Understanding the Substrate Selectivity and the Product Regioselectivity of Orf2-Catalyzed Aromatic Prenylations†

Guanglei Cui, Xue Li, and Kenneth M. Merz, Jr.*

Department of Chemistry and Quantum Theory Project, 2328 New Physics Building, P.O. Box 118435, University of Florida, Gainesville, Florida 32611-8435

Received October 5, 2006; Revised Manuscript Received December 1, 2006

ABSTRACT: Orf2, a recently identified prenyltransferase of aromatic natural products, displays relaxed substrate selectivity and interesting product regioselectivity. This gives rise to the opportunity to engineer the active site to tune the functionality of terpenoids for therapeutic applications. The structural basis of substrate binding has been determined, but the source of the observed substrate selectivity and product regioselectivity cannot be completely understood on the basis of the static picture that the crystal structures of Orf2 and its complexes afford. The electron density and *B*-factors of the substrates, particularly those of 1,6-dihydroxynaphthalene, suggest significant conformational fluctuation in the Orf2 binding site. We thoroughly explored the binding of 1,6-dihydroxynaphthalene and quantitatively evaluated the relative free energies of three binding states that we identified in terms of a two-dimensional potential of mean force. The available experimental orientation, which gives the major prenylated product of 1,6-dihydroxynaphthalene, corresponds to the global free energy minimum. Two alternative binding states were identified on the calculated free energy surface, and both are readily accessible at 300 K. The alternative binding conformations were extracted from the potential of mean force calculation and were subjected to further validation against the experimental X-ray diffraction data using a refinement protocol supplemented with a hybrid quantum mechanical and molecular mechanical energy function. The agreement was excellent as indicated by the R and R_{free} factors that were comparable to that obtained for the published orientation using a similar protocol. These binding states are the origin of the selectivity and regioselectivity in Orf2-catalyzed aromatic prenylations. Our analyses also suggest that Ser214 and Tyr288, forming hydrogen bonds with the alternative binding states of 1,6-dihydroxynaphthalene and flaviolin, are good candidates for site-directed mutagenesis, and changing them to, for example, their hydrophobic counterparts would affect the substrate selectivity and product regioselectivity.

Isoprenoids or terpenoids make up a large class of naturally occurring organic compounds with tremendous chemical and structural diversity. They are organic materials produced in the HMG-CoA (3-hydroxy-3-methylglutaryl-CoA) reductase pathway (*1*, *2*) or the MEP/DOXP (2-C-methyl-D-erythritol 4-phosphate/deoxyxylulose phosphate) pathway (*3*) having molecular structures containing carbon backbones made up of isoprene units. Isoprenoids are ubiquitous across a wide range of organisms, such as eubacteria, archaea, algae, plants, animals, and fungi. Plant isoprenoids have been long appreciated for their "aromatic" qualities, such as those found in citral, menthol, camphor, etc., and have been commonly used in traditional herbal remedies (*4*). It was recently discovered that certain isoprenoids and their derivatives exhibit desirable pharmaceutical characteristics, such as antimicrobial, antioxidant, antiinflammatory, antiviral, and anticancer effects (*5*−*14*). More importantly, natural isoprenoids have low cellular toxicity and good membrane permeability, which make them ideal drug template compounds.

Several thousand natural terpenoids have been isolated and classified on the basis of their origin, the number of isoprene units that they contain, and the roles that they play in their parent organisms. The biosynthesis and the chemical diversity of terpenoids have always been of great interest to biochemists and medicinal and organic chemists. A number of terpenoid synthase families have been categorized and extensively studied, including monoterpene, diterpene, triterpene, and sesquiterpene cyclases. The advances in the research of terpenoid biosynthesis have been summarized in a recent review by Christianson (*15*), in which the structural foundation of various terpenoid cyclases was described in great detail. A common trait of these enzymes is the essential magnesium cluster and the aspartate rich magnesium binding motifs, which implies a certain similarity in their reaction mechanisms. In addition to terpenoid synthase, prenyltransferase contributes to terpenoid biosynthesis as well. One such example is Orf2 (*16*), a newly identified 300-residue prenyltransferase from *Streptomyces*. Although the natural substrate is unknown, Orf2 accepts a great variety of aromatic compounds, such as dihydroxynaphthalenes, flaviolin, and 4-hydroxyphenylpyruvate, and

† This work was supported by a NIH Grant (GM44974) to K.M.M.

* To whom correspondence should be addressed: Department of Chemistry and Quantum Theory Project, 2328 New Physics Building #92, P.O. Box 118435, University of Florida, Gainesville, FL 32611-8435. Telephone: (352) 392-6973. Fax: (352) 392-8722. E-mail: merz@qtp.ufl.edu.

10.1021/bi062076z CCC: $37.00
Published on Web 01/18/2007

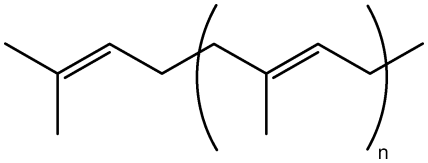

FIGURE 1: Chemical structure of a prenyl group. *N* can be 0 (dimethylallyl), 1 (geranyl), 2 (farnesyl), 3 (geranylgeranyl), etc.

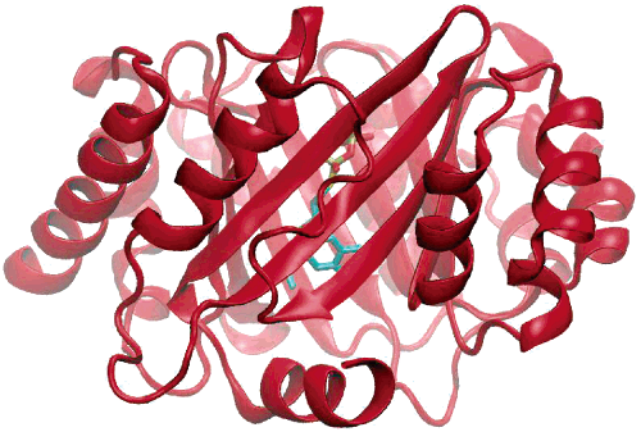

FIGURE 2: PT-barrel of Orf2 (red ribbons) and the bound substrates, GSPP and 1,6-DHN (shown in licorice colored by atom types) from 1ZB6.

attaches geranyl (C_{10}) or farnesyl groups (C_{15}) to them (Figure 1). The geranylated products can then be further combined, modified, cyclized, and transferred by downstream enzymes. This certainly explains, at least in part, the apparent chemical diversity in plant and animal terpenoids. It also opens the opportunity of exploiting the relaxed substrate specificity of Orf2, possibly through protein engineering, in simplifying the synthesis of terpenoids in drug discovery (*17*, *18*). In addition to the unusual substrate specificity, Orf2 also displays interesting regioselectivity in the prenylated products. Those of 1,6-dihydroxynaphthalene (1,6-DHN) have been characterized by both mass spectroscopy and ^{1}H nuclear magnetic resonance analyses as *trans*-2-geranyl 1,6-DHN and *trans*-5-geranyl 1,6-DHN with a yield ratio of 10:1.

The three-dimensional crystal structures of Orf2 and the Orf2 substrate complexes provided the initial clues to the observed variability in the observed reactants and products. Unlike two other members, farnesyltransferase (FTase) and geranylgeranyltransferase (GGTase), of the prenyltransferase family whose structures have been determined (*19*−*21*), Orf2 adopts a novel α/β barrel fold [termed PT-barrel by the authors (Figure 2)] that results in a spacious binding site for geranyl diphosphate (GPP) and aromatic substrates. A magnesium ion, required for enzyme activity, is located near the diphosphate group of GPP and bound to the carboxylate of Asp62, the α-phosphate of GPP, and four water molecules. The commonly found trinuclear magnesium cluster and magnesium cluster binding motifs in terpenoid synthase were not observed in any of the crystal structures that were determined, which are summarized in Table 1. Although Orf2 is an Asp/Glu rich protein, most of the aspartate and glutamate residues are distributed on the outer α-barrel, and only two aspartate residues reside near where the magnesium and diphosphate are bound, Asp62 and Asp110. In 1ZB6, the carboxylate of Asp110 is 3.1 Å from a magnesium-bound water molecule, providing additional support for the metal binding possibly through hydrogen bond interactions. Geranyl *S*-thiolodiphosphate (GSPP), a GPP analogue, was used to allow the determination of the ternary complexes. The GSPP substitution and the binding of the aromatic substrates have little impact on the side chain conformations of the residues in the binding site. When they are superimposed, the all-atom root-mean-square deviation (rmsd) of any two Orf2 structures (residues 10−300) is no greater than 0.5 Å.

We noted that 1,6-DHN and flaviolin have rather different conformations in the deposited structures (Figure 3). In 1ZB6, 1,6-DHN is oriented so that C2 and C5 are 4.1 and 6.7 Å from C1 of GSPP, respectively, which can only explain the formation of the major product. In 1ZDW, the closest atom of flaviolin is ∼6 Å from C1 of GSPP. How the minor prenylated 1,6-DHN and prenylated flaviolin products are formed cannot be directly understood from the two crystal conformations. One possibility is that the aromatic substrates are not tightly bound and can have significant fluctuations in their active site orientation(s). The *B*-factors of 1,6-DHN and flaviolin were inspected as well as their electron density maps ($2F_o - F_c$) downloaded from the Uppsala Electron Density Server (*22*). Unlike the residues that are located in the PT-barrel, both 1,6-DHN and flaviolin have much higher *B*-factors, on average 65.68 and 48.63 Å^2, respectively, similar to those of the surface residues (Figure 4). The electron density contour plots (Figure 5) at a density value of 1.5σ reveal the lack of clearly defined density for a unique 1,6-DHN conformation, which is a strong indication of substantial fluctuation in substrate binding. At the same contour level, however, flaviolin can be clearly placed with little ambiguity. As we examined the lower-density contour levels in 1ZB6 (Figure 5), a rather large volume of space is observed and it is difficult to explain the shape of this volume by just using a single 1,6-DHN orientation.

This unusual substrate binding observed in Orf2 deserves further investigation in the pursuit of a definitive answer to the questions raised above. We quantitatively characterized the binding of 1,6-DHN by computing the potential mean force of binding as a function of two independent variables, the distances between the reaction centers. Two alternative binding conformations were identified and validated through our QM/MM X-ray refinement approach (*23*). The calculations offer a convincing interpretation on the observed substrate selectivity and product regioselectivity in the Orf2-catalyzed 1,6-DHN prenylation, which may be used as a general guide for rationalizing the outcome of the prenylation of other similar aromatic substrates. In the following sections, we describe the details of the methods that we used in this study and present and discuss our results.

METHODS

Unless noted otherwise, all molecular dynamic (MD) calculations in this study were performed using the *sander* module in the AMBER8 modeling suite (*24*); all quantum mechanical calculations were performed using Gaussian'03 (*25*).

MD Simulation of the Orf2 Ternary Complex with GPP and 1,6-DHN. The small variation in the observed structure of Orf2 among different crystal structures makes it relatively straightforward to build a model for the reactive Orf2 ternary complex with GPP and 1,6-DHN. We combined the coor-

Table 1: Available Crystal Structures of Orf2 Complexes

PDB entry	mean *B*-factor	resolution (Å)	bound substrate(s)
1ZCW	29.40	2.25	geranyl diphosphate
1ZDY	16.70	1.44	*N*-[tris(hydroxymethyl)methyl]-3-aminopropanesulfonic acid
1ZB6	45.60	1.95	1,6-dihydroxynaphthalene and geranyl *S*-thiolodiphosphate
1ZDW	29.00	2.02	flaviolin and geranyl *S*-thiolodiphosphate

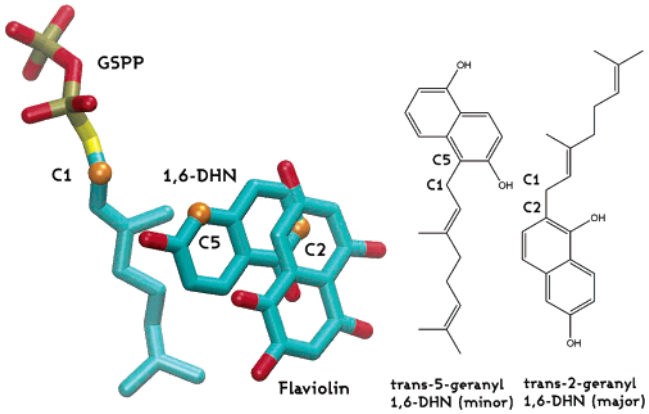

FIGURE 3: Crystal binding orientations (left) of 1,6-DHN and flaviolin observed in the deposited 1ZB6 and 1ZDW structures and two-dimensional sketches of the prenylated 1,6-DHN (right). The C2 and C5 atoms of 1,6-DHN and the C1 atom of GSPP are shown as orange spheres.

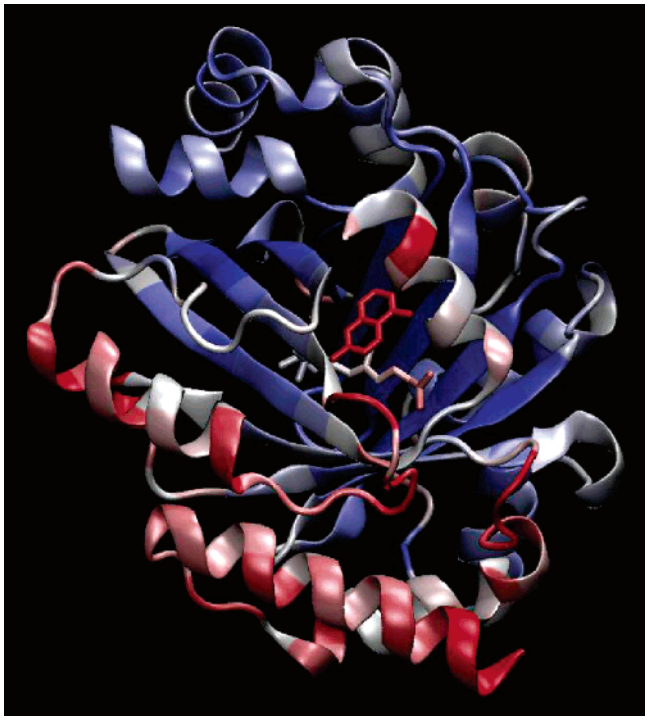

FIGURE 4: Crystal structure of 1ZB6 colored by *B*-factor values. The color scale (from blue to white to red) has a range from 23.0 to 67.0 Å^2. 1,6-DHN is shown with thick cylinders with an average *B*-factor of 65.68 Å^2.

dinates of 1ZCW, the Orf2 binary complex with GPP, with those of 1,6-DHN from 1ZB6 by simple protein alignment. The coordinates of solvent and the magnesium ion in 1ZCW were also carried into this initial model, while the nitrate ion was left out. The AMBER *parm99* force field (*26*) was used to model the enzyme, and the AMBER *gaff* force field (*27*) for small organic compounds was used to model 1,6-DHN, whose atomic charges were computed using a two-stage restrained electrostatic potential [RESP (*28*, *29*)] fitting procedure. The force field parameters for prenyl diphosphates were developed in our previous study of the farnesyltransferase ternary complex (*30*) and were transferred directly to build geranyl diphosphate, which was modeled as being fully deprotonated bearing three negative charges. This ternary complex was then solvated in a rectangular periodic box with each side at least 9 Å from the closest solute atoms. The three-site constrained water model TIP3P (*31*) was used in all simulations. A sodium ion was added to neutralize the system. Long-range electrostatic interactions were treated with the particle mesh Ewald [PME (*32*)] method, and the default settings (8 Å cutoff for the real-space nonbond interactions and a reciprocal space grid spacing of approximately 1 Å) were used. All bonds with hydrogen atoms were constrained using SHAKE (*33*), and the translation of the whole system inside the periodic box was removed. The whole system (33 571 atoms) was first relaxed to clear possible close contacts with a few short, weakly restrained, energy minimizations, and then the system temperature was slowly increased to 300 K over 50 ps of restrained MD simulation with a 1 fs time step. Concomitantly, the restraining force was gradually removed before the following constant NPT production run. The density, potential energy, and volume of the system were examined to ensure a properly relaxed system. During the 16 ns production run, a 2 fs time step was used and the structure snapshots were saved every 2 ps.

Two-Dimensional Potential of Mean Force Calculation. The conformational preference of 1,6-DHN confined in the Orf2 binding site was calculated as the potential of mean force [PMF (*34*)] of two distances, namely, from the C1 atom of GPP to the C2 and C5 atoms of 1,6-DHN, and will be called D_1 and D_2, respectively. The ranges of D_1 and D_2 were determined from the 16 ns MD calculation to be 3.55−10.64 and 4.29−12.37 Å, respectively. The area defined was then partitioned into a 10 × 10 regular grid. The grid points that were never sampled during the 16 ns MD simulation were not considered in the PMF calculation, which resulted in a total of 55 "valid" grid points. The region around each valid grid point was sampled in a MD simulation restrained by a two-dimensional umbrella biasing potential, which was harmonic and centered at the grid point. Additional sampling points were introduced at locations where the statistical errors were not acceptable. The same biasing strength (20.0 kcal mol^{-1} Å^{-2}) was used for all biased MD simulations, each of which was equilibrated for 500 ps followed by data collection for an additional 500 ps. Distances sampled were saved every 20 fs. The weighted histogram analysis method [WHAM (*35*)] was used to reconstruct the unbiased D_1 and D_2 distribution and calculate the free energy profile for the range of $D_1 \in [3.0, 9.5]$ and $D_2 \in [3.0, 11.0]$. Twenty-five bins were used for each dimension, and the convergence criterion of the WHAM analysis was 0.001 kcal. The

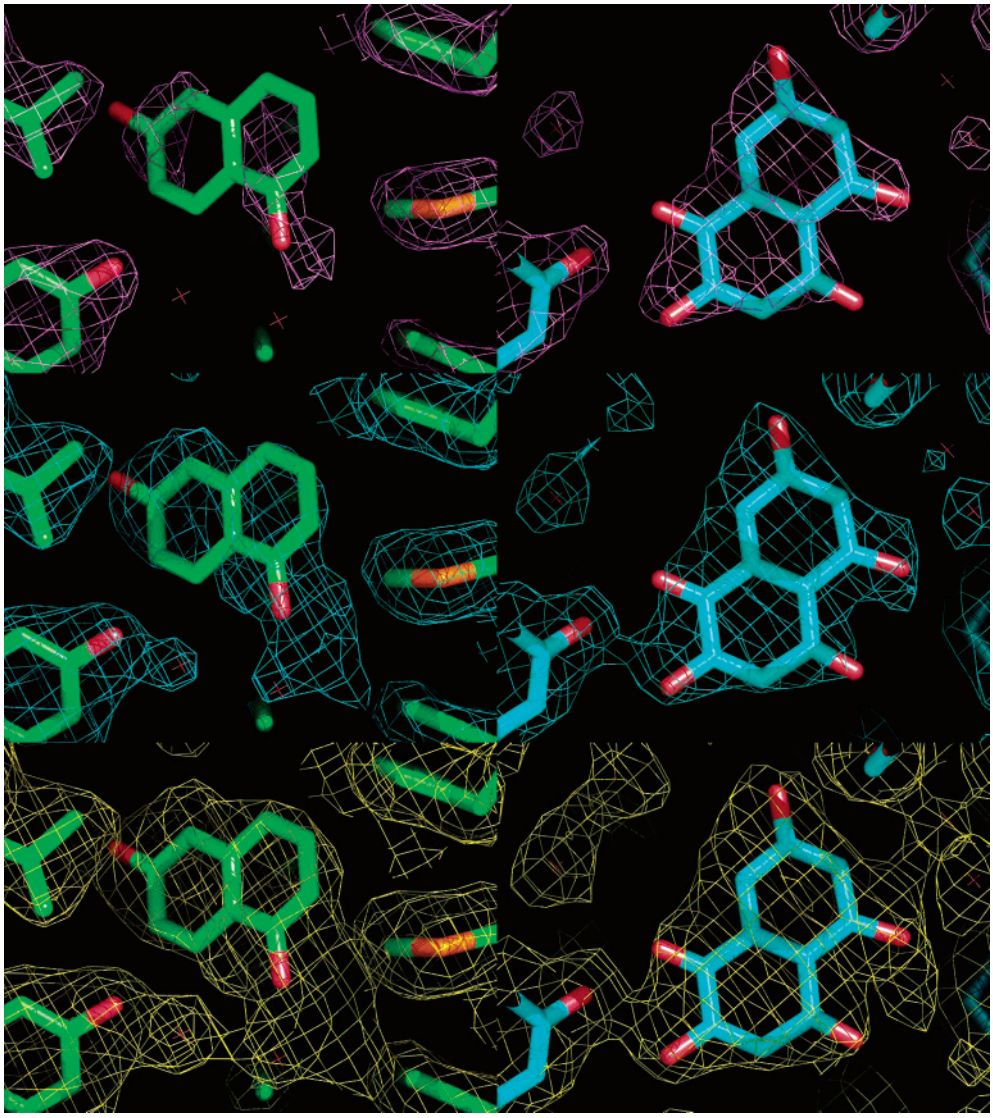

FIGURE 5: Electron density contour plots ($2F_o - F_c$) for 1,6-DHN (left) and flaviolin (right). The contour values are 1.5σ (high), 1.0σ (medium), and 0.5σ (low) from the top to the bottom.

statistical uncertainty was estimated by a Monte Carlo bootstrap analysis.

CNS and QM/MM X-ray Refinements. The PMF and the 1ZB6 structure suggested alternative binding conformers for 1,6-DHN, and these were further validated and refined using the crystallography and NMR system (CNS) (*36*) and a hybrid quantum mechanical/molecular mechanical (QM/MM) X-ray refinement approach that we have developed (*23*). The methodology and implementation of our QM/MM X-ray refinement approach can be found in detail elsewhere (*23*). To briefly recapitulate, the essence of this method is replacing the simple classical molecular mechanic energy term in the target function of CNS with a more sophisticated and physical quantity represented by the combination of the AMBER all-atom force field *parm99* for general protein residues and semiempirical quantum mechanical Hamiltonians for chosen residues that need special focus, for example, metal ions and nonpeptidic substrates, the reliable force field parameters of which are not readily available. The X-ray component and the associated weighting factors in the target function of CNS are kept intact.

In all the refinements carried out in this study, the complete set of 22 923 reflections between the resolution limits of 1.95 and 29.74 Å obtained from the Protein Data Bank were considered, 1154 of which were used for validation. The alternative 1,6-DHN conformers identified from the free energy calculation were introduced into 1ZB6 by simple superposition using the protein backbone after the original 1,6-DHN conformer, and certain crystal water molecules were removed to avoid steric clashes. Both the CNS and QM/MM X-ray refinements adapted the minimization protocol provided by the CNS program suite. A two-stage refinement protocol was established through experimentation, in which the structure of any binding conformation was first optimized with the CNS program suite using the built-in MM energy function and the widely accepted Engh and Huber parameter set (*37*). Subsequently, the results of the previous CNS calculations were used as the initial structures and were minimized with the AMBER MM energy function to relax the hydrogen atoms while keeping the heavy atoms fixed. In the following QM/MM X-ray refinement calculations, the residues chosen to be treated quantum mechanically include Mg^{2+}, GSPP, 1,6-DHN, and all the protein side chains and solvent molecules that are in direct contact through either hydrogen bonding, metal coordination, or van der Waals interactions. This created a neutral QM region of 666 atoms.

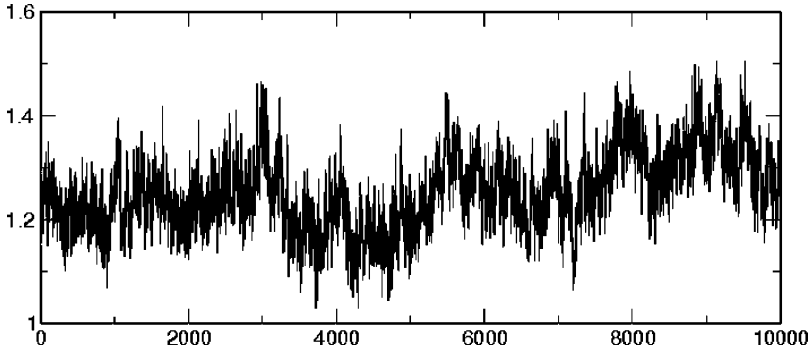

FIGURE 6: Root-mean-square deviation (angstroms) of the backbone of Orf2 over time (picoseconds).

Divide-and-Conquer (*38*−*42*), a linear-scaling semiempirical technique, was employed to efficiently compute the QM energy at the PM3 (*43*, *44*) level of theory. In both the CNS and QM/MM refinements, three different X-ray weighting factors (wa), 0.01, 0.2, and 1.0, were selected, allowing us to evaluate the effect of weighting factors on the quality of the final refinements.

RESULTS AND DISCUSSION

In the deposited crystal structure 1ZB6, a single conformation was assigned to 1,6-DHN, which accounts for the formation of the major prenylated product, even though the deposited electron density, in our estimation, suggests that other binding conformations are also possible. The distribution of all possible binding conformations is determined by the interplay of protein−substrate, protein−solvent, and substrate−solvent interactions. To better understand the Orf2 substrate selectivity and explain the observed regioselectivity in the prenylated product, we first characterized the binding landscape and then evaluated the probability of this distribution. The approach that we use is based on the principles of statistical mechanics and the techniques of numerical simulations.

MD Simulation of the Orf2 Ternary Complex with GPP and 1,6-DHN. To characterize and retrieve the distribution of binding of 1,6-DHN, the time evolution of the ternary complex of Orf2 with GPP and 1,6-DHN was calculated for a total duration of 16 ns starting with the model complex structure that we built from crystal structures 1ZB6 and 1ZCW. When our simulation results are compared to those of 1ZCW, the backbone conformation of the protein was well-preserved during the MD simulation with an average rmsd of 1.3 Å for the last 10 ns (Figure 6). A snapshot from the MD simulation is shown in Figure 7. The magnesium ion maintained a stable octahedral configuration, coordinated by four solvent molecules located in the same plane, the α-phosphate of GPP, and the side chain of Asp62. The diphosphate group of GPP strongly interacted with the side chains of several residues, including Lys119, Lys169, Arg228, Tyr216, and Lys284. Lys169 and Arg228 are situated near the edge of the β-barrel and were solvent-exposed in the deposited structure. During the MD simulation, they were attracted by the negative charge on the diphosphate and formed salt bridges to further stabilize the bound GPP. The geranyl group of GPP remained mostly extended, but unlike what was observed in a previous MD simulation of the FTase ternary complex where the rotation of backbone torsion angles in farnesyl diphosphate (FPP) was seriously hindered, the backbone of GPP was more flexible.

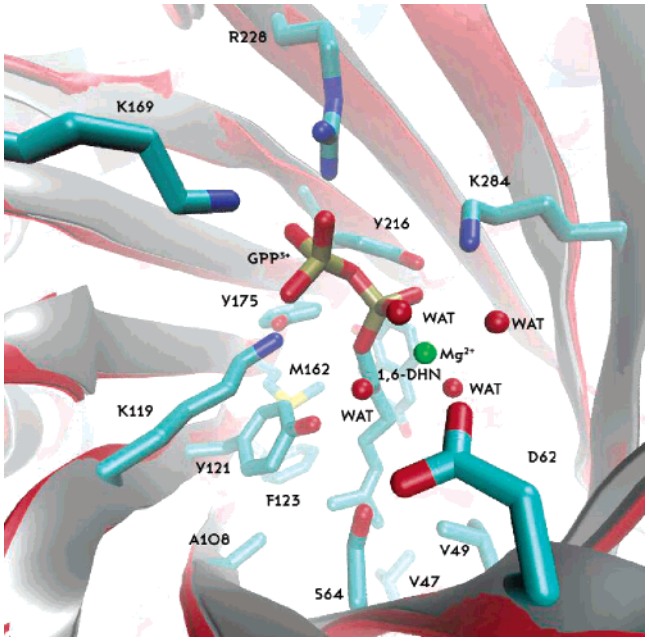

FIGURE 7: Snapshot of the MD simulation of the Orf2 ternary complex looking down the PT-barrel. The protein backbone is shown as ribbons (red) superimposed on top of the crystal conformation (gray). GPP and the magnesium ion are shown as licorice as well as 1,6-DHN, proximal protein side chains (within 3 Å), and four Mg-coordinating solvent molecules.

In contrast to the relatively stable fold, the 16 ns MD simulation sampled a variety of different orientations for the bound 1,6-DHN, which we characterized using distances D_1 and D_2, defined previously. The ranges of D_1 and D_2 sampled during the 16 ns MD simulation are plotted in Figure 8, which can be grouped into three clusters, an indication of three different binding states. We term these clusters/states S1, S2, and S3 in the following discussion. The crystal binding conformation (shorter D_1 and longer D_2) is located in cluster S1. During the MD simulation that we started in cluster S1, the substrate first explored similar conformations that are within the same cluster and then migrated into clusters S2 and S3. Snapshots from the MD simulation that highlight the transitions among these binding states are provided in the Supporting Information. Conformations with a shorter D_2 and a longer D_1 can be found in cluster S3, suggesting that this binding state may lead to the minor prenylated product. Cluster S2 is significantly less populated than clusters S1 and S3 and is possibly an intermediate state connecting S1 and S3.

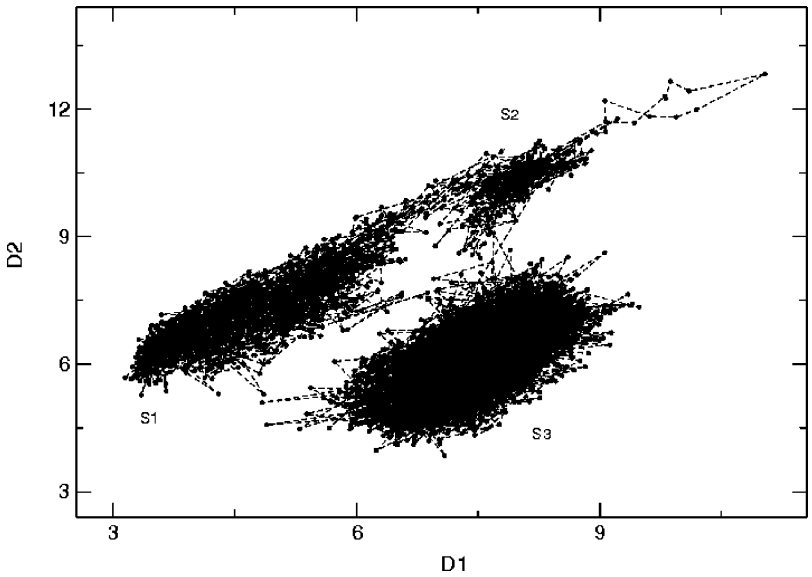

FIGURE 8: Distribution of D_1 and D_2 (in angstroms) sampled during the 16 ns MD simulation of the Orf2 ternary complex.

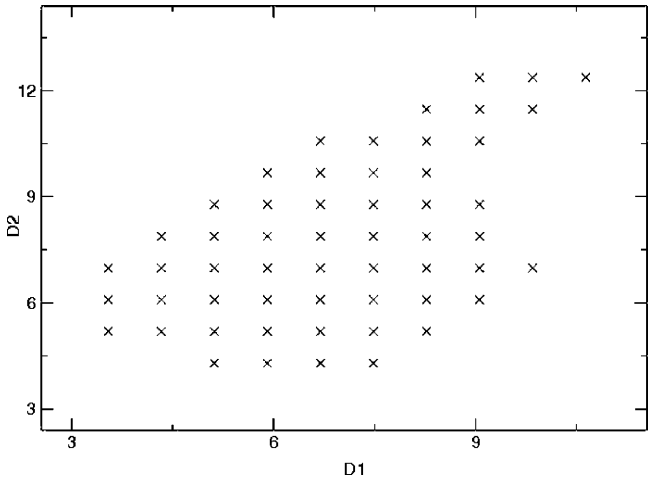

FIGURE 9: Centers of sampling of the initial 55 umbrella MD simulations.

Relative Free Energy Surface of 1,6-DHN Binding. To have a quantitative appreciation of these three binding states and the transitions among them, we computed the two-dimensional potential of mean force as a function of D_1 and D_2 with the umbrella sampling technique followed by a WHAM analysis. Calculating free energies of conformational changes is generally difficult in biological systems due to the fact that the reaction coordinates of conformational changes are usually complicated and are not known a priori. In the case described here, we were fortunate to have observed a trajectory from which the distribution of the substrate was retrieved and adequately characterized by two independent variables, D_1 and D_2. However, the need to use two variables to discriminate different conformational states of the substrate definitely increases computational costs. The initial sampling consisted of 55 umbrella MD simulations carried out on a regular two-dimensional grid with a grid spacing of 0.7 and 0.8 Å in D_1 and D_2, respectively (Figure 9). However, we found that the overlap of the D_1 and D_2 distributions of adjacent simulations was not sufficient (data not shown) to yield a smooth PMF surface with low statistical uncertainty. One simple solution is to increase the sampling resolution by using a finer grid (*45*). We first doubled the amount of sampling by placing additional umbrella points at the centers of the grid cells in Figure 9. This greatly reduced the statistical error in the areas of low free energies. To reduce the errors in the barrier regions and regions near the boundaries of the binding site, particularly in the region that may lead to the minor prenylated product, we further doubled the amount of sampling by placing umbrella points at the midpoints of the edges of the grid cells. Altogether, more than 200 umbrella MD simulations were carried out for the final PMF calculation and more than 100 ns of sampling was used in the WHAM analysis.

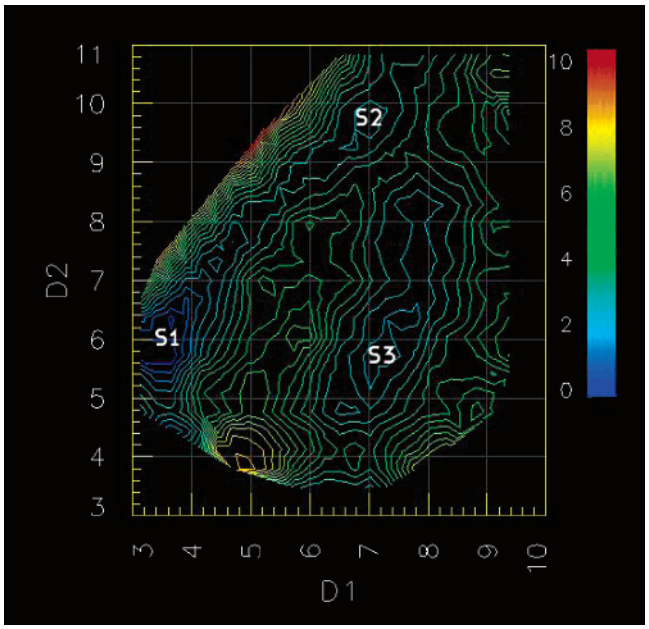

FIGURE 10: Potential of mean force (kilocalories per mole) as a function of D_1 and D_2 (angstroms) shown as a 20-level contour plot and colored by the free energy values.

The computed PMF surface is shown in Figure 10 as a 20-level contour plot. The statistical errors are less than 0.5 kcal/mol in most of the regions except the bottom portion of cluster S3 where D_2 is ~3.5 Å or less. The previously proposed binding states, S1, S2, and S3, are clearly outlined in the contour plot by well-defined free energy minima. S1,

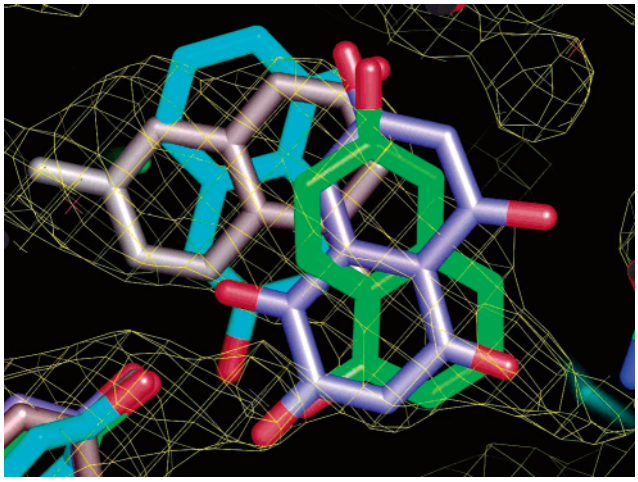

FIGURE 11: Snapshots of the S2 (green) and S3 (cyan) binding states predicted from the calculated potential of mean force superimposed with 1ZB6 (gray) and 1ZDW (purple). The electron density contour of 1ZB6 at the level of 0.5σ is shown as a yellow mesh.

the most stable binding state, is approximately 2.3 kcal/mol more favorable than S2 and S3. S1 and S2 are directly connected by a small energy barrier, which suggests S2 should be readily accessible from S1, consistent with the frequent transitions between them observed in the 16 ns MD simulation. It may be possible to use a single variable to describe both S1 and S2 as the path connecting them is almost one-dimensional. S1 and S3 are separated by a much larger and higher barrier, and the probability of direct barrier crossing (the shortest path) is very low. Hence, the low-energy transition pathways between S1 and S3 must proceed through S2.

The MD snapshots corresponding to the S2 and S3 binding states were randomly picked from two umbrella points whose centers were the closest to the lowest points in the two free energy basins and were superimposed with crystal structure 1ZB6 and the low-electron density contour (Figure 11). In both snapshots, the naphthalene ring fits nicely into the volume enclosed by the electron density contour. Both S2 and S3 are stabilized by forming hydrogen bonds with the side chains of Ser214 and Tyr288. In crystal structure 1ZDW, the conformation of flaviolin is similar to that of S2 (Figure 11) and is also hydrogen bonded to the side chains of Ser214 and Tyr288. The free energy basin where S3 is located is rather wide and flat in the direction of D_2. Conformations with an even shorter D_2 [near (7.0, 3.5)] are only 0.4 kcal/mol higher in free energy compared to S3 and can be approached by crossing a small barrier (approximately 2.0 kcal/mol relative to S3) from S3, which may lead to the formation of the minor prenylated 1,6-DHN product. Multiple barrier crossing events captured during the 16 ns MD simulation revealed that the S1 ↔ S2 and S2 ↔ S3 transitions involve primarily the rotation of 1,6-DHN around the normal of the naphthalene plane accompanied by the translocation of the center of the molecular mass. MD snapshots that highlight these transitions are provided in the Supporting Information.

CNS and QM/MM X-ray Refinement Calculations. To establish the best representation of the alternative binding conformations and critically assess the credibility of our potential of mean force study, the two selected MD snapshots were subjected to experimental validation against the observed X-ray diffraction data using a refinement protocol supplemented with a hybrid QM/MM energy function. The reported R and R_{free} factors for 1ZB6 are 0.233 and 0.263, respectively. The calculated R and R_{free} factors are compiled in Table 2 for three X-ray weighting factors. Overall, the two-stage QM/MM refinement protocol produced structures with comparable agreement with the X-ray diffraction data at high X-ray weighting factors. Due to the nature of the maximum likelihood refinement (MLR) algorithm, the agreement was weakened progressively as smaller weighting factors were used. However, we noted a remarkable difference between the two refinement methods in that the refinements using a more physical and realistic QM/MM energy function display less dependence on the input from the X-ray diffraction measurements. The three different conformations of the ternary complex are of similar quality, even though S1 appears to have the best agreement (lowest R and R_{free} values) followed by S3 and S2. Therefore, the alternative conformations identified in the PMF calculation are both acceptable models for the Orf2 ternary complex. Distances D_1 and D_2 in the CNS and QM/MM refined conformations were measured and are given in Table 2. In general, all three refined conformations were shifted slightly away from the locations of the free energy minima on the PMF surface (Figure 10). This is not surprising because the potential of mean force calculation was carried out on a solvated Orf2 ternary complex with the real substrates at 300 K, in which only the AMBER molecular mechanical force field was used to describe the interactions in the system, while the refined conformations correspond to local minima on the combined potential energy landscape that are shaped by a physical energy term and restraints from the X-ray diffraction measurements. The difference in D_1 and D_2 between the two refinement methods is significant at smaller X-ray weighting factors but gradually diminishes as larger weights are used. Although the detailed procedure and setup of the original CNS refinement of 1ZB6 were not reported, we estimated that an X-ray weighting factor of 1.5 is necessary to achieve the optimized R and R_{free} values. The fact that our QM/MM combined refinement method is less dependent on the diffraction data is potentially an important

Table 2: CNS and QM/MM X-ray Refinement of S1, S2, and S3 Conformers

conformer	refinement protocol	X-ray weight	R	R_{free}	D_1 (Å)	D_2 (Å)
S1	QM/MM	0.01	0.2540	0.2674	3.96	7.09
		0.2	0.2419	0.2629	3.97	7.12
		1.0	0.2290	0.2628	4.01	7.21
	CNS	0.01	0.3735	0.4015	5.03	8.17
S2		0.2	0.2606	0.3004	4.53	7.71
		1.0	0.2307	0.2754	4.10	7.17
	QM/MM	0.01	0.2604	0.2894	6.89	9.82
		0.2	0.2432	0.2798	6.78	9.73
		1.0	0.2285	0.2734	6.80	9.79
	CNS	0.01	0.3690	0.4021	8.15	10.48
		0.2	0.2617	0.3015	7.53	10.28
		1.0	0.2320	0.2763	7.10	10.11
S3	QM/MM	0.01	0.2496	0.2795	5.91	4.04
		0.2	0.2414	0.2749	5.81	3.96
		1.0	0.2283	0.2699	5.87	3.97
	CNS	0.01	0.3709	0.4018	7.36	5.27
		0.2	0.2642	0.3057	6.95	4.56
		1.0	0.2315	0.2777	6.42	4.20

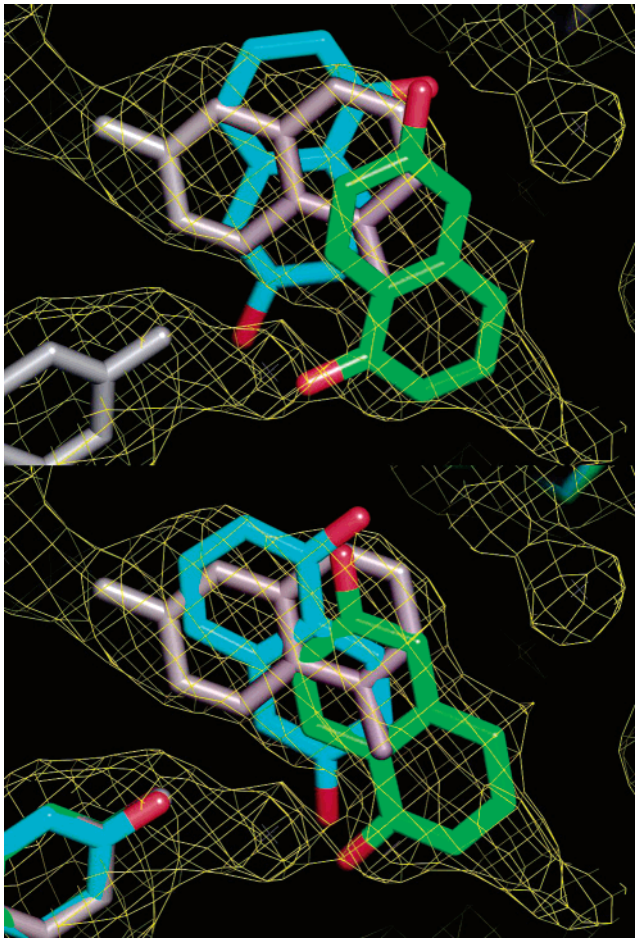

FIGURE 12: Snapshots of the S2 (green) and S3 (cyan) binding states from the PMF calculation (top) and results from the QM/MM X-ray optimization (bottom) superimposed on top of crystal structure 1ZB6 (gray) and the electron density contour at 0.5σ.

advantage over traditional methods based on simple energetic descriptions, which may be particularly well suited in applications such as evaluating possible side chain protonation states and hydrogen bonding possibilities at critical locations in the system of interest. The final conformations of the QM/MM refined S2 and S3 binding states are shown in Figure 12

CONCLUSIONS

Unlike most enzymes that are known for being efficient, highly specific, and highly selective biological catalysts, Orf2 can be considered as a "reaction chamber" for many small aromatic substrates, displaying not only reduced substrate selectivity but also interesting regioselectivity of the prenylated products. The notion of a reaction chamber gives rise to the opportunity to engineer the active site of Orf2 to adjust the functionality of terpenoids for potentially novel therapeutic applications. The underlying three-dimensional framework was revealed in a set of recently determined crystal structures; however, in this study, the mechanism of substrate selectivity and product regioselectivity in Orf2 prenylation was further elucidated. We thoroughly explored the binding site of Orf2 and quantitatively evaluated the relative free energies of several binding states of 1,6-DHN in terms of a two-dimensional potential of mean force. The deposited substrate conformation corresponds to the most stable binding state, but the others are certainly accessible at 300 K. Given the moderate free energy barriers that separate them, it would not be unreasonable to assume that the substrate binds in a single conformation and rapidly equilibrates S1 and S3 (possibly via S2). Consequently, the observed substrate electron density is the thermodynamic average of the contributions of all these binding states. The binding site of Orf2 is narrow in one dimension, which is consistent with the observed substrate preference, but spacious in the others to admit aromatic molecules of different sizes, possibly up to three or four fused rings. The conformations of the other two binding states that we identified and validated against experimental X-ray diffraction data not only explained the electron density profile at low contour values and the high *B*-factors of the substrate but also offered a plausible interpretation of the regioselectivity in the prenylated products. The product yields are dependent upon the thermodynamics of binding, the incubation time, and the relative reactivities. The thermodynamics of substrate binding is determined by the interplay of substrate−protein, protein−solvent, and substrate−solvent interactions. For 1,6-DHN, a 10:1 product ratio (S1 product to S3 product) was observed when the concentration of the substrate was ~160-fold in excess of that of the enzyme and both were incubated at 298 K for 4 h. We estimate that the concentration ratio of the two competing conformations is close to 100:1 from a 2.7 kcal/mol free energy difference between S1 and S3, assuming a constant concentration of the ternary complex. On the basis of the Curtin-Hammett principle, we speculate that the prenylation of S3 is a faster reaction with a lower reaction barrier height relative to that of S1. Ser214 and Tyr288 are likely the moderators of substrate binding, the mutations of which may affect the selectivity and regioselectivity, and therefore are good candidates for site-directed mutagenesis experiments. Flaviolin is bound similarly to the S2 state of 1,6-DHN, also hydrogen bonded to the side chains of Ser214 and Tyr288. However, the electron density of 1ZDW is better defined around flaviolin, indicating a single preferred conformation. This is possibly due to the fact that an extra hydrogen bond can be formed between the carbonyl group of flaviolin and Tyr288, which favors this orientation and reduces the degree of conformational fluctuation.

Detailed and reliable analysis of substrate and receptor interactions is crucial in many aspects of current enzyme research, such as reaction mechanism studies, protein active site engineering, and drug development. The approaches that we used to identify and evaluate different states of binding are quantitative, robust, and computationally efficient. We expect that computational approaches of a similar nature will become a valuable component of the tool set for computer-aided rational molecule design.

SUPPORTING INFORMATION AVAILABLE

Computed dynamics of 1,6-DHN within the Orf2 binding site and the transitions among three binding states made into an animation for better apprehension. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Spurgeon, S. L., and Porter, J. W. (1981) in *Biosynthesis of Isoprenoid Compounds* (Porter, J. W., and Spurgeon, S. L., Eds.) pp 1, John Wiley and Sons, New York.

2. Qureshi, N., and Spurgeon, S. L. (1981) in *Biosynthesis of Isoprenoid Compounds* (Porter, J. W., and Spurgeon, S. L., Eds.) pp 47, John Wiley and Sons, New York.
3. Rohmer, M. (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants, *Nat. Prod. Rep. 16*, 565−574.
4. Piironen, V., Lindsay, D., Miettinen, T., Toivo, J., and Lampi, A. (2000) Plant sterols: Biosynthesis, biological function and their importance to human nutrition, *J. Sci. Food Agric. 80*, 939−966.
5. Santos, F. A., and Rao, V. S. N. (1998) Inflammatory edema induced by 1,8-cineole in the hindpaw of rats: A model for screening antiallergic and anti-inflammatory compounds, *Phytomedicine 5*, 115−119.
6. Blanco-Colio, L. M., Tunon, J., Martin-Ventura, J. L., and Egido, J. (2003) Anti-inflammatory and immunomodulatory effects of statins, *Kidney Int. 63*, 12−23.
7. Grosser, N., Hemmerle, A., Berndt, G., Erdmann, K., Hinkelmann, U., Schurger, S., Wijayanti, N., Immenschuh, S., and Schroder, H. (2004) The antioxidant defense protein heme oxygenase 1 is a novel target for statins in endothelial cells, *Free Radical Biol. Med. 37*, 2064−2071.
8. Chowdhury, S. A., Kishino, K., Satoh, R., Hashimoto, K., Kikuchi, H., Nishikawa, H., Shirataka, Y., and Sakagami, H. (2005) Tumor-specificity and apoptosis-inducing activity of stilbenes and flavonoids, *Anticancer Res. 25*, 2055−2063.
9. Jahangir, T., Khan, T. H., Prasad, L., and Sultana, S. (2005) Alleviation of free radical mediated oxidative and genotoxic effects of cadmium by farnesol in Swiss albino mice, *Redox Rep. 10*, 303−310.
10. Soria-Mercado, I. E., Prieto-Davo, A., Jensen, P. R., and Fenical, W. (2005) Antibiotic terpenoid chloro-dihydroquinones from a new marine actinomycete, *J. Nat. Prod. 68*, 904−910.
11. Zhou, Y. D., Kim, Y. P., Mohammed, K. A., Jones, D. K., Muhammad, I., Dunbar, D. C., and Nagle, D. G. (2005) Terpenoid tetrahydroisoquinoline alkaloids emetine, klugine, and isocephaeline inhibit the activation of hypoxia-inducible factor-1 in breast tumor cells, *J. Nat. Prod. 68*, 947−950.
12. Boucher, K., Chad, S. S., Sharma, P., Hauschka, P. V., and Solomon, K. R. (2006) HMG-CoA reductase inhibitors induce apoptosis in pericytes, *Microvasc. Res. 71*, 91−102.
13. Hwang, D. R., Chang, C. W., Lien, T. W., Chen, W. C., Tan, U. K., Hsu, J. T. A., and Hsieh, H. P. (2006) Synthesis and antiviral activity of a series of sesquiterpene lactones and analogues in the subgenomic HCV replicon system, *Bioorg. Med. Chem. 14*, 83−91.
14. Jahangir, T., Khan, T. H., Prasad, L., and Sultana, S. (2006) Farnesol prevents Fe-NTA-mediated renal oxidative stress and early tumour promotion markers in rats, *Hum. Exp. Toxicol. 25*, 235−242.
15. Christianson, D. W. (2006) Structural Biology and Chemistry of the Terpenoid Cyclases, *Chem. Rev. 106*, 3412−3442.
16. Kuzuyama, T., Noel, J. P., and Richard, S. B. (2005) Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products, *Nature 435*, 983−987.
17. Botta, B., Delle Monache, G., Menendez, P., and Boffi, A. (2005) Novel prenyltransferase enzymes as a tool for flavonoid prenylation, *Trends Pharmacol. Sci. 26*, 606−608.
18. Koehl, P. (2005) Relaxed specificity in aromatic prenyltransferases, *Nat. Chem. Biol. 1*, 71−72.
19. Taylor, J. S., Reid, T. S., Terry, K. L., Casey, P. J., and Beese, L. S. (2003) Structure of mammalian protein geranylgeranyltransferase type-I, *EMBO J. 22*, 5963−5974.
20. Zhang, H., Seabra, M. C., and Deisenhofer, J. (2000) Crystal structure of Rab geranylgeranyltransferase at 2.0 angstrom resolution, *Struct. Folding Des. 8*, 241−251.
21. Park, H. W., Boduluri, S. R., Moomaw, J. F., Casey, P. J., and Beese, L. S. (1997) Crystal structure of protein farnesyltransferase at 2.25 angstrom resolution, *Science 275*, 1800−1804.
22. Kleywegt, G., Harris, M., Zou, J., Taylor, T., Wahlby, A., and Jones, T. (2004) The Uppsala Electron-Density Server, *Acta Crystallogr. D60*, 2240−2249.
23. Yu, N., Hayik, S. A., Wang, B., Liao, N., Reynolds, C. H., and Merz, K. M., Jr. (2006) Assigning the protonation states of the key aspartates in β-secretase using QM/MM X-ray structure refinement, *J. Chem. Theory Comput. 2*, 1057−1069.
24. Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., et al. (2004) *AMBER 8*.
25. Frisch, M. J., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2003) *Gaussian'03*, Gaussian Inc., Wallingford, CT.
26. Wang, J. M., Cieplak, P., and Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem. 21*, 1049−1074.
27. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general AMBER force field, *J. Comput. Chem. 25*, 1157−1174.
28. Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995) Application of the Multimolecule and Multiconformational Resp Methodology to Biopolymers-Charge Derivation for DNA, RNA, and Proteins, *J. Comput. Chem. 16*, 1357−1377.
29. Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges-the Resp Model, *J. Phys. Chem. 97*, 10269−10280.
30. Cui, G., Wang, B., and Merz, K. (2005) Computational studies of the farnesyltransferase ternary complex-Part I: Substrate binding, *Biochemistry 44*, 16513−16523.
31. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) *J. Chem. Phys. 79*, 926−935.
32. Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald-an N.Log(N) Method for Ewald Sums in Large Systems, *J. Chem. Phys. 98*, 10089−10092.
33. Allen, M. P., and Tildesley, D. J. (1987) *Computer Simulation of Liquids*, Clarendon Press, Oxford, U.K.
34. Daan Frenkel, B. S. (2002) *Understanding Molecular Simulation From Algorithms to Applications*, 2nd ed., Vol. 1, Academic Press, New York.
35. Kumar, S., Bouzida, D., Swendsen, R., Kollman, P., and Rosenberg, J. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. 1. the method, *J. Comput. Chem. 13*, 1011−1021.
36. Brunger, A. T. P. D. A., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D*, 905−921.
37. Engh, R., and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein-structure refinement, *Acta Crystallogr. A47*, 392−400.
38. Yang, W. (1991) Direct calculations of electron-density in density-functional theory, *Phys. Rev. Lett. 66*, 1438−1441.
39. Yang, W., and Lee, T. (1995) A density-matrix divide-and-conquor approach for electronic-structure calculations of large molecules, *J. Chem. Phys. 103*, 5674−5678.
40. Dixon, S. L., and Merz, K. M. (1996) Semiempirical molecular orbital calculations with linear system size scaling, *J. Chem. Phys. 104*, 6643−6649.
41. Lee, T., York, D., and Yang, W. (1996) Linear-scaling semiempirical quantum calculations for macromolecules, *J. Chem. Phys. 105*, 2744−2750.
42. Dixon, S. L., and Merz, K. M. (1997) Fast, accurate semiempirical molecular orbital calculations for macromolecules, *J. Chem. Phys. 107*, 879−893.
43. Stewart, J. (1989) Optimization of parameters for semiempirical methods. 1: Method, *J. Comput. Chem. 10*, 209−220.
44. Stewart, J. (1989) Optimization of parameters for semiempirical methods. 2: Applications, *J. Comput. Chem. 10*, 221−264.
45. Roux, B. (1995) The calculation of the potential of mean force using computer simulations, *Comput. Phys. Commun. 91*, 275−282.

BI062076Z