JCTC Journal of Chemical Theory and Computation

Structural Fluctuations in Enzyme-Catalyzed Reactions: Determinants of Reactivity in Fatty Acid Amide Hydrolase from Multivariate Statistical Analysis of Quantum Mechanics/Molecular Mechanics Paths

Alessio Lodola,*,† Jitnapa Sirirak,‡ Natalie Fey,‡ Silvia Rivara,† Marco Mor,† and Adrian J. Mulholland*,‡

Dipartimento Farmaceutico, Università degli Studi di Parma, 43124 Parma, Italy, and Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom

Received May 19, 2010

Abstract: The effects of structural fluctuations, due to protein dynamics, on enzyme activity are at the heart of current debates on enzyme catalysis. There is evidence that fatty acid amide hydrolase (FAAH) is an enzyme for which reaction proceeds via a high-energy, reactive conformation, distinct from the predominant enzyme−substrate complex (Lodola et al. *Biophys. J.* **2007**, *92*, L20−22). Identifying the structural causes of differences in reactivity between conformations in such complex systems is not trivial. Here, we show that multivariate analysis of key structural parameters can identify structural determinants of barrier height by analysis of multiple reaction paths. We apply a well-tested quantum mechanics/molecular mechanics (QM/MM) method to the first step of the acylation reaction between FAAH and oleamide substrate for 36 different starting structures. Geometrical parameters (consisting of the key bond distances that change during the reaction) were collected and used for principal component analysis (PCA), partial least-squares (PLS) regression analysis, and multiple linear regression (MLR) analysis. PCA indicates that different "families" of enzyme−substrate conformations arise from QM/MM molecular dynamics simulation and that rarely sampled, catalytically significant conformational states can be identified. PLS and MLR analyses allowed the construction of linear regression models, correlating the calculated activation barriers with simple geometrical descriptors. These analyses reveal the presence of two fully independent geometrical effects, explaining 78% of the variation in the activation barrier, which are directly correlated with transition-state stabilization (playing a major role in catalysis) and substrate binding. These results highlight the power of statistical approaches of this type in identifying crucial structural features that contribute to enzyme reactivity.

Introduction

Enzymes exhibit a wide range of internal motions, some of which are essential for their activity.[1,2] The relationship among these conformational changes, structural fluctuations, and enzyme catalysis is a current central issue in enzymology.[3−5] Debates on the possible role of dynamics in enzyme catalysis are very heated. Different authors have suggested different definitions of the term "dynamics". A rigorous definition is that dynamical effects are any corrections to transition-state theory predictions, due to, e.g., recrossing of the barrier: the indications are that such effects on the rates of enzyme-catalyzed reactions are relatively small.[6] There is compelling evidence that transition-state theory provides a good basis

* Corresponding author e-mail: alessio.lodola@unipr.it (A.L.); adrian.mulholland@bristol.ac.uk (A.J.M.).

† Università degli Studi di Parma.

‡ University of Bristol.

10.1021/ct100264j
Published on Web 08/03/2010

for understanding chemical reactions in enzymes[7−9] and other catalysts.[10] Other authors identify the effects of dynamics with structural fluctuations that do not "drive" reaction, but may be crucial in permitting efficient reaction. Such fluctuations are subsumed into the free energy of activation in terms of transition-state theory. Conformational changes are an essential part of many enzyme catalytic cycles,[11] allowing, for example, efficient product release or preventing loss or hydrolysis of intermediates. The ability of many enzymes to undergo specific structural changes rapidly appears to have been evolved as an essential part of their overall catalytic efficiency. Subtle changes in the enzyme active site geometry can be crucial for efficient reaction.[12] They may also have a functional role, for example in enhancing substrate specificity. In some cases, local fluctuations of the enzyme active site could couple to substrate fluctuations, significantly reducing the barrier heights.[13] In other examples, fluctuations between different conformations of the enzyme−substrate complex could simply reflect a relatively flat free energy surface along a specific reaction coordinate.[14] However, it is also possible that a distinct high-energy (lowly populated) conformation of the enzyme−substrate complex may dominate its reactivity. Fatty acid amide hydrolase (FAAH) appears to be such a case.[15]

Detailed insight into the functioning of enzymes can be obtained from molecular simulations.[16,17] Simulations of enzyme−substrate complexes (e.g., by applying molecular mechanics (MM) force fields) can provide insight into some potentially important determinants of reactivity, such as the proximity of reacting groups or the conformational behavior of substrates.[11,18] However, for a detailed understanding of a catalytic process, both the identification of transition-state (TS) structures and their energies relative to the reactants are required. A well-established approach is hybrid quantum mechanics/molecular mechanics (QM/MM) methods,[19,20] which combine the simplicity and speed of the MM treatment of the protein structure with the flexibility and power of a quantum chemical treatment, thus allowing computational modeling of bond breaking and making in enzymes.[21]

In the QM/MM approach, a small region of the active site (where the reaction happens) is treated by a QM electronic structure method. This interacts with the protein and solvent environment, which are described by an empirical MM force field.[22] Most QM/MM studies of enzyme-catalyzed reactions[23] make use of a protein−substrate complex, built starting from the crystallographic coordinates of a protein−ligand structure (e.g., from the Protein Data Bank (PDB)). A refinement procedure, including addition of hydrogen atoms and missing side chains, solvation (by adding water molecules), and thermal equilibration of the resulting complex by molecular dynamics (MD) simulations, is usually carried out before the catalytic process is modeled.[24] Once the structure has been prepared, a QM/MM potential, in combination with a conformational sampling algorithm, can be applied to identify a TS for the reaction and to estimate its activation free energy.[25] However, this approach can be computationally demanding, particularly when a biochemical process is complex (e.g., composed of several chemical steps) and/or multiple possible mechanisms need to be explored. In some of these cases, a suitable strategy to investigate enzyme-catalyzed reactions is the "adiabatic mapping" approach,[26] where the potential energy surface (PES) of a reaction is calculated by a simple energy minimization along an approximate reaction coordinate.[27] With this approach, only a single starting structure is employed to explore the PES, with the assumption that this structure is a reasonable representation of the conformational space of the reacting enzyme−substrate complex. Adiabatic mapping has been useful in elucidating several enzyme mechanisms,[28] but is not always appropriate (e.g., for reactions involving large changes of solvation or movements of charge).[29,30] In enzymes for which adiabatic mapping is useful (i.e., those reactions for which a potential energy surface is similar to the free energy surface), such calculations must be performed and analyzed carefully: for example, the selection of an erroneous starting structure can lead to a structurally or chemically incorrect mechanism being modeled.[31] Geometrical fluctuations of the active site can significantly affect the overall energetic barrier, suggesting that the use of different protein conformations as starting geometries is important.[32] As emphasized in recent literature,[23,31] in QM/MM potential energy surface calculations several "representative" transition-state structures should be considered along with their corresponding minima. The resulting multiple PESs allow detailed analysis of the structural features that affect the barrier height of an enzyme-catalyzed reaction. While this approach does not give complete configurational sampling, it will at least partly capture the influence of the conformational diversity of the environment on the reaction. An expedient way to generate a selection of configurations is to take snapshots from an MD trajectory and use them as starting structures in subsequent QM/MM optimizations.[23,32]

FAAH is an important enzyme in the central nervous system, responsible for the (deactivating) hydrolysis of the endocannabinoids and other bioactive lipid amides.[33] We have previously identified a likely mechanism for the first step of the acylation reaction of FAAH with oleamide (OA) using QM/MM methods employing a limited number (four) of starting structures.[34] This mechanism is also supported by QM/MM Monte Carlo simulations[35] and kinetic experiments.[36] Here, we examine in detail conformational effects in FAAH, building on our multiple QM/MM reaction path approach, by extracting 36 snapshots from an MD trajectory as starting points for modeling the FAAH−OA reaction at the PM3-CHARMM22 level.[37,38] Although semiempirical methods are known to overestimate many energy barriers,[39] including FAAH,[34] calculations at this level of theory have been shown to be useful in many cases both for comparing different reaction mechanisms[40] and for modeling multiple reaction pathways with several substrates.[41] Here, the focus is on a comparison of barriers for the same mechanism, and such relative barriers will be considerably more accurate.[15] Thus, the present approach should allow a detailed exploration of the effect of structural fluctuations on the calculated energy barrier. We focused our investigation on the key points of the PES (substrate complex and TS) and analyzed the variation of geometrical parameters (consisting of the

key distances and angles that change during the course of the reaction) for all starting geometries and TS structures. To extract meaningful information from QM/MM calculations, we have employed a range of statistical approaches: (a) principal component analysis (PCA),[42] (b) partial least-squares (PLS) regression analysis,[43,44] (c) multiple linear regression (MLR) analysis.[45] These are widely used statistical techniques[46] in quantitative structure−activity relationships (QSARs).[47−49] Although PCA and/or PLS has been applied to the field of QM calculations on small model systems,[50−52] to the best of our knowledge this is the first time where several multivariate statistical techniques have been employed to relate the changes in energy to distances and angles in QM/MM calculations of an enzyme. The analyses revealed the presence of two fully independent geometrical effects directly correlated with substrate binding and TS stabilization controlling the rate of the reaction. These results highlight the power of statistical approaches in identifying crucial features contributing to enzyme catalysis.

Methods: Multivariate Data Analysis

PCA, PLS, and MLR statistical models were generated to ascertain how the structural parameters of the key species in the reaction change from snapshot to snapshot and how these changes affect the energy of the system. PCA is a statistical method for reducing the amount of data to be analyzed by exploiting the correlated nature of the variables within a data set.[53] Linear combinations of the correlated variables are derived such that the majority of the variance of the original data can be described by a few orthogonal components.[42] PLS also relies on linear combinations of the original variables but differs from PCA in that it employs a least-squares regression analysis step to relate the extracted components to a response, in this case activation energies. Furthermore, in PLS the linear combinations of variables are chosen to maximize the correlation between the extracted components and the response.[43] The new components (called latent variables) capture the essential information of the original variables, but this approach can be applied without the typical drawbacks associated with using correlated descriptors in multiple linear regression.[44] MLR is the classical linear regression analysis based on using the original **X**-matrix and **Y**-vector to build an equation model which satisfies the least-squares principle.[54]

All the generated statistical models used 15 geometric descriptors obtained from the QM/MM-optimized FAAH−oleamide complexes. These descriptors consisted of the key distances that change significantly during the course of the reaction. All descriptors were mean centered and scaled to unit variance because the numerical values of the descriptors vary significantly. This gives each variable the same opportunity of influencing the PCA/PLS model. The goodness of the correlation both in the PLS and MLR models can be reported using the coefficient of determination (R^2, where values close to 1 indicate a good model) and the standard error. The greater the number of variables compared to observations in a model, the more likely it is that a strong chance correlation may be found and that the model is overfitted. To determine the significance of the models

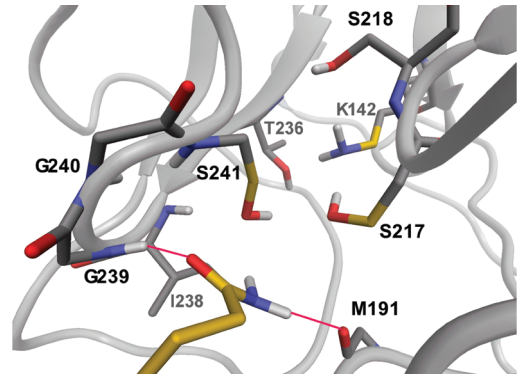

Figure 1. Representative QM/MM-optimized structure of the FAAH−oleamide complex (namely, snapshot 6 taken from the QM/MM MD simulation). Carbon atoms included in the quantum mechanical region are shown in yellow, and all the other carbons are shown in gray. Hydrogen bonds between OA and FAAH are also indicated.

generated, we applied the leave-one-out (LOO) cross-validation procedure,[55] where one observation is left out of the model-building process and its activation barrier is subsequently predicted by the reduced observation model. This is repeated until every observation has been left out and predicted at least once. A cross-validated regression coefficient, Q^2, can then be calculated and compared to R^2. The closer the R^2 value to the Q^2 value, the greater the confidence in the model one can have. A PCA/PLS component was considered significant if the Q^2 value was larger than 0.

PCA and PLS studies were performed with the SIMCA-P+ software.[56] MLR calculations were performed with Microsoft Excel97, employing the built-in statistical functions and automated macro procedures to determine the empirical value for regression coefficients and statistical parameters.

Results and Discussion

General Description of the Reaction. Multiple structures of the FAAH−OA enzyme−substrate complex were obtained from a QM/MM MD simulation (Supporting Information) based on the crystallographic coordinates of rat FAAH.[57] The complexes extracted from the MD trajectory showed an arrangement of the reactants similar to that previously published by our group[58] and other groups[35,59] (Figure 1). The long lipophilic chain of the OA substrate was accommodated in the acyl chain binding site, while its polar head pointed toward the cytosolic access channel of the enzyme. The amide group shows polar interactions within the FAAH active site: while the carbonyl oxygen was anchored to the oxyanion hole, forming a hydrogen bond with Gly239, the oleamide NH group formed an additional hydrogen bond with the backbone CO group of Met191. With few exceptions (vide infra), the members of the catalytic triad were found to be well oriented to initiate the catalytic process, with the bridging Ser217 accepting a hydrogen bond from Ser241 and donating another one to Lys142. Lys142 formed two further hydrogen bonds, one with Ser218 and the other with Thr236, suggesting that, in agreement with mutagenesis,[60] Ser218 is important for catalysis.

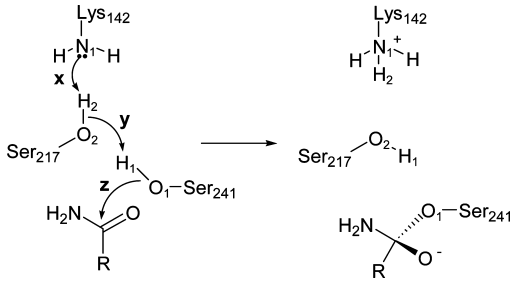

Figure 2. First step of oleamide hydrolysis catalyzed by FAAH: activation of Ser241 (**X**, **Y**) and formation of the tetrahedral adduct (**Z**). Labels are consistent with the definition of the reaction coordinates.

Starting from each structure of the enzyme−substrate complex, the first step of the acylation reaction was simulated by applying an adiabatic mapping strategy based on a hybrid PM3-CHARMM22 potential (Supporting Information). This approach has been used successfully in previous studies of the FAAH-catalyzed reaction involving amide[34] and ester[58] substrates and a carbamate inhibitor.[61]

Mutagenesis and kinetic experiments[62] in conjunction with simulations[15,34,35,58] suggested that the probable mechanism of the first step of the acylation reaction involves the activation of Ser241 through the abstraction of a proton by the neutral Lys142, via the bridging residue Ser217 (**X**, **Y**), and the nucleophilic addition of Ser241 to the carbonyl carbon of oleamide (**Z**), generating the tetrahedral complex (Figure 2).[34]

PESs were calculated by restraining two reaction coordinates, named R_x and R_y, which account for all the key chemical processes. The first coordinate $R_x = d[O_2, H_1] - d[O_1, H_1] - d[O_1, C]$ includes the proton abstraction from Ser241 performed by Ser217 (**Y**) and the nucleophilic attack (**Z**) led by Ser241 on the carbonyl carbon C. The second reaction coordinate $R_y = d[O_2, H_2] - [N_1, H_2]$ describes the proton transfer between Ser217 and the neutral Lys142 (**X**). A detailed description of the calculations is reported in the Supporting Information.

The mechanism is the same for all the FAAH−OA snapshots. The shape of the PESs and the relative positions of the stationary points are not affected by the geometry of the starting points. As examples, four PM3-CHARMM22 PESs are shown in Figure 3.

The first step of the acylation reaction can be described as follows: neutral Lys142 initiates reaction by accepting a proton from Ser217, leading to the formation of **2**. In turn,

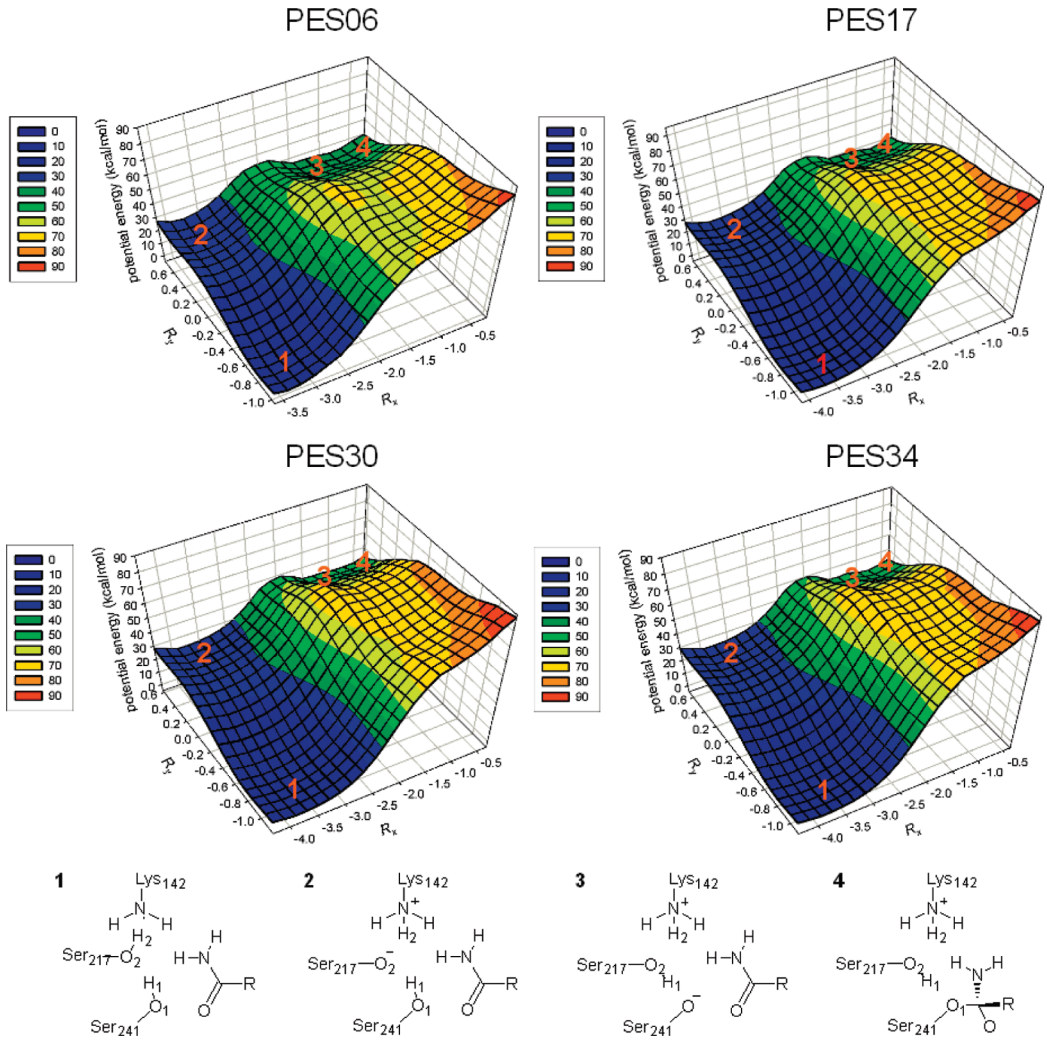

Figure 3. Representative PESs from four distinct snapshots (namely, 6, 17, 30, and 34) taken from QM/MM MD simulation. **1**−**4** are the minimum structures identified along the reaction pathway.

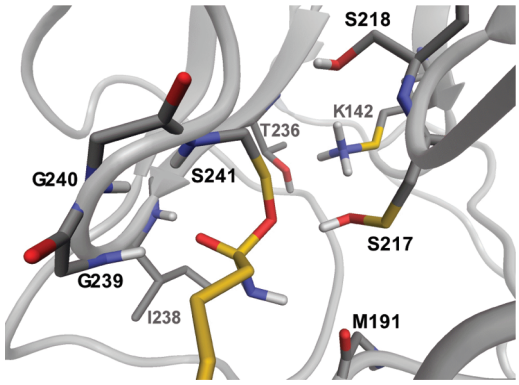

Figure 4. Representative TI of the FAAH−oleamide system (generated from snapshot 6). Bonds involving carbon atoms treated by quantum mechanics are shown in yellow, and all the other carbon atoms are shown in gray.

Ser217 deprotonates Ser241 to form the nucleophilic Ser241 anion (**3**). This short-lived intermediate **3** prompts the subsequent nucleophilic attack, leading to the tetrahedral intermediate (TI) **4**. The PESs indicate that the highest barrier is for deprotonation of Ser241. The absence of a significant (between 0.5 and 1 kcal mol^{-1}) barrier between **3** and **4** indicates that TI formation and Ser241 deprotonation are effectively concerted.

Visual inspection of the structures along points **3** and **4** showed that, as the nucleophilic attack proceeds, the hybridization of the reacting carbonyl carbon changes from sp^2 to sp^3 and the distances between the carbonyl oxygen and the NH groups of the residues forming the oxyanion hole become smaller, improving the stabilization of the TI. The resulting protonated Lys142 is caged in a tight network of hydrogen bonds involving the proton shuttle Ser217 and two peripheral residues of the active site, namely, Thr236 and Ser218 (Figure 4).

PCA and PLS Analysis. The activation barriers, extracted from the PESs, are given in Table 1 for the 36 snapshots. In the same table, geometric parameters (15 variables total) for the FAAH−OA structures employed for the statistical analysis are also reported. These descriptors were selected on the basis of interactions observed in the enzyme−substrate complex and in the TI (Figures 1 and 4). They include distances between the carbonyl oxygen of OA and the N−H backbone groups belonging to the oxyanion hole residues (I238, G239, G240, S241), the N−H of OA and the amide binding site (M191), the Lys142 polar hydrogens and their environment (S218, T236), and the reactant atoms such as the carbonyl double bond distance CO, the nucleophile attacking distance O1C, and others (N1O1, H1O1, H2O2,

Table 1. Energy Barriers, Absolute Energies, and Geometrical Descriptors Taken from FAAH−Oleamide Complexes Employed in the Multivariate Analysis

snapshot	E_{att}[a]	energy[a]	I238[b]	G239[b]	G240[b]	S241[b]	M191[b]	S218[b]	T236[b]	CO[b]	O1C[b]	N1O1[b]	H1O1[b]	H2O2[b]	H1O2[b]	H2N1[b]	N1O2O1[c]
1	45.97	−1068.67	3.158	1.913	3.429	2.819	1.983	2.173	2.001	1.228	3.171	4.504	0.950	0.951	2.198	2.050	98.605
2	53.66	−1440.92	2.047	1.811	3.326	2.549	1.966	2.244	2.088	1.234	2.877	3.566	0.953	0.951	3.920	2.539	57.316
3	46.47	−1218.24	2.592	1.815	3.280	2.589	1.932	2.196	2.015	1.230	2.432	4.707	0.943	0.952	1.987	1.968	109.553
4	45.43	−1253.98	2.793	1.831	3.310	2.714	2.041	2.178	1.987	1.230	2.461	4.778	0.943	0.951	1.988	1.958	111.174
5	47.06	−1231.37	3.034	1.868	3.322	2.702	2.012	2.219	2.019	1.228	2.977	4.504	0.946	0.951	2.113	2.006	99.056
6	43.08	−1282.47	2.808	1.823	3.453	2.819	1.959	2.209	1.996	1.228	2.554	4.591	0.944	0.951	2.017	1.994	104.389
7	45.17	−1119.30	2.905	1.849	3.254	2.659	1.983	2.166	1.982	1.229	2.908	4.664	0.944	0.950	2.053	2.098	104.952
8	49.61	−1198.66	3.015	1.847	3.270	2.701	1.983	2.163	2.017	1.229	3.111	4.589	0.946	0.949	2.116	2.025	103.042
9	49.51	−1135.02	3.049	1.854	3.071	2.690	1.945	2.176	2.042	1.229	2.946	4.619	0.947	0.950	2.124	2.038	103.297
10	50.37	−1250.90	3.026	1.868	3.269	2.674	2.064	2.137	2.022	1.228	2.956	4.687	0.946	0.951	2.103	2.215	102.695
11	46.48	−1301.06	3.178	1.928	3.433	2.730	1.991	2.160	2.042	1.227	3.058	4.487	0.949	0.951	2.151	2.090	96.908
12	48.68	−1266.21	3.198	1.927	3.310	2.722	1.998	2.148	2.028	1.227	3.158	4.558	0.949	0.952	2.147	2.125	98.346
13	44.35	−1160.14	2.159	1.787	3.449	2.677	1.916	2.120	2.001	1.232	2.430	4.748	0.948	0.951	2.100	2.103	107.314
14	48.28	−1277.85	3.029	1.920	3.323	2.730	1.952	2.124	2.010	1.227	2.952	4.608	0.949	0.952	2.158	2.192	98.033
15	44.35	−1141.59	2.469	1.805	3.363	2.719	1.923	2.110	1.994	1.230	2.443	4.805	0.943	0.951	1.982	2.011	111.551
16	44.55	−1254.56	2.614	1.823	3.318	2.664	1.968	2.117	2.008	1.230	2.449	4.904	0.944	0.951	2.007	2.016	113.395
17	46.06	−1243.52	3.010	1.850	3.284	2.721	2.033	2.123	1.975	1.228	2.958	4.514	0.946	0.951	2.108	2.072	98.241
18	48.28	−1115.84	3.043	1.873	3.354	2.854	2.148	2.109	1.972	1.229	3.037	4.664	0.949	0.951	2.191	2.273	98.862
19	46.83	−1093.53	2.926	1.850	3.279	2.678	2.001	2.152	2.022	1.229	2.969	4.371	0.946	0.951	2.112	2.074	94.078
20	44.03	−1263.99	2.542	1.805	3.234	2.606	1.995	2.166	2.044	1.231	2.423	4.709	0.942	0.951	1.981	1.977	108.280
21	48.27	−1102.41	2.958	1.852	3.280	2.733	2.099	2.176	2.071	1.229	2.987	4.582	0.947	0.950	2.118	2.029	100.490
22	47.34	−1059.90	3.069	1.870	3.280	2.774	2.108	2.219	2.018	1.229	3.021	4.422	0.947	0.951	2.114	2.015	96.461
23	49.23	−1066.58	3.214	1.911	3.325	2.809	1.899	2.131	2.013	1.228	3.078	4.797	0.947	0.951	2.630	2.271	98.566
24	48.71	−1109.99	3.150	1.900	3.392	2.816	2.108	2.115	2.000	1.227	3.121	4.587	0.948	0.951	2.188	2.104	98.657
25	52.28	−1180.44	3.202	1.911	3.224	2.821	2.074	2.188	2.048	1.229	3.351	4.382	0.948	0.952	2.173	2.151	92.389
26	50.76	−1083.32	3.120	1.889	3.301	2.784	2.060	2.147	2.011	1.228	3.184	4.477	0.947	0.952	2.131	2.134	95.874
27	46.41	−1275.03	2.712	1.819	3.239	2.648	1.978	2.132	2.025	1.230	2.669	4.732	0.945	0.951	2.035	2.070	106.764
28	49.95	−1228.80	3.061	1.873	3.302	2.732	2.025	2.154	2.029	1.229	3.046	4.363	0.947	0.951	2.074	2.092	94.544
29	50.05	−1078.36	3.157	1.894	3.309	2.753	2.030	2.185	2.033	1.228	3.120	4.355	0.948	0.951	2.098	2.068	94.382
30	48.69	−1059.09	3.099	1.892	3.371	2.781	1.979	2.166	2.005	1.228	3.097	4.411	0.948	0.952	2.148	2.126	93.797
31	48.25	−1053.67	3.177	1.925	3.416	2.801	1.937	2.153	2.012	1.228	3.165	4.420	0.948	0.952	2.163	2.112	93.955
32	50.10	−1006.53	3.247	1.941	3.420	2.839	1.960	2.188	2.060	1.227	3.201	4.351	0.949	0.952	2.179	2.084	92.384
33	49.36	−1118.60	3.133	1.898	3.329	2.750	2.038	2.179	2.043	1.228	3.138	4.313	0.948	0.951	2.148	2.053	92.505
34	49.70	−1117.11	3.135	1.885	3.292	2.747	2.047	2.194	2.070	1.228	3.136	4.327	0.949	0.951	2.162	2.070	92.449
35	49.22	−1148.48	3.041	1.868	3.267	2.713	2.088	2.135	2.024	1.229	3.059	4.499	0.948	0.952	2.117	2.108	96.923
36	52.19	−1260.25	3.124	1.889	3.217	2.714	2.417	2.181	2.039	1.231	3.144	4.189	0.960	0.972	1.785	1.776	100.373
av	48.02	−1174.07	2.950	1.868	3.321	2.729	2.018	2.162	2.021	1.229	2.924	4.522	0.947	0.952	2.162	2.086	98.878
SD	2.50	95.21	0.289	0.041	0.060	0.071	0.091	0.033	0.027	0.001	0.268	0.233	0.003	0.004	0.325	0.119	9.276

[a] Energies in kilocalories per mole. [b] Distances in angstroms. [c] Angles in degrees.

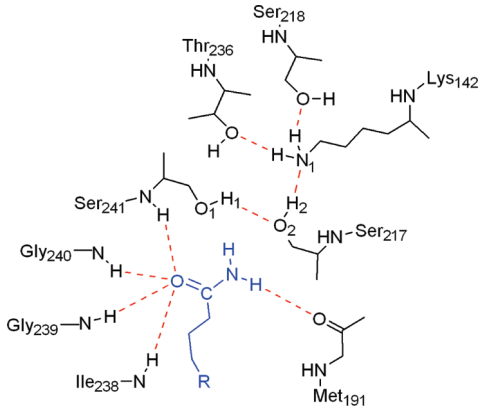

Figure 5. Two-dimensional representation of oleamide (in blue) within the FAAH active site (black line). Polar interactions (hydrogen bonds) are highlighted with red dotted lines. Atom labels are consistent with the reaction coordinate as well as with variable names included in the statistical analysis.

H1O2, H2N1; see Figure 5). The angle formed by Lys142 nitrogen, Ser217 oxygen, and Ser241 oxygen (N1O2O1) was also considered. In the same table the mean values of the collected variables are reported, together with their standard deviations (SDs).

The same geometrical descriptors were also collected for the TS structures identified on the PESs (Table S1, Supporting Information). Comparison of the standard deviations indicates that the differences between the enzyme−substrate complex geometries are in general larger than those between the corresponding transition-state structures, with a few exceptions (H1O1 and H2O2 descriptors). This also suggests that the accessible conformational space of the FAAH−OA system is wider at the enzyme−substrate complex than at the TS, at least along the specified variables. For these reasons, multivariate analyses were limited to the first family of structures, i.e., the enzyme−substrate complexes.

PCA was applied to the geometrical descriptors reported in Table 1. PCA produces a summary of the data structure showing how the observations (here the snapshots) are related to each other and highlights any deviating (outlier) observations. The advantage of PCA is that the information contained in the original **X**-matrix (here constituting 15 descriptors) can be concentrated in 2−3 principal components, representing truly independent effects.[42] Here, the first two principal components (called $t[1]$ and $t[2]$) describe approximately 58% of the total variance in the 15 descriptors. The model has a negative cross-validated Q^2, indicating that some inconsistency may be present in the data set.[46] Indeed, the score plot of the first two components reported in Figure 6 indicates the presence of one outlier, namely, snapshot 2, which lies outside Hotelling's T^2 confidence ellipse.[63]

The analysis of the raw data suggests that snapshot 2 differs from the other snapshots in that Ser241 (H1) does not form a hydrogen bond with Ser217 (O2) (H1O2 = 3.920 Å), an interaction which is crucial for proton transfer between them. Visual inspection of the FAAH−OA complex showed that, in snapshot 2, Ser217 adopts a different conformation at the binding site, with χ_1 in an *anti* conformation rather than in the *gauche* (+) conformation as observed for all other

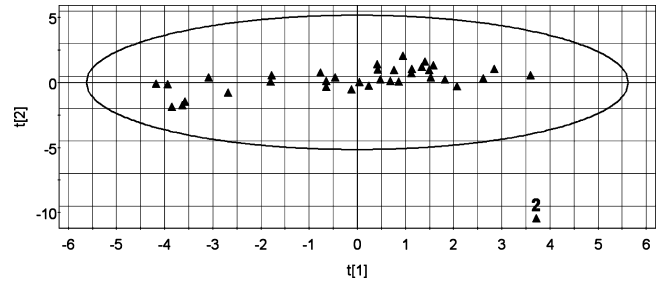

Figure 6. PCA score plot of component 1, $t[1]$, against component 2, $t[2]$. The curve displayed is Hotelling's ellipsoid, describing the area in which an observation can be expected to fall with a probability of 95%.

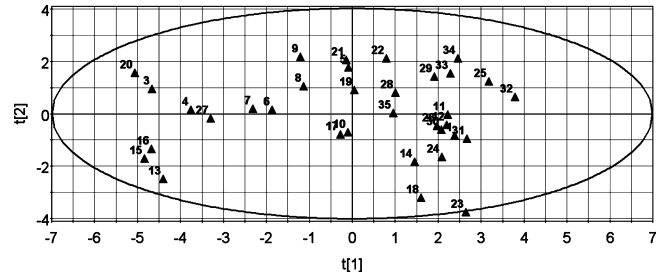

Figure 7. PCA score plot of component 1, $t[1]$, against component 2, $t[2]$, for the second data set (snapshots 2 and 36 excluded).

starting geometries. This difference is also mirrored by a N1O2O1 value of 57.31°, far from the mean value of the series for this angle (mean ± SD = 98.88 ± 9.28°). Snapshot 2 was thus excluded from the data set as further multivariate analysis would essentially only explain the differences between this structure and all the others. To detect more moderate outliers, which are not powerful enough to shift the model plane and therefore to show up as outliers in score plots, we looked at the observation distance to the generated model in **X**-space, calculating DmodX, also known as the residual error.[46] Snapshot 36 can be identified as a moderate outlier and was also excluded from further analysis. In this case (snapshot 36), analysis of the raw data suggests a different accommodation for the $-NH_2$ group of OA, as the interaction with the carbonyl oxygen of Met191 (M191 = 2.417 Å) is not optimal compared to that observed in all other snapshots (mean value ± SD = 2.018 ± 0.061 Å). A final PCA was performed with a data set composed of the remaining 34 observations and 15 variables ($N = 34$; $X = 15$). The observations, reported in a new score plot (Figure 7), are more evenly distributed when compared to the previous data set, indicating a more homogeneous **X**-matrix.

The performance of the PCA is reported in Table 2. The first component explains 48% of the total variance, while the second one extracts an additional 16% of information. The introduction of other components is not justified as both the third and fourth PCs have a negative Q^2. Thus, the best model has a cumulative (cum) R^2 of 0.64 with a positive cross-validated Q^2(cum) of 0.40, indicating that descriptive and predictive powers are satisfactory to perform further statistical analysis.

The loadings plot (Figure 8) allows interpreting the score plot further, as it displays the relationship among the

Table 2. Summary of the PCA Analysis[a]

comp no.	R^2(X)	R^2(X)(cum)	Q^2	Q^2(cum)
1	0.480	0.480	0.387	0.387
2	0.158	0.638	0.025	0.402
3	0.093	0.730	−0.075	0.357
4	0.071	0.802	−0.140	0.292

[a] R^2(X) refers to the variance in the descriptor matrix described by the PCA models. Q^2 refers to the cross-validated R^2(X) for the PCA model. The comp no. is the number of the respective principal component.

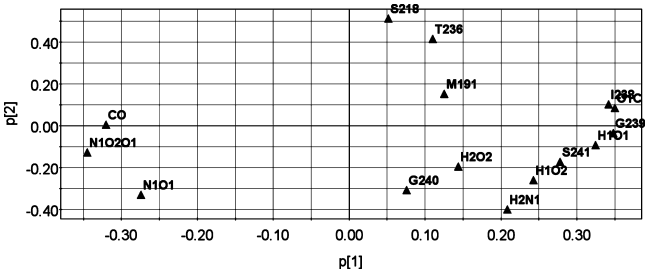

Figure 8. PCA loading plot of the first two PCs of the data set where snapshots 2 and 36 were excluded.

variables.[42] The loadings are the weights of the original variables when calculating the principal component scores. Score and loading plots are complementary; thus, a pattern observed in the score plot can be interpreted by looking along that direction in the loading plot. On the loading plot, descriptors found close to each other in both dimensions are highly correlated (e.g., I238 and O1C). Those that are similar on just one component are still correlated, but to a lesser extent (e.g., S218 and G240), whereas those at opposite ends of the components are inversely correlated (O1C and CO). Descriptors found at the extreme ends of the *x* or *y* axes have the most significant impact on the component that defines that axis, whereas unimportant variables are around the origin of the plot. Here, many of the descriptors have large positive or negative loadings (∼0.3) on the *x* axis, indicating that, as expected, the descriptors in the data set are highly correlated.

The first principal component encodes information mainly related to the accommodation of the amide group of OA into the FAAH active site. The most important variables are (i) with positive loadings, O1C (the nucleophile attacking distance) and I238 and G239 distances, describing two (of four) hydrogen bonds formed by the carbonyl oxygen with the oxyanion hole, and (ii) with negative loadings, the N1O2O1 angle and the CO distance, measuring the length of the carbonyl double bond of oleamide. These five variables were also highly correlated, with absolute values of their correlation coefficients often higher than 0.8 and never lower than 0.65 (Table S2, Supporting Information). Also S241, the third anchor point in the oxyanion hole, N1O1, the distance between the Lys142 nitrogen and the nucleophile oxygen, and H1O1 showed large loadings on the first component, but they showed slightly lower correlation coefficients with the variables discussed above and with each other (Table S2). Thus, these eight variables describe a concerted movement within the active site, resulting in deeper accommodation of OA into the oxyanion hole, approaching the nucleophile and polarization of the carbonyl bond. In fact, shorter hydrogen bonds in the oxyanion hole (I238, G239, S241) correspond to shorter O1C distances and to a longer and possibly more polarized carbonyl bond. As a consequence, the distance O1C can be considered representative of all these concerted movements.

Table 3. Summary of the PLS Analysis[a]

comp no.	R^2(X)	R^2(X)(cum)	R^2(Y)	R^2(Y)(cum)	Q^2	Q^2(cum)
1	0.473	0.473	0.628	0.628	0.582	0.582
2	0.114	0.587	0.148	0.776	0.219	0.674
3	0.124	0.711	0.019	0.795	−0.128	0.641
4	0.061	0.772	0.014	0.809	−0.374	0.605

[a] R^2(X) refers to the variance in the descriptor matrix described by the PLS models. R^2(Y) refers to the variance in the energy explained by the PLS model. Q^2 refers to the cross-validated R^2(Y) for the PLS model. The comp no. refers to the component number.

Looking at the score plot, it is possible to note the presence of two observation groups with respect to the O1C descriptor: all the snapshots with high O1C distances (e.g., longer than 2.95 Å) have positive *t*[1] and are located on the right-hand side of the plot, whereas those with smaller O1C values are on the left. The second component explains 16% of the total variation in the descriptors and is dominated by the positive effect of S218 and T236, accounting for the hydrogen bond network stabilizing the Lys142 side chain nitrogen. It is interesting to observe that these two variables present low intercorrelation ($r = 0.43$), and they are barely correlated with other variables having significant loadings on the second component, such as H2N1 and G240 (see Table S2, Supporting Information). These variables seem therefore to bring a common portion of information, but their variation is much less concerted than that described by the first component.

PLS is a regression technique for modeling the relationship between multiple *X* and *Y* variables, useful in the case of a complex **X**-descriptor matrix, where traditional MLR cannot be applied due to the presence of highly correlated variables.[43] PLS tries to extract only statistically relevant information contained in the **X**-matrix (in that way it is similar to PCA) and to use such information to build regression models.[43] This means that, in PLS, a small group of new independent variables (called "latent variables") will be generated to explain the **Y**-space. In the current investigation, the activation energy (E_{att}) was used as the response variable. It changes significantly among the 36 observations, with an absolute range of variation of ∼11 kcal mol^{-1}, corresponding to 22% of the E_{att} mean value. An exploratory regression analysis reveals that E_{att} does not depend on the absolute energy of the system (data not shown). This indicates that the QM/MM protocol applied is robust with respect to small changes in the composition of the system.

Table 3 reports the PLS models for the data set consisting of 34 observations, 1 response variable, and 15 descriptors ($N = 34$; $Y = 1$; $X = 15$) where the overall R^2 and Q^2 statistics change as a function of the model complexity. The first latent variable describes 47% (R^2(X)) of the information contained in the **X**-matrix and correlates with **Y**, giving an R^2(Y) of 0.63 and Q^2 of 0.58. The introduction of a second latent variable slightly improves the model, as both the

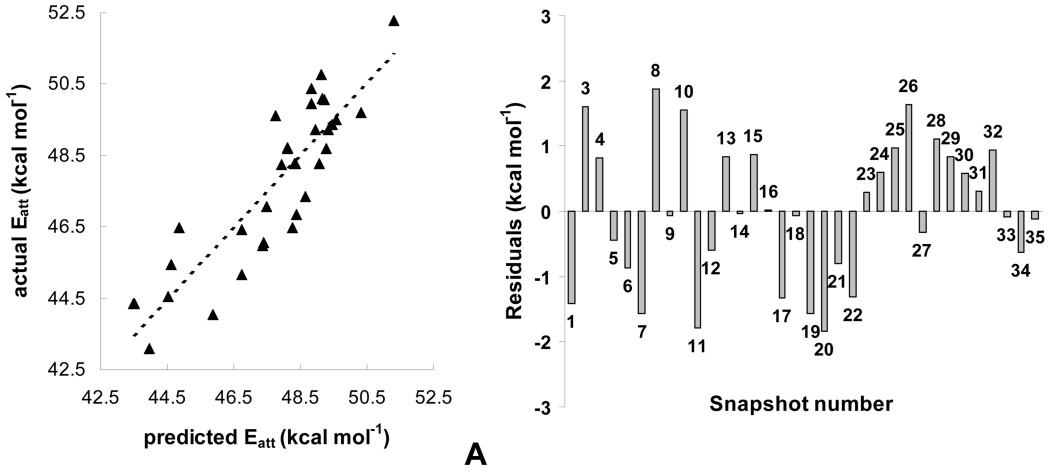

Figure 9. Observed E_{att} values vs those calculated by the PLS model (A, left) and residuals, calculated as the difference between the actual E_{att} and calculated E_{att} (B, right).

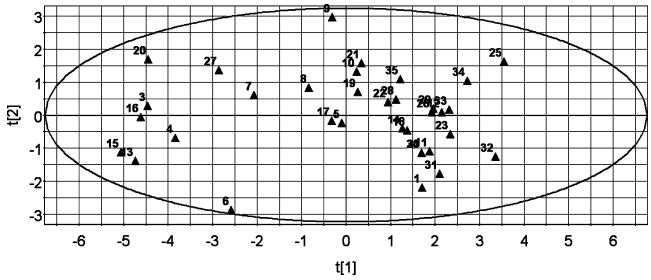

Figure 10. PLS score plot of the two first latent variables of the data set.

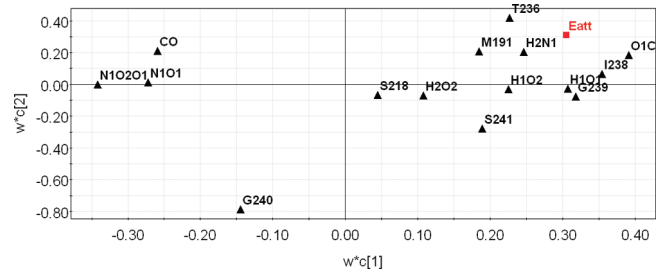

Figure 11. *W*C* weight plot for latent variables 1 and 2. The *Y* variable (E_{att}) is marked in red. This plot shows both the *X* weights (*W**) and *Y* weights (*C*), and thus the correlation between the original **X**-matrix of geometrical descriptors and *Y* (E_{att}).

cumulative $R^2(\mathbf{Y})$ and Q^2 increase to values of 0.78 and 0.67, respectively. Here, two components appear appropriate as Q^2 starts to decrease when additional latent variables are introduced.

The performance of this two-variable model is fairly good, with 78% of the variation in E_{att} explained by the PLS equation, giving an acceptable root-mean-square error (RMSE) of 1.098 kcal mol^{-1}. Analysis of the residuals (the difference between the actual E_{att} and calculated E_{att}, Figure 9) indicates that some snapshots (e.g., 3, 8, 10, 11, 20, and 26) are poorly predicted. However, these structures are homogeneous observations in the PCA space (Figure 7); therefore, we chose to retain them in the present analysis.

The score plot derived from the 15-descriptor PLS models (Figure 10) does not differ significantly from that derived from the PCA model built using the same data set. The presence of two groups of observations can be detected in this case as also discussed for PCA, particularly along the first latent variable (t[1]). The similarity and differences between the PLS and PCA components can be understood by an analysis of the PLS weight plot (*W*C* plot, Figure 11), corresponding to the loading plot in PCA. However, the weight plot in PLS differs somewhat from the loading plot in PCA in that the response variable (E_{att}) is also displayed. It is thus possible to determine the effect of a particular descriptor on E_{att} on the basis of their spatial relationship in the PLS weight plot.[43]

The most important descriptors for the first latent variable, positively correlated with E_{att}, are the nucleophile attacking distance O1C and I238 distance, describing the polar interaction between the Ile238 NH backbone and the carbonyl oxygen of OA. On the other hand, the N1O2O1 variable, describing the angle formed among the Lys142 basic nitrogen, Ser217 hydroxyl oxygen, and Ser241 hydroxyl oxygen is negatively correlated with the activation barrier. (e.g., the higher the angle value, the lower the barrier).

With respect to PCA, some differences also arise from the second latent variable, which is now dominated by G240, describing the polar interaction distance between the OA carbonyl oxygen and the NH backbone of the oxyanion hole residue Gly240, and to a lesser extent by T236, describing the hydrogen bond distance between the Thr236 hydroxyl hydrogen and Lys142 side chain nitrogen. Finally, compared to PCA, S218 loses its influence on the model, being located close to the origin of the *W*C* plot.

A complementary way to recognize the impact of the original variables on the PLS model is to calculate the VIP (variable importance in the projection), summarizing the importance of the *X* variables in both the descriptor and response spaces. VIPs are squared functions of the PLS weights, *W**, taking into account the amount of explained *Y* variance in each dimension. Descriptors with a VIP larger than 1 (e.g., O1C, G240) are the most influential for modeling the observed variation in E_{att} (Figure 12), consistent with what was observed in the *W*C* weight plot for latent variables 1 and 2 (Figure 11).

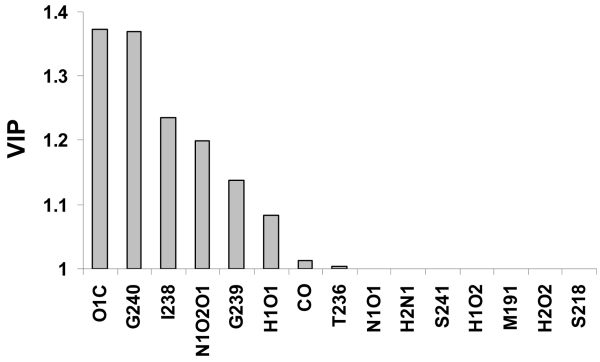

Figure 12. VIP values calculated for the descriptors employed in the PLS analysis. To highlight the most significant variables, the VIP scale starts from 1.

In our previous investigation on conformational fluctuations in FAAH catalysis, we identified N1O1 as the crucial determinant for reactivity.[15] With the current data set, this descriptor plays a negligible role in affecting the barrier of the acylation (shown by its small VIP value (<1)). This apparent discrepancy appears to be due to the different method employed to generate the enzyme−substrate complexes. Indeed, inclusion of these structures (renamed snapshots 37−40) in the principal component analysis ($N = 38$; $X = 15$) shows that these enzyme−substrate complexes are outliers, placed at the border of Hotelling's T^2 confidence ellipsoid (Figure S3, Supporting Information). PCA reveals that the main sources of variation between snapshots 37−40 and all the others (Figure S4, Supporting Information) reside both in accommodation of the amide group of OA in the FAAH active site (as revealed by the loading values of G239 and O1C along the first component) and in bond distances involving hydrogen atoms crucial for the reaction (e.g., H1O2, H2O2, H2N1 for the first component and H1O1 for the second one).

MLR Analysis and Chemical Interpretation of the Models. The interpretation of the latent variables generated in a PLS regression may be difficult, due to the necessity to combine the original descriptors to form derived variables.[64]

An MLR analysis was therefore also performed, employing the data set composed of the 15 original descriptors and 34 homogeneous observations ($N = 34$; $Y = 1$; $X = 15$), exploring linear relationships between the response data and smaller subsets of descriptors. However, it should be considered that MLR presents a higher risk of oversimplification, giving emphasis on a specific variable, or overfitting, by inclusion of noisy variables.

Among the 15 possible univariate models, O1C gives the best correlation with E_{att}:

$$E_{att} = (28.369 \pm 2.550) + (6.601 \pm 0.865)\text{O1C} \quad (1)$$

$$n = 34 \quad R^2 = 0.646 \quad s = 1.354$$
$$F = 58.1 \ (p = 1.1 \times 10^{-8}) \quad Q^2 = 0.602$$

The positive relationship between E_{att} and O1C indicates that the activation barrier is lowered if the nucleophilic oxygen of Ser241 and the carbonyl carbon of OA are in close proximity. However, current and previous[34] calculations

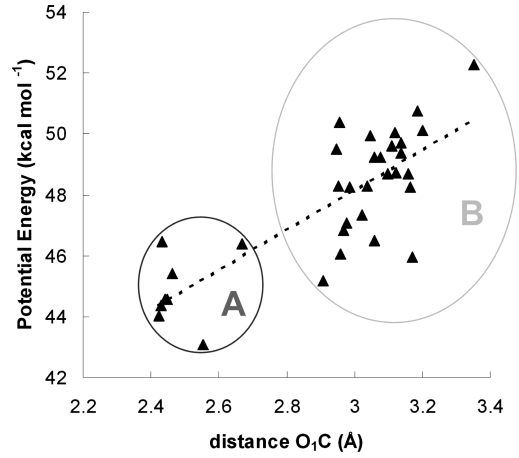

Figure 13. Calculated activation energy E_{att} (kcal mol^{-1}) vs the nucleophile attack distance O1C (Å).

indicated that the TS for the first step of the acylation corresponded to the deprotonation of Ser241 rather than to the nucleophilic attack. It is likely that the closeness of O_1 and C, induced by a coming together of the substrate with the S241 nucleophile oxygen, facilitates the key proton transfer of the reaction. It is possible that the more the substrate is pushed toward S241, the more it resembles the TS structure, leading to a lower activation barrier (vide infra).

The scatter plot of E_{att} vs O1C shows the presence of two different clusters in the data set (Figure 13). All the starting structures where the O1C distance is shorter than 2.8 Å belong to cluster A, as is the case for snapshots 3, 4, 6, 13, 15, 16, 20, and 27. These generally have a lower energy barrier than members of cluster B (where the O1C distance exceeds 2.8 Å), which contains the remaining 26 observations.

Although this simple linear regression mainly explains the difference in reactivity between these two clusters rather than highlighting differences among structures within any single group, it gives important insights for the reaction. Visual inspection of the 34 TS structures for all the calculated PESs (with a proton being transferred from Ser241 to Ser217) shows that, on average, the O1C distance measures 2.035 ± 0.025 Å (mean ± SD). Conformational fluctuations of the OA carbonyl in the FAAH active site influence the catalytic process: in the substrate complex, a O1C distance close to that in the TS structure is associated with a lower activation barrier.

As already discussed, O1C is significantly correlated with other geometrical descriptors describing a concerted process. In fact, other related variables gave significant linear correlations with E_{att}, even if with lower descriptive and predictive power.

PLS analysis indicated the presence of another effect accounting for the differences in the calculated barriers. Among all the possible 105 bivariate linear regression equations, only two models have satisfactory descriptive and predictive power, bringing independent information (determinant of the correlation matrix >0.9), namely, eqs 2 and 3. In both cases, the nucleophilic attack distance O1C remains an important variable for modeling the variance of the activation barrier: its regression coefficient does not change significantly from one equation to the others.

$$E_{att} = (55.838 \pm 9.379) + (6.564 \pm 0.773)\text{O1C} - (8.250 \pm 2.733)\text{G240} \quad (2)$$

$$n = 34 \quad R^2 = 0.726 \quad s = 1.209 \quad F = 41\ (p = 2.0 \times 10^{-9}) \quad Q^2 = 0.674$$

$$E_{att} = (-13.571 \pm 18.538) + (5.99 \pm 0.857)\text{O1C} + (21.663 \pm 9.495)\text{T236} \quad (3)$$

$$n = 34 \quad R^2 = 0.696 \quad s = 1.274 \quad F = 35\ (p = 1.1 \times 10^{-8}) \quad Q^2 = 0.637$$

Equation 2 shows a negative relationship between E_{att} and G240. The Gly240 NH group is one of the anchor points (in the oxyanion hole) for the carbonyl oxygen of OA within the FAAH active site; thus, the G240 descriptor can be considered as an indirect measure of the tightness of binding to the substrate. The analysis indicates that the shortening of this distance in the substrate complex significantly increases the barrier heights for the reaction.

In eq 3, the positive correlation between E_{att} and T236 indicates that shortening of the hydrogen bond distance between Lys142 (acting as the hydrogen bond donor) and the Thr236 side chain oxygen (acting as the hydrogen bond acceptor) in the substrate (Michaelis) complex lowers the barrier height for the reaction. It could be speculated that a tighter interaction between these two residues would result in an indirect increase of the nitrogen basicity of Lys142. Such a change would accelerate the reaction, consistent with the role of Lys142, acting as a general base in the early phase of the catalysis.[34−36]

As expected, the two variables included with O1C in eqs 2 and 3 are those having the largest absolute values in the second latent variable of the PLS model, with consistent signs. On the other hand, the low correlation between the variables G240 and T236 ($r = -0.27$) indicates that they contain additional effects that may introduce noise in the MLR models, as well as the common information leading to their projection into the latent variable. In fact, the coefficient/standard error ratio for T236 in eq 2 is only barely significant in a t test ($p = 0.030$), while the coefficient of G240 in eq 2 is more significant ($p = 0.005$).

It is important to recognize that structural fluctuations associated with reaction could take place in any environment and so are not necessarily of any catalytic benefit. Conformational effects of the type identified here (e.g., fluctuations of the O1C distance) could be observed for a comparable ("reference") reaction in solution. In such a case, the fluctuation itself would not be necessarily a source of catalysis,[6] but part of the "complete" reaction coordinate, able to better describe the whole catalytic process. In enzymes where reaction may take place via distinct, high-energy conformations, definition of the entire reaction coordinate may be important: it is possible that the ability to efficiently undergo fluctuations leading to reactive conformations may be associated with substrate specificity. In any case, a full description of a reaction within an enzyme requires characterization of motions associated with the reaction.

Conclusions

We report here the first example of multivariate analysis of enzyme reactivity, applied to an extended data set of FAAH−oleamide complexes extracted from a QM/MM MD simulation. A total of 36 independent PESs for the first step of the acylation reaction were calculated and analyzed to investigate the relationship between the geometry of the starting structure and the barrier height of the reaction. A total of 15 geometrical descriptors were collected from the active site of the 36 starting structures and employed for multivariate PCA and PLS and MLR analysis to assess the impact of conformational fluctuations of the active site on the calculated activation barrier. The variable selection strategy was to some extent based on our knowledge of the FAAH−substrate system and its catalytic mechanism. However, some general insights resulting from the present analysis may be transferable to other catalytic systems, e.g., the importance of considering distances describing polar interactions between the enzyme and the substrate or between members of the catalytic residues.

Results from PCA indicated that different "families" of enzyme−substrate conformations arise from QM/MM MD simulation and that a rarely sampled conformational state (e.g., snapshot 2) can be identified along a trajectory. Sampling is a crucial point of enzyme reaction simulation, as an inappropriate choice of starting structure could lead to an incorrect mechanism being modeled. A proper equilibration of the starting structure is therefore required, and multivariate PCA could help in selecting starting points which are different from each other in the space of catalytically relevant geometrical parameters.

Once outliers are identified and excluded from the data set, the relationship between the activation barrier and geometrical parameters can be investigated. PLS analysis revealed the presence of two distinct and fully independent geometrical effects, explaining ∼78% of the response variable (i.e., the barrier). MLR confirms the results obtained with PLS, indicating that conformational fluctuations associated with the nucleophile attacking distance (O1C), substrate binding (G240), or stabilization of the general base Lys142 (T236) play a significant role in determining the activation barrier. Conformational fluctuations of the active site affecting these crucial distances, also of a few fractions of an angstrom, may considerably influence the calculated reaction barrier, contributing up to 20% of the variation from the mean value.

From a technical point of view, our findings highlight the importance of using a large number of starting structures in calculations of potential energy surfaces for enzyme reactions and confirm that structural fluctuation is an essential part of the process of the reaction within the enzyme. Furthermore, the use of multiple starting structures can capture protein motions (usually neglected in energy-minimization studies of enzymatic processes), along the specified reaction coordinate, identifying catalytically relevant enzyme−substrate

complexes and giving a rough estimate of the energy required to preorganize the active site for catalysis (for instance, looking at the standard deviation of the calculated barriers).

The results warn against simplistic analyses of structural effects, because multiple subtle inter-related structural factors affect the barrier height and shape for reactions in enzymes. The use of multivariate statistical approaches is crucial for explaining quantitatively the differences in the calculated barrier for different starting structures. Detailed statistical analyses of the type applied here should be useful in analyses of other enzymes to mine potentially hidden conformational effects. This should be considered when the effects of protein dynamics on enzyme reactions are analyzed.

Acknowledgment. A.J.M. thanks the Engineering and Physical Sciences Research Council (EPSRC) for support: A.J.M. is an EPSRC Leadership Fellow. The Thai Government is also acknowledged for funding (J.S.).

Supporting Information Available: Computational details for QM/MM simulations, correlation matrix, and supplementary figures for PCA. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Karplus, M.; Gao, Y. Q.; Ma, J.; van der Vaart, A.; Yang, W. Protein Structural Transitions and Their Functional Role. *Philos. Trans. R. Soc. London, A* **2005**, *363*, 331–355.

(2) Min, W.; English, B. P.; Luo, G.; Cherayil, B. J.; Kou, S. C.; Xie, X. S. Fluctuating Enzymes: Lessons from Single-Molecule Studies. *Acc. Chem. Res.* **2005**, *38*, 923–931.

(3) Qian, H.; Shi, P. Z. When Does the Michaelis-Menten Equation Hold for Fluctuating Enzymes? *J. Phys. Chem. B* **2009**, *113*, 2225–2230.

(4) Gorfe, A. A.; Lu, B.; Yu, Z.; McCammon, J. A. Enzymatic Activity versus Structural Dynamics: The Case of Acetylcholinesterase Tetramer. *Biophys. J.* **2009**, *97*, 897–905.

(5) Liu, Y. H.; Konermann, L. Conformational Dynamics of Free and Catalytically Active Thermolysin Are Indistinguishable by Hydrogen/Deuterium Exchange Mass Spectrometry. *Biochemistry* **2008**, *47*, 6342–6351.

(6) Olsson, M. H.; Parson, W. W.; Warshel, A. Dynamical Contributions to Enzyme Catalysis: Critical Tests of a Popular Hypothesis. *Chem. Rev.* **2006**, *106*, 1737–1756.

(7) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science* **2004**, *303*, 186–195.

(8) Olsson, M. H.; Mavri, J.; Warshel, A. Transition State Theory Can Be Used in Studies of Enzyme Catalysis: Lessons from Simulations of Tunnelling and Dynamical Effects in Lipoxygenase and Other Systems. *Philos. Trans. R. Soc. London, B* **2006**, *361*, 1417–1432.

(9) Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H. J. High-Accuracy Computation of Reaction Barriers in Enzymes. *Angew. Chem., Int. Ed.* **2006**, *45*, 6856–6859.

(10) Honkala, K.; Hellman, A.; Remediakis, I. N.; Logadottir, A.; Carlsson, A.; Dahl, S.; Christensen, C. H.; Nørskov, J. K. Ammonia Synthesis from First-Principles Calculations. *Science* **2005**, *307*, 555–558.

(11) Pentikäinen, U.; Pentikäinen, O. T.; Mulholland, A. J. Cooperative Symmetric to Asymmetric Conformational Transition of the Apo-Form of Scavenger Decapping Enzyme Revealed by Simulations. *Proteins* **2008**, *70*, 498–508.

(12) Thorpe, I. F.; Brooks, C. L. Conformational Substates Modulate Hydride Transfer in Dihydrofolate Reductase. *J. Am. Chem. Soc.* **2005**, *127*, 12997–13006.

(13) Karplus, M.; McCammon, J. A. Dynamics of Proteins: Elements and Function. *Annu. Rev. Biochem.* **1983**, *53*, 263–300.

(14) Villà, J.; Warshel, A. Energetics and Dynamics of Enzymatic Reactions. *J. Phys. Chem. B* **2001**, *105*, 7887–7907.

(15) Lodola, A.; Mor, M.; Zurek, J.; Tarzia, G.; Piomelli, D.; Harvey, J. N.; Mulholland, A. J. Conformational Effects in Enzyme Catalysis: Reaction via a High Energy Conformation in Fatty Acid Amide Hydrolase. *Biophys. J.* **2007**, *92*, L20–L22.

(16) van der Kamp, M. W.; Mulholland, A. J. Computational Enzymology: Insight into Biological Catalysts from Modeling. *Nat. Prod. Rep.* **2008**, *25*, 1001–1014.

(17) Cavalli, A.; Carloni, P.; Recanatini, M. Target-Related Applications of First Principles Quantum Chemical Methods in Drug Design. *Chem. Rev.* **2006**, *106*, 3497–3519.

(18) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652.

(19) Senn, H. D.; Thiel, W. QM/MM Studies of Enzymes. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182–187.

(20) Warshel, A. Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.

(21) Friesner, R. A.; Guallar, V. *Ab Initio* Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Methods for Studying Enzymatic Catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389–427.

(22) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.

(23) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.

(24) Mulholland, A. J. Computational Enzymology: Modelling the Mechanisms of Biological Catalysts. *Biochem. Soc. Trans.* **2008**, *36*, 22–26.

(25) Hu, H.; Yang, W. Free Energies of Chemical Reactions in Solution and in Enzymes with *ab Initio* Quantum Mechanics/Molecular Mechanics Methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573–601.

(26) Mulholland, A. J. The QM/MM Approach to Enzymatic Reactions. In *Theoretical Biochemistry*; Eriksson, L. A., Ed.; Elsevier: Amsterdam, 2001; pp 597−653.

(27) Cui, Q.; Karplus, M. Quantum Mechanical/Molecular Mechanical Studies of the Triosephosphate Isomerase-Catalyzed Reaction: Verification of Methodology and Analysis of Reaction Mechanisms. *J. Phys. Chem. B* **2002**, *106*, 1678–1698.

(28) Lodola, A.; Woods, C. J.; Mulholland, A. J. Applications and Advances of QM/MM Methods in Computational Enzymology. In *Annual Reports in Computational Chemistry*; Wheeler, R. A., Spellmeyer, D. C., Eds.; Elsevier: Amsterdam, 2008; Vol. 4, Chapter 9, pp 155−169.

(29) Kamerlin, S. C.; Haranczyk, M.; Warshel, A. Progress in *ab Initio* QM/MM Free-Energy Simulations of Electrostatic Energies in Proteins: Accelerated QM/MM Studies of pKa, Redox Reactions and Solvation Free Energies. *J. Phys. Chem. B* **2009**, *113*, 1253–272.

(30) Bowman, A. L.; Ridder, L.; Rietjens, I. M.; Vervoort, J.; Mulholland, A. J. Molecular Determinants of Xenobiotic Metabolism: QM/MM Simulation of the Conversion of 1-Chloro-2,4-dinitrobenzene Catalyzed by M1-1 Glutathione S-Transferase. *Biochemistry* **2007**, *46*, 6353–6363.

(31) Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. On Possible Pitfalls in ab Initio Quantum Mechanics/Molecular Mechanics Minimization Approaches for Studies of Enzymatic Reactions. *J. Phys. Chem. B* **2005**, *109*, 15645–15650.

(32) Zhang, Y.; Kua, J.; McCammon, J. A. Influence of Structural Fluctuation on Enzyme Reaction Energy Barriers in Combined Quantum Mechanical/Molecular Mechanical Studies. *J. Phys. Chem. B* **2003**, *107*, 4459–4463.

(33) Piomelli, D. The Molecular Logic of Endocannabinoid Signalling. *Nat. Rev. Neurosci.* **2003**, *4*, 873–884.

(34) Lodola, A.; Mor, M.; Hermann, J. C.; Tarzia, G.; Piomelli, D.; Mulholland, A. J. QM/MM Modelling of Oleamide Hydrolysis in Fatty Acid Amide Hydrolase (FAAH) Reveals a New Mechanism of Nucleophile Activation. *Chem. Commun.* **2005**, *439*, 9–4401.

(35) Tubert-Brohman, I.; Acevedo, O.; Jorgensen, W. L. Elucidation of Hydrolysis Mechanisms for Fatty Acid Amide Hydrolase and Its Lys142Ala Variant via QM/MM Simulations. *J. Am. Chem. Soc.* **2006**, *128*, 16904–16913.

(36) McKinney, M. K.; Cravatt, B. F. Evidence for Distinct Roles in Catalysis for Residues of the Serine-Serine-Lysine Catalytic Triad of Fatty Acid Amide Hydrolase. *J. Biol. Chem.* **2003**, *278*, 37393–37399.

(37) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods. 2. Applications. *J. Comput. Chem.* **1989**, *10*, 221–264.

(38) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. A* **1998**, *102*, 3586–3616.

(39) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J. Computational Enzymology. *Chem. Commun.* **2010**, 2354–2372.

(40) Hermann, J. C.; Ridder, L.; Mulholland, A. J.; Höltje, H. D. Identification of Glu166 as the General Base in the Acylation Reaction of Class A β-Lactamases through QM/MM Modeling. *J. Am. Chem. Soc.* **2003**, *125*, 9590–9591.

(41) Ridder, L.; Mulholland, A. J.; Vervoort, J.; Rietjens, I. M. C. M. Correlation of Calculated Activation Energies with Experimental Rate Constants for an Enzyme Catalyzed Aromatic Hydroxylation. *J. Am. Chem. Soc.* **1998**, *120*, 7641–7642.

(42) Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52.

(43) Wold, S.; Sjostrom, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.

(44) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735–743.

(45) Draper, N. R.; Smith, H. *Applied Regression Analysis*, 2nd ed.; Wiley: New York, 1980.

(46) Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Trygg, J.; Wikstrom, C.; Wold, S. *Multi- and MegaVariate Data Analysis—Basic Principles and Applications*, 2nd ed.; Umetrics AB: Umeå, Sweden, 2006.

(47) Mor, M.; Rivara, S.; Lodola, A.; Plazzi, P. V.; Tarzia, G.; Duranti, A.; Tontini, A.; Piersanti, G.; Kathuria, S.; Piomelli, D. Cyclohexylcarbamic Acid 3′- or 4′-Substituted Biphenyl-3-yl Esters as Fatty Acid Amide Hydrolase Inhibitors: Synthesis, Quantitative Structure−Activity Relationships, and Molecular Modeling Studies. *J. Med. Chem.* **2004**, *47*, 4998–5008.

(48) Mor, M.; Lodola, A.; Rivara, S.; Vacondio, F.; Duranti, A.; Tontini, A.; Sanchini, S.; Piersanti, G.; Clapper, J. R.; King, A. R.; Tarzia, G.; Piomelli, D. Synthesis and Quantitative Structure−Activity Relationship of Fatty Acid Amide Hydrolase Inhibitors: Modulation at the N-Portion of Biphenyl-3-yl Alkylcarbamates. *J. Med. Chem.* **2008**, *51*, 3487–3498.

(49) Valitutti, G.; Duranti, A.; Lodola, A.; Mor, M.; Piersanti, G.; Piomelli, D.; Rivara, S.; Tontini, A.; Tarzia, G.; Traldi, P. Correlation between Energetics of Collisionally Activated Decompositions, Interaction Energy and Biological Potency of Carbamate FAAH Inhibitors. *J. Mass Spectrom.* **2007**, *42*, 1624–1627.

(50) Tantanak, D.; Limtrakul, J.; Gleeson, M. P. Probing the Structural and Electronic Factors Affecting the Adsorption and Reactivity of Alkenes in Acidic Zeolites Using DFT Calculations and Multivariate Statistical Methods. *J. Chem. Inf. Model.* **2005**, *45*, 1303–1312.

(51) Gleeson, D. Application of QM Simulations and Multivariate Analysis in the Study of Alkene Reactivity in the Zeolite H-ZSM5. *J. Chemom.* **2008**, *22*, 372–377.

(52) Fey, N.; Guy Orpen, A.; Harvey, J. N. Building Ligand Knowledge Bases for Organometallic Chemistry: Computational Description of Phosphorus(III)-Donor Ligands and the Metal−Phosphorus Bond. *Coord. Chem. Rev.* **2009**, *253*, 704–722.

(53) Chatfield, C.; Collins, A. J. *Introduction to Multivariate Analysis*; Chapman and Hall: London, 1980.

(54) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters*; Wiley: New York, 1978.

(55) Cramer, R. D., III.; Bunce, J. D.; Patterson, D. E. Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.

(56) *SIMCA-P+*, version 11.0; Umetrics AB: Umeå, Sweden, 2005.

(57) Bracey, M. H.; Hanson, M. A.; Masuda, K. R.; Stevens, R. C.; Cravatt, B. F. Structural Adaptation in a Membrane Enzyme That Terminates Endocannabinoid Signaling. *Science* **2002**, *298*, 1793–1796.

(58) Lodola, A.; Mor, M.; Sirirak, J.; Mulholland, A. J. Insights into the Mechanism and Inhibition of Fatty Acid Amide Hydrolase from Quantum Mechanics/Molecular Mechanics (QM/MM) Modeling. *Biochem. Soc. Trans.* **2009**, *37*, 363–367.

(59) Acevedo, O.; Jorgensen, W. L. Advances in Quantum and Molecular Mechanical (QM/MM) Simulations for Organic and Enzymatic Reactions. *Acc. Chem. Res.* **2010**, *43*, 142–151.

(60) Patricelli, M. P.; Cravatt, B. F. Clarifying the Catalytic Roles of Conserved Residues in the Amidase Signature Family. *J. Biol. Chem.* **2000**, *275*, 19177–18184.

(61) Lodola, A.; Mor, M.; Rivara, S.; Christov, C.; Tarzia, G.; Piomelli, D.; Mulholland, A. J. Identification of Productive Inhibitor Binding Orientation in Fatty Acid Amide Hydrolase (FAAH) by QM/MM Mechanistic Modelling. *Chem. Commun.* **2008**, 214–216.

(62) McKinney, M. K.; Cravatt, B. F. Structure and Function of Fatty Acid Amide Hydrolase. *Annu. Rev. Biochem.* **2005**, *74*, 411–432.

(63) Hotelling's ellipsoids are equidensity contours resulting from multivariate normal distributions. They describe the area in which an observation can be expected to fall with a certain probability (e.g., 95%).

(64) Mansson, R. A.; Welsh, A. H.; Fey, N.; Orpen, A. G. Statistical Modeling of a Ligand Knowledge Base. *J. Chem. Inf. Model.* **2006**, *46*, 2591–2600.

CT100264J