Binding of flexible ligands to proteins: Valproic acid and its interaction with cytochrome P450cam

ARTICLE *in* INTERNATIONAL JOURNAL OF QUANTUM CHEMISTRY · MARCH 1993

Impact Factor: 1.43 · DOI: 10.1002/qua.560480718

6 AUTHORS, INCLUDING:

Paul R. Ortiz de Montellano
University of California, San…

479 PUBLICATIONS **17,579**
CITATIONS

SEE PROFILE

Binding of Flexible Ligands to Proteins: Valproic Acid and Its Interaction with Cytochrome P450cam

YAN-TYNG CHANG AND GILDA H. LOEW
Molecular Research Institute, 845 Page Mill Road, Palo Alto, California 94304

ALLAN E. RETTIE, THOMAS A. BAILLIE, AND PAMELA R. SHEFFELS
Department of Medicinal Chemistry, School of Pharmacy, University of Washington, Seattle, Washington 98195

PAUL R. ORTIZ DE MONTELLANO
Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446

Abstract

A combined theoretical and experimental study of the binding and interaction of valproic acid (VPA) with the bacterial cytochrome P450cam enzyme and the determination of regio- and stereoselective hydroxylation product distribution was performed. From the experiments, C4—OH VPA was found to be the predominant hydroxylation product with a small amount of C5—OH VPA formed. The experimental stereoselectivity of hydroxylation was 2R4S > ~ 2S4R > 2R4R > ~ 2S4S and 2S5 > ~ 2R5. The overall goals of the theoretical study were twofold: (1) to characterize as completely as possible, using energy optimization and molecular dynamics simulations, the interactions of flexible ligands with their target proteins, and (2) to determine the extent to which these results could be used to develop criteria to predict or explain the experimentally observed regio- and stereoselectivity of hydroxylation of the flexible ligands. Among the useful insights into the behavior of flexible ligands upon binding to their target proteins obtained are (1) a large change in conformation occurs for many conformers of VPA upon binding to P450cam, (2) low-energy conformers of VPA do not necessarily lead to optimum interactions with the target protein, and (3) the most favorable mode of interaction of this flexible ligand with the protein binding site has been identified and found to be a result of strong electrostatic interactions between VPA and both Tyr96 and Asp297. For the study of the hydroxylated VPA products, the challenging aspect of this problem was to determine criteria for weighing the contribution of each of the possible protein–ligand complexes. To this end, various possibilities were examined and compared with the experimental results. No single complex was found to reproduce the observed experimental regio- and stereoselectivity. This result indicates that more than one bioactive form of VPA contributes to its oxidation. Results most consistent with experiment are obtained by using the interaction energy of the protein–ligand complex as a criterion for including its contribution to product formation. Although there are remaining disparities between predicted and observed product distributions, the combined theoretical and experimental effort has led to insights into the modes of interaction of this flexible ligand that lead to its observed product specificity.

Introduction

The theoretical study of the interrelationship between flexible ligands and their target proteins is important due to its increasing application in computer-aided

International Journal of Quantum Chemistry: Quantum Biology Symposium 20, 161–180 (1993)
 CCC 0360-8832/93/010161-20

drug design and risk assessment. In many cases, the lack of the three-dimensional structures of the proteins confines the investigation to the flexible ligands themselves. Usually, a thorough conformational search of a number of active and inactive ligand analogs is performed *in vacuo* or in an environment that mimics solvent effects. Comparison among the analogs is then carried out to deduce the possible bioactive conformer(s) [1]. However, the question of how the protein environment affects the conformation of the bioactive conformer(s) after the ligand diffuses into the protein and binds to the active site cannot be assessed without the knowledge of the receptor structure.

If the protein structure is known, the interaction between it and the ligand can be probed directly. For many years, computer graphics have been widely used to manually dock ligands into the binding site of proteins using the criterion of a "hand-and-glove" steric fit. This approach of determining initial complex geometries appears adequate when dealing with rigid ligands. However, two complications arise when docking a flexible analog: (1) What initial conformation(s) shall be used? and (2) What orientations lead to the highest protein–ligand affinity?

Recently, an automated docking method [2] for flexible ligands has been developed by Goodsell and Olson. In this method, the global minimum of the interaction energy of the protein–flexible ligand complex is obtained by simulated annealing. The best conformation and orientation of the ligand in the active site are thus found simultaneously at the end of the simulation. Despite the potential usefulness of this procedure for finding the global minimum, several weaknesses lessen its reliability. Specifically, the simple potential form and the static protein geometry used to make it computationally feasible limits the reliability of this method. In addition, in many cases, it is not clear that the most stable ligand–protein complex is the only one contributing to the observed properties of interest.

In the work presented here, we have systematically probed the relationship between the stable conformations of a highly flexible ligand, valproic acid (VPA), in its free state and its behavior as a ligand in the binding site of the oxidative metabolizing enzyme, cytochrome P450cam [3]. This system was chosen because the metabolism of VPA, a widely used antiepileptic agent, and other aliphatic acids, widely used as industrial solvents, by P450s lead to toxic products [4]. In addition, in a collaborative effort, the oxidation of VPA by cytochrome P450cam, an isozyme with a known 3-D structure [5], has been studied and the regio- and stereoselectivity of the hydroxylation product examined. The ultimate goal of these studies is to develop strategies for the prediction of the product distribution of flexible ligands. We have taken the first step in this direction by addressing the question of which conformers and orientations of VPA are primarily involved in interactions with P450cam that lead to the regio- and stereoselectivity and by comparisons of these predictions with the experimental results. As shown in Figure 1, there are 15 possible sites of hydroxylation of VPA assumed to proceed by H-abstraction by the ferryl oxygen followed by fast radical recombination [3,6–9]. Experimental studies reported here reveal regioselectivity of the C-hydroxy VPA products, i.e., C4—OH $\gg$ C5—OH, with no measurable C2—OH or C3—OH product formation. In addition, stereoselective C-hydroxy products were formed among the possible stereo-

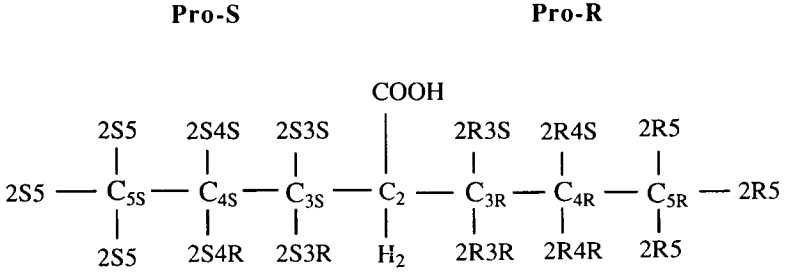

Figure 1. The definition of the stereoisomers of C-hydroxy VPA. Each label represents the stereoisomer formed if the specific hydrogen atom is abstracted by the active ferryl oxygen atom of the P450cam.

isomers shown in Figure 1. Previous theoretical studies that we performed indicate that electronic criteria for product formation are not sufficient to account for the observed regio- and stereoselectivity [10]. Thus steric interaction between VPA and P450cam must play an important role in modulating hydroxylated product distribution.

The specific questions addressed in the theoretical study presented here are (1) the extent to which conformations found in the free state are modified in the bound state, (2) whether the relative energy of a free-state conformer is a determinant of the stability of the complex formed by it, (3) to what extent the dynamically averaged interaction energy between the ligand and the protein is different from its energy-optimized one, (4) whether more than one protein–ligand complex contributes to the regio- and stereoselective product formation observed, and (5), if so, what are the best criteria to use in determining the relative contributions of the many protein–ligand complexes to the observed product distribution?

Theoretical Methods

A. Determination of Conformational Profile of the Free Ligand

In a previous theoretical study that we performed of the teratogenicity of valproic acid the low-energy conformations of VPA *in vacuo* were calculated. The CHARMm force field implemented in QUANTA [11] was used in this earlier investigation and a nested torsional angle grid scan was carried out for the six torsional angles defined in Figure 2. A total of 486 starting conformers were generated and then minimized using CHARMm. The minimized structures were then clustered based on torsional angle similarity. Any two conformers with all of their torsional angles differing by no more than 90° were considered similar. This procedure resulted in a clustering of 74 unique conformers that were used in the present study.

To test the effect of different semiempirical and empirical methods on the relative energies of the conformations, single-point energy calculations using the AM1 method [12] and an alternate force field in AMBER [13] were carried out for each of the 74 conformers. In a further step, the geometry of each conformer was reop-

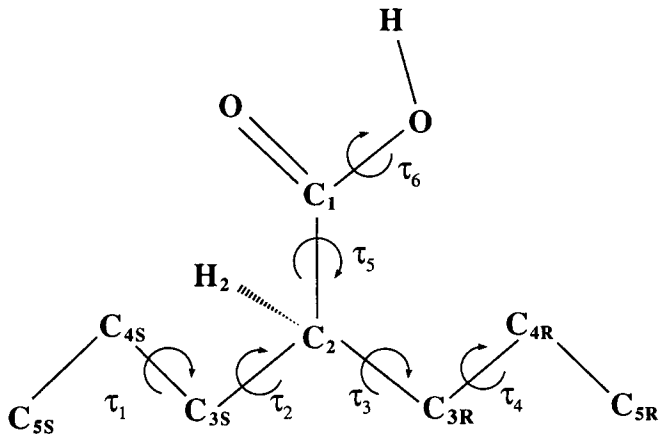

Figure 2. VPA and the six torsional angles used for conformational search and analysis. τ_1: C_{5S}—C_{4S}—C_{3S}—C_2; τ_2: C_{4S}—C_{3S}—C_2—C_{3R}; τ_3: C_{3S}—C_2—C_{3R}—C_{4R}; τ_4: C_2—C_{3R}—C_{4R}—C_{5R}; τ_5: O=C_1—C_2—H_2; τ_6: O=C_1—O—H.

timized *in vacuo* using these two methods. For consistency, the 72 unique conformations thus found with the AMBER force field were used since the molecular dynamics study of the P450cam–VPA complex were performed with the AMBER programs.

B. Docking of VPA to P450cam

An extended binding site model [14] of the cytochrome P450cam enzyme developed previously in our laboratory was used for studying the P450cam–VPA binding and interaction. Based on the 3-D structure of the adamantanone-bound P450cam enzyme, 87 amino acids in the vicinity of the active site together with seven structural waters were included in the model. In this study, the putative catalytically active state of P450cam with a ferryl oxygen atom at a distance of 1.7 Å from the iron of the heme unit was used to study the oxidation of VPA by P450cam.

Since the active site pocket is large enough to accommodate the longest extension of the all-*trans* conformation of VPA (see Fig. 2), many ligand orientations are possible. However, the existence of a carbonyl oxygen in VPA as part of the polar group, similar to the polar group of camphor, which is the native ligand of the P450cam, allows an electrostatic recognition of these groups in both ligands with the Tyr96–OH group of P450cam. Only two orientations for VPA were found to retain this recognition. As shown in Figure 3, using one of the conformers as an example, these two initial orientations chosen for docking VPA differ by a rigid rotation of 180° such that the second orientation is just the "upside-down" version of the first orientation.

Each of the 72 unique conformers was docked manually into the active site of P450cam in these two orientations, using the graphic program MIDAS [15]. The

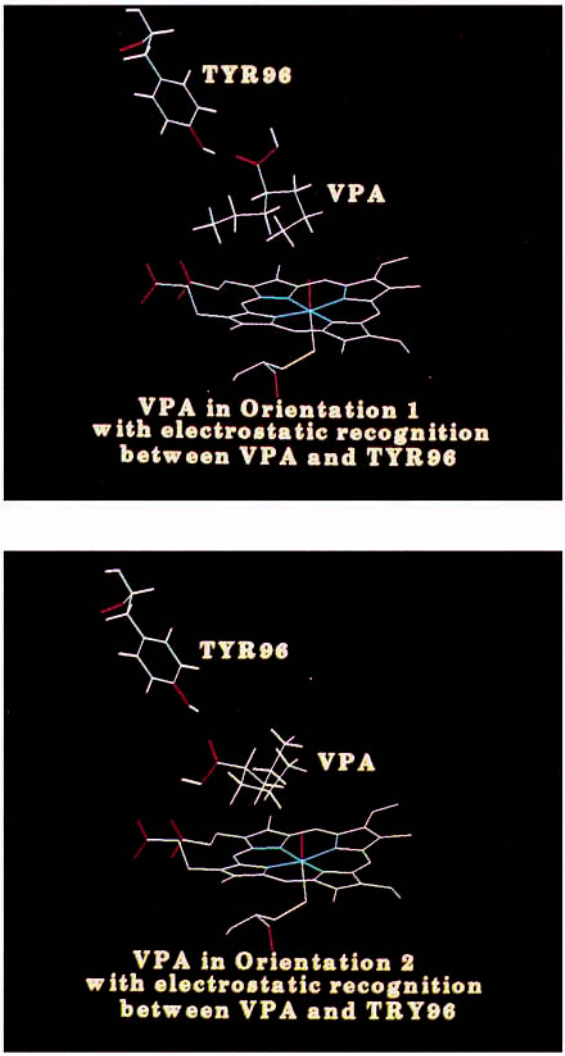

Figure 3. The two relatively different initial docking orientations of VPA in the P450cam binding site: (a) orientation 1; (b) orientation 2, obtained by a rigid rotation of the same conformer in orientation 1 by 180°. Both orientations allow an electrostatic recognition between the C=O of VPA with the OH group of Tyr96.

144 P450cam–VPA complexes thus generated were then optimized using the AMBER 3.0a program package. The six torsional angles of VPA bound in each optimized complex were examined and compared to the free conformer to evaluate the extent of any conformational change, i.e., an induced fit by the ligand in the binding site pocket of the protein. The interaction energy between VPA and P450cam in each complex was calculated with the ANAL program of the AMBER package.

C. *Molecular Dynamics Simulations*

In all of the molecular dynamics (MD) simulations, since the 87 amino acids forming the active site are not contiguous, a harmonic force constant of 100 kcal/Å^2 was applied to the backbone (N,C_α,C) atoms of the residues of P450cam to keep the active site intact. A distance-dependent dielectric $\varepsilon = r$ was used since no water solvent was included in the simulation. Each of the optimized complexes was heated gradually to reach room temperature (298 K) and equilibrated in a total period of 5 ps. A 10 ps MD simulation was then carried out. During the simulation, the Cartesian coordinates of all atoms were saved at 0.2 ps intervals, resulting in 50 snapshots for each trajectory. The total interaction energy between VPA and P450cam as well as the contributions from the electrostatic and the van der Waals terms to the total interaction energy of each snapshot from each trajectory were calculated. The energy decomposition analysis was included to gain a better understanding of what kind of interaction determines the mode of VPA binding to the P450cam enzyme. A modified version of the ANAL program that reads in the coordinates snapshot by snapshot was used for this purpose.

Two sets of MD simulations were carried out: In the first, the 12 optimized P450cam–VPA complexes from each orientation with the best interaction energy were chosen for short MD simulations. The results showed that the time-averaged (over 10 ps) interaction energies of the complexes in the initial orientation 1 were less favorable, in general, than those of the complexes in the initial orientation 2. Therefore, in a second study, all the 72 P450cam–VPA-optimized complexes generated using the initial orientation 2 were subjected to a 5 ps heating-equilibration and a 10 ps MD simulation. The dynamic behavior of the six torsional angles of VPA was monitored during the 10 ps simulation. In addition, the coordinates of all atoms of the 50 snapshots obtained during the 10 ps simulation were saved and the interaction energy of each snapshot analyzed. It is this second set of data that was used for all subsequent analyses.

D. *Hydroxylation Product Distribution*

Each of the 10 ps MD simulations resulting from the initial orientation 2 was used to predict the regio- and stereoproduct distribution of VPA for each of the P450cam–VPA complexes. To this end, the distance between the active ferryl oxygen to each of the 15 candidate hydrogens was calculated from the stored structures. In this study, an O—H distance of 2.7 Å was used as a cutoff for determining whether the hydrogen atom of VPA was close enough to be abstracted by the ferryl oxygen. This distance was chosen based on the sum of the van der Waals radius of the oxygen and hydrogen atoms. Also recorded are the 15 O· · ·C—H angles that are measures of the linearity of the oxygen attack. An O· · ·C—H angle less than 45° was considered effective for abstracting the hydrogen atoms.

Experimental

Preliminary studies using unlabeled VPA showed that this compound was metabolized by P450cam to C4—OH, C5—OH VPA, and a terminal desaturated Δ^4-

VPA. C3—OH was detected in only trace amounts. Since our focus was the hydroxylation product distribution in this study, the experimental detection of Δ^4-VPA is not discussed here.

The P450cam used in this study was prepared as described by Ortiz de Montellano et al. [16] and was stored at −70°C in the presence of camphor to stabilize the protein. By using a ^{13}C label at the terminal carbon of the pro-*R* or pro-*S* side of VPA, it was possible to establish the absolute configuration at C_2 of each of the observed metabolites. (*R*)- and (*S*)-[5-^{13}C]VPA were thus prepared by synthesis, as described previously [17]. Incubations of [^{13}C]VPA (0.3 m*M*) with KCl (100 m*M*) and P450cam (2.1 μM) in potassium phosphate buffer (pH 7.0, 50 m*M*) were performed. Following equilibration at 30°C for 2–3 min, NADH (1.0 m*M*) was added to initiate metabolism. The samples were then stored overnight at 4°C to allow for precipitation of proteins and cyclization of C4—OH—VPA to the corresponding γ-lactone form. The samples were then centrifuged at 4°C, and the supernatants were extracted with ethyl acetate. After treatment with BSTFA to form trimethylsilyl (TMS) derivatives, aliquots of the final solutions were taken directly for analysis with gas chromatography-mass spectrometry (GC–MS). Metabolites were identified on the basis of their mass spectra and GC retention characteristics. C5—OH VPA was converted to TMS derivatives and C4—OH VPA cyclized to [^{13}C]4-OH-VPA-γ-lactone before analysis. Under the GC–MS conditions employed, the retention times (min:s) of the compounds of interest were as follows: 2*R*4*S*/2*S*4*R* diastereoisomer of C4-OH VPA, 11:35; 2*R*4*R*/2*S*4*S* diastereoisomer of C4-OH VPA, 11:52; and C5-OH VPA (TMS ether, TMS ester), 21:38. For quantitative analyses, the following sets of ions were monitored with GC-MS: [^{13}C]5-OH-VPA bis-TMS, *m/z* 185 and *m/z* 186; [^{13}C]4-OH-VPA-γ-lactone, *m/z* 100 and *m/z* 101. The measured ion currents were corrected for background signals.

The absolute stereochemistry (at the asymmetric C_2 position) of each metabolite of [^{13}C]VPA was thus determined by measuring the relative abundance of each pair of diagnostic ions. Since the relative ion currents at *m/z* 100 and 101 reported on the relative proportions of 2*R* and 2*S* lactones in each chromatographic peak, it was possible to deduce the corresponding proportions of molecules of lactone of 4*R* vs. 4*S* configuration. By this approach, metabolism of [^{13}C]VPA to C4—OH VPA was defined completely in terms of stereoselectivity of oxidation at both C_2 and C_4. A more detailed description of the experiments is to be presented elsewhere [18].

Results and Discussion

A. Optimized Conformers of VPA in Vacuo

The energies of the 74 CHARMm-optimized conformers *in vacuo* are shown in Figure 4(a) in increasing order. Conformer 1 has the lowest energy of −4.0 kcal/mol, whereas conformer 74 has the highest energy, 9.0 kcal/mol. Figure 4(b) and (c) shows the AMBER and AM1 energies for these CHARMm-optimized conformations. The rank order of energies obtained with CHARMm is not maintained

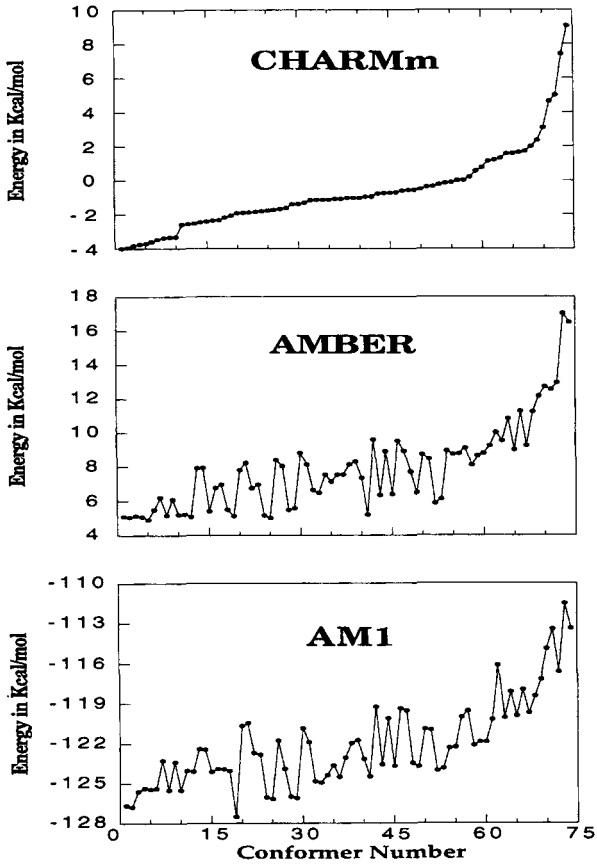

Figure 4. The energies of each of the 74 CHARMm-optimized conformers obtained from three different methods: (a) CHARMm; (b) AMBER; (c) AM1.

using these two methods. For instance, the lowest-energy conformer predicted from AM1 is conformer 19, which has an all-*trans* conformation as shown in Figure 2, whereas that from AMBER is conformer 25. The highest-energy conformer is conformer 73 from both methods, instead of conformer 74 found in CHARMm.

After geometry reoptimization with AM1 and AMBER, further changes in the rank order of the energies were observed. Conformer 19 remains as the lowest-energy one and conformer 71 is the highest from the AM1-optimized results. Conformers 12 and 70 are the lowest- and the highest-energy conformers from the AMBER calculation, respectively. The new *in vacuo* conformational energies obtained with AMBER are shown in the first column of Table II accompanied by their conformer numbers.

By inspecting the six torsional angle profiles among the AM1-optimized conformers, 72 unique conformations were obtained using a smaller angle of 60° (instead of 90°) for the similarity test with two conformers missing because they were each

identical to one retained. Comparing the AM1-optimized conformations with the initial CHARMm conformers, 65 of them remain very similar to their corresponding CHARMm conformers. Only nine conformers have one or more torsional angles deviating from their counterpart by more than 40°. Conformers 73 and 74 undergo the most change with two angles, in each case deviating by more than 100°. Conformer 69 has its τ_4 moved by about 90° from CHARMm to AM1. Similar behavior was observed in comparing the AMBER-optimized conformations with the corresponding CHARMm conformers. Seventy-two unique conformations were found and 68 of them remain similar in conformation to their CHARMm counterpart.

The observation that the rank order of the conformational energy varies from one method to another indicates the weakness of using this energy as the primary guide in choosing the bioactive form of a flexible ligand. Thus, all 72 AMBER-optimized conformers were used in the study of P450cam–VPA interaction.

B. Energy-optimized P450cam–VPA Complexes

When each of the 72 VPA conformers was docked in each of the two orientations, very little conformational change of VPA was observed after energy-minimization. Thus, the induced fit of the ligand in the protein–ligand complex does not move it too far away from its starting configuration and the conformation of VPA and its orientation in the active site are largely retained at this stage. However, the extent to which each conformer interacts with P450cam varies significantly from conformer to conformer. For the 72 unique conformers with orientation 1, the range of this interaction energy is from −13 to −24 kcal/mol. A similar range (−15 to −24 kcal/mol) was observed for the 72 conformers in orientation 2. Therefore, based on the optimized complexes, there is no basis of selecting either of the two initial orientations. As an illustration, the conformer number and the van der Waals, electrostatic, and total interaction energies of the 12 optimized complexes with the best interaction energy in each orientation are shown in columns 1–4 of Table I.

C. Molecular Dynamics Simulations of Selected P450cam–VPA Complexes with the Best Interaction Energies from Both Initial Orientations

Each of the 12 optimized P450cam–VPA complexes from each initial docking orientation were then subjected to MD simulations. During the 10 ps of simulation at 298 K, the Cartesian coordinates of all atoms were recorded and the interaction energy was calculated every 0.2 ps. To have a simple comparison of the effect of dynamic behavior on the binding strength for the different complexes, time-averaged (over 10 ps) quantities of the van der Waals, electrostatic, and total interaction energies were calculated and the results are listed in the last three columns of Table I. As shown in Table I(b), conformers 13, 22, 27, 39, 70, and 71 in orientation 2 have strong MD averaged electrostatic interaction energy ($|\langle \text{Eees} \rangle| > 7$ kcal/mol), leading to a more favorable interaction between VPA and P450cam. This behavior is not seen in the MD simulations for any of the 12 complexes in the initial orientation 1. Thus, after MD simulations, the initial orientation 2 could be selected as the one leading to favorable interactions.

TABLE I. Energy analysis of the P450cam–VPA complexes.

Conformer	Evdw	Eees	Eint	⟨Evdw⟩	⟨Eees⟩	⟨Eint⟩
(a) Orientation 1						
41	−17.6	−5.5	−23.6	−15.7	−4.2	−20.1
67	−17.4	−5.0	−22.8	−15.7	−4.6	−20.5
45	−16.5	−5.5	−22.5	−15.4	−6.4	−22.1
6	−16.8	−5.5	−22.4	−15.3	−5.9	−21.5
35	−16.1	−5.5	−22.1	−15.8	−3.4	−19.4
56	−16.3	−5.1	−22.0	−16.0	−5.6	−21.9
11	−16.3	−4.8	−21.6	−15.5	−5.9	−21.7
26	−15.6	−5.2	−21.2	−14.5	−4.6	−19.7
69	−16.1	−4.6	−21.0	−15.5	−6.7	−22.5
50	−17.6	−3.2	−20.9	−16.2	−3.8	−20.3
51	−17.6	−3.2	−20.9	−15.2	−4.9	−20.4
43	−14.6	−5.8	−20.7	−15.8	−4.0	−20.1
(b) Orientation 2						
27	−16.2	−6.9	−23.6	−15.4	−9.8	−25.5
42	−16.3	−6.0	−23.4	−15.6	−6.3	−22.1
13	−16.8	−6.1	−23.2	−15.1	−8.5	−24.2
7	−16.6	−5.8	−22.9	−15.1	−6.6	−21.9
29	−16.9	−5.0	−22.5	−15.4	−4.3	−20.5
54	−15.9	−5.8	−22.3	−14.9	−6.3	−21.8
11	−15.2	−6.1	−22.1	−14.4	−7.0	−21.7
70	−15.6	−5.4	−22.1	−15.4	−11.3	−27.3
39	−16.2	−5.4	−22.0	−15.7	−7.5	−23.9
22	−16.1	−5.6	−22.0	−14.6	−8.3	−25.5
71	−16.3	−4.3	−21.0	−15.5	−7.1	−23.1
50	−15.1	−5.4	−21.0	−15.3	−6.3	−19.5

Columns 2–4 are for the energy-minimized structures and columns 5–7 are obtained from averaged quantities of a 10 ps MD simulation. (a) Complexes with initial orientation 1; (b) complexes with initial orientation 2.

D. Molecular Dynamics Simulations of All 72 P450cam–VPA Complexes in Initial Orientation 2

To more completely explore the dynamic effects, all 72 complexes in orientation 2 were subjected to MD simulations. As in the optimized complexes, the 72 MD averaged conformers of VPA were compared with their initial ligand free form to determine if the dynamic behavior increased the extent of the difference between the free and bound forms of the ligand. For each conformer, the deviation in the torsional angles between the optimized conformation *in vacuo* and the averaged conformation from the 10 ps MD simulation was examined. The result for each of the six torsional angles is given in Figure 5 as a function of the conformer number. The largest deviation is seen for the 6th torsional angle, O—H—C=O, for many conformers, and the least deviation is for the third torsional angle, C_{3S}—C_2—C_{3R}—C_{4R}. Modest deviations are seen for the other four torsional angles. If a 60° cutoff for every torsional angle is used as the similarity criterion between the VPA *in*

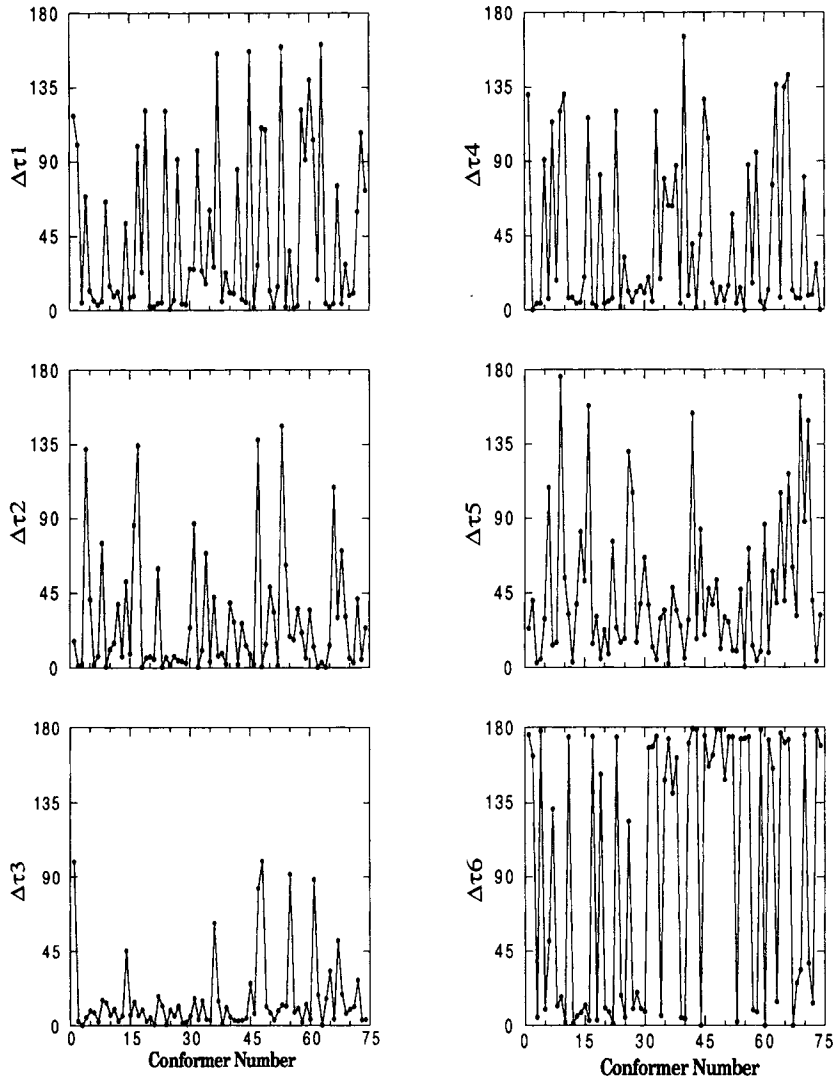

Figure 5. The change in the six torsional angles (defined in Fig. 2) when each of the VPA conformers found *in vacuo* was placed in the P450cam binding site and subjected to a 5 ps heating-equilibration and a 10 ps simulation. The deviation is calculated as the difference between the torsional angle *in vacuo* and the averaged quantity obtained from a 10 ps MD simulation.

vacuo and the dynamically averaged VPA in the P450cam binding site, only 13 conformers remain similar in these two environments. This is another argument against selecting the bioactive forms based on a study that is performed of the free ligand, since the "bioactive conformers" chosen in the free form may change their conformations dramatically when binding to the protein.

TABLE II. Comparison of the interaction energy, Eint, between VPA and the P450cam enzyme after energy optimization, heating-equilibration, and 10 ps of MD simulation.

Conformer	Eint(M)	Eint(H)	⟨Evdw⟩	⟨Eees⟩	⟨Eint⟩
70* (7.7)	−22.1	−22.3	−15.4	−11.3	−27.3
46* (3.9)	−18.2	−22.1	−15.5	−11.0	−27.1
5* (0.1)	−17.7	−26.6	−16.7	−8.4	−25.8
26* (2.9)	−15.8	−26.7	−14.8	−10.1	−25.5
27* (2.4)	−23.6	−29.2	−15.4	−9.8	−25.5
22* (1.8)	−22.0	−25.8	−14.6	−8.3	−25.5
34* (3.1)	−21.8	−26.8	−16.0	−8.5	−25.2
40* (2.1)	−21.6	−25.0	−15.2	−8.7	−24.6
16* (1.5)	−20.5	−26.0	−14.7	−9.1	−24.5
8* (1.5)	−20.0	−24.8	−14.9	−8.9	−24.5
13* (2.3)	−23.2	−26.0	−15.1	−8.5	−24.2
30* (3.3)	−19.5	−22.5	−15.7	−7.7	−24.0
53* (1.4)	−17.3	−26.4	−14.8	−8.5	−24.0
39* (2.9)	−22.0	−23.5	−15.7	−7.5	−23.9
72* (6.9)	−20.9	−25.8	−15.8	−7.3	−23.6
12* (−0.1)	−20.6	−24.3	−14.5	−8.4	−23.6
18* (0.2)	−18.7	−23.1	−15.1	−7.9	−23.6
44* (2.8)	−19.1	−21.1	−15.4	−7.5	−23.6
9* (1.1)	−18.6	−23.4	−15.9	−7.2	−23.5
74 (1.8)	−16.8	−24.1	−14.9	−7.7	−23.1
71* (7.6)	−21.0	−20.5	−15.5	−7.1	−23.1
14* (2.7)	−20.6	−24.2	−14.4	−7.9	−23.0
73* (2.3)	−21.0	−24.3	−16.3	−6.4	−22.9
21 (1.2)	−19.2	−18.7	−13.7	−8.5	−22.8
1 (1.0)	−20.0	−19.4	−14.8	−7.2	−22.5
62* (5.2)	−21.0	−23.8	−14.7	−7.4	−22.5
56 (2.6)	−19.1	−24.4	−15.3	−6.5	−22.4
42 (3.8)	−23.4	−22.8	−15.6	−6.3	−22.1
51 (3.6)	−19.4	−23.3	−16.4	−5.6	−22.1
7 (1.1)	−22.9	−22.8	−15.1	−6.6	−21.9
66 (5.8)	−20.5	−20.9	−15.5	−6.2	−21.9
41 (0.5)	−18.5	−22.8	−14.6	−6.8	−21.9
17 (1.7)	−17.6	−19.7	−15.1	−7.9	−21.8
52 (1.4)	−20.6	−19.8	−14.9	−6.4	−21.8
54 (3.0)	−22.3	−22.5	−14.9	−6.3	−21.8
45 (1.6)	−18.4	−18.0	−15.6	−5.9	−21.7
11 (0.1)	−22.1	−24.9	−14.4	−7.0	−21.7
31 (2.5)	−17.8	−19.0	−15.4	−5.8	−21.7
32 (1.9)	−19.7	−24.0	−15.7	−5.7	−21.7
36 (3.1)	−19.8	−21.5	−15.5	−5.5	−21.5
47 (4.3)	−21.4	−22.4	−15.3	−5.9	−21.5
33 (1.8)	−18.8	−18.6	−15.7	−5.4	−21.5
65 (3.6)	−20.0	−17.1	−14.8	−6.2	−21.5
67 (3.8)	−19.7	−16.8	−16.7	−4.4	−21.4
4 (0.4)	−19.9	−21.4	−15.5	−5.4	−21.2
59 (3.6)	−18.7	−22.1	−14.9	−5.8	−21.2
15 (0.2)	−18.7	−21.7	−15.2	−5.1	−21.2
23 (1.9)	−19.7	−18.6	−14.3	−6.5	−21.1
69 (3.4)	−20.0	−20.0	−15.2	−5.3	−20.9

TABLE II. (*Continued*)

Conformer	Eint(M)	Eint(H)	⟨Evdw⟩	⟨Eees⟩	⟨Eint⟩
37 (2.3)	−18.1	−20.8	−14.4	−5.8	−20.8
2 (1.0)	−19.9	−20.2	−13.8	−6.3	−20.6
35 (3.2)	−18.8	−23.3	−15.7	−4.9	−20.6
61 (3.8)	−20.5	−18.0	−14.7	−5.5	−20.6
55 (2.9)	−20.8	−16.2	−14.8	−5.5	−20.5
29 (1.3)	−22.5	−20.0	−15.4	−4.3	−20.5
57 (2.7)	−19.4	−16.9	−15.6	−4.3	−20.2
43 (1.6)	−18.4	−18.9	−14.3	−5.6	−20.2
50 (4.0)	−21.0	−21.6	−15.3	−6.3	−19.5
49 (1.6)	−17.6	−19.1	−14.8	−4.4	−19.5
38 (2.9)	−21.5	−16.6	−15.2	−3.9	−19.5
10 (1.4)	−19.8	−14.8	−13.9	−4.8	−19.2
28 (1.2)	−20.0	−20.8	−14.8	−3.8	−18.9
6 (0.8)	−17.3	−20.5	−16.1	−2.8	−18.9
19 (1.2)	−14.9	−19.9	−12.7	−5.8	−18.8
63 (4.2)	−18.9	−19.4	−15.6	−2.5	−18.4
24 (1.0)	−15.0	−18.8	−16.5	−1.7	−18.2
20 (1.1)	−19.8	−19.8	−14.1	−3.7	−18.2
25 (0.9)	−17.9	−16.9	−15.4	−2.3	−18.0
68 (5.0)	−17.6	−18.4	−16.3	−1.7	−18.0
58 (3.3)	−20.3	−18.5	−15.3	−2.5	−17.8
60 (3.8)	−20.4	−18.6	−16.4	−0.7	−17.2
3 (0.5)	−18.6	−15.2	−15.3	−1.3	−16.7

⟨Evdw⟩ represents the van der Waals contribution to the interaction energy and ⟨Eees⟩ is the electrostatic interaction energy. Total interaction energies obtained from the optimized complexes are indicated with (M) and those obtained at the end of the 5 ps heating-equilibration are indicated with (H). Quantities in ⟨ ⟩ are for the averaged results from 10 ps MD simulations. Also listed in the first column are the conformational energy of each conformer *in vacuo* as obtained from AMBER. All energies are in kcal/mol. The complexes that reside in a preferred orientation similar to that shown in Figure 6(A) during the MD simulations are indicated with an *.

Table II lists the components of the interaction energies obtained for all 72 complexes at three different stages of modeling. The three stages are (1) after the optimization, (2) after the 5 ps heating-equilibration, and (3) after the 10 ps MD simulation. The total interaction energies of each complex after optimization and at the end of the heating-equilibration period are shown in the second and third columns of Table II. The time-averaged van der Waals, electrostatic, and total interaction energies obtained from the 10 ps MD simulation stage are listed in columns 4–6. This table is arranged in the order of decreasing MD-averaged total interaction energy (i.e., $|\langle \text{Eint} \rangle|$) of the 72 complexes.

There are many useful observations about flexible ligand behavior not generally addressed from the results presented in Table II. The first is that there is no relationship between the relative energy of a VPA conformer in the free state and the extent of its interaction with P450cam in the bound complex. This lack of correlation may be seen by comparing the conformational energy shown in parentheses in

column 1 and the averaged total interaction energy in column 6 of Table II. We see, e.g., that the conformer with the highest energy *in vacuo* (conformer 70 from AMBER) has the strongest interaction with P450cam in the bound complex, whereas conformer 3, a low-energy conformer, has the worst interaction, and, in general, there is no relationship between these two properties. This result reinforces the earlier suggestion that conformational energy *in vacuo* should not be used as a criterion for choosing the bioactive conformer of a flexible ligand. A second observation from this table is that about ⅔ of the complexes were found to have better interaction energy even after just the heating-equilibration process. This result demonstrates the importance of heating-equilibration for the complex to overcome the local barriers in order to move out of a local minimum. This process also provides a dynamic sampling of the configuration space.

The energy decomposition during the dynamics simulations provides additional insight into the origin of the rank order of the interaction between VPA and P450cam. As seen in column 4 of Table II, the contribution from the van der Waals interaction varies very little, between ~ -13 and ~ -16.5 kcal, from conformer to conformer. The small variation in the average van der Waals energy is not surprising since the P450cam binding site is composed mostly of hydrophobic residues and VPA also contains two hydrophobic propyl branches. Therefore, most VPA conformations should find good van der Waals interactions with the binding site. By contrast, comparison of the MD-averaged electrostatic energies of the 72 complexes shows a striking variation from -0.7 to -11.3 kcal/mol.

To understand the origin of this large variation, the motion of VPA inside the active site was analyzed and three qualitatively different preferred positions of VPA in the active site were found. These three different positions are shown in Figure 6 using one specific conformer that remains in each position as an example. In Figure 6(A), VPA is oriented in a way that not only preserves the electrostatic recognition between Tyr96–OH and the carbonyl oxygen of VPA, but also allows an additional strong electrostatic interaction between the anionic Asp297 and the acidic hydrogen of VPA. Conformers that prefer this local position usually are more "compact" due to steric constraints. In Figure 6(B) and (C), VPA occupies positions such that its backbone is either perpendicular or parallel to the two propionate groups of the heme unit. In these two positions, the electrostatic interaction between Tyr96 and VPA can be retained, but no strong additional interaction between Asp297 and VPA is observed. Furthermore, the conformers that prefer these two positions are more likely to have extended geometry. This analysis led to the conclusion that those complexes (indicated with an * in Table II) with a preferred orientation similar to that shown in Figure 6(A) have the best interaction energies.

E. Regio- and Stereoselective Hydroxylation Product Distribution

The experimental results for regio- and stereoselective hydroxylation of VPA by P450cam are shown in Table III. We see from this table that C4—OH VPA is the major metabolite with some C5—OH and little or no C3—OH and C2—OH formed. Also given is the stereoselectivity relative to the two chiral positions C_2

and C_4 for the C4—OH products and relative to C_2 for the C5—OH products. These results posed an experimental challenge because VPA is not a tightly coupled ligand for P450cam and the total amount of hydroxylated product was small. In addition, a mass spectrometric technique was necessary to determine the regio- and stereoselectivity of the metabolites.

This system also posed the formidable theoretical challenge of whether it is possible to predict the regio- and stereoselective hydroxylation product distribution for such a flexible ligand as VPA that forms so many different stable complexes with P450cam. The electronic criterion used to select preferred H-atom sites is the stability of the radical formed by H-atom abstraction by the ferryl oxygen. The steric criterion used to select H-atom sites for abstraction is a ferryl oxygen–hydrogen distance less than 2.7 Å and an O· · ·C—H angle less than 45°. For every snapshot retained in the MD simulations of each complex, the distance and angle were calculated for each H atom. The preferred positions of hydroxylation were obtained directly from the ratios of the frequency of preference at given sites. Although these criteria are based on sound mechanistic evidence and are relatively clear-cut and simple to evaluate, the extent to which each complex contributes to the product formation is not. Moreover, another confounding complication particularly applicable to short-chain aliphatic hydroxylation is that the competing distances are extremely sensitive to very small changes in conformation. Given these complexities, the questions are: (1) To what extent can we reproduce the experimental regio- and stereoselectivity? and (2) What is the best criterion for deciding which complex contributes most to the observed product distribution? To this end, various possibilities were examined and the results were compared with the experimental observations shown in Table III.

First, the extent to which the 12 energy-optimized complexes [listed in Table I(a)] in orientation 1 chosen for the MD simulations could account for the regio- and stereoselectivity was examined. The results are listed in Section I of Table III and they do not agree well with the experiments. Specifically, C5—OH VPA was found to be the most sterically favored product with the production rate of C3—OH similar to that of C4—OH VPA. However, the C5 radical intermediate is found to be the least stable from a previous AM1 calculation [16], and if both the electronic and steric effects are taken into account, the formation of C5—OH VPA is not as likely. However, even with this requirement, the calculated regio- and stereoselectivity based on these 12 complexes do not agree with the experiments.

The next set of complexes examined were the 12 most favorable complexes for the initial orientation 2 obtained from the energy-optimized results. The results using the MD simulations of these 12 complexes are listed in Section II. Compared to the results for orientation 1, the predicted regio- and stereoselectivity agree more with the experiments for these complexes in orientation 2. C3—OH VPA is still found, but to a much lesser extent. The predicted stereoselectivity for the C5—OH isomers and for the C4—OH products are quite similar to the experimental results. The conformers in this orientation are more likely to have stronger interaction with P450cam (as seen in Table I) and more closely reproduce the observed ex-

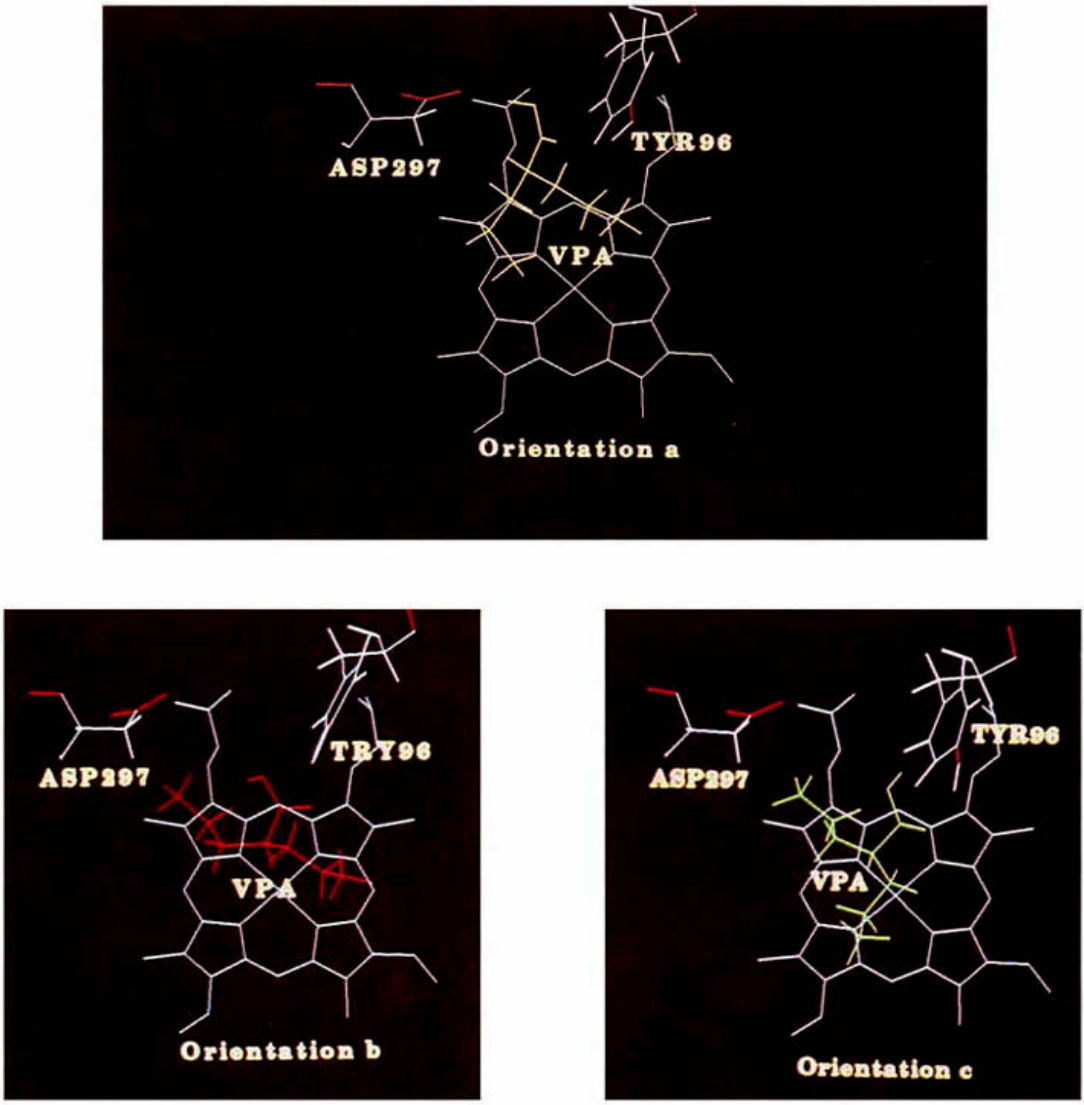

Figure 6. Three different positions that VPA occupies within the active site pocket of P450cam. All three positions are viewed from the top of the heme unit.

perimental regio- and stereoproduct distributions. Thus, one conclusion that can be reached from these studies is that the initial orientation 2 is the dominant one.

To further explore the contributions from this orientation, all 72 complexes in the initial orientation 2 were examined and the results are shown in Section III of Table III. If all complexes are used, the calculated regio- and stereoselectivity are not in as good agreement with the experiments as if only the 12 complexes are used as presented in Section II. To more fully explore the dynamic effect in the most

favorable complexes in orientation 2, 16 complexes with the best MD-averaged total interaction energies (see Table II) in this orientation were each subjected to another 40 ps of MD simulations. For all 16 complexes, the VPA resides in a position similar to that shown in Figure 6(A), although the conformations are quite different. Each of the sixteen 50 ps MD simulations were again analyzed for the regio- and stereo-product formation. The results are shown in Section IV of Table III. If both the results from these MD simulations and the study of the stabilities of the radical intermediates were taken into account, good agreement with the experiments in the regio- and stereoselective product distribution is obtained, except for the small amount of C3—OH formation predicted. The combined results indicate that the P450cam–VPA complexes with better interaction energy make the major contributions to the product formation.

Very different product distributions are found for each of the sixteen 50 ps MD simulations. It is clear that no one P450cam–VPA complex is able to account for the observed regio- and stereoselectivity. Therefore, we conclude that the observed product distribution is a result of VPA in many different conformers and that the

TABLE III. Experimental and calculated regio- and stereoselectivities of the hydroxylated VPA products catalyzed by the cytochrome P450cam enzyme.

	Regioselectivity	Stereoselectivity
(a) Experimental results		
	C4—OH ≫ C5—OH	2*S*5 > ~ 2*R*5
		2*R*4*S* > ~ 2*S*4*R* > 2*R*4*R* > ~ 2*S*4*S*
	No C3—OH	
(b) Theoretical predictions		
I	C5—OH(225)	2*S*5(99) < 2*R*5(126)
	C4—OH(166)	2*R*4*S*(61) > 2*S*4*S*(47) > 2*R*4*R*(39) > 2*S*4*R*(19)
	C3—OH(151)	2*R*3*S*(99) > 2*R*3*R*(38) > 2*S*3*S*(14) > 2*S*3*R*(0)
	C2—OH(0)	
II	C5—OH(182)	2*S*5(94) > 2*R*5(88)
	C4—OH(204)	2*R*4*S*(98) > 2*S*4*R*(47) > 2*S*4*S*(34) > 2*R*4*R*(25)
	C3—OH(67)	2*S*3*R*(29) > 2*S*3*S*(15) > 2*R*3*R*(14) > 2*R*3*S*(9)
	C2—OH(0)	
III	C5—OH(1068)	2*S*5(614) > 2*R*5(454)
	C4—OH(714)	2*S*4*R*(270) > 2*S*4*S*(233) > 2*R*4*S*(148) > 2*R*4*R*(63)
	C3—OH(670)	2*S*3*R*(344) > 2*R*3*R*(143) > 2*S*3*S*(101) > 2*R*3*S*(82)
	C2—OH(58)	
IV	C5—OH(1998)	2*S*5(1109) > 2*R*5(889)
	C4—OH(892)	2*S*4*R*(384) > 2*R*4*S*(326) > 2*S*4*S*(101) > 2*R*4*R*(81)
	C3—OH(270)	2*R*3*R*(149) > 2*S*3*R*(102) > 2*S*3*S*(17) > 2*R*3*S*(2)
	C2—OH(0)	

(I) Overall statistics from 12 complexes with initial orientation 1, 10 ps MD simulation used for each complex. (II) Overall statistics from 12 complexes with initial orientation 2, 10 ps MD simulation used for each complex. (III) Overall statistics with all 72 complexes with initial orientation 2, 10 ps MD simulation used for each complex. (IV) From 16 complexes that have best interaction energy from the MD simulation. Fifty picosecond simulation was performed for each of the 16 complexes.

protein–ligand interaction energy is a reasonable criterion for determining the complexes that contribute to the product formation. The fact that many conformers can be bioactive shows that one should not use only the global minimum as in the automated docking method when considering the interaction between a flexible ligand and its receptor.

Conclusion

A combined theoretical and experimental study of the binding and interaction of valproic acid (VPA) with the bacterial P450cam enzyme and the subsequent regio- and stereoselective hydroxylation product distribution was performed. In the first part of the theoretical studies, 72 unique VPA conformers were found using AMBER and all were used for the study of the P450cam–VPA interaction. The results obtained provide a number of useful insights into the behavior of flexible ligands upon binding to their target proteins. For example, most conformers of VPA undergo a large change in conformation when they bind to P450cam, as determined by comparisons of the MD-averaged torsional angles of the bound form with the free-form conformers. In addition, we have shown that low-energy conformers of VPA do not necessarily lead to optimum interactions with the target protein. In fact, there is no relationship between the energy of a free-form ligand and the MD-averaged intermolecular interaction energy of the complex resulting from binding of the free-form ligand with its target protein. We have also demonstrated the importance of MD simulations in identifying the preferred binding position of flexible ligands. The conformation and orientation of the ligand obtained in the energy-minimized complex may change during the MD simulations, leading in most cases to a lower-energy, more favorable complex. In the preferred position identified, there are strong electrostatic interactions between VPA and both Tyr96 and Asp297.

The second part of this study addressed the regio- and stereoselectivity of hydroxylation products of VPA. Experimentally, C4—OH VPA was found to be the predominant hydroxylation product with small amounts of C5—OH VPA formed. The stereoselectivity of hydroxylation found was $2R4S > \sim 2S4R > 2R4R > \sim 2S4S$ and $2S5 > \sim 2R5$. In the computational studies, we identified several key reasons why prediction of product distribution of flexible ligands can be a very challenging task. More than one form of the protein–ligand complex contributes to its oxidative metabolism; the conformational behavior of the free ligand is not a good criterion for ruling out the contribution of a given conformer to the product formation, nor is the energy of the optimized complex. In addition, the geometric criteria used for determining product selectivity for short-chain aliphatic hydroxylation are among the most sensitive to small fluctuations in bound protein–ligand conformations. Furthermore, an important aspect of the interaction and reaction between ligands and proteins that we did not address in this paper is that the ultimate quantity associated with the binding affinity is the Gibbs free energy of binding. There have been a few attempts in the literature to calculate $\Delta G_{\text{binding}}$ empirically using the crystallographic structure of the binding complex [19,20].

However, so far, there is no simple way of incorporating all contributions to $\Delta H_{\text{binding}}$ and $\Delta S_{\text{binding}}$ systematically into the computational search of the configuration of the protein–flexible ligand complex. These combinations of factors make the problem a maximum challenge of the predictive capabilities of computational studies. The best criteria found thus far for including the contribution of a given complex to product formation is the optimum interaction energy of the protein–ligand complex after MD simulation, which still leads to remaining disparities between predicted and observed product distributions. However, the combined theoretical and experimental effort has allowed the identification of the preferred modes of interaction among the many possible complexes for this flexible ligand with P450cam, a goal that is not yet possible for either method alone.

Acknowledgments

The support from the Environmental Protection Agency Grant #CR-818677-01-0 for Y.T.C. and G.H.L. for this work is gratefully acknowledged. A.E.R., T.A.B., and P.R.S. are supported by NIH Grant #GM32165. The support from NIH Grant #GM25515 for P.O.M. is also acknowledged. We also wish to acknowledge the support of the Pittsburgh Supercomputing Center where the computations were performed on a Cray-YMP machine.

Bibliography

[1] (a) C. Cometta-Morini and G. H. Loew, J. Comput. Aided Mol. Design **5**, 335 (1991). (b) J. J. Perez, H. O. Villar, and G. H. Loew, *Ibid.* **6**, 175 (1992).
[2] D. S. Goodsell and A. J. Olson, Proteins Struct. Funct. Genet. **8**, 195 (1990).
[3] P. R. Ortiz de Montellano, in *Cytochrome P450: Structure, Mechanism, and Biochemistry*, P. R. Ortiz de Montellano, Ed. (Plenum Press, New York, 1986), pp. 217–271.
[4] (a) A. E. Rettie, A. W. Rettenmeier, W. N. Howald, and T. A. Baillie, Science **235**, 890 (1987). (b) A. E. Rettie, M. Boberg, A. W. Rettenmeier, and T. A. Baillie, J. Biol. Chem. **263**, 13733 (1988). (c) P. R. Ortiz de Montellano, Trends Pharmacol. Sci. **10**, 354 (1989). (d) M. J. Eadie, W. D. Hooper, and R. G. Dickinson, Med. Toxicol. **3**, 85 (1988).
[5] T. L. Poulos, in *Cytochrome P450: Structure, Mechanism, and Biochemistry*, P. R. Ortiz de Montellano, Ed. (Plenum Press, New York, 1986), pp. 505–523.
[6] R. E. White, M.-B. McCarthy, K. D. Egeberg, and S. G. Sligar, Arch. Biochem. Biophys. **228**, 493 (1984).
[7] M. H. Gelb, D. C. Heimbrook, P. Malkonen, and S. G. Sligar, Biochemistry **21**, 370 (1982).
[8] S. G. Sligar, M. H. Gelb, and D. C. Heimbrook, in *Microsomes, Drug Oxidations and Drug Toxicology*, R. Sato and R. Kato, Eds. (Wiley-Interscience, New York, 1982), p. 155.
[9] G. H. Loew, J. R. Collins, B. T. Luke, A. Waleh, and A. Pudzianowski, Enzyme **36**, 54 (1986).
[10] J. R. Collins, D. L. Camper, and G. H. Loew, J. Am. Chem. Soc. **113**, 2736 (1991).
[11] Polygen Corporation, Inc., CHARMm Version 2.1 (Waltman, MA, 1989).
[12] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, J. Am. Chem. Soc. **107**, 3902 (1985).
[13] U. C. Singh, P. K. Weiner, J. W. Caldwell, and P. A. Kollman, AMBER UCSF Version 3.0a (Dept. of Pharmaceutical Chemistry, University of California, San Francisco, 1986).
[14] J. A. Fruetel, J. R. Collins, D. L. Camper, G. H. Loew, and P. R. Ortiz de Montellano, J. Am. Chem. Soc. **114**, 6987 (1992).

[15] T. E. Ferrin, The MIDAS Display System. J. Mol. Graph. **6**, 13 (1988).
[16] P. R. Ortiz de Montellano, J. A. Fruetel, J. R. Collins, and G. H. Loew, J. Am. Chem. Soc. **113**, 3195 (1991).
[17] D. J. Porubek, H. Barnes, G. P. Meier, L. J. Theodore, and T. A. Baillie, in *Synthesis and Applications of Isotopically Labelled Compounds-1988,* T. A. Baillie and J. R. Jones, Eds. (Elsevier, Amsterdam, 1989), pp. 367–374.
[18] A. E. Rettie, T. A. Baillie, P. R. Sheffels, Y. T. Chang, G. H. Loew, and P. R. Ortiz de Montellano, in preparation.
[19] K. P. Murphy, D. Xie, K. C. Garcia, L. M. Amzel, and E. Freire, Proteins Struct. Funct. Genet. **15**, 113 (1993).
[20] J. Novotny, R. E. Bruccoleri, and F. A. Saul, Biochemistry **28**, 4735 (1989).

Received March 14, 1993