironi, I. G. (1996) *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Biomos & Hochschulverlag AG an der ETH Zurich, Groningen & Zurich.

(26) Kendall, R. A., Aprà, E., Bernholdt, D. E., Bylaska, E. J., Dupuis, M., Fann, G. I., Harrison, R. J., Ju, J., Nichols, J. A., Nieplocha, J., Straatsma, T. P., Windus, T. L., and Wong, A. T. (2000) High performance computational chemistry: an overview of NWChem a distributed parallel application. *Computer Phys. Comm. 128*, 260-283.

(27) Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc. 117*, 5179-5197.

(28) Tobias, D. J. (2001) Electrostatics calculations: recent methodological advances and applications to membranes. *Curr. Opin. Struct. Biol. 11*, 253-261.

(29) Beck, D. A. C., Armen, R. S., and Daggett, V. (2005) Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. *Biochem. 44*, 609-616.

(30) Radisky, E. S., Lee, J. M., Lu, C. J. K., and Koshland, D. E. (2006) Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proc. Natl. Acad. Sci. USA 103*, 6835-6840.

(31) Ng, M. H., Johnston, S., Wu, B., Murdock, S., Tai, K., Fangohr, H., Cox, S. J., Essex, J. W., Sansom, M. S. P., and Jeffreys, P. (2006) BioSimGrid: Grid-enabled biomolecular simulation data storage and analysis. *Future Generation Comput. Sys. 22*, 657-664.

(32) Karplus, M., Gelin, B. R., and McCammon, J. A. (1980) Internal dynamics of proteins: short time and long time motions of aromatic sidechains in PTI. *Biophys. J. 32*, 603-618.

(33) Date, C. J. (2000) *An introduction to database systems*, Addison Wesley Longman, Reading, Massachusetts.

(34) Dobson, P., and Doig, A. (2005) Predicting enzyme class from protein structure without alignments. *J. Mol. Biol. 345*, 187-199.

(35) White, S. H. (2004) The progress of membrane protein structure determination. *Prot. Sci.*

13, 1948-1949.
(36) Bracey, M. H., Hanson, M. A., Masuda, K. R., Stevens, R. C., and Cravatt, B. F. (2002) Structural adaptations in a membrane enzyme that terminates endocannabinoid signaling. *Science 298*, 1793-1796.
(37) Gower, M., Cohen, J., Phillips, J., Kufrin, R., and Schulten, K. (2006) in *Proceedings of the 2006 TeraGrid Conference*.
(38) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD - Visual Molecular Dynamics. *J. Molec. Graph. 14*, 33-38.
(39) Merritt, E. A., and Bacon, D. J. (1997) Raster3D: Photorealistic molecular graphics. *Methods Enzymol. 277*, 505-524.

Table 1: Trajectories used in this study. The treatments of long-range electrostatics are marked "cutoff" or "Ewald". The BioSimGrid identification code is of the form "BioSimGrid_GB-OXF_11", where "GB-OXF" is an ISO 3166-2 location code, and "11" a serial number.

Trajectory	**Description**	**Duration / ns**	**ID code**
OMPLA1	unliganded monomer, Ewald	10	GB-OXF_16
OMPLA2	unliganded dimer, Ewald	10	GB-OXF_17
OMPLA3	liganded dimer, Ewald	10	GB-OXF_15
OMPLA4	unliganded monomer, cutoff	5.7	GB-OXF_9
OMPLA5	unliganded dimer, cutoff	5	GB-OXF_10
OMPLA6	liganded dimer, cutoff	5	GB-OXF_14
AChEε	unliganded, ε-protonated	15	GB-OXF_11
Fas-AChEε	fasciculin 2 bound, ε-protonated	5	GB-OXF_12
AChEδ	His-447 δ-protonated	5	GB-STH_1966
PagP (charged)	His-33 δ,ε-protonated	3	GB-OXF_88
PagP (neutral)	His-33 δ-protonated	4	GB-OXF_31
OmpT (PME)	particle-mesh Ewald	10	GB-OXF_95
OmpT (cutoff)	cutoff	10	GB-OXF_94
OmpT (short)	4 × 2.5 ns trajectories with cutoff	10	GB-OXF_91 to GB-OXF_96

Figures

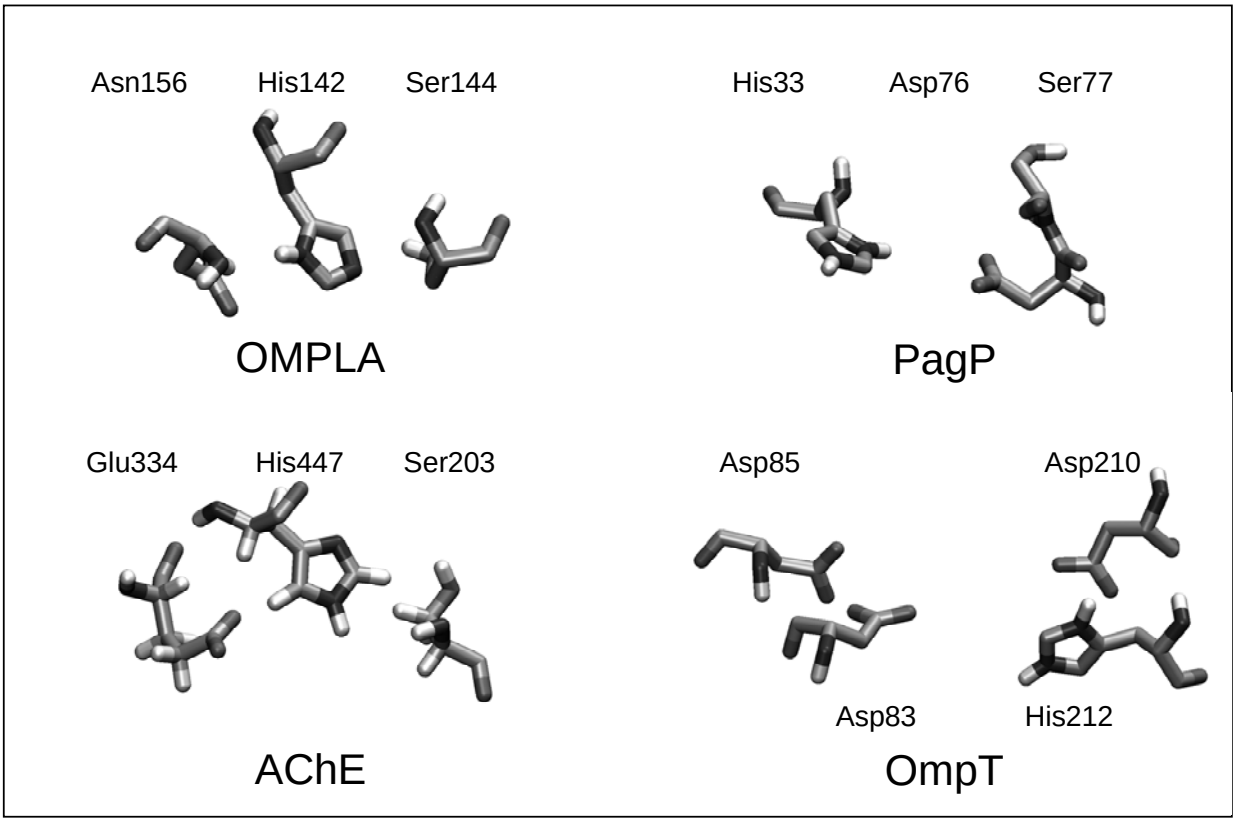
Figure 1: Structures (from snapshots near the beginning of the trajectories) of the active sites of the enzymes included in this study, showing key residues in the active sites of OMPLA and of AChE, and in the proposed active sites of PagP and OmpT. Diagrams generated using VMD (*38*).

Figure 2: Schematics of the active sites studied. Black arrows indicate the Asx|Glx:C–His:CG distance measured (d_1); and grey arrows the His:CG–Ser:OG distance (d_2).

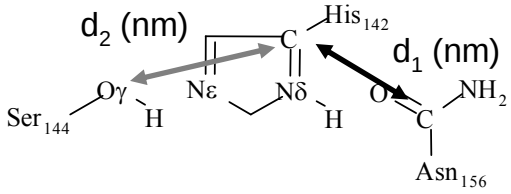
Figure 3: **A.** Scatter-plot of the occurrence of the active site geometry defined by distance pair (d_1,d_2) during the simulations listed in Table 1. In the case of the OMPLA dimer simulations, measurements from both active sites are marked using the same symbol. Note that OMPLA4–6 are sampled every 10 ps, rather than at 1 ps as in the other trajectories. **B.** Average of the distance pair measurements for each full nanosecond, with data-points of adjacent nanoseconds connected by lines.

Figure 4: The catalytic-triad positions for the active-site "template", sampling frames at each full nanosecond from the trajectories OMPLA3, Fas-AChE, and the PagP(charged). Frames are superimposed their His residues. The clusters from the both of the OMPLA3 monomers and from the "His-unproductive" Fas-AChE trajectory are labelled. Also labelled are the some of the sparsely-placed PagP atoms. Asx are in yellow; Glx in orange; Ser (oxygen atoms) in pink (inhibitor covalently-bound for OMPLA) and red (free); His in cyan (carbon) and blue (nitrogen). Rendered using VMD (*38*) and Raster3D (*39*).

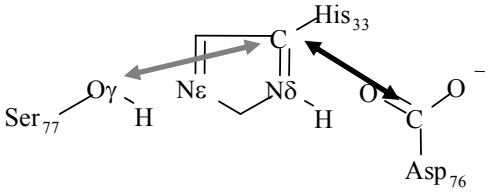
Tai et al., Fig. 1



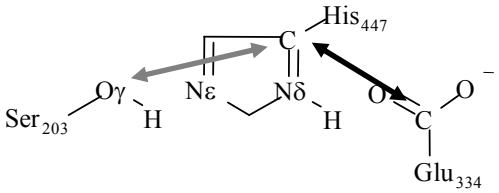
Tai et al., Fig. 2



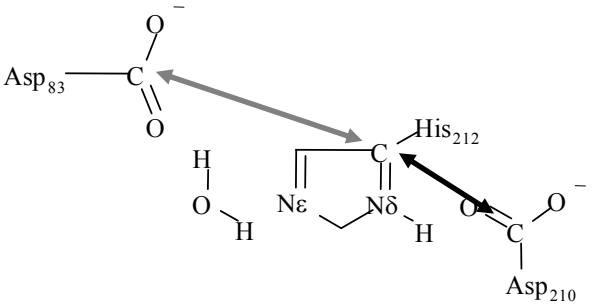
OMPLA



PagP

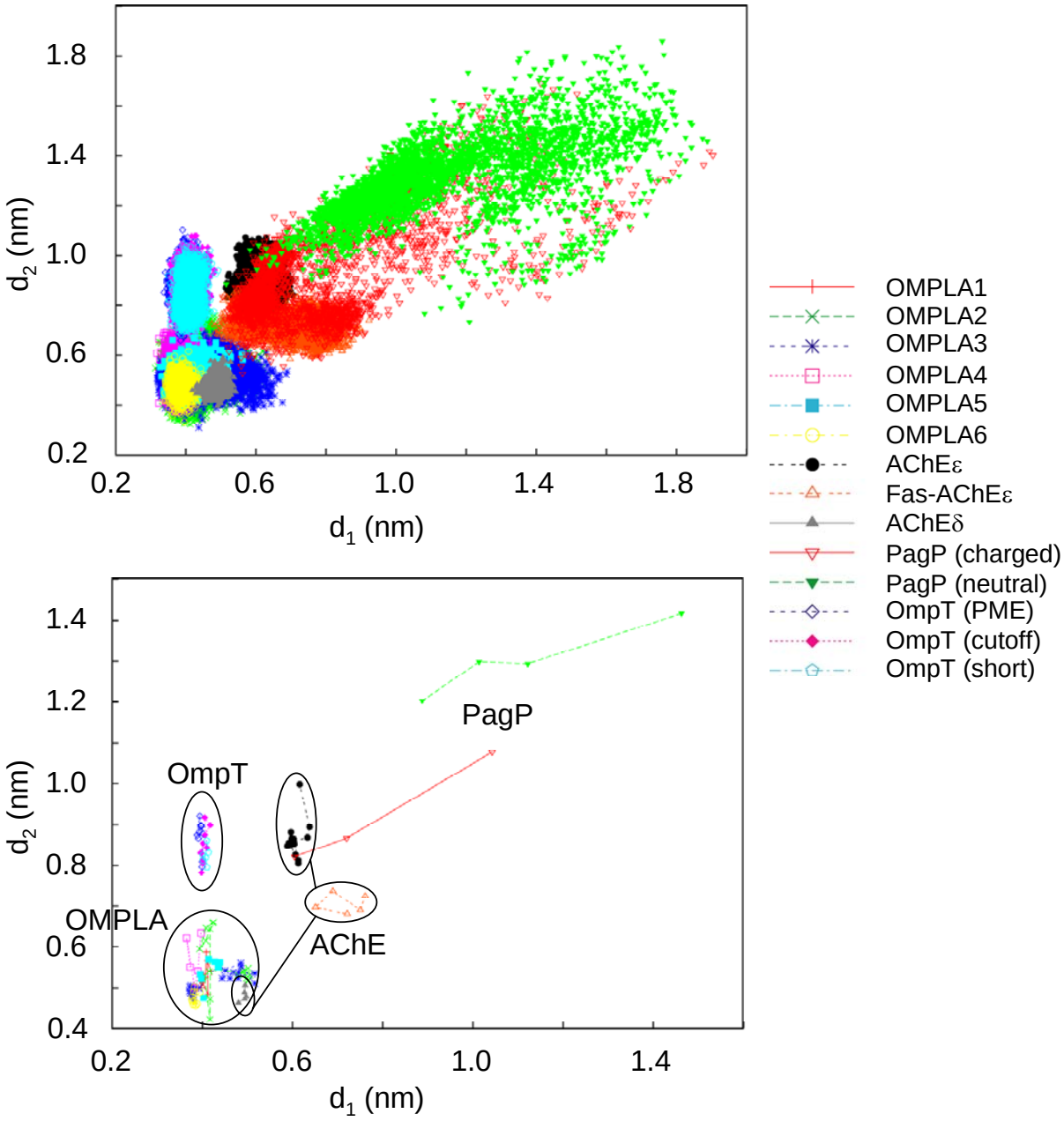


AChE



OmpT

Tai et al., Fig. 3



Tai et al., Fig. 4