Probing the Structures of Leishmanial Farnesyl Pyrophosphate Synthases: Homology Modeling and Docking Studies

Prasenjit Mukherjee,[‡] Prashant V. Desai,[†,‡] Anuradha Srivastava,[§] Babu L. Tekwani,[§] and Mitchell A. Avery*,[‡,§,‖]

Department of Medicinal Chemistry, School of Pharmacy, National Center for Natural Products Research, and Department of Chemistry and Biochemistry, University of Mississippi, University, Mississippi 38677

Leishmania donovani and *Leishmania major* farnesyl pyrophosphate synthase (*Ld*FPPS and *Lm*FPPS) are potential targets for the development of antileishmanial therapy. The protein sequence for *Ld*FPPS was recently elucidated in our laboratory. Highly refined homology models were generated using the protein sequences of *Ld*FPPS and the closely related *Lm*FPPS enzyme. A ligand-refined model of *Lm*FPPS with a bound bisphosphonate ligand was generated using restraint-guided molecular mechanics followed by quantum mechanics/molecular mechanics refinement. The ligand-refined model of *Lm*FPPS was further validated through extensive pose validation, enrichment, and other docking studies involving known bisphosphonate inhibitors. The model was able to explain the critical binding site interactions and site-directed mutagenesis data obtained from experimental studies on related FPPS enzymes. The ligand-refined model in conjunction with the validated docking protocol could be utilized in the future for structure-based virtual screening and rational drug design studies against these targets.

INTRODUCTION

Leishmaniasis, a collective group of parasitic diseases caused by the *Leishmania* species is prevalent in over 88 countries worldwide and exposes around 350 million people annually to the risks of infection.[1] Visceral leishmaniasis or Kala Azar, found in tropical countries, is the most lethal form of the disease and is caused by *Leishmania donovani*. If untreated, visceral leishmaniasis may cause mortality rates as high as 90% within 4 to 24 months of the infection.[2] Several epidemic outbreaks of visceral leishmaniasis have been reported over the past 10 years,[3] with the majority of them occurring in the tropical countries such as Sudan, Brazil, India, Bangladesh, and Afghanistan. Cutaneous and mucocutaneous leishmaniases are the other variants of this disease and are caused by *Leishmania major* and *Leishmania mexicana*, respectively. An estimated 2 million new infections are caused every year due to cutaneous and visceral leishmaniases.[1] A newer cause of concern against these diseases is their development as opportunistic infections in immunocompromised patients, such as those suffering from AIDS.

The pentavalent antimonials[4–6] such as Pentosam (sodium stibogluconate) and Glucatime (meglumine antimoniate) have been widely used as first-line therapies for leishmanial infections. These molecules show a long history of usage and have a high degree of efficacy. The two primary concerns against their usage are their high levels of toxicity and

development of drug resistance, which has been reported to be as high as 40% in certain geographical areas.[6] Other less toxic but expensive alternatives include a liposomal formulation of the antibiotic amphoterecinB (AmBisome),[7,8] Miltefosine (hexadecylphosphocholine),[8] and Flucanazole.[9,10] The development of novel antileishmanial therapies with improved toxicity and resistance profiles are therefore of imminent importance. A majority of the world population susceptible to leishmaniasis resides in the less developed countries. It is therefore essential that the novel antileishmanial therapies be highly cost-effective and made accessible to patients with limited financial means.

Bisphosphonates[10,11] (Figure 1a) are an important class of drugs with proven efficacy in the treatment of various human bone-related diseases such as osteoporosis,[12] metastatic bone disease,[13] hypocalcaemia,[14] and Paget's disease.[15] The mode of action of these bisphosphonate drugs[16–18] involves the inhibition of the human farnesyl pyrophosphate synthase (FPPS), an enzyme from the mevalonic acid pathway[19,20] present in the osteoclasts and macrophages of the bone tissue. Recently, some of these bisphosphonates have been found to exhibit antiparasitic activity. Risendronate (Actonel), alendronate (Fosamax), and pamidronate (Aredia) have shown *in vitro* cell-based activity against a range of protozoal parasites including *Trypanosoma cruzi*, *Trypanosoma brucei*, *Leishmania donovani*, and *Leishmania major*. A parasitological cure of visceral and cutaneous leishmaniasis caused by *L. donovani* and *L. major* has been documented in the mouse model of the disease.[21,22] The antiparasitic activity of these compounds has been attributed to the inhibition of the FPPS enzyme of the protozoal mevalonic acid pathway. On the basis of these studies, inhibition of the parasitic FPPS can be considered as a viable target for development of novel antileishmanial therapies.

* Corresponding author phone: 662-915-5879, fax: 662-915-5638, e-mail: mavery@olemiss.edu.
† Current address: Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, Indiana 46285.
‡ Department of Medicinal Chemistry, School of Pharmacy.
§ National Center for Natural Products Research.
‖ Department of Chemistry and Biochemistry.

Leishmanial Farnesyl Pyrophosphate Synthases

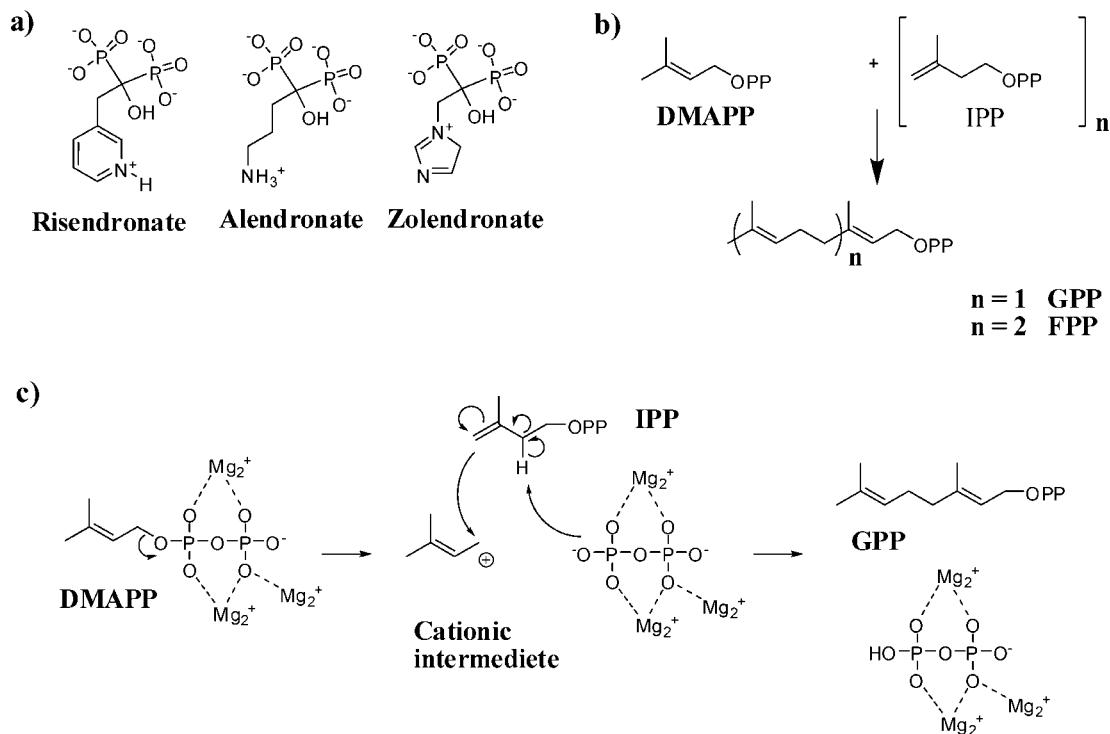*J. Chem. Inf. Model.*, *Vol. 48, No. 5, 2008* **1027**

Figure 1. (a) Bisphosphonate drugs, which are currently marketed as therapies for bone-related diseases. (b) The FPPS enzyme catalyzes the condensation of two molecules of IPP with one molecule of DMAPP to produce FPP as the final product. (c) A schematic showing the magnesium-mediated reaction scheme of the FPPS enzyme.

The mevalonic acid pathway is ubiquitously present in all animals, ranging from mammals and extending to bacteria. This pathway is responsible for the synthesis of various sesquiterpenes or isoprenoids, built of three isoprene units. The sesquiterpenes generated through the mevalonic acid pathway function as precursors for various other biosynthetic pathways, leading to the formation of dolichols, ubiquinones, prenylated proteins, and various sterols such as cholesterol. The prenyl transferase, FPPS, is a critical enzyme in this pathway and catalyzes the condensation of one molecule of dimethyl allyl pyrophosphate (DMAPP) with two molecules of isopentenyl pyrophosphate (IPP) to form farnesyl pyrophosphate (FPP) (Figure 1b). The reaction occurs in a two-step process, wherein in the first step one molecule of DMAPP (C_5 isoprene) is condensed with one molecule of IPP (C_5 isoprene) to generate geranyl pyrophosphate (GPP), a C_{10} isoprenoid. In the second step, geranyl pyrophosphate is condensed with another unit of IPP to generate FPP, a C_{15} isoprenoid. The enzyme has two binding sites—the DMAPP binding site binds one molecule of DMAPP, GPP, or FPP during the various stages of the catalytic process, while the IPP binding site binds a molecule of IPP. The enzyme's catalytic activity[23] (Figure 1c) requires the presence of magnesium ions, which coordinate to the oxygen atoms of DMAPP's pyrophosphate group. The bisphosphonates bind to the DMAPP binding site and mimic the interactions undergone by the native substrate DMAPP, thereby acting as competitive inhibitors of this site.

The three-dimensional protein structure of the leishmanial FPPS would therefore be an ideal starting point for the design and development of novel chemical leads against leishmaniasis. Crystal structures of FPPS from various species ranging from humans to *Escherichia coli* have been solved both in the apo form as well as with cocrystallized ligands.

The structure of a leishmanial FPPS is, however, yet to be reported, and in its absence, homology models[24–27] of these enzymes built from crystal structures of close structural homologues would serve as ideal starting points for drug discovery. Through a collaborative effort, the protein sequence of the *Ld*FPPS was recently solved in our laboratory. Utilizing the newly elucidated sequence of *Ld*FPPS and a previously deposited protein sequence of *Lm*FPPS, homology models of these two proteins were built. A ligand-refined model of the *Lm*FPPS enzyme containing a bound bisphosphonate inhibitor risendronate (RIS) (Figure 1a) and IPP was generated. Quantum mechanical/molecular mechanical (QM/MM) refinement is an ideal tool for modeling the geometries of enzyme active sites. However, for the QM/MM calculations to converge, the starting protein structure should have reasonable geometry and preferably in a stable local minimum. A restraint-guided molecular mechanics refinement followed by QM/MM refinement was therefore utilized to generate the ligand-refined model. The predictive ability of the model was further validated through extensive docking studies involving multiple bisphosphonate ligands active against the *Lm*FPPS. The study provides insights into the requirements for the ligand binding process as well as structural features of the ligands responsible for biological activity against this enzyme. The ligand-refined model and the validated docking formalism could be utilized in structure-based virtual screening and rational drug design efforts against these targets.

METHODS

Computational Resources. The computational studies were carried out on an 8 processor SGI Origin 350 server equipped with R16000 chipsets and 2 GB of memory.

Sequence alignment was achieved using clustalw.[28] Homology modeling was carried out using Modeller 7.0,[29] applying the CHARMM v.22 force field through its interface with InsightII (Accelrys Inc., San Diego, CA). Geometric and local area consistency of the models were evaluated using Procheck[30] and Profiles-3D[31] (InsightII) as well as the Matchmaker[32] (Sybyl 6.9; Tripos Inc., St. Louis, MO) and Ramachandran plot analysis.[30] Molecular mechanics refinement of the protein structures was carried out within the Macromodel 9.1 (Schrödinger, LLC, Portland, OR) molecular mechanics package, while the QM/MM refinement was carried out using Qsite 4.0[33,34] (Schrodinger, LLC, Portland, OR). GOLD 3.1[35,36] (CCDC, Cambridge, U.K.) and the Cscore[37] module of Sybyl 6.9 were utilized for the docking studies. Graphics were generated using Pymol[38] and Sybyl 6.9.

Homology Modeling. The protein sequence of *Ld*FPPS (GenBank accession no. Q0GKD7) was generated from the nucleotide sequence of the *Ld*FPPS gene, which was cloned and characterized in our laboratory. The protein sequence of *Lm*FPPS (GenBank accession no. Q4QBL1) was retrieved from the SWISS-PROT and TrEMBL database of the ExPASy Molecular Biology Server.[39] The target sequences are 97% identical to each other with nine residues differing between the two. The sequences were submitted to NCBI-BLAST,[40] and a search of the Protein Data Bank was carried out to identify close structural homologues for which crystal structures were available. The closest sequence homologues for *Ld*FPPS and *Lm*FPPS for which crystal structures are available were the FPPS from *T. cruzi* (63% identity),[41] *T. brucei* (63% identity),[42] human (32% identity),[43] avian (chicken; 30% identity),[44] *E. coli* (15% identity),[45] *Staphylococcus aureus* (15% identity), and *Shigella flexneri* (14% identity), with the first four being the more logical choices.

Since one of our aims was to model an enzyme with ligands bound to both the substrate sites, we considered crystal structures showing a similar occupancy pattern. The crystal structures from chicken and *Tb*FPPS were not our first choice since neither of them showed occupancy of the IPP binding site. The structure from the *Hu*FPPS (2F8Z, PDB code) structure could be a good reference since it had occupancy for both the sites. However, there were two noticeable differences between the *Tc*FPPS and *Hu*FPPS structures. FPPS enzymes from across the species show a conserved BXB motif in the C terminus consisting of two basic residues. In the apo enzyme, the C terminus is solvent-exposed, but in the case of cocrystal structures with the ligand bound to the IPP site, this segment appears to fold back into the site, forming interactions with the ligand pyrophosphate group. The orientation of the BXB motif in case of the *Hu*FPPS structure is slightly different from that of the *Tc*FPPS structure. In *Tc*FPPS, both the basic residues, Arg360 and Lys362, form side-chain interactions with the pyrophosphate moiety of the ligand. On the other hand, in *Hu*FPPS, only the first basic residue of the BXB motif corresponding to Arg360 (*Tc*FPPS numbering) undergoes interaction with the ligand, while the second residue corresponding to Lys362 (*Tc*FPPS numbering) points away from the binding site. Another basic residue corresponding to Lys48 (*Tc*FPPS numbering) changes its rotameric state (compared to the *Tc*FPPS structure) and substitutes for the missing interaction of the second basic residue of the BXB

motif. In the present case, the orientation of the Lys362 residue from the *Tc*FPPS ligand-bound structure was considered appropriate since it is phylogenetically more related to the target sequence. Secondly, there is an 11-member loop region present in protozoal FPPS's (also present in the generated models) but absent in the higher-order structures from human and chicken. The presence of this loop in the case of *Ld*FPPS and *Lm*FPPS sequences further supported the use of the *Tc*FPPS structure for model building.

Two ligand-bound crystal structures of *Tc*FPPS,[41] 1YHL and 1YHM (PDB codes), have been elucidated. In 1YHL, RIS is bound to the DMAPP binding site while a molecule of DMAPP is bound to the IPP binding site. The authors have used a high concentration of DMAPP in the crystallization process and suggest this as the likely reason for this atypical binding. The second structure, 1YHM, has another bisphosphonate, alendronate, bound to the DMAPP site while a molecule of IPP is bound to the IPP binding site. The application of either crystal structure in model building has certain advantages and disadvantages. The bisphosphonate ligands bound to the DMAPP site in both the crystal structures show similar spatial orientation as well as interaction patterns. In the 1YHL structure, the atypical binding of DMAPP to the IPP binding sites leads to altered rotameric states of certain IPP site residues compared to the 1YHM structure. While the 1YHM structure is ideal for modeling the IPP binding site, the coordinates for some of the residues, including some residues within the conserved DDXXD motifs of the DMAPP site and water molecules involved in coordination geometry, are missing.

Therefore, to evaluate the suitability of either structure, two model sets were generated for *Ld*FPPS and *Lm*FPPS each, using (i) 1YHL only—model set 1 or (ii) 1YHM only—model set 2 as the reference structure. Protein sequence alignment (Figure 2) was carried out in clustalw applying the BLOSUM 62 substitution matrix. Model building was carried out using Modeller 7.0 through its interface with InsightII. Within Modeller, the sequence alignment is used to generate a series of geometrical constraints for the target sequence. The constraints are then converted into probability density functions (PDFs) and included in an objective function. The function is iteratively refined through a process of conjugate gradient minimization and simulated annealing techniques to generate the final model. Overall, PDF violations and protein structure check tools such as Procheck, Profiles-3D, Matchmaker, and Ramachandran plot analysis were utilized in the model selection (Table 1). A total of 10 models of "high" quality (refinement using simulated annealing) were generated for each model set. In the case of both FPPSs, the best-ranked model (based on PDF violations) from set 1 had a slightly better overall structure check profile as compared to the top-ranking model from set 2 and was selected for further refinement. In the case of *Lm*FPPS, overall structure check results for the top-ranked model from each model set were comparable. Unlike the top-ranked model from model set 1, the top-ranked model from model set 2 had different rotameric states for the binding site residues Asp99 and Arg107 as compared to the crystal structure, wherein they form a stabilizing salt-bridge interaction. Since the modeling of the DMAPP site was of primary importance, the best protein model from set 1 (modeled on the basis of 1YHL) was selected for further refinement.
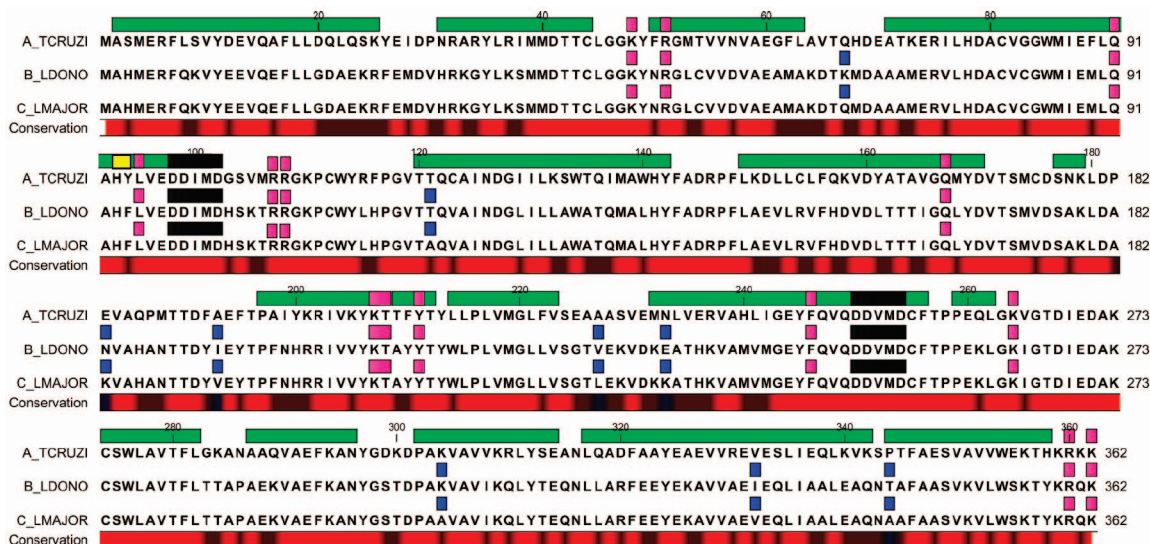
LEISHMANIAL FARNESYL PYROPHOSPHATE SYNTHASES

J. Chem. Inf. Model., Vol. 48, No. 5, 2008 **1029**

Figure 2. Sequence alignment of *Tc*FPPS, *Ld*FPPS, and *Lm*FPPS. The residues marked in black represent the two DDXXD motifs, the first aspartate-rich motif (FARM) towards the N terminus and the second aspartate-rich motif (SARM) towards the C terminus. The residues marked in magenta represent binding site residues other than the DDXXD motifs. The three-residue BXB motif is located on the C terminus and consists of two basic residues which form a part of the binding site. The fourth and the fifth residues upstream of the FARM are the chain-length regulation residues (marked in yellow). In *Tc*FPPS, it consists of a His-Tyr pair but is changed to a His-Phe pair in the cases of *Lm*FPPS and *Ld*FPPS. The residues differing between *Ld*FPPS and *Lm*FPPS are marked in blue. The green regions represent the major α-helical domains.

Table 1. Protein Structure Analysis and PDF Violations for the Best Models Obtained from Each Model Set for *Ld*FPPS and *Lm*FPPS

	*Ld*FPPS		*Lm*FPPS	
model set	1	2	1	2
PDF violation[a]	1274.75	1387.76	1265.39	1363.88
$\varphi-\psi$ violations[b]	0.6	0	0.6	0
Procheck[c]	0.6	1.2	0.6	0.6
Profiles-3D[d]	150.01	145.02	140.45	146.63
Matchmaker[e]	−0.16	−0.13	−0.15	−0.12

[a] Obtained from Modeller output. [b] Percentage of residues lying in the disallowed region. [c] Percentage of total residues with a bond length or bond angle violation of >0.05 Å or >10°, respectively, from reference values. [d] The minimum and maximum possible scores for a model are 74.23 and 164.97, respectively. [e] In kT units.

LdFPPS Optimized Potential for Liquid Simulations/Generalized Born Surface Atom (OPLS/GBSA) Refined Model. The best-ranked *Ld*FPPS model from Modeller was further refined using molecular mechanics. Hydrogens were added to the model using the PPREP script (Schrodinger, LLC, Portland, OR). Molecular mechanics refinement was carried out in Macromodel using the OPLS 2001 force field.[46] The GBSA implicit water treatment[47] with extended nonbonded cutoffs (electrostatic 20 Å, vdw 8Å) and a dielectric constant of 1 were utilized for the refinement protocol. We choose the OPLS force field since it has been extensively used for molecular mechanics[48,49] and QM/MM[50,51] simulation of macromolecular systems. This choice also helped us in maintaining continuity in force field usage for the molecular mechanics and QM/MM stages of the refinement protocol. A detailed description of the multistep minimization protocol is provided in the Supporting Information.

LmFPPS QM/MM Ligand-Refined Model. The protocol utilized for the generation of the *Lm*FPPS QM/MM ligand-refined model is shown in Figure 3. The crude ligand complex was generated by copying the coordinates of the key ligand molecules from the *Tc*FPPS structures, 1YHL and 1YHM, after C$_\alpha$-based superimposition with the Modeller-generated protein model. RIS, three Mg^{2+} ions, and eight structural waters which complete the octahedral geometry of the metal centers were copied from 1YHL, while the coordinates of IPP were copied from the 1YHM structure. In accordance with recent experimental work,[42] each phosphate group of the bisphosphonate moiety of RIS was modeled in the dianionic state, while the terminal phosphate group of the IPP ligand was modeled in the monoanionic state. Hydrogen atoms were added using the PPREP utility, maintaining the overall charge neutrality of the system. The acidic residues Asp98, Asp102, and Asp250, which form direct interactions with the three Mg^{2+} ions in the DMAPP binding site, were modeled in the charged state. The basic residues Lys48, Arg51, Arg108, Arg360, and Arg362 in the IPP binding site and the residues Arg107, Arg207, and Lys264 in the DMAPP binding site were modeled with a positive charge. The rotameric state of three residues, Arg51, Arg108, and Lys362, in the IPP binding site were not optimally oriented for interaction with the ligand and were therefore manually modified to form suitable interactions with the ligand in accordance with what is seen in the reference crystal structure. The pre-refined complex, generated in this fashion, was then submitted to molecular mechanics refinement.

Molecular Mechanics Refinement. Molecular mechanics refinement of the protein model was carried out to attain a relaxed protein structure with reasonable active site geometry suitable for QM/MM refinement. The point charge model of metal centers in molecular mechanics often leads to inaccuracies and distortions in the representation of metal coordination geometries. Several approaches[52−54] have been proposed to counter these inaccuracies ranging from the use of geometric constraints to cationic dummy atom approaches incorporating force field-based parametrizations. In our case,
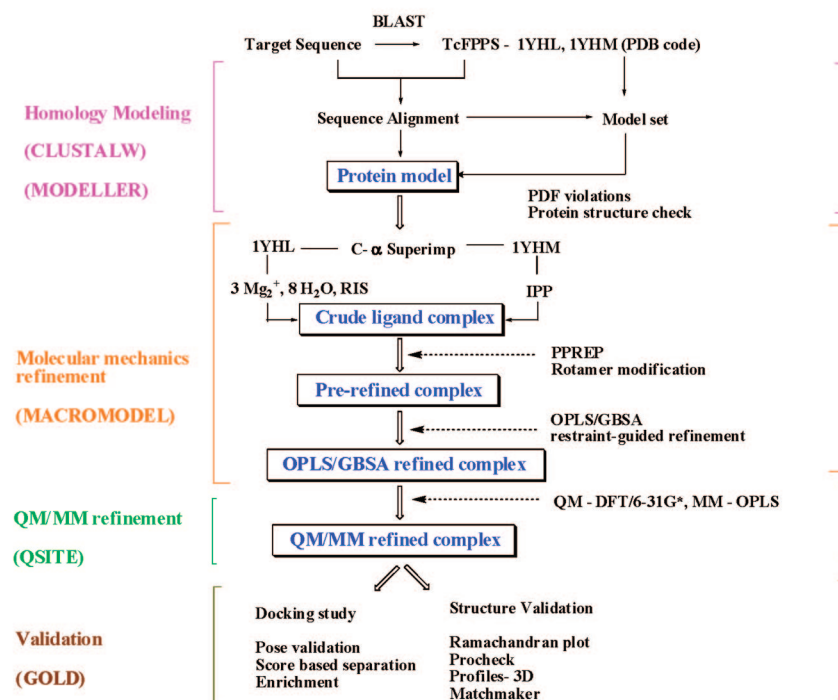
Model refinement and validation methodology

Figure 3. Flowchart depicting the protocols applied for the generation of the *Lm*FPPS QM/MM ligand-refined model. The relevant modeling software or programs utilized at each stage of development are shown on the left outside the square brackets.

Table 2. Definition of Zones Utilized in the Molecular Mechanics Refinement of the *Lm*FPPS QM/MM Ligand-Refined Model

zones	definition
1	RIS, IPP, 8H$_2$O, 3Mg^{2+}
2	residues located within 10 Å of zone 1
3	all residues located outside of zone 2

a distance-restraint-guided refinement was utilized to generate a molecular-mechanics-refined structure with reasonable metal coordination geometry for QM/MM refinement. The OPLS/GBSA potential setup used for the refinement of the *Ld*FPPS model was also used in the refinement of the *Lm*FPPS QM/MM ligand-refined model. The crude ligand complex was divided into three zones (Table 2) and a multistep minimization was carried out. The process of minimization originated from the binding site (zone one) of the protein and was gradually extended to include the full protein. The detailed description for this multistep refinement is provided in the Supporting Information.

A molecular dynamics simulation was carried out for the structure obtained from the restraint-guided minimization to release any local strains introduced by the restrained minimization protocol. The simulation was carried out using the OPLS/GBSA potential setup. A time step of 2 fs was applied, and SHAKE was used to constrain bonds to hydrogen atoms. A total of 23 flat-bottom distance restraints (from the restrained-guided minimization stage) were applied with a weak force constant of 25 kJ/mol·Å^2 and a ±0.2 Å half-width of the potential well to conserve the key binding interaction and metal coordination geometries. The definition of the restraint set is same as that used for the restraint-guided minimization (details provided in the Supporting Information). The residue backbone atoms were also constrained using a very low force constant of 10 kJ/mol·Å^2 to

prevent undue movements, specially the C-terminal region which forms a part of the binding site. The system was equilibrated for 40 ps at 300 K, after which a production run of 200 ps was carried out. The energy profiles of the dynamics snapshots were collected and monitored to see that a stable trajectory had been attained. The final snapshot from the dynamics simulation was extracted and minimized using the OPLS/GBSA potential setup. The minimization protocol involved 5000 steps of Polak Rabiere conjugate gradient with a convergence criterion of 0.01 kJ/mol·Å. The structure was initially minimized with the same positional constraints and distance restraints setup as used in the dynamics simulation. The positional constraints and distance restraints were removed sequentially to obtain the fully relaxed OPLS/GBSA refined complex.

QM/MM Refinement. The QM/MM refinement of the model was carried out in Qsite 4.0, which utilizes the Jaguar ab initio quantum mechanics program for the QM treatment and the Impact molecular mechanics program for the treatment of the MM region. In QM/MM calculations, the computational complexity increases exponentially with the number of atoms involved in the calculation. Hence, one needs to find the best compromise between model size and computational complexity to get the desired results. Since the metal centers in our model are located within the DMAPP binding site, the QM region was defined around this site. RIS, three chelating Mg^{2+} ions, and the eight coordinating water molecules were considered within the QM region. Since the interactions of the binding site residues are through the side chains, the QM/MM interface was defined between the amino acid C$_\alpha$–C$_\beta$ atoms. The acidic side chains Asp98, Asp102, and Asp250 were included in the QM region. The basic residues Arg107, Lys207, and Lys264, which undergo charge−charge interactions with the two phosphate groups

LEISHMANIAL FARNESYL PYROPHOSPHATE SYNTHASES

J. Chem. Inf. Model., Vol. 48, No. 5, 2008 **1031**

of the ligand's bisphosphonate moiety, as well as Thr208 and Tyr211, which interact with the cationic nitrogen and the aromatic ring system, were treated using quantum mechanics. The QM refinement was carried out using density functional theory (DFT) applying the Becke-3−Lee–Yang-–Parr (B3LYP) functional. The electronic structure of the QM region was described using the 6-31G* split-valence basis set. The self-consistent field and geometric convergence criteria were set at their default values. The QM region consisted of 200 atoms and utilized 1234 basis functions. The remaining portion of the protein was treated using molecular mechanics. Within Qsite, the molecular mechanics calculation is handled through the Impact (Levy and co-workers[55]) program, which utilizes the OPLS2001 force field. The molecular mechanics region is adiabatically minimized at the end of each quantum mechanical optimization cycle, thereby allowing for faster convergence. The minimization protocol was set to 1000 steps of conjugate gradient with an energy and gradient convergence criterion of 1×10^{-7} kcal/mol·Å and 0.01 kcal/mol·Å, respectively. No positional or distance constraints were applied during the refinement process. QM/MM potential treatments similar to ours have been reported to provide favorable results[50,56–58] in geometry optimization problems. The choice of the 6-31G* basis set was supported by a number of literature precedents,[59–62] where it has been found to produce acceptable results in computationally expensive QM/MM calculations for QM systems containing phosphorus and magnesium atoms. The resulting QM/MM refined complex was utilized for all the docking studies.

Dimeric Model Generation. Dimeric *Ld*FPPS and *Lm*-FPPS models were generated using the *Tc*FPPS structure 1YHL (PDB code) as a reference. The model generation was carried out using the "Superimpose_aln" functionality of InsightII and used the sequence alignment between the targets (*Ld*FPPS and *Lm*FPPS) and the reference (*Tc*FPPS). Individual monomeric units of the target enzymes (*Ld*FPPS OPLS/GBSA refined model, *Lm*FPPS QM/MM ligand-refined model) were superimposed onto the monomeric units of the reference using a C_α-based superimposition generated from the sequence alignment.

Docking Study. Docking validation studies of the *Lm*FPPS QM/MM ligand-refined model were carried out using Gold 3.1. The protein for docking calculations was generated from the QM/MM ligand-refined model. The protein coordinates were converted to the mol2 format (Sybyl), and the proper assignment of residue names and atom types was carried out. The RIS pose from the QM/MM ligand-refined model was also converted into the mol2 format and used for the docking study. Atom and bond type definitions of the phosphate groups and aromatic cationic nitrogen centers were defined as per the definitions given in the GOLD manual for difficult groups.

The binding site for the pose validation study was defined using the RIS pose, and docking calculations were performed at the automatic GA-200% accuracy settings. Two scoring functions, Goldscore (GL) and Chemscore (C), were used for the docking calculations, and 10 poses were obtained for each run with no early termination criteria. The heavy-atom root mean square deviation (rmsd) of the docked poses was calculated in reference to the RIS pose from the *Lm*FPPS QM/MM ligand-refined model. A multiligand docking study

Table 3. Protein Structure Validation Parameters for the *Ld*FPPS OPLS/GBSA Refined and *Lm*FPPS QM/MM Ligand-Refined Models

	*Ld*FPPS[a]	*Lm*FPPS[b]
C_α-rmsd[c]	0.1879	1.3372
Matchmaker[d]	−0.14	−0.15
Profiles-3D[e]	152.61	146.86
Procheck[f]	0.0	0.6
$\varphi-\psi$ occupancy[g]	0.6	0.9

[a] OPLS/GBSA refined model. [b] QM/MM ligand-refined model. [c] Calculated in reference with *T. cruzi* FPPS structure 1YHL. [d] Unit is in kT. [e] The minimum and maximum possible scores for a model are 74.23 and 166.97, respectively. [f] Percentage of total residues with a bond length or bond angle violation of >0.05Å or >10°, respectively, from reference values. [g] Percentage of residues lying in the disallowed region.

was conducted using a set of 31 ligands which were modeled in Sybyl 6.9 and minimized using the conjugate gradient method to a rms gradient of 0.01 kcal/mol·Å, applying the Gasteiger−Huckel partial charge method and the Tripos force field. The prepared ligands were submitted for docking using the protocol utilized in the pose validation study. The poses were rescored using five different scoring functions Fscore (F), Dscore (D), PMFscore (PMF), Gscore (G), and C available within the Cscore module of Sybyl 6.9. The best-ranked pose for every ligand was selected using the individual scoring functions and utilized for the analysis. Comparison of the scoring functions was carried out after normalization using range scaling (eq 1).

$$S_{n(\text{rangescaled})} = (S_n - S_{\min})/(S_{\max} - S_{\min}) \qquad (1)$$

Two enrichment studies were conducted using data sets comprised of 14 (active) + 995 (decoys) and 9 (active) + 995 (decoys) molecules. The active molecules were obtained from the data set of 31 compounds, while the 995 dummy inactive molecules were obtained from the ZINC database. The molecules were docked using the same docking parameter set as that in the previous step. A total of 10 poses were collected per molecule and were scored by GL as well as the Cscore functions. For each molecule, the best-ranked pose out of the 10 possible poses was selected using each of the scoring functions and used for comparison of enrichment.

RESULTS AND DISCUSSION

Protein Structure Validation. The geometrical and structural consistency of the *Ld*FPPS and *Lm*FPPS models were evaluated at the intermittent and the final stage of model refinement using various protein structure checking tools. Results for the *Ld*FPPS OPLS/GBSA refined and the *Lm*FPPS QM/MM ligand-refined models are reported in Table 3. C_α-based superimposition rmsd values of the *Ld*FPPS and *Lm*FPPS models are 0.1879 and 1.3372 Å, respectively, with their reference *Tc*FPPS structure (1YHL, PDB code) suggesting that the protein structural topology had been preserved from the reference to the target structure. Procheck analysis of the protein structures was carried out to measure the geometrical accuracy of bond lengths and bond angles and showed that only the *Lm*FPPS model had slightly higher violations for 0.6% of the residues. Profiles-3D overall self-compatibility scores for both the models were much higher than the lowest possible scores conceivable for such models. No major misfolded regions were present
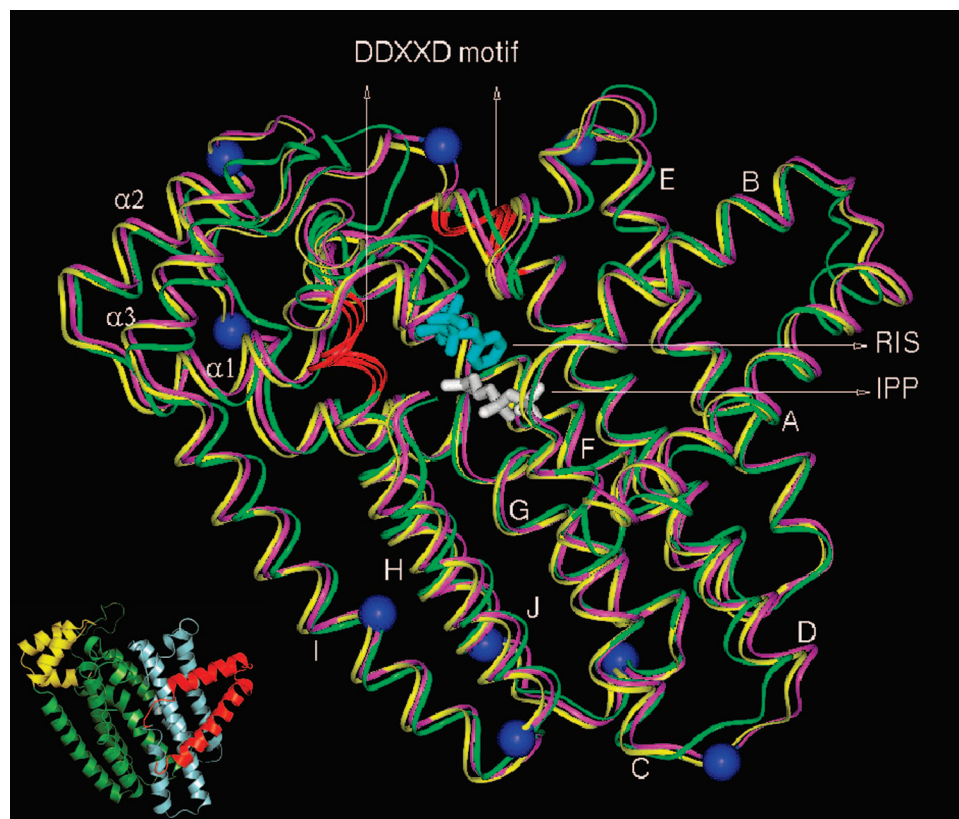
Figure 4. Ribbon representation showing the C_α-based alignment of the *Tc*FPPS (yellow; 1YHL, PDB code), *Ld*FPPS (magenta), and *Lm*FPPS (green) structures. Blue CPKs represent C_α atoms of residues differing between *Ld*FPPS and *Lm*FPPS. The regions colored in red represent the two conserved DDXXD motifs FARM and SARM. RIS (cyan) and IPP (white) are shown in tube representation and define the binding site region. (Inset) A cartoon representation of the *Lm*FPPS model showing the topological regions 1 (red), 2 (blue), and 3 (green, yellow). Region 1 consists of helices A and B forming an α hairpin, which is arranged orthogonally to the other regions. Region 2 is made up of the helices C, D, E, and a portion of F, which connects regions 2 and 3. Helices C, D, and E are arranged antiparallely, while F attains an angular vector and connects to region 3. Region 3 is made up of G, H, α1, α2, α3, I, and J (α1, α2, and α3 refer to three small helices). Helices G, H, I, and J (colored green in inset) are antiparallely arranged and align with region 2. Helices α1, α2, and α3 (colored yellow in inset) are also antiparallely arranged but are orthogonal to the other helices of regions 2 and 3. FARM is located in region 2 on the C terminus of helix D, while SARM is located in region 3 around the C-terminal region of helix H. The DMAPP and IPP binding sites are located between regions 2 and 3.

near the binding site. A fairly negative value was obtained for both of the models in the Matchmaker analysis, suggesting good geometric accuracy. A Ramachandran plot analysis (Figure S1a,b, Supporting Information) of $\varphi-\psi$ angle distribution showed that only 0.6 and 0.9% of the residues of the *Ld*FPPS and *Lm*FPPS models, respectively, were located in the disallowed regions. In summary, the above-mentioned results suggested that reasonable 3D protein models could be generated for both of the enzyme isoforms.

Protein Sequence and Structural Topology Analysis. A sequence homology of 63% between the target and the reference sequence led to a satisfactory sequence alignment (Figure 2) showing conservation of the key DDXXD motifs (FARM, first aspartate-rich motif; SARM, second aspartate-rich motif), C-terminus BXB motif, and the binding site residues. Minor differences were observed in the constitution of the chain-length regulation residues between the *Pf*FPPS and *Ld*FPPS/*Lm*FPPS. Nine residues differ between the sequences of *Ld*FPPS and *Lm*FPPS, and none of these residues are close to the binding site and are located in the peripheral regions of the protein. A list of these differing residues along with their possible interactions is provided in the Supporting Information (Table S2). The modeled protein structures of *Ld*FPPS and *Lm*FPPS (Figure 4) consist of 13 helical regions and can be classified in accordance with

the other enzymes from this family as α-helical proteins. Among the characteristic features, a subsystem consisting of three small helices, α1, α2, and α3, is observed within region 3 of the two proteins, which is characteristic of FPPS enzymes from protozoal and higher-order species and is not found in bacterial FPPSs. A visible kink is present within helix G (T208−T212, *Tc*FPPS numbering), which accommodates for the differences in the orientation of helix F and the antiparallel region comprising helices G, H, I, and J.

In the apo form, the two loop regions following the conserved DDXXD motifs (loops between helices D/E and H/α1) are in a slightly open conformation. In the event of ligand binding, the loop regions demonstrate induced-fit effects and come down closer into the binding site to stabilize the ligand pose through side-chain interactions. Arg107 and Arg108 from the D/E loop interact with the ligands present in the DMAPP and IPP site, respectively, while Lys264 present in the H/α1 loop interacts with the ligand present in the DMAPP binding site. It has been proven experimentally that the binding of a ligand to the DMAPP site is enhanced if the IPP binding site is also occupied. The binding of IPP triggers an induced-fit effect in the C-terminus region which contains the conserved BXB motif. The reorganization and stabilization of the C terminus during IPP binding decreases the overall entropy of the enzyme active site and adds to
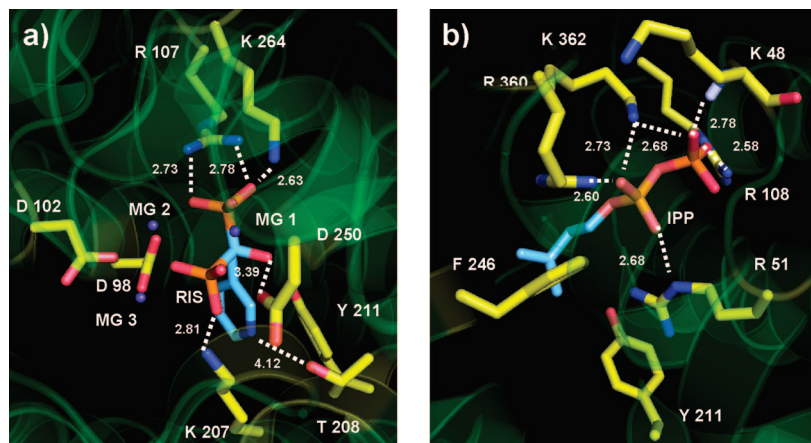
Figure 5. The (a) DMAPP and (b) IPP binding sites of the *Lm*FPPS QM/MM ligand-refined model. The carbon atoms of the ligands RIS and IPP are colored cyan, while those of the protein residues are colored yellow. The RIS phosphate groups pointing towards and away from us are referred to as the frontal and distal groups, respectively. The phosphate oxygen atoms coordinating MG1 are referred to as axial oxygens. The polar ligand–protein interactions are shown in white dotted lines.

the energetic stabilization of the complex. Additionally, IPP itself undergoes hydrophobic interactions with the DMAPP site ligand and provides further stabilization. We generated our models from the *Tc*FPPS structures, where both of these sites were occupied, allowing us to model the bound form of the enzyme as opposed to its apo form. Modeling of the bound form is also more relevant from the perspective of drug design. Since the *Ld*FPPS and *Lm*FPPS sequences vary only by nine residues, which are located in the peripheral regions of the protein, ligand-based refinement of the model was carried out only for the *Lm*FPPS structure.

Crude dimeric models of *Ld*FPPS and *Lm*FPPS (Figure S2a,b, Supporting Information) were constructed to identify the residues of the monomeric units involved in the formation of the dimeric interface and the critical intersubunit interactions responsible for the energetic stabilization of the dimeric structure. The dimeric interface between the two enzyme units is made up of the loop regions between helices A/B, D/E, and F/G as well as the helices B, D, E, F, and G. The interactions at the dimeric interface are of both hydrophobic and polar nature. Intersubunit salt bridges form a critical component of the energetic stabilization obtained through dimerization. In case of the reference *Tc*FPPS structure, four possible intersubunit salt bridges (Table S1, Supporting Information) could be identified. In case of *Ld*/*Lm*FPPS, the corresponding residues for two of the salt bridges, Asp29–Arg202 and Arg32–Glu194, are conserved from *Tc*FPPS. The Lys205 (Glu27–Lys205 salt bridge) and Arg34 (Arg34–Asp181 salt bridge) residues are replaced by valine and glycine, respectively, in both *Ld*/*Lm*FPPS (Figure S2c,d, Supporting Information). On the other hand, in *Ld*/*Lm*FPPS, a salt-bridge interaction is possible between the residues Arg25 and Asp158. This interaction is not possible in *Tc*FPPS, where both Arg25 and Asp158 are replaced by lysine residues. Among the nine residues differing between the *Ld*FPPS and *Lm*FPPS sequence (Table S2, Supporting Information), two residues, 121 and 183, are located at the dimeric interface. While residue 121 does not show any interaction in either case, Asn183 (backbone carbonyl oxygen) in *Ld*FPPS shows a possible intersubunit hydrogen-bonding interaction with His116 (N[δ]–H) from the other monomer. In the case of *Lm*FPPS, the corresponding residue is a lysine which does not show an intersubunit interaction.

***Lm*FPPS QM/MM Ligand-Refined Model.** Significant information on enzyme inhibitors is available for *Lm*FPPS and could be utilized in extensive docking-based validation studies. Since *Lm*FPPS and *Ld*FPPS are closely related, ligand-based refinement and docking-based validation were carried out only for the *Lm*FPPS model. The ligand-refined model of *Lm*FPPS showing occupancy of both the DMAPP and IPP binding sites was constructed using RIS and IPP. The DMAPP binding site has been widely targeted from the perspective of drug design. Therefore, in addition to overall structure refinement, an accurate representation of the interactions within this site and the adjoining IPP site are of paramount importance. Within the DMAPP site, the three magnesium ions undergo key interactions with the two phosphate groups of the bisphosphonate ligand. They also interact with two acidic residues, Asp98 and Asp102, from the FARM and one residue, Asp250, from the SARM, thereby forming a system of bridged interactions which plays a central role in stabilizing the ligand binding pose. The octahedral geometry of the metal centers is completed by the structural waters.

Binding Site Analysis. The DMAPP and IPP binding sites (Figure 5) of *Lm*FPPS are highly conserved with FPPS isoforms from other species. The enzyme catalyzes the C–C bond formation between the two substrates DMAPP and IPP to form the final product FPP. The metal-bridged network plays important roles in both the stabilization of the substrate binding poses and induction of the reaction process. As discussed earlier, binding of the substrates/ligands to the binding site induces certain conformational changes and helps in the formation of the metal octahedral geometry. The bisphosphonates bind to the DMAPP binding site wherein the phosphate group mimics the pyrophosphate moiety of DMAPP and undergoes similar interactions with the metal centers. The positive charge on the nitrogen-containing bisphosphonates is thought to mimic the cationic intermediate (Figure 1c) derived from DMAPP during the enzymatic reaction. The DMAPP and IPP binding sites of the *Lm*FPPS model are shown in Figure 5. In this view, the IPP binding site is located to the right of the DMAPP site.

DMAPP Site. The binding pose of RIS in the DMAPP binding site of *Lm*FPPS is shown in Figures 5a and 6. The two phosphate groups of the ligand were modeled in the
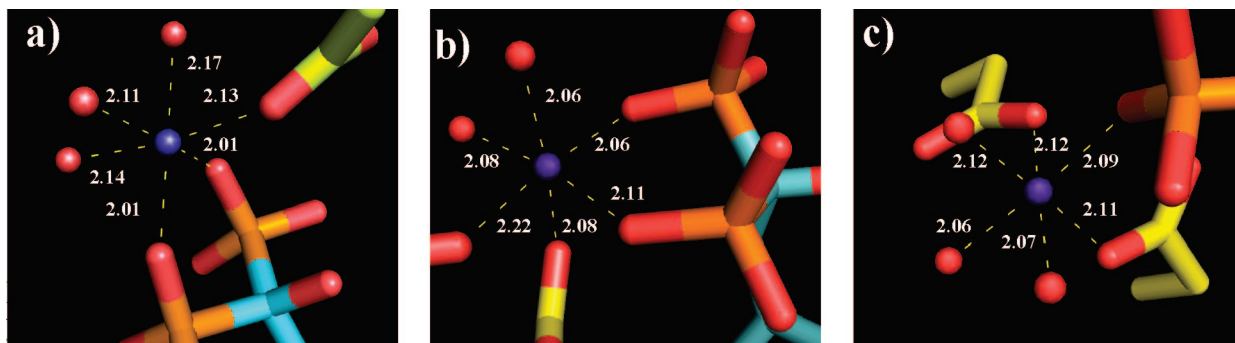
Figure 6. The metal–ligand and metal−protein interactions for (a) MG1, (b) MG2, and (c) MG3 in the *Lm*FPPS QM/MM ligand-refined model.

dianionic state and are seen to undergo several charge–charge interactions with the metal centers as well as positively charged amino acid side chains. The axial oxygens of the distal and frontal phosphate groups coordinate MG1 strongly with Mg^{2+}−O distances of 2.01 Å. The equatorial oxygens of the two phosphate groups pointing away from the IPP site coordinate MG2 with Mg^{2+}−O distances of 2.06 and 2.11 Å. MG3 forms a lone coordination with an equatorial oxygen of the frontal phosphate at a Mg^{2+}−O distance of 2.09 Å. The distal phosphate group oxygens also interact with the side chains of Arg107 and Lys264. The two guanidino nitrogens of Arg107 form bifurcated interactions (N−O distances, 2.73 and 2.78 Å) with the two equatorial oxygens of the distal phosphate group, while the Lys264 nitrogen interacts (N−O distance, 2.63 Å) with the axial oxygen of the frontal phosphate group. The equatorial oxygen of the frontal phosphate group interacts (N−O distance, 2.81 Å) strongly with the side chain of Lys207. The hydroxyl group of the ligand forms a hydrogen bond (O−O distance, 3.39 Å) with the carboxylate side chain of Asp250. The alkyl headgroup projects toward the floor of the binding site and undergoes interactions of both a polar and hydrophobic nature. The positively ionized nitrogen center can interact with Thr208 and Tyr211 side-chain hydroxyl groups. The aromatic side chain of Tyr211 forms a hydrophobic interaction with the pyridine ring of the ligand. The alkyl headgroup of IPP provides further stabilization by undergoing hydrophobic interaction with the aromatic ring of RIS.

The three magnesium ions in octahedral geometry (Figure 6) are critical to the orientation of the bisphosphonate ligand. In addition to the two phosphate oxygens, MG1 is coordinated to the carboxyl oxygen of Asp250 (Mg^{2+}−O distance, 2.13 Å), which is a critical residue of the SARM, and to three water molecules completing the octahedral geometry (Mg^{2+}−O distances, 2.17, 2.11, and 2.14 Å). MG2 is coordinated to the carboxylate side chains of Asp 98 (Mg^{2+}−O distance, 2.08 Å) and Asp102 (Mg^{2+}−O distance, 2.22 Å) located on the FARM. The residues Asp98 and Asp102 are involved in interactions with both MG2 and MG3. Asp98 undergoes a bridged interaction with the two carboxylate oxygens coordinating to the MG2 and MG3. The interaction is different in the case of Asp102, where one of the carboxylate oxygens interacts with both of the magnesium cations. The octahedral geometry of MG2 is completed by two water molecules (Mg^{2+}−O distances, 2.06 and 2.08 Å). MG3 is coordinated to three water molecules, two carboxylate side chains, and a lone phosphate oxygen. The water oxygens are at 2.12, 2.06, and 2.07 Å from the magnesium

center. Asp98 and Asp102 side-chain oxygens coordinate MG3 with Mg^{2+}−O distances of 2.11 and 2.12 Å, respectively. According to a study[63] of 32 protein structures (resolution < 1.65 Å) containing metal−phosphate oxygen coordination, the range of distances for Mg^{2+}−O (phosphate) is 1.79−2.22 Å with a mean of 2.06 Å. Similarly, the mean Mg^{2+}−O (water) coordination distance from 269 protein complexes from the PDB was found to be 2.09 ± 0.08 Å.[64] The mean distance for Mg^{2+}−O (carboxylate, monodendate), obtained from 43 protein complexes from the PDB, has been pegged at 2.08 ± 0.08 Å. The values vary over a wider range when the interactions become bidendate or bridged. A comparison of our final model to these reference values suggested that we had been able to attain reasonable geometrical accuracy corroborating with experimental results.

IPP Site. The C−C−O bond angle of the IPP binding pose in the *Tc*FPPS structure 1YHM is 97°, which is deviated from the mean tetrahedral angle of 109°. This allows the alkyl headgroup of IPP and the pyrophosphate moiety to attain a specific conformation inside the binding site. In our modeled structure, the IPP molecule has been fully relaxed through the course of model building; hence, a slightly altered conformation is generated, although the major interactions are still preserved (Figure 5b). The pyrophosphate moiety of IPP is observed to make multiple charge−charge interactions with several positively charged side chains of Arg51, Arg108, Arg360, and Lys362 (bifurcated interaction). It also forms interactions with the backbone NH of Lys48 and side chain of Gln91. The alkyl headgroup of the ligand is stabilized through a number of hydrophobic interactions with the side chains of Tyr211, Leu95, and Tyr246. In the final model, the terminal methylene group of IPP is 3.26 Å from the hydroxyl group of RIS.

Mutation Data. Mutation studies carried out on the *rat*FPPS[65] and *Sc*FPPS[66] (*Saccharomyces cerevisiae*) have provided us with a clearer picture of the amino acids critical to the catalytic activity of the enzyme. These residues are located in or near the DMAPP and IPP binding sites of the enzyme. They are either directly involved in interactions with the substrates during catalytic processing or are critical in maintaining the binding site integrity. The *Lm*FPPS QM/MM ligand-refined model is able to explain these mutations in terms of interactions with the bound ligands RIS and IPP. Since bisphosphonates undergo the same pattern of binding site interactions as the native substrate DMAPP, the loss of these interactions would rationalize the loss in enzyme activity upon mutation of the relevant DMAPP site residues. Aspartate to alanine mutation of the residues, corresponding
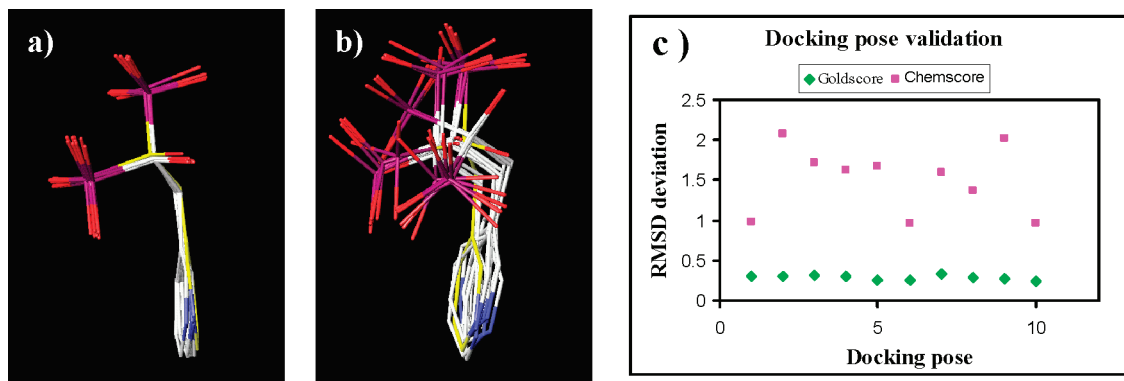
Figure 7. Binding poses obtained for RIS using the scoring functions (a) GL, (b) C, and (c) their heavy atom rmsd distribution in comparison to the reference pose from the *Lm*FPPS QM/MM ligand-refined model.

to Asp98, Asp99, and Asp102 from the FARM as well as Asp250, Asp251, and Asp254 from the SARM, were found to reduce catalytic activity. While Asp98, Asp102, and Asp250 are directly involved in the metal-bridged network critical for the stabilization of the ligand pose, Asp99 and Asp251 form salt bridges with Arg107 and Lys264, respectively, providing stability to the binding site. Asp254 is involved in a hydrogen-bonding interaction with the backbone NH of Thr267, stabilizing the H/α1 loop which forms part of the DMAPP binding site. Mutation of the residues corresponding to Arg107Gln and Arg108Gln were also found to reduce catalytic activity. These two residues are involved in interactions with the bisphosphonate moiety of the DMAPP site ligand, RIS, and the pyrophosphate moiety of IPP, respectively, and would lead to the destabilization of the binding poses. Two other mutations, Arg360Ala and Lys264Ala, were also shown to mildly affect the enzyme turnover rate. While Lys264 undergoes interaction with the bisphosphonate moiety of RIS, the Arg360 residue interacts with the β-phosphate moiety of IPP. Mutation studies[67] on the bulky phenylalanine chain-length regulation residues of the *Av*FPPS enzyme to smaller residues such as serine and alanine have led to a change in the product specificity and generation of products with longer chain lengths. In the case of *Ld/Lm*FPPS, the chain-length regulation residues are a histidine−phenylalanine pair located at the end of a channel at the bottom of the DMAPP site. They are reachable only by FPP or ligands of equivalent length and perform a regulatory role by providing steric repulsion and preventing the chain-length propagation of the final product from extending beyond three isoprene units.

Docking Study. The predictive ability of a homology model can be validated by using it in a docking study to calculate the binding poses for a series of active molecules, which are then evaluated on the basis of their interaction and geometric profile by a scoring metric and classified in accordance with their biological activity. As part of this evaluation, a pose validation study of the DMAPP site ligand, RIS bound to the *Lm*FPPS model, was carried out using the Gold docking program. Two different scoring functions, GL and C, were evaluated for their accuracy in the calculation of the docking poses. Heavy atom rmsd of the 10 docking solutions (Figure 7a,b) with the original ligand pose (*Lm*FPPS QM/MM ligand-refined model) were calculated and plotted (Figure 7c). GL was clearly the better performer of the two, with the rmsd of all the solutions lying in a narrow range of 0.24−0.32 Å, while those for C ranged from 0.96 to 2.07

Å. Comparison of the docked poses in Figure 7a,b suggests that, in the case of GL, the scoring function was able to accurately evaluate the positioning of the ligand's phosphate oxygens in respect to the magnesium cations so as to satisfy the octahedral geometry of the metal centers. On the other hand, the C function was unable to evaluate the geometrical accuracy of the placement of the ligand's phosphate oxygens in respect to the metal centers and generated poses which did not satisfy the required coordination geometry. This led to a relatively poorer performance for C compared to the GL function. The cationic center on the pyridine ring was also placed in close vicinity of the actual pose interacting with Thr208 (Figure 7a) in the case of the GL poses. Since no constraints were applied during the docking process and the binding pose of the ligand could be replicated independently using a docking program, it might be postulated that the binding site of the model had been suitably refined.

The second portion of the evaluation focused on the docking of a set of 31 bisphosphonates (including RIS; Figure S3, Supporting Information)[68] which inhibit the *Lm*FPPS. The activity range of these molecules spanned 2.8 log orders, and they were divided into three groups on the basis of their biological activity for the study. The highly active (H) group had 14 molecules (including RIS) with activity spanning from 7.0 to 8.0 pK_i. The eight medium actives (M) rangxed from 6.25 to 6.99 pK_i, while the nine low actives (L) ranged from 5.13 to 6.24 pK_i. The molecules from the low-active (L) group share some of the pharmacophoric features with the molecules present in the highly active group (H), thereby providing a tougher challenge for the scoring function in achieving separation between the two groups. The relative performance of the six different scoring functions is shown in Figure 8 and also tabulated in Table 4. In Figure 8, the horizontal spread of each bar depicts the mean ± standard deviation range of the range-scaled scores for the docked ligands from each activity group. Our main aim was to attain a docking-score-based separation of the molecules from the groups H and L. Scores for group M molecules were expected to partially overlap with those from the H and L groups. On the basis of the mean scores for each of the activity groups (Table 4), GL, F, and PMF seemed to be the better performers in comparison to C, D, and G. It is evident that, in the case of both F and GL, group H molecules were scored highly with a narrow spread. On the other hand, group L molecules were scored less accurately and with a wider spread in both cases, leading to a partial overlap with the score ranges of group H molecules.
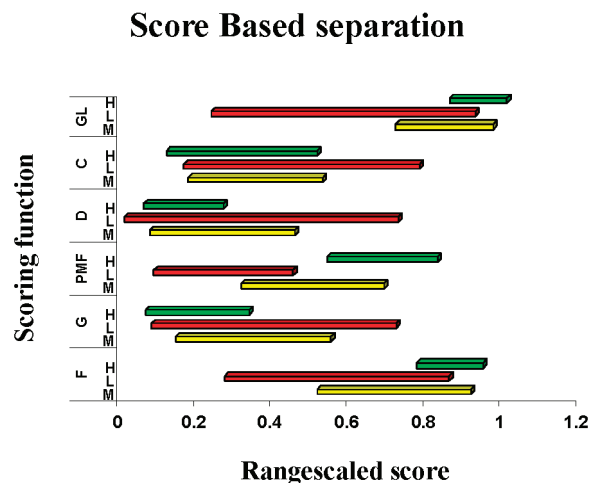
Score Based separation

Figure 8. A graph showing the separation of the H and L group molecules on the basis of range-scaled scores. The three categories: H is colored in green, M in yellow, and L in red. The scoring functions on the *y* axis are F, G, PMF, D, C, and GL. The extents of the bars depict mean ± standard deviation.

Table 4. Mean and Standard Deviations of the Range-Scaled Scores Obtained for Groups H, M, and L[a]

scoring function	mean	SD	scoring function	mean	SD
H		0.864	0.086		0.168	0.104
M	F	0.717	0.200	D	0.268	0.189
L		0.569	0.292		0.371	0.358
H		0.204	0.135		0.320	0.196
M	G	0.350	0.202	C	0.355	0.175
L		0.403	0.319		0.477	0.308
H		0.687	0.144		0.937	0.075
M	PMF	0.505	0.187	GL	0.849	0.127
L		0.272	0.182		0.585	0.344

[a] The evaluated scoring functions are F, G, PMF, D, C, and GL.

PMF was the best performer, where the score ranges for the groups H and L did not overlap with each other and there was a clear separation of the two groups.

Two enrichment studies were carried out using molecules from groups H (first run) and L (second run) mixed with a data set of decoys. In the first run (Figure 9a), F and GL followed by D and PMF were the better performers in terms of retrieving the group H molecules with a ~100% retrieval at the 5% database screened mark. C and G showed a poor retrieval of group H molecules in this range. In the second run (Figure 9b), F, GL, and D showed a high retrieval of the group L molecules, while PMF showed a comparatively lower retrieval at the 5% database screened mark. Similar to the previous run, G showed a very low retrieval in this run, suggesting that it is totally unresponsive in identifying the molecules from both groups H and L. C showed a higher retrieval of the group L molecules as compared to the group H molecules from the previous run. The interaction of the bisphosphonate moiety with the metal centers and the basic amino acids is a critical pharmacophoric feature required for binding to the DMAPP site, and all of the compounds used in the enrichment studies possess this feature. Therefore, the variation in the activity between the group H and L molecules can primarily be attributed to the presence or absence of features other than the bisphosphonate moiety. Comparison of the two enrichment plots suggests that, in the cases of F, D, and GL functions, the enrichment is mainly dependent on the interactions undergone by the bisphosphonate moiety
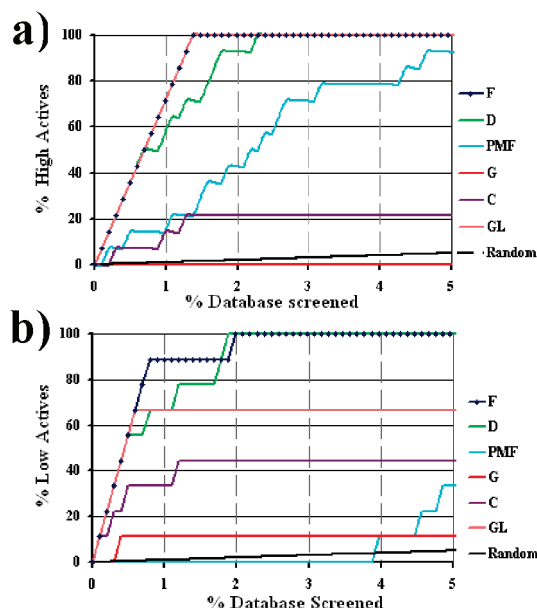

Figure 9. Enrichment plots for the two enrichment studies conducted using molecules from groups (a) H and (b) L. The evaluated scoring functions are F, G, PMF, D, C, and GL. The plots for F and GL overlap with each other in the first enrichment study. A marker (blue diamond) was therefore added in the case of F. G does not show any retrieval in the first enrichment study, and hence the plot overlaps with the baseline.

in the ligands. On the other hand, PMF shows sensitivity to the presence or absence of these "other" features in addition to the bisphosphonate moiety interactions and attains a higher enrichment for group H as compared to the group L molecules. On the basis of these observations, we can conclude that the modeled structure could be used to develop a predictive docking formalism: Gold docking utilizing the GL function for the calculation of docking poses and rescoring followed by pose selection using PMF. This protocol would prove highly useful and applicable in future virtual screening efforts. Furthermore, since no ligand-specific constraints were applied during the docking process and the scoring was entirely based on the interaction profile, the current protocol may be aptly used to dock ligands with varying scaffolds.

In the present study, a number of scoring functions were evaluated for their ability to reproduce the modeled pose of the RIS ligand as well as for the ranking of poses of different ligands in accordance with their activity profile. The evaluated scoring functions can be broadly classified into three categories: empirical, GL, C, and F; force-field-based, D and G; and knowledge-based, PMF. The pose validation studies conducted on the RIS pose using the GL and C functions showed that GL was the better performer of the two. Previous studies conducted by the developers using the CCDC/Astex Test Set[69] as well as independent comparative studies conducted by scientists at GlaxoSmithKline[70] have shown that GL performs well in the prediction of binding poses in proteins containing metal ions. These results corroborate with what is seen in the present case. One possible reason for the better performance of the GL function may be attributed to the fact that it performs very well when interactions of a directional nature, such as hydrogen bonding and so forth, are present between the ligand and the protein. On the other hand, C is known to perform relatively well when hydro-
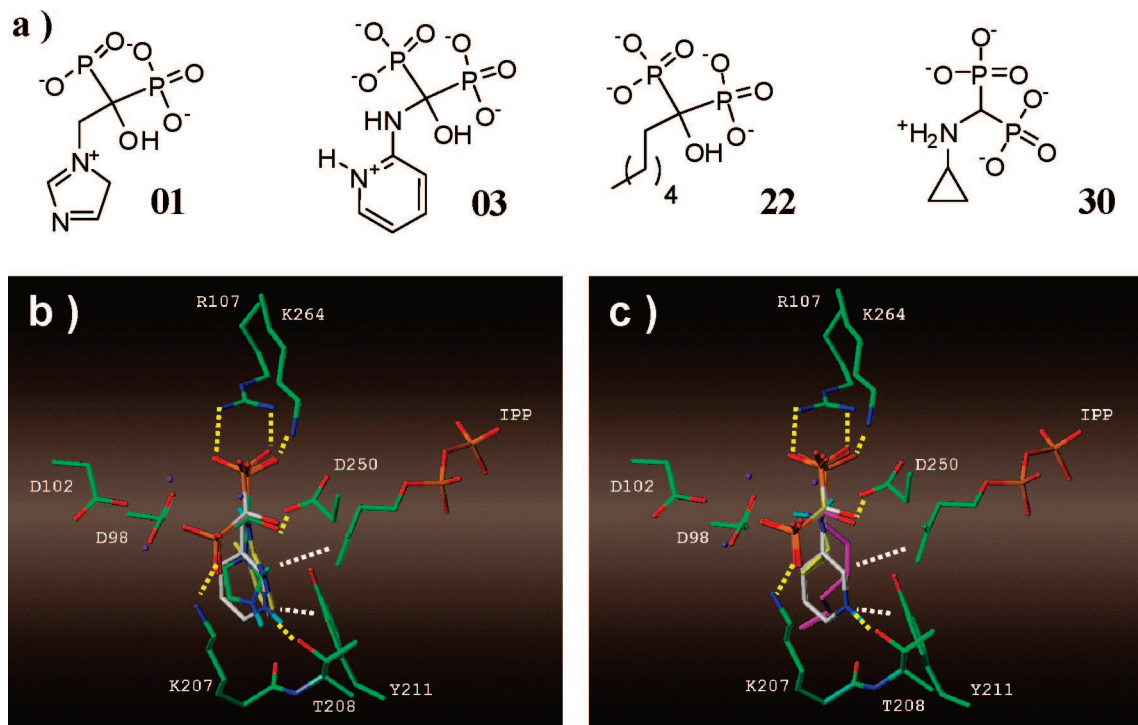
Figure 10. (a) Structures of the bisphosphonates **01**, **03**, **22**, and **30** used in the docking study. (b) Binding pose of the highly active ligands **01** (green) and **03** (yellow) superimposed with that of RIS (white). (c) Binding pose of the low-active ligands **22** (magenta) and **30** (yellow) superimposed with that of RIS (white). The major polar interactions between the ligand and the protein are shown in yellow dotted lines, while the hydrophobic interactions are shown in white dotted lines.

phobic interactions appear to primarily drive the ligand-binding process. In the present study, positioning of the bisphosphonate ligand shows a strong dependence on the interactions of the phosphate oxygens with the metal centers as well as the basic side chains. These interactions are highly directional and, hence, are better evaluated by the GL function.

The enrichment study involved the evaluation of different scoring functions for their ability to rank the docked poses in accordance with their biological activity. In this evaluation, the PMF function shows the maximum sensitivity toward the structural changes in the ligands. It is interesting to notice that, among the scoring functions, PMF[71] is the only knowledge-based scoring function. This type of scoring function derives statistical preferences as potentials for calculating the protein–ligand atom pair interactions. On the other hand, empirical and force-field-based methods have specific terms dedicated toward the calculation of the various biophysical components of protein–ligand binding such as intermolecular interactions, solvation, entropy, enthalpy, and so forth. Usually, the terms are calculated using a simplified approach, which may result in the introduction of errors in the calculation of one or more of the terms and may lead toward a bias for a specific term. This becomes evident in our study, wherein the F, D, and GL functions rank the high actives as well as the low actives in the same range because these functions appear to overestimate the contributions from the interactions of the phosphate groups and, hence, may not be sensitive to the fluctuations introduced by the interactions offered by the alkyl headgroup. The PMF function treats the individual components of binding in an implicit fashion and eliminates the bias generated by the strong polar interactions of the ligand phosphate groups. It is therefore able to differentiate the molecules belonging to

the H and L groups. Previous results on the use of the PMF scoring function have shown that it performs well in correlating binding data to the docking score in metal-containing enzymes such as stromelysin[72] and in reliably ranking actives[73] having weak binding affinity, which provides further support to our conclusions.

The predicted binding poses for two of the highly active (**01**, **03**) and low active molecules (**22**, **30**) (Figure 10a) considered in the docking study are depicted in Figure 10b,c. The binding poses of these molecules are comparable to that of the reference RIS pose obtained from the *Lm*FPPS QM/MM ligand-refined model. The most critical binding interactions of these molecules appear to be mediated through the bisphosphonate moiety. In all of the poses, the bisphosphonate moiety of the ligand appears to be placed accurately, satisfying the octahedral geometry of the metal centers. The binding pose of **01** (zolendronate) is similar to that of RIS with the terminal five-membered ring overlapping with the pyridine nucleus of RIS. The terminal positively ionizable nitrogen center points in the direction of the residue Thr208. This binding pose also appears to corroborate closely with the cocrystallized pose of zolendronate from the *Hu*FPPS structure (2F8Z, PDB code). The other highly active molecule, **03**, has a pyridine nucleus with a positively ionizable center at the *ortho* position. It also attains a pose similar to the reference RIS pose from the *Lm*FPPS QM/MM ligand-refined model with the two aromatic side chains overlapping each other. The pyridine nucleus is connected to the bisphosphonate moiety through an aniline linker, and the substituted aniline group lies in plane with the ring system, suggesting a possibility for the delocalization of the positive charge between the two nitrogen centers. The positively ionizable center at the *ortho* position seems to point in the direction of Thr208.

In the case of the two low actives, **22** and **30**, the molecules lack one or more of the structural features present in the highly active molecules. Apart from the metal–ligand interactions, the presence of a hydrophobic group attached to the bisphosphonate moiety and the presence of a positively ionizable center are two of the other pharmacophoric features known to enhance the binding affinity of the ligands.[68] In the case of **22**, the longer alkyl side chain provides the necessary hydrophobic interactions but lacks a positively ionizable center, which accounts for its poorer binding affinity. The large number of rotatable bonds present in this molecule may enhance ligand entropy and affect the binding affinity of the ligand. In the case of **33**, a positively ionizable center is present, but the smaller cyclopropyl substitution provides a weaker hydrophobic interaction in comparison to the aromatic rings present in the highly active molecules **01**, **03**, and RIS. It should be mentioned that there are a few additional hydrogen-bonding sites such as the backbone carbonyl oxygens of Lys207 and Phe94 which may be utilized in future drug design efforts to enhance the binding affinity of the ligands.

Comparison of the Human and Leishmanial FPPS. *Hu*FPPS shares a 33% sequence identity with *Ld*FPPS and *Lm*FPPS, and the overall structural topology of the two proteins are very similar. The *Ld*FPPS and *Lm*FPPS models have a C_α-based superimposition rmsd of 2.51 and 2.86 Å, respectively, with the *Hu*FPPS structure (2F8Z, PDB code) (Figure S2e, Supporting Information). The helical regions of the proteins are highly conserved, and differences between the two structures are noticed primarily in the loop regions. Minor differences could be noticed in the chain length of the loop regions between the helices A/B, C/D, and I/J, which, in the case of *Hu*FPPS, are shorter by approximately 1, 4, and 2 residues, respectively, as compared to *Ld*FPPS/ *Lm*FPPS. A major difference arises in the loop region connecting helices F and G. This loop region interacts with the other monomer and forms part of the dimeric interface. In the case of the *Ld*FPPS and *Lm*FPPS, this loop region is longer by 11 residues as compared to the *Hu*FPPS structure.

The DMAPP binding site of *Lm*FPPS is highly conserved with that of *Hu*FPPS. The very fact that the bisphosphonates were originally discovered as inhibitors of *Hu*FPPS and are also active in the parasitic enzymes suggests that the knowledge gained so far against the human enzyme could be now be applied in the design against the leishmanial FPPSs. The two DDXXD motifs which are critical toward enzyme catalysis are oriented similarly in the DMAPP binding site. The other basic residues which interact with the bisphosphonate moiety as well as the residues which interact with the alkyl headgroup and the positively ionizable center are also conserved.

The IPP binding site is also highly conserved between the three enzymes, except for the terminal BXB motif, which is oriented differently in *Ld*FPPS/*Lm*FPPS, as compared to what is seen in the *Hu*FPPS structure. The C-terminal lysine from the BXB motif in the case of *Hu*FPPS is rotated out of the pocket. Instead, Lys48 (*Ld*/*Lm*FPPS numbering), which undergoes only a backbone interaction with IPP in the plasmodial structures, rotates inward and substitutes for the lysine by forming side-chain interactions with the oxygens from the α- and β-phosphate of IPP.

The only major difference between the human and the leishmanial FPPSs lie in the region where the chain-length regulation residues (Figure S2f, Supporting Information) are located. When processing the normal substrates DMAPP and IPP, these residues provide steric bulk and prevent the alkyl headgroup of the final product FPP from growing beyond three isoprene units. In the case of *Hu*FPPS, the chain-length regulation residues are two phenylalanines, while in the case of *Ld*FPPS/*Lm*FPPS, they are a histidine–phenylalanine (His93, Phe94) pair. These residues are located at the end of a long channel that is situated at the bottom of the DMAPP site. In our docking studies, only the bisphosphonate ligands with long alkyl side chains could reach up to these residues. While both of these residues are hydrophobic in the case of *Hu*FPPS, the first of the two residues in the case of *Ld*FPPS/ *Lm*FPPS is a histidine (His93), which is stabilized through possible interactions with Glu97 and Asp160. In the case of *Hu*FPPS, the Glu97 and Asp160 residues are replaced by a methionine and serine, respectively. It remains to be seen whether this minor difference could be utilized to attain any selectivity over the human isoform.

CONCLUSION

In the present study, homology models of *Ld*FPPS and *Lm*FPPS were generated using *Tc*FPPS as a reference structural homologue. Restraint-guided molecular mechanics refinement followed by QM/MM refinement was utilized to generate a ligand-refined model of *Lm*FPPS with a bisphosphonate, RIS and IPP, bound to the DMAPP and IPP binding sites, respectively. The final model showed a conservation of the overall protein structural topology with the other enzymes from the same family. The highly refined binding site of the *Lm*FPPS QM/MM ligand-refined model showed a number of key interactions which are conserved with the available ligand-bound crystal structures of other FPPS isoforms. The model could also be used to explain the results of mutation studies carried out on the binding site (including residues of the DDXXD motif) and the chain-length regulation residues in related FPPS isoforms. The metal coordination geometry of the enzyme's DMAPP site, which contains RIS coordinated to three Mg^{2+} ions, was found to be in accordance with what is observed in a number of experimentally determined protein structures.

The predictive ability of the *Lm*FPPS QM/MM ligand-refined model was assessed through a rigorous docking study. The modeled pose of RIS could be replicated independently through docking using the GL function and was suggestive of the structural accuracy of the model's binding site. A score-based active/inactive separation and enrichment study was conducted using a set of known bisphosphonate-based *Lm*FPPS inhibitors and helped in the elucidation of a predictive docking protocol. The best predictive ability was obtained by using the PMF function to rescore docked poses obtained by using the GL function. Analysis of the docked poses provided important insights into the key structural features of the ligands which are responsible for their biological activity. Since *Ld*FPPS and *Lm*FPPS are very similar, the inferences drawn from docking studies on one enzyme can be easily extrapolated to the other isoform. The validated docking protocol would prove to be a useful tool in any structure-based design efforts against these targets.

LEISHMANIAL FARNESYL PYROPHOSPHATE SYNTHASES

J. Chem. Inf. Model., Vol. 48, No. 5, 2008 **1039**

It is also worth mentioning that the use of bisphosphonates for bone-related diseases is highly favored since they are selectively taken up into the phagocytic or endocytic cells of the bone, such as osteoclasts and macrophages, and the remaining plasma concentration is rapidly eliminated through the kidneys. While the use of bisphosphonates as antileishmanials has been validated at the levels of *in vitro* parasite cultures and *in vivo* animal infection models, their use in actual therapy apart from human bone-related diseases may still present several pharmacokinetic issues. For instance, the current nitrogen-containing bisphosphonates have an oral bioavailability of around 1%, and they do not reach the efficacious drug concentrations in plasma or soft tissues required for producing a therapeutic effect.[43] The highly refined model along with the validated docking and scoring algorithms could be utilized to conduct structure-based virtual screening against commercially available databases to identify hits with novel scaffolds and favorable efficacy as well as pharmacokinetic profiles.

Supporting Information Available: Ramachandran plots of the *Ld*FPPS/*Lm*FPPS models, images of the dimeric models of *Ld*FPPS/*Lm*FPPS, C_α superimposition of *Lm*FPPS with *Tc*FPPS and *Hu*FPPS, a list of possible intersubunit salt-bridge interactions for the *Ld*FPPS/*Lm*FPPS dimeric models, a list of residues differing between *Ld*FPPS/*Lm*FPPS, structures and activities of the *Lm*FPPS inhibitors used in the docking study, and the details of the minimization protocols used for generating the *Ld*FPPS and *Lm*FPPS models. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Desjeux, P. The increase in risk factors for leishmaniasis worldwide. *Trans. R. Soc. Trop. Med. Hyg.* **2001**, *95* (3), 239–243.

(2) Bora, D. Epidemiology of visceral leishmaniasis in India. *Natl. Med. J. India* **1999**, *12* (2), 62–68.

(3) World Health Organization: Magnitude of the Problem. http://www.who.int/leishmaniasis/burden/magnitude/burden_magnitude/en/index.html (accessed Feb 2008).

(4) Croft, S. L.; Yardley, V. Chemotherapy of leishmaniasis. *Curr. Pharm. Des.* **2002**, *8* (4), 319–342.

(5) Roberts, W. L.; McMurray, W. J.; Rainey, P. M. Characterization of the antimonial antileishmanial agent meglumine antimonate (glucantime). *Antimicrob. Agents Chemother.* **1998**, *42* (5), 1076–1082.

(6) Grogl, M.; Thomason, T. N.; Franke, E. D. Drug resistance in leishmaniasis: its implication in systemic chemotherapy of cutaneous and mucocutaneous disease. *Am. J. Trop. Med. Hyg.* **1992**, *47* (1), 117–126.

(7) Adler-Moore, J.; Proffitt, R. T. AmBisome: liposomal formulation, structure, mechanism of action and pre-clinical experience. *J. Antimicrob. Chemother.* **2002**, *49* (Suppl 1), 21–30.

(8) Escobar, P.; Yardley, V.; Croft, S. L. Activities of hexadecylphosphocholine (miltefosine), AmBisome, and sodium stibogluconate (Pentostam) against Leishmania donovani in immunodeficient scid mice. *Antimicrob. Agents Chemother.* **2001**, *45* (6), 1872–1875.

(9) Alrajhi, A. A.; Ibrahim, E. A.; De Vol, E. B.; Khairat, M.; Faris, R. M.; Maguire, J. H. Fluconazole for the treatment of cutaneous leishmaniasis caused by Leishmania major. *N. Engl. J. Med.* **2002**, *346* (12), 891–895.

(10) Geusens, P.; McClung, M. Review of risedronate in the treatment of osteoporosis. *Expert Opin. Pharmacother.* **2001**, *2* (12), 2011–2025.

(11) Pistevou-Gombaki, K.; Eleftheriadis, N.; Sofroniadis, I.; Makris, P.; Kouloulias, V. Palliative treatment of painful bone metastases from non-Hodgkin lymphoma with disodium pamidronate. *J. Exp. Clin. Cancer Res.* **2002**, *21* (3), 429–432.

(12) Delmas, P. D. Treatment of postmenopausal osteoporosis. *Lancet* **2002**, *359* (9322), 2018–2026.

(13) Coleman, R. E. Bisphosphonates: clinical experience. *Oncologist* **2004**, *9* (Suppl. 4), 14–27.

(14) Berenson, J. R. Treatment of hypercalcemia of malignancy with bisphosphonates. *Semin. Oncol.* **2002**, *29* (6 Suppl. 21), 12–18.

(15) Langston, A. L.; Ralston, S. H. Management of Paget's disease of bone. *Rheumatology (Oxford)* **2004**, *43* (8), 955–959.

(16) van Beek, E.; Pieterman, E.; Cohen, L.; Lowik, C.; Papapoulos, S. Farnesyl pyrophosphate synthase is the molecular target of nitrogen-containing bisphosphonates. *Biochem. Biophys. Res. Commun.* **1999**, *264* (1), 108–111.

(17) Bergstrom, J. D.; Bostedor, R. G.; Masarachia, P. J.; Reszka, A. A.; Rodan, G. Alendronate is a specific, nanomolar inhibitor of farnesyl diphosphate synthase. *Arch. Biochem. Biophys.* **2000**, *373* (1), 231–241.

(18) Dunford, J. E.; Thompson, K.; Coxon, F. P.; Luckman, S. P.; Hahn, F. M.; Poulter, C. D.; Ebetino, F. H.; Rogers, M. J. Structure-activity relationships for inhibition of farnesyl diphosphate synthase in vitro and inhibition of bone resorption in vivo by nitrogen-containing bisphosphonates. *J. Pharmacol. Exp. Ther.* **2001**, *296* (2), 235–242.

(19) Fisher, J. E.; Rodan, G. A.; Reszka, A. A. In vivo effects of bisphosphonates on the osteoclast mevalonate pathway. *Endocrinology* **2000**, *141* (12), 4793–4796.

(20) Rogers, M. J.; Gordon, S.; Benford, H. L.; Coxon, F. P.; Luckman, S. P.; Monkkonen, J.; Frith, J. C. Cellular and molecular mechanisms of action of bisphosphonates. *Cancer* **2000**, *88* (12 Suppl.), 2961–2978.

(21) Yardley, V.; Khan, A. A.; Martin, M. B.; Slifer, T. R.; Araujo, F. G.; Moreno, S. N.; Docampo, R.; Croft, S. L.; Oldfield, E. In vivo activities of farnesyl pyrophosphate synthase inhibitors against Leishmania donovani and Toxoplasma gondii. *Antimicrob. Agents Chemother.* **2002**, *46* (3), 929–931.

(22) Rodriguez, N.; Bailey, B. N.; Martin, M. B.; Oldfield, E.; Urbina, J. A.; Docampo, R. Radical cure of experimental cutaneous leishmaniasis by the bisphosphonate pamidronate. *J. Infect. Dis.* **2002**, *186* (1), 138–140.

(23) Cornforth, J. W.; Cornforth, R. H.; Popjak, G.; Yengoyan, L. Studies on the biosynthesis of cholesterol. XX. Steric course of decarboxylation of 5-pyrophosphomevalonate and of the carbon to carbon bond formation in the biosynthesis of farnesyl pyrophosphate. *J. Biol. Chem.* **1966**, *241* (17), 3970–3987.

(24) Patny, A.; Desai, P. V.; Avery, M. A. Ligand-supported homology modeling of the human angiotensin II type 1 (AT(1)) receptor: insights into the molecular determinants of telmisartan binding. *Proteins* **2006**, *65* (4), 824–842.

(25) Sabnis, Y.; Rosenthal, P. J.; Desai, P.; Avery, M. A. Homology modeling of falcipain-2: validation, de novo ligand design and synthesis of novel inhibitors. *J. Biomol. Struct. Dyn.* **2002**, *19* (5), 765–774.

(26) Sabnis, Y. A.; Desai, P. V.; Rosenthal, P. J.; Avery, M. A. Probing the structure of falcipain-3, a cysteine protease from Plasmodium falciparum: comparative protein modeling and docking studies. *Protein Sci.* **2003**, *12* (3), 501–509.

(27) Singh, N.; Cheve, G.; Avery, M. A.; McCurdy, C. R. Comparative protein modeling of 1-deoxy-D-xylulose-5-phosphate reductoisomerase enzyme from Plasmodium falciparum: a potential target for antimalarial drug discovery. *J. Chem. Inf. Model.* **2006**, *46* (3), 1360–1370.

(28) EBI Toolset: ClustalW2. http://www.ebi.ac.uk/clustalw (accessed Feb 2008).

(29) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234* (3), 779–815.

(30) RCSB Validation Server (V11.1A). http://deposit.pdb.org/validate (accessed Feb 2008).

(31) Luthy, R.; Bowie, J. U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **1992**, *356* (6364), 83–85.

(32) Godzik, A.; Kolinski, A.; Skolnick, J. Topology fingerprint approach to the inverse protein folding problem. *J. Mol. Biol.* **1992**, *227* (1), 227–238.

(33) Philipp, D. M. F.; Richard, A. Mixed Ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. *J. Comput. Chem.* **1999**, *20* (14), 1468–1494.

(34) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J. Comput. Chem.* **2000**, *21* (16), 1442–1457.

(35) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.

(36) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245* (1), 43–53.

(37) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graphics Modell.* **2002**, *20* (4), 281–295.

(38) PyMOL Home Page. http://pymol.sourceforge.net/ (accessed Feb 2008).

(39) Bairoch, A.; Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucleic Acids Res.* **1997**, *25* (1), 31–36.

(40) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.

(41) Gabelli, S. B.; McLellan, J. S.; Montalvetti, A.; Oldfield, E.; Docampo, R.; Amzel, L. M. Structure and mechanism of the farnesyl diphosphate synthase from Trypanosoma cruzi: implications for drug design. *Proteins* **2006**, *62* (1), 80–88.

(42) Mao, J.; Mukherjee, S.; Zhang, Y.; Cao, R.; Sanders, J. M.; Song, Y.; Zhang, Y.; Meints, G. A.; Gao, Y. G.; Mukkamala, D.; Hudock, M. P.; Oldfield, E. Solid-state NMR, crystallographic, and computational investigation of bisphosphonates and farnesyl diphosphate synthase-bisphosphonate complexes. *J. Am. Chem. Soc.* **2006**, *128* (45), 14485–14497.

(43) Rondeau, J. M.; Bitsch, F.; Bourgier, E.; Geiser, M.; Hemmig, R.; Kroemer, M.; Lehmann, S.; Ramage, P.; Rieffel, S.; Strauss, A.; Green, J. R.; Jahnke, W. Structural basis for the exceptional in vivo efficacy of bisphosphonate drugs. *ChemMedChem* **2006**, *1* (2), 267–273.

(44) Tarshis, L. C.; Yan, M.; Poulter, C. D.; Sacchettini, J. C. Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-A resolution. *Biochemistry* **1994**, *33* (36), 10871–10877.

(45) Hosfield, D. J.; Zhang, Y.; Dougan, D. R.; Broun, A.; Tari, L. W.; Swanson, R. V.; Finn, J. Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. *J. Biol. Chem.* **2004**, *279* (10), 8526–8529.

(46) Jorgensen, W. L. M.; David, S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225–11236.

(47) Still, W. C. T.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127–6129.

(48) Li, X.; Jacobson, M. P.; Friesner, R. A. High-resolution prediction of protein helix positions and orientations. *Proteins* **2004**, *55* (2), 368–382.

(49) Felts, A. K.; Gallicchio, E.; Wallqvist, A.; Levy, R. M. Distinguishing native conformations of proteins from decoys with an effective free energy estimator based on the OPLS all-atom force field and the Surface Generalized Born solvent model. *Proteins* **2002**, *48* (2), 404–422.

(50) Gherman, B. F.; Goldberg, S. D.; Cornish, V. W.; Friesner, R. A. Mixed quantum mechanical/molecular mechanical (QM/MM) study of the deacylation reaction in a penicillin binding protein (PBP) versus in a class C beta-lactamase. *J. Am. Chem. Soc.* **2004**, *126* (24), 7652–7664.

(51) Gherman, B. F.; Lippard, S. J.; Friesner, R. A. Substrate hydroxylation in methane monooxygenase: quantitative modeling via mixed quantum mechanics/molecular mechanics techniques. *J. Am. Chem. Soc.* **2005**, *127* (3), 1025–1037.

(52) Oelschlaeger, P.; Klahn, M.; Beard, W. A.; Wilson, S. H.; Warshel, A. Magnesium-cationic dummy atom molecules enhance representation of DNA polymerase beta in molecular dynamics simulations: improved accuracy in studies of structural features and mutational effects. *J. Mol. Biol.* **2007**, *366* (2), 687–701.

(53) Pang, Y. P. Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins* **2001**, *45* (3), 183–189.

(54) Pang, Y. P.; Xu, K.; Yazal, J. E.; Prendergas, F. G. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Sci.* **2000**, *9* (10), 1857–1865.

(55) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005**, *26* (16), 1752–1780.

(56) Diaz, N.; Suarez, D.; Sordo, T. L. Quantum chemical study on the coordination environment of the catalytic zinc ion in matrix metalloproteinases. *J. Phys. Chem. B* **2006**, *110* (47), 24222–24230.

(57) Guallar, V.; Jacobson, M.; McDermott, A.; Friesner, R. A. Computational modeling of the catalytic reaction in triosephosphate isomerase. *J. Mol. Biol.* **2004**, *337* (1), 227–239.

(58) Kravitz, J. Y.; Pecoraro, V. L.; Carlson, H. A. Quantum Mechanics/ Molecular Mechanics Calculations of the Vanadium Dependent Chloroperoxidase. *J. Chem. Theory Comput.* **2005**, *1* (6), 1265–1274.

(59) Cheng, Y.; Zhang, Y.; McCammon, J. A. How does the cAMP-dependent protein kinase catalyze the phosphorylation reaction: an ab initio QM/MM study. *J. Am. Chem. Soc.* **2005**, *127* (5), 1553–1562.

(60) Lin, P.; Pedersen, L. C.; Batra, V. K.; Beard, W. A.; Wilson, S. H.; Pedersen, L. G. Energy analysis of chemistry for correct insertion by DNA polymerase beta. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (36), 13294–13299.

(61) Alberts, I. L.; Wang, Y.; Schlick, T. DNA polymerase beta catalysis: are different mechanisms possible. *J. Am. Chem. Soc.* **2007**, *129* (36), 11100–11110.

(62) Wang, L.; Yu, X.; Hu, P.; Broyde, S.; Zhang, Y. A water-mediated and substrate-assisted catalytic mechanism for Sulfolobus solfataricus DNA polymerase IV. *J. Am. Chem. Soc.* **2007**, *129* (15), 4731–4737.

(63) Table 1 Metal Phosphate Interactions in Metalloproteins. http://tanna.bch.ed.ac.uk/mphos/mphos3wt.htm (accessed Feb 2008).

(64) Harding, M. M. Small revisions to predicted distances around metal sites in proteins. *Acta Crystallogr., Sect. D* **2006**, *62* (Pt 6), 678–682.

(65) Joly, A.; Edwards, P. A. Effect of site-directed mutagenesis of conserved aspartate and arginine residues upon farnesyl diphosphate synthase activity. *J. Biol. Chem.* **1993**, *268* (36), 26983–26989.

(66) Song, L.; Poulter, C. D. Yeast farnesyl-diphosphate synthase: site-directed mutagenesis of residues in highly conserved prenyltransferase domains I and II. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91* (8), 3044–3048.

(67) Tarshis, L. C.; Proteau, P. J.; Kellogg, B. A.; Sacchettini, J. C.; Poulter, C. D. Regulation of product chain length by isoprenyl diphosphate synthases. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93* (26), 15018–15023.

(68) Sanders, J. M.; Gomez, A. O.; Mao, J.; Meints, G. A.; Van Brussel, E. M.; Burzynska, A.; Kafarski, P.; Gonzalez-Pacanowska, D.; Oldfield, E. 3-D QSAR investigations of the inhibition of Leishmania major farnesyl pyrophosphate synthase by bisphosphonates. *J. Med. Chem.* **2003**, *46* (24), 5171–5183.

(69) Gold Performance as a Function of Protein Type. http://www.ccdc.cam.ac.uk/support/documentation/gold/3_2/doc/portable_html/gold_portable-3-214.html (accessed Feb 2008).

(70) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49* (20), 5912–5931.

(71) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42* (5), 791–804.

(72) Ha, S.; Andreani, R.; Robbins, A.; Muegge, I. Evaluation of docking/ scoring approaches: a comparative study based on MMP3 inhibitors. *J. Comput.-Aided Mol. Des.* **2000**, *14* (5), 435–448.

(73) Muegge, I.; Martin, Y. C.; Hajduk, P. J.; Fesik, S. W. Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. *J. Med. Chem.* **1999**, *42* (14), 2498–2503.