articles

molecular pharmaceutics

Structure and Specificity of a Human Valacyclovir Activating Enzyme: A Homology Model of BPHL

Insook Kim,[†] Gordon M. Crippen,[‡] and Gordon L. Amidon*,[†]

Department of Pharmaceutical Sciences and Department of Medicinal Chemistry, College of Pharmacy, The University of Michigan, 428 Church Street, Ann Arbor, Michigan 48109-1065

Received March 15, 2004

Abstract: Biphenyl hydrolase-like (BPHL) protein is a novel serine hydrolase which has been identified as human valacyclovirase (VACVase), catalyzing the hydrolytic activation of valine ester prodrugs of the antiviral drugs acyclovir and ganciclovir as well as other amino acid ester prodrugs of therapeutic nucleoside analogues. The broad specificity for nucleoside analogues as parent drugs suggests that BPHL may be particularly useful as a molecular target for prodrug activation. In order to develop an initial structural view of the specificity of BPHL, a homology model of BPHL based on the crystal structure of 2-hydroxy-6-oxo-7-methylocta-2,4-dienoate hydrolase was developed using the Molecular Operating Environment package (Chemical Computing Group, Montreal, Quebec), evaluated for its stereochemical quality and identification of free cysteines, and used in a molecular docking study. The BPHL model has residues S122, H255, and D227 comprising the putative catalytic triad in proximity and potential charge−charge interaction sites, M52 or D123 for the α-amino group. The model also suggested that the structural preference of BPHL for hydrophobic amino acyl promoieties and its limited activity for the secondary alcohol substrates may be attributed to the hydrophobic acyl-binding site formed by residues I158, G161, I162, and L229, and the spatial constraint around the catalytic site by a loop on one side, the active serine and histidine on the other side, and L53 and L179 on top. In addition, the broad specificity for nucleoside analogues may be due to the relatively less constrained nucleoside-binding site opening toward the entrance of the substrate-binding pocket. The homology model of BPHL provides a basis for further investigation of the catalytic and active site residues, can account for the observed structure activity profile of BPHL, and will be useful in the design of nucleoside prodrugs.

Keywords: Homology model; human valacyclovirase, biphenyl hydrolase-like; BPHL; prodrugs; α/β-hydrolase fold; substrate specificity; hydrolysis; prodrug activation; nucleoside analogue; antiviral; anticancer; drug delivery; prodrug design

Introduction

The reliable activation of prodrugs is of critical importance for effecting the pharmacological activity of the parent drug. Yet, often the enzyme(s) responsible for activation is unidentified. Recently, we have identified a novel prodrug activating enzyme, biphenyl hydrolase-like protein (BPHL), as a human valacyclovirase (VAVCase) that catalyzes the hydrolytic activation of ester prodrugs of the nucleoside analogues valacyclovir and valganciclovir.[1] Further, we have shown that BPHL hydrolyzes amino acid ester prodrugs of

* To whom correspondence should be addressed: College of Pharmacy, The University of Michigan, 428 Church St., Ann Arbor, MI 48109-1065. Tel: 734-764-2440. Fax: 734-763-6423. E-mail: glamidon@umich.edu.

[†] Department of Pharmaceutical Sciences.

[‡] Department of Medicinal Chemistry.

(1) Kim, I.; Chu, X.-Y.; Kim, S.; Provoda, C. J.; Lee, K.-D.; Amidon, G. L. Identification of a Human Valacyclovirase: Biphenyl Hydrolase-Like Protein as Valacyclovir Hydrolase. *J. Biol. Chem.* **2003**, *278*, 25348−25356.

10.1021/mp049959+ CCC: $27.50
Published on Web 09/24/2004

various other nucleoside analogues, floxuridine, gemcitabine, zidovudine, and 2-bromo-5,6-dichloro-1-(β-D-ribofuranosyl) benzimidazole.[2] BPHL exhibits a preference for hydrophobic amino acids such as valine, phenylalanine, and proline as a promoiety irrespective of the nucleoside leaving group, and prefers the primary alcohol, 5′-esters to the secondary alcohol, 3′-esters as leaving groups. In addition, BPHL is highly selective for an L-isomer amino acid promoiety over a corresponding D-isomer promoiety and exhibits greater stereoselectivity for the valine isomers than for the phenylalanine isomers.[2] BPHL is an intracellular enzyme, which was purified from a crude mitochondrial fraction of Caco-2 cells[1] and homologous to several peptide sequences from a previously identified rat liver valacyclovirase.[1,29] There are no reports of a secreted form of BPHL. High expression of BPHL in human liver and kidney and relatively low expression in small intestine, heart, and skeletal muscle suggest a potential role in detoxification of xenobiotics[3] as well as in prodrug activation in these organs. BPHL is a novel serine esterase whose activity was significantly inhibited by serine hydrolase inhibitors such as Pefabloc SC,[1] and diisofluorophosphate[3] as well as by a free cysteine modifying agent *p*-chloromercuribenzoic acid (PCMB) but not by a cysteine hydrolase inhibitor E-64 nor a reducing agent dithiothreitol.[1]

BPHL does not appear to have significant similarity with other sequences in the human genome; however, the homology database search has revealed potential orthologues in

(2) Kim, I.; Song, X.; Vig, B. S.; Mittal, S.; Shin, H.-C.; Lorenzi, P. J.; Amidon, G. L. A Novel Nucleoside Prodrug Activating Enzyme: Substrate Specificity of Biphenyl Hydrolase-like Protein. *Mol. Pharm.* **2004**, *1*, 117−127.

(3) Puente, X. S.; Lopez-Otin, C. Cloning and Expression Analysis of a Novel Human Serine Hydrolase with Sequence Similarity to Prokaryotic Enzymes Involved in the Degradation of Aromatic Compounds. *J. Biol. Chem.* **1995**, *270*, 12926−12932.

(4) Puente, X. S.; López-Otin, C. The PLEES proteins: a family of structurally related enzymes widely distributed from bacteria to humans. *Biochem. J.* **1997**, *322*, 947−949.

(5) Fushinobu, S.; Saku, T.; Hidaka, M.; Jun, S.-Y.; Nojiri, H.; Yamane, H.; Shoun, H.; Omori, T.; Wakagi, T. Crystal structures of a meta-cleavage product hydrolase from Pseudomonas fluorescens IP01 (CumD) complexed with cleavage products. *Protein Sci.* **2002**, *11*, 2184−2195.

(6) Cuff, J. A.; Clamp, M. E.; Siddiqui, A. S.; Finlay, M.; Barton, G. J. Jpred: A Consensus Secondary Structure Prediction Server. *Bioinformatics* **1998**, *14*, 892−893.

(7) McGuffin, L. J.; Bryson, K.; Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **2000**, *16*, 404−405.

(8) Deleage, G.; Blanchet, C.; Geourjon, C. Protein structure prediction. Implications for the biologist. *Biochimie* **1997**, *79*, 681−686.

(9) King, R. D.; Sternberg, M. J. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci*. **1996**, *5*, 2298−2310.

(10) Garnier, J.; Gibrat, J.-F.; Robson, B. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* **1996**, 540−553.

(11) Guermeur, Y. Combinaison de classifieurs statistiques, Application a la prediction de structure secondaire des proteines. Ph.D. Thesis, 1997.

(12) Rost, B. PHD: predicting 1D protein structure by profile based neural networks. *Methods Enzymol.* **1996**, *266*, 525−539.

(13) Frishman, D.; Argos, P. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng.* **1996**, *9*, 133−142.

(14) Levin, J. M.; Robson, B.; Garnier, J. An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS Lett.* **1986**, *205*, 303−308.

(15) Geourjon, C.; Deleage, G. SOPM: a self-optimized method for protein secondary structure prediction. *Protein Eng.* **1994**, *7*, 157−164.

(16) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.

(17) Chou, K.-C.; Howe, W. J. Prediction of the Tertiary Structure of the [beta]-Secretase Zymogen. *Biochem. Biophys. Res. Commun.* **2002**, *292*, 702−708.

(18) Daines, R. A.; Pendrak, I.; Sham, K.; Van Aller, G. S.; Konstantinidis, A. K.; Lonsdale, J. T.; Janson, C. A.; Qiu, X.; Brandt, M.; Khandekar, S. S.; Silverman, C.; Head, M. S. First X-ray Cocrystal Structure of a Bacterial FabH Condensing Enzyme and a Small Molecule Inhibitor Achieved Using Rational Design and Homology Modeling. *J. Med. Chem.* **2003**, *46*, 5−8.

(19) Howarth, N. M.; Purohit, A.; Robinson, J. J.; Vicker, N.; Reed, M. J.; Potter, B. V. L. Estrone 3-Sulfate Mimics, Inhibitors of Estrone Sulfatase Activity: Homology Model Construction and Docking Studies. *Biochemistry* **2002**, *41*, 14801−14814.

(20) Goettig, P.; Groll, M.; Kim, J.-S.; Huber, R.; Brandstetter, H. Structures of the tricorn-interacting aminopeptidase F1 with different ligands explain its catalytic mechanism. *EMBO J.* **2002**, *21*, 5343−5352.

(21) Hofmann, B.; Tolzer, S.; Pelletier, I.; Altenbuchner, J.; van Pee, K. H.; Hecht, H. J. Structural investigation of the cofactor-free chloroperoxidases. *J. Mol. Biol.* **1998**, *279*, 889−900.

(22) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

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

(24) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.

(25) Clauser, K. R.; Baker, P. R.; Burlingame, A. L. Role of accurate mass measurement (±10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* **1999**, *71*, 2871.

(26) Halgren, T. A. The Merck Force Field, I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(27) McGuffin, L.; Jones, D. T. Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics* **2003**, *19*, 874−881.

(28) Heikinheimo, P.; Goldman, A.; Jeffries, C.; Ollis, D. Of barn owls and bankers: a lush variety of α/β hydrolases. *Structure* **1999**, *7*, R141−R146.

(29) Burnette, T. C.; Harrington, J. A.; Reardon, J. E.; Merrill, B. M.; Miranda, P. d. Purification and Characterization of a Rat Liver Enzyme That Hydrolyzes Valaciclovir, the L-Valyl Ester Prodrug of Acyclovir. *J. Biol. Chem.* **1995**, *270*, 15827−15831.

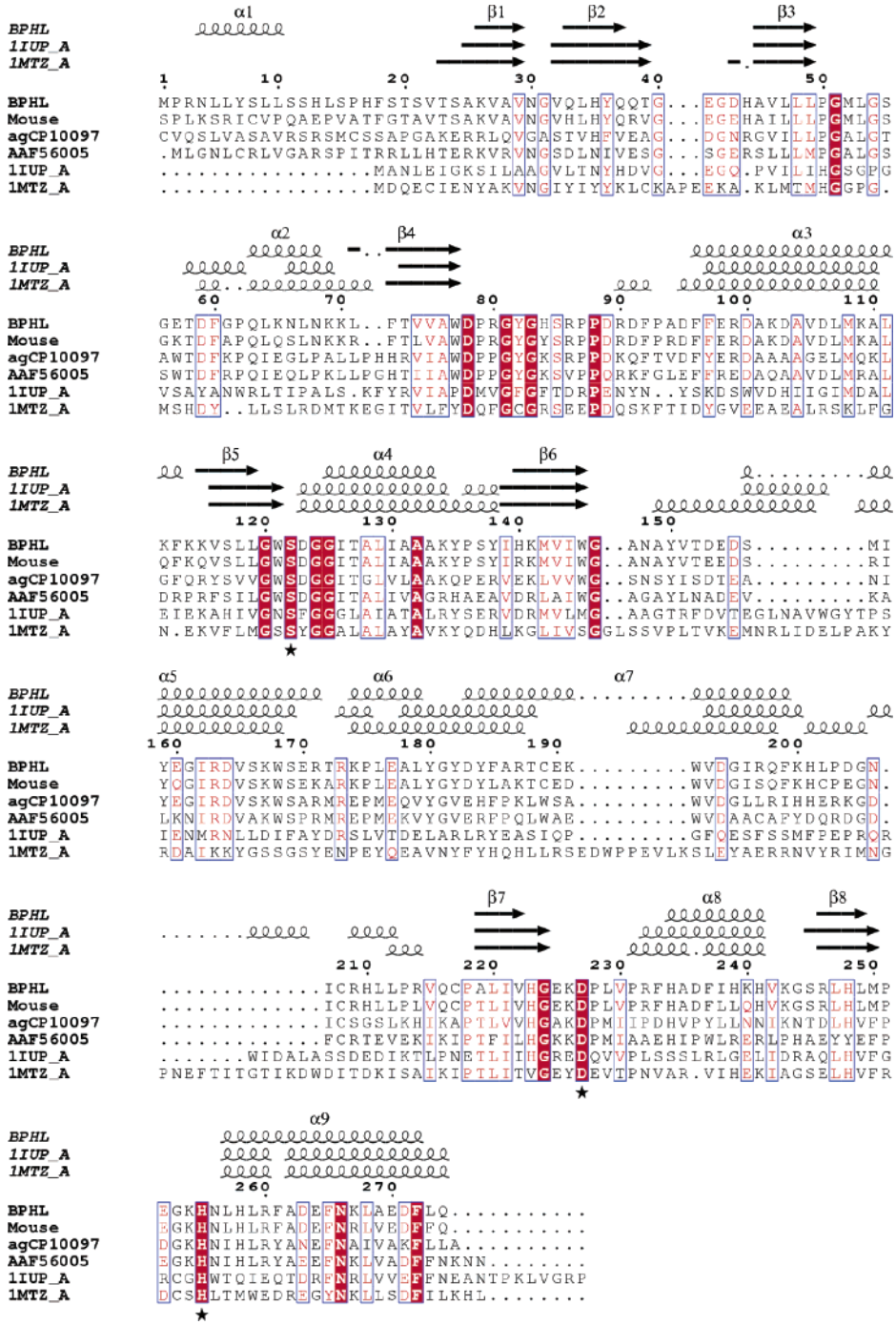

Figure 1. Multiple sequence alignment between BPHL and its homologues. Secondary structures of 1IUP.A and 1MTZ.A and the predicted secondary structure of BPHL are shown on top. Helices are presented as coils and strands as arrows. The catalytic residues are marked with stars. BPHL: biphenyl hydrolase-like (NP004323). Mouse: biphenyl hydrolase-like (AAH23146). agCP10097: hypothetical protein (EAA07406, *Anopheles gambiae str.* PEST). AAF56005: CG5377-PA (*Drosophila melanogaster*).

other eukaryotes such as mouse, zebrafish, fruitfly, and mosquito.[1] BPHL was named on the basis of the sequence similarity (~30% similarity) to microbial biphenyl hydrolases such as 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (BPHD) although BPHL did not hydrolyze biphenyls.[3] The multiple sequence alignment of BPHL and other similar sequences suggested the putative catalytic triad composed of S122-H255-D227 and the serine hydrolase signature motif, G-X-S-X-G.[3,4]

In this study, a homology model of BPHL was constructed based on the crystal structure of a meta-cleavage product hydrolase, 2-hydroxy-6-oxo-7-methylocta-2,4-dienoate hydrolase (PDB code: 1IUP.A)[5] to gain structural insights into the substrate specificity of BPHL. The substrate-binding site

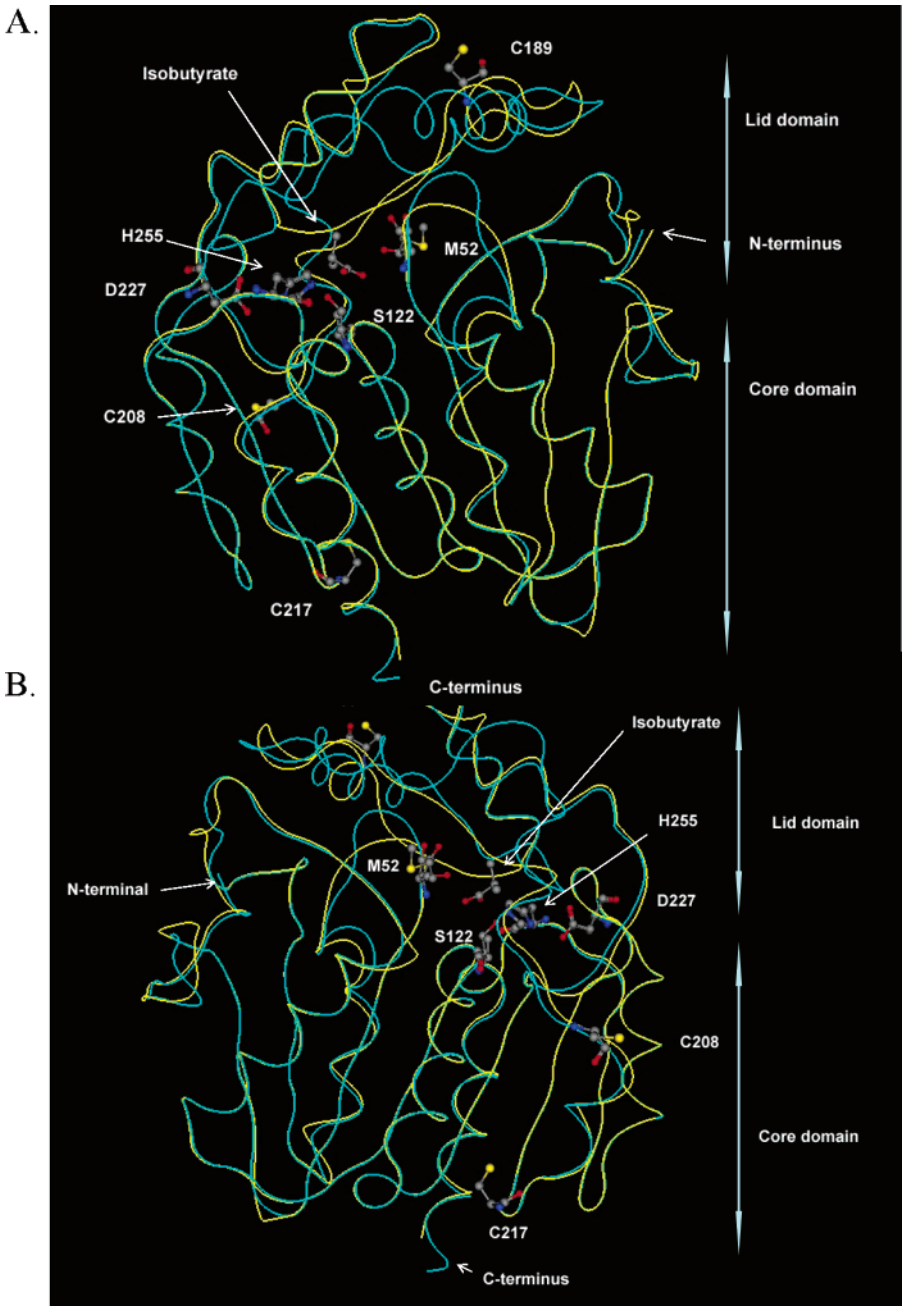

Figure 2. The superposition of a backbone trace of BPHL (in yellow) on the crystal structure of CumD complexed with a product ligand, isobutyrate (I1UP.A, in blue): (A) view from the entrance of the binding site; (B) view from the opposite side.

and specific residues likely involved in substrate specificity and catalysis were identified.

Experimental Section

Secondary Structure Prediction. The secondary structure of BPHL was predicted by JPred[6] and PSI-Pred,[7] and the prediction of a consensus secondary structure was performed via the NPS@ Web server (http://npsa-pbil.ibcp.fr)[8] using the DSC,[9] GOR4,[10] HNN,[11] PHD,[12] Predator,[13] SIMPA96,[14] and SOPM[15] methods.

Homology Modeling of BPHL. The sequence of BPHL was obtained from the nonredundant (nr) protein database (NCBI accession number NP004323) and PSI-Blast search against the nr protein database (ref 16; Blosum 62, gap penalty 11-1, $E < 0.0005$ for iterations) via the NCBI website (http://www.ncbi.nlm.nih.gov) identified 2-hydroxy-6-oxo-7-methylocta-2,4-dienoate hydrolase from *Pseudomonas fluorescens* (CumD; NCBI accession number: BAA12150) as the best scoring sequence with an X-ray structure. Out of four crystal structures of CumD with different ligands, BPHL models from S19 to Q274 were built on the tertiary structure of S103A mutant CumD complexed with isobutyrate (PDB code: 1IUP.A).[5] All computations were performed with the Molecular Operating Environment (MOE) package.[17−19]

Briefly, a secondary structure driven multiple sequence alignment was performed (MOE-Align; Blosum 62 substitu-

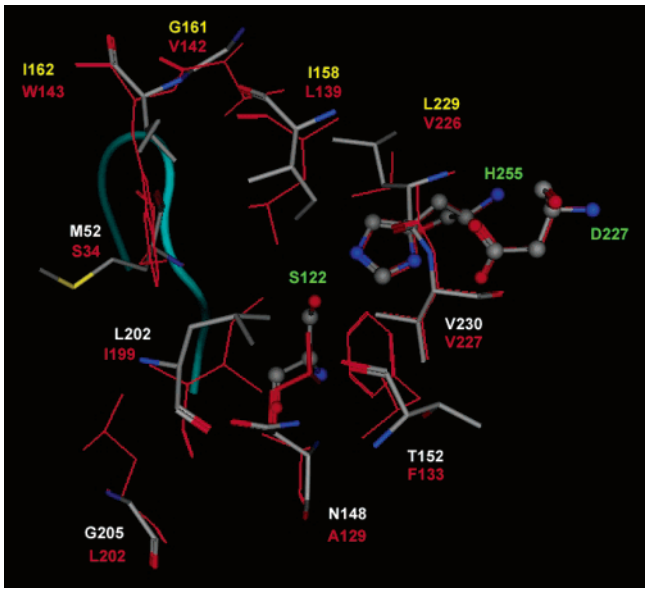

Figure 3. Superposition of the homology model of BPHL on CumD. The BPHL model was superposed on CumD (1IUP.A). Numbers are based on BPHL, and hydrogen atoms are not shown. The catalytic triad is shown in a ball-and-stick mode (number in green), and the loop forming one side of the catalytic site is shown in blue ribbon. BPHL residues corresponding to CumD residues involved in the recognition of the isopropyl group of the ligand and in the formation of the deeper space of the D-part are labeled in yellow and white, respectively.

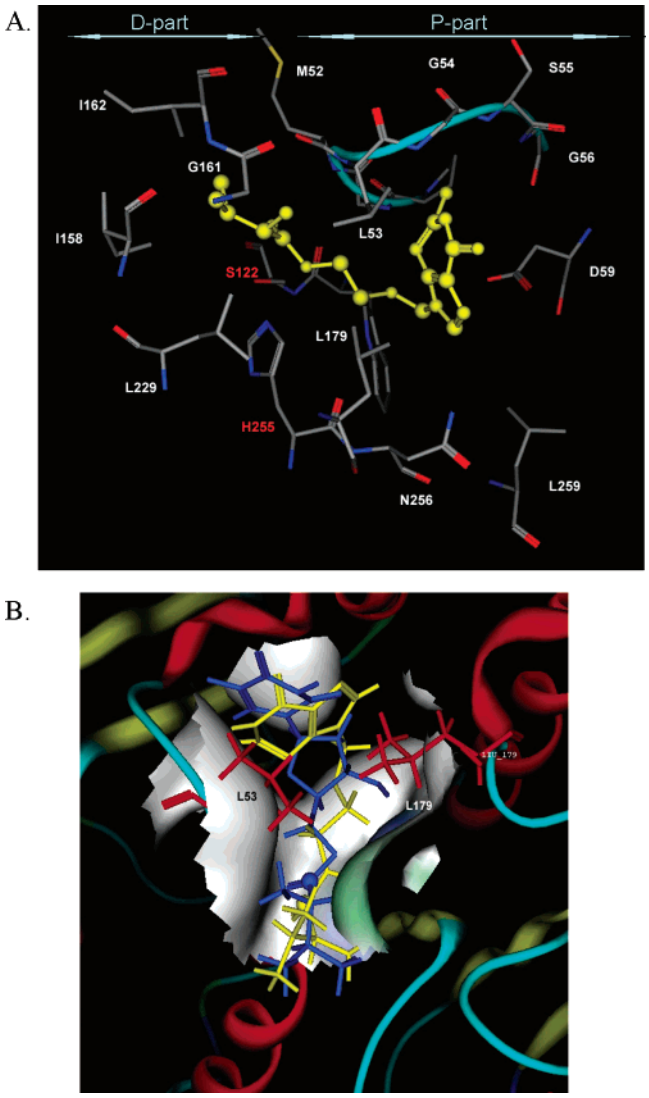

Figure 4. Molecular docking of ligands to BPHL model. (A) Residues in proximity of the docked L-VACV in BPHL. (B) L-VACV (in yellow) and L-Val-FUdR (in blue) are shown on the molecular surface of the substrate-binding pocket, L53 and L179 (in red) are shown in the stick mode, and carbonyl carbons of L-VACV and L-Val-FUdR are shown as a ball.

tion matrix, gap penalty 11-1) among BPHL and similar sequences including sequence-only eukaryotic homologues (accession number: AAH23146, AAF56005, and EAA07406) and putative structural homologues 1IUP.A, proline iminopeptidase (PDB code 1MU0),[20] and chloroperoxidase (PDB code 1A8S)[21] (Figure 1). Multiple sequence alignment was used to identify conserved residues which may be structurally important. The atomic coordinates of the protein structures were obtained from the PDB database.[22] Conserved residues among all aligned sequences were identified, and among them residues whose Cα positions were within 1 Å root mean square distance (rmsd) among the crystal structures upon superposition were fixed. The initial geometry for the BPHL model was taken from that of 1IUP.A after the multiple sequence alignment. Where residue identity was conserved between the template and the model, all coordinates were copied; otherwise, only backbone coordinates were used. However, the coordinates of the active serine were copied from the aligned 1A8S, since the active serine in 1IUP.A was mutated to alanine. The models were constructed using the potential energy model AMBER94[23] with the solvation term. The first 10 intermediate models with different loop candidates and side chain rotamers were generated, and the best intermediate model was subjected to energy minimization with an rms gradient test of 1 to produce a final model. The generated models were visually inspected for the proper geometry of the catalytic triad, and the stereochemical quality of the final model was evaluated for bond angle, length, and dihedral (MOE-Protein Report, Whatif, and Procheck (http://biotech.ebi.ac.uk:8400/)[24]). The selected final model was used for structure analysis and molecular docking studies.

Identification of Free Cysteines. The activity of BPHL was completely abolished by PCMB, while a reducing agent, dithiothreitol, did not significantly alter BPHL activity,[1] which suggested the importance of free cysteines but not a disulfide bond on BPHL activity. BPHL has three cysteines, C189, C208, and C217, and possibly one disulfide bond could be present and at least one free cysteine should be present. In the BPHL model, nonetheless, three cysteines were located at least 14 Å apart from each other, and two of them, C208 and C217, were positioned on the core domain while C189 was on the lid domain, and all three of them were located on the solvent accessible surface, suggesting

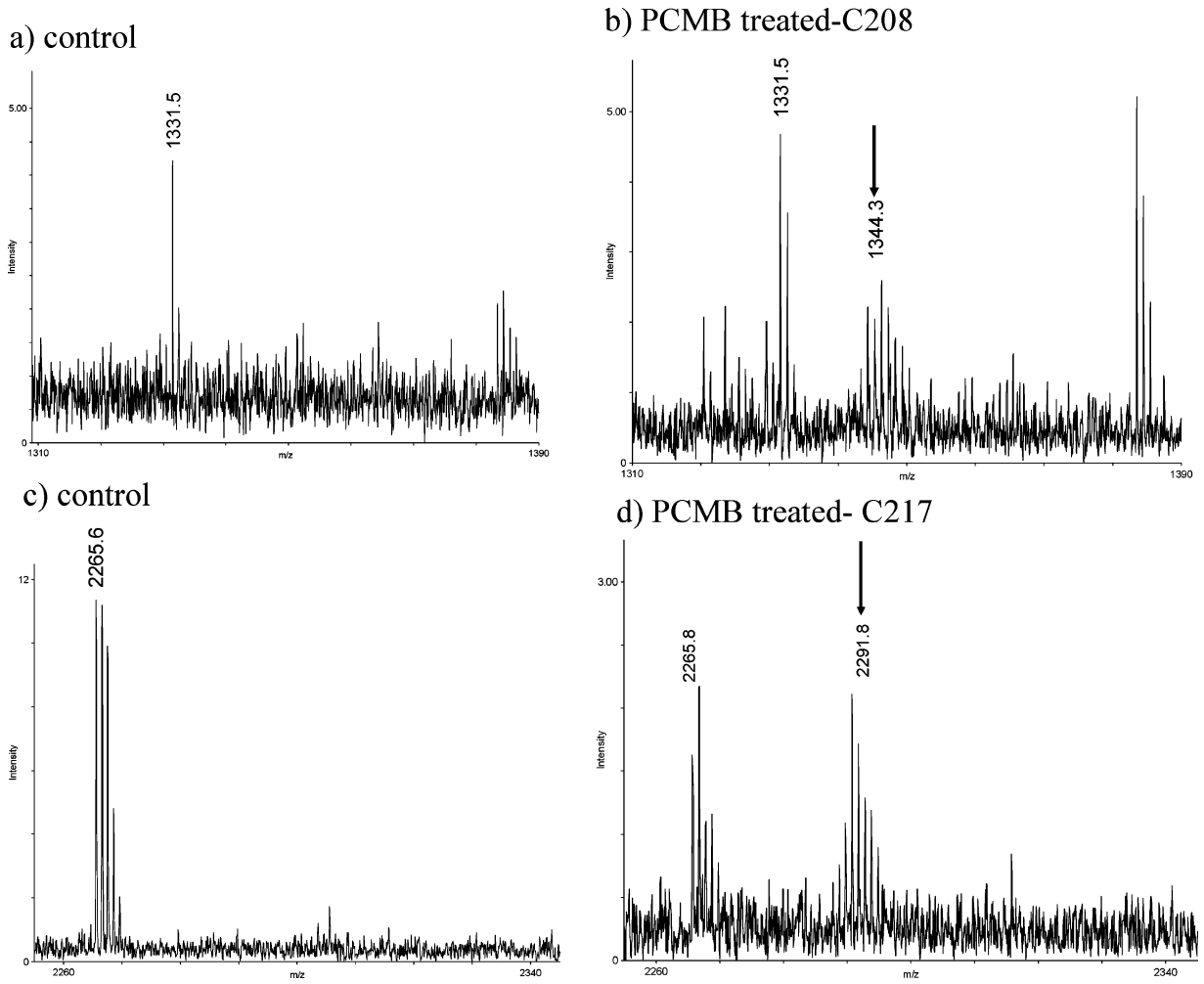

Figure 5. Mass spectra of peptide ions of BPHL containing a cysteine: mass spectra of peptide ions of BPHL in control (a, c) and samples treated with PCMB (b, d); (b) the peptide with modified C208 and (d) the peptide with the modified C217 are indicated by arrows.

that the presence of an intramolecular disulfide bond is unlikely (Figure 2). This hypothesis was tested by the identification of free cysteines using a free thiol modifier, PCMB. Briefly, the recombinant BPHL was prepared as described previously.[1] Approximately 10 pmol of BPHL (310 ng) was incubated with 0.3 mM of PCMB in 100 μL reaction volume at 37 °C for 20 min. Unreacted chemicals were removed by 80% acetone at −20 °C overnight, followed by centrifugation for 10 min at 4 °C. The pellet was rinsed with 100 μL of prechilled acetone three times and air-dried. The BPHL without chemical labeling served as a control. The sample was digested with the sequence grade-modified trypsin (Promega, Madison, WI) at 1:50 w/w ratio (enzyme: protein) in 50 μL of fresh 100 mM ammonium bicarbonate for 18 h at 40 °C. The digestion was quenched with 1% (final) formic acid. Peptides were collected, pooled, and dried in a vacuum centrifuge. The digest was redissolved in aqueous 0.1% trifluoroacetic acid and purified by micro ZipTip C18 pipet tips (Millipore, Billerica, MA) following the manufacturer's protocol. Molecular masses of peptides were determined by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) using α-cyano-4-hydroxycinnamic acid (5 mg/mL) in 50% (v/v) acetonitrile, 0.1% TFA as a matrix. In brief, the peptides were eluted with 1 μL of 70% acetonitrile with 0.1% TFA directly from the ZipTip onto the MALDI target with matrix solution and air-dried. MALDI-TOF-MS was performed on an Applied Biosystems 4700 TOF/TOF (Toronto, Canada) at the Michigan Proteome Consortium at the University of Michigan. The spectra were scanned in the range between *m*/*z* 800 and 4000. The tandem mass spectrometry product ion spectra were recorded on the same instrument for selected peptides. The mass of peptide digests was predicted by ProteinProspector (http://prospector.ucsf.edu/),[25] and the spectra were analyzed by Moverz (Proteometrics, LLC, New York, NY).

Molecular Docking of Substrates into the Active Site of BPHL. The structures of all compounds were modeled using the MOE-Molecule Builder. The initial structures were minimized using molecular mechanics with the MMFF94 force field.[26] A test molecule was manually positioned to the putative active site similarly to the isobutyrate bound in CumD by superposition of the BPHL model on 1IUP.A. Docking of a substrate such as D/L-valacyclovir (D-VACV

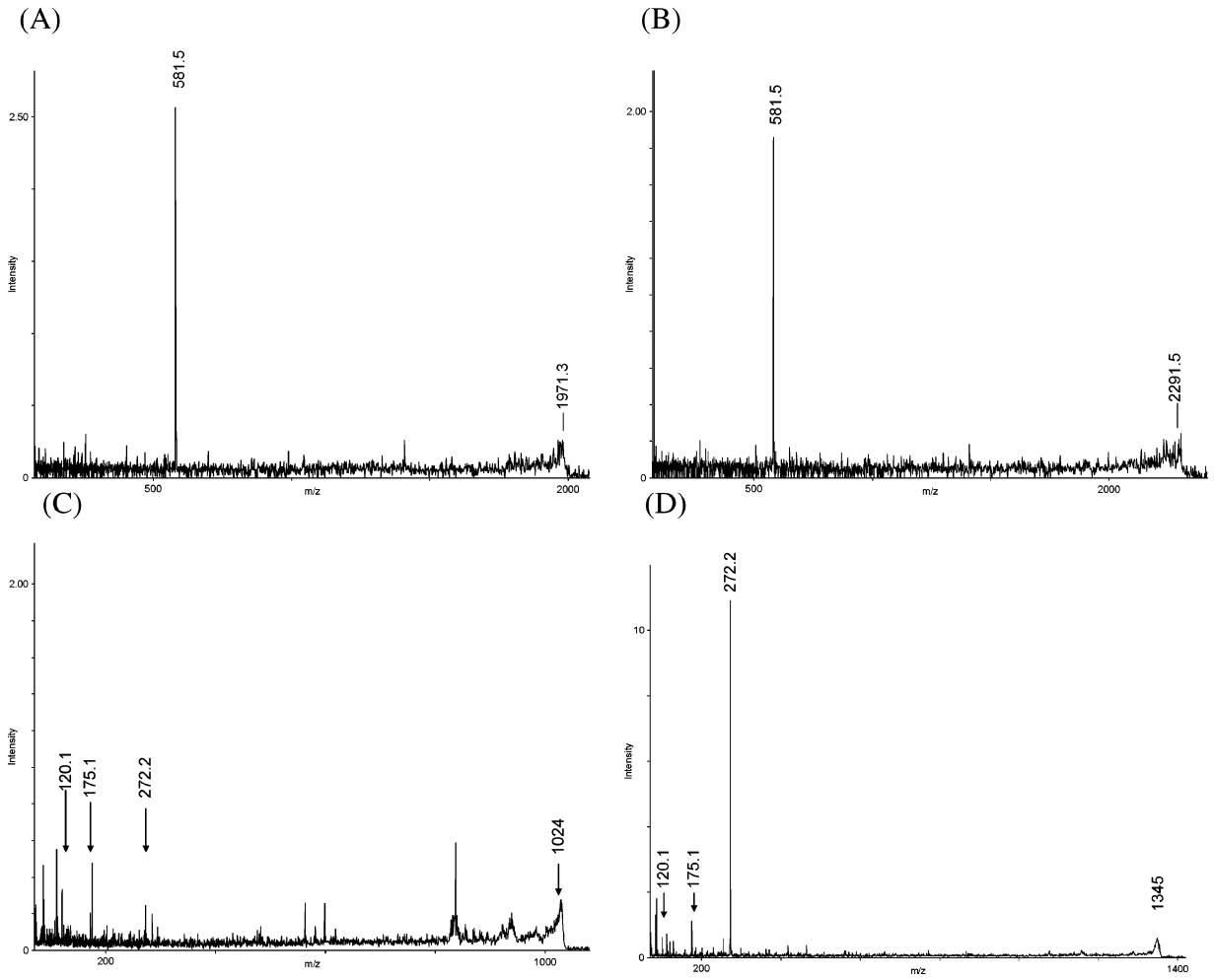

Figure 6. Tandem MS analysis of peptide fragments with a cysteine residue. Peptide fragments with a cysteine C217 and C208 (A and C, respectively) and a modified C217 and C208 (B and D, respectively) were analyzed by tandem MS. Representative peaks used for comparison between the control and the modified sample were marked.

and L-VACV) and D/L-5′-valyl floxuridine ester (D-Val-FUdR and L-Val-FUdR) to (rigid body) BPHL was carried out by simulated annealing with the MMFF94 force field[26] with the solvation term (distance dependent dielectric constant) by MOE-DOCK. A docking box was designated around a manually positioned prodrug with 50 × 50 × 50 grid points and a 0.375 Å grid spacing.

Results

Secondary Structure Prediction and Template Search. Initial analysis, consistent with the previous prediction,[3,4] suggested that BPHL be a member of the α/β-hydrolase fold protein family based on the sequence similarity with other α/β-hydrolase fold enzymes, secondary structure predictions, and fold recognition. BPHL was predicted to have secondary structural elements composed of eight β-strands and five intervening α-helices with a helix-only domain inserted between 6β and 7β and an additional helix at the N-terminus (Figure 1). In addition, a fold-recognition algorithm, mGenTHREADER[27] suggested that BPHL exhibits significant similarity to the α/β-hydrolase fold family (data not shown).

PSI-Blast search identified CumD, proline iminopeptidase, and chloroperoxidase as potential structural homologues. Although the sequence homology between BPHL and the selected CumD falls in the "twilight zone" with a sequence identity of 20−30%, the predicted secondary structure elements of BPHL were well overlaid with that of CumD, proline iminopeptidase, and chloroperoxidase, all of which belong to the α/β-hydrolase fold family (Figure 1). In addition to the predicted global fold resemblance, CumD was considered a reasonable template since it is a serine hydrolase with the conserved catalytic triad Ser-His-Asp and a serine hydrolase signature motif G-X-S-X-G, as is BPHL, suggesting that the geometry of the catalytic site of BPHL is likely similar to that of CumD, although a functional similarity between BPHL and CumD is unknown.[5,28]

Homology Modeling of BPHL. The final model has a normal stereochemistry, and 90% of the residues were in generously allowed regions of the Ramachandran plot (77% in core region). Overall the homology model of BPHL was well superposed on the template CumD. The rmsd between BPHL and CumD over 254 pairs of superimposed Cα positions was 1.07 Å. While the N-terminal 18 amino acid

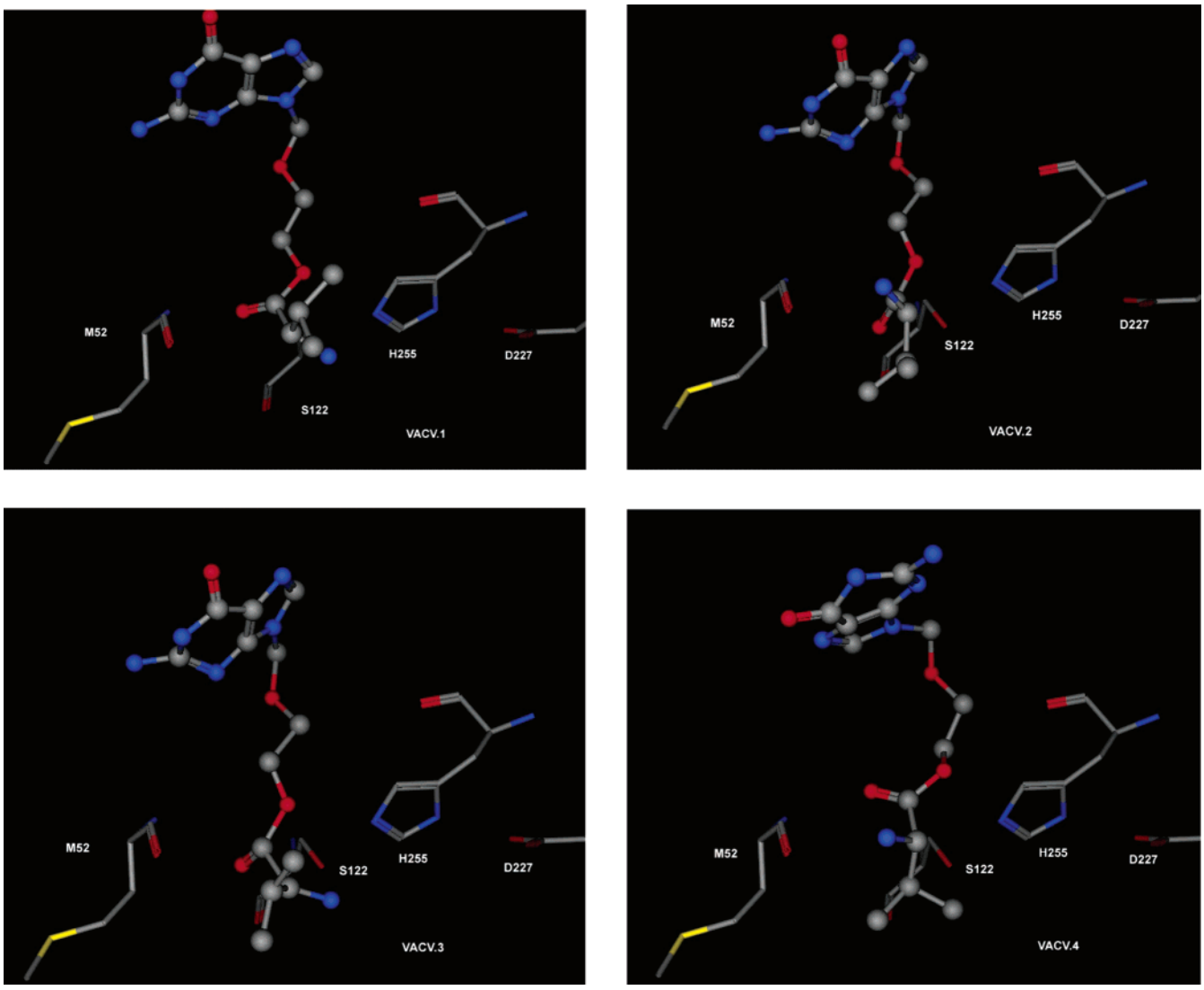

Figure 7. Suggested binding modes of L-VACV to BPHL.

region was not modeled due to lack of a corresponding template region, it was predicted to have an α-helix structure (Figure 1). The structure of BPHL can be divided into two domains as proposed for the CumD structure, the core domain (residues 19−152 and 201−274), which takes a well-conserved α/*β*-hydrolase fold, and the lid domain (residues 153−200), which is composed of mainly α helices (Figure 2).[5] As expected from the well-conserved characteristics of the α/*β*-hydrolase fold family, the core domain of BPHL was modeled consistently, while the lid domain was modeled with higher variability. The rmsd for the core domain was 0.47 Å, with 215 pairs of Cα atoms, while that for lid domain was 2.26 Å, with 47 pairs of Cα atoms. For the lid domain, the rmsd over 28 pairs of superposed Cα positions of the N-terminal region of the lid domain of BPHL corresponding to α4 and α7 of CumD was 1.18 Å, whereas the rmsd of Cα positions for the C-terminal half of the lid domain region from Y182 to K200 of BPHL was 3.3 Å. The higher deviation on the lid domain can be attributed to 6 and 9 amino acid long gaps introduced in the sequence alignment.

Active Site of BPHL. Three residues predicted to comprise the catalytic triad, S122, H255, and D227, are positioned in proximity and well superposed on that of CumD (1IUP.A) (Figure 3). The putative catalytic triad is located on the core domain at the hinge region between the core and the lid domains (Figure 2). As shown in other α/*β*-hydrolase fold proteins,[28,30] the active S122 is on a sharp turn between *β*5 and α3, the so-called "nucleophile elbow", and H255 and D227 are positioned on a loop following *β*7 and *β*8, respectively. The substrate-binding site of BPHL was inferred by the superposition of the BPHL model on 1IUP.A complexed with a product ligand, isobutyrate. The binding site for the product ligands isobutyrate or acetate in CumD (1IUP.A and 1IU0, respectively) was consistent.[5]

The substrate-binding pocket of BPHL is a gorge or groove type and can be divided into two parts by the S122 residue as suggested for CumD, the proximal and the distal parts to the entrance, P-part and D-part, respectively (Figure 4A).[5] By the superposition of BPHL on CumD, residues likely to be involved in the formation of the D-part of BPHL were

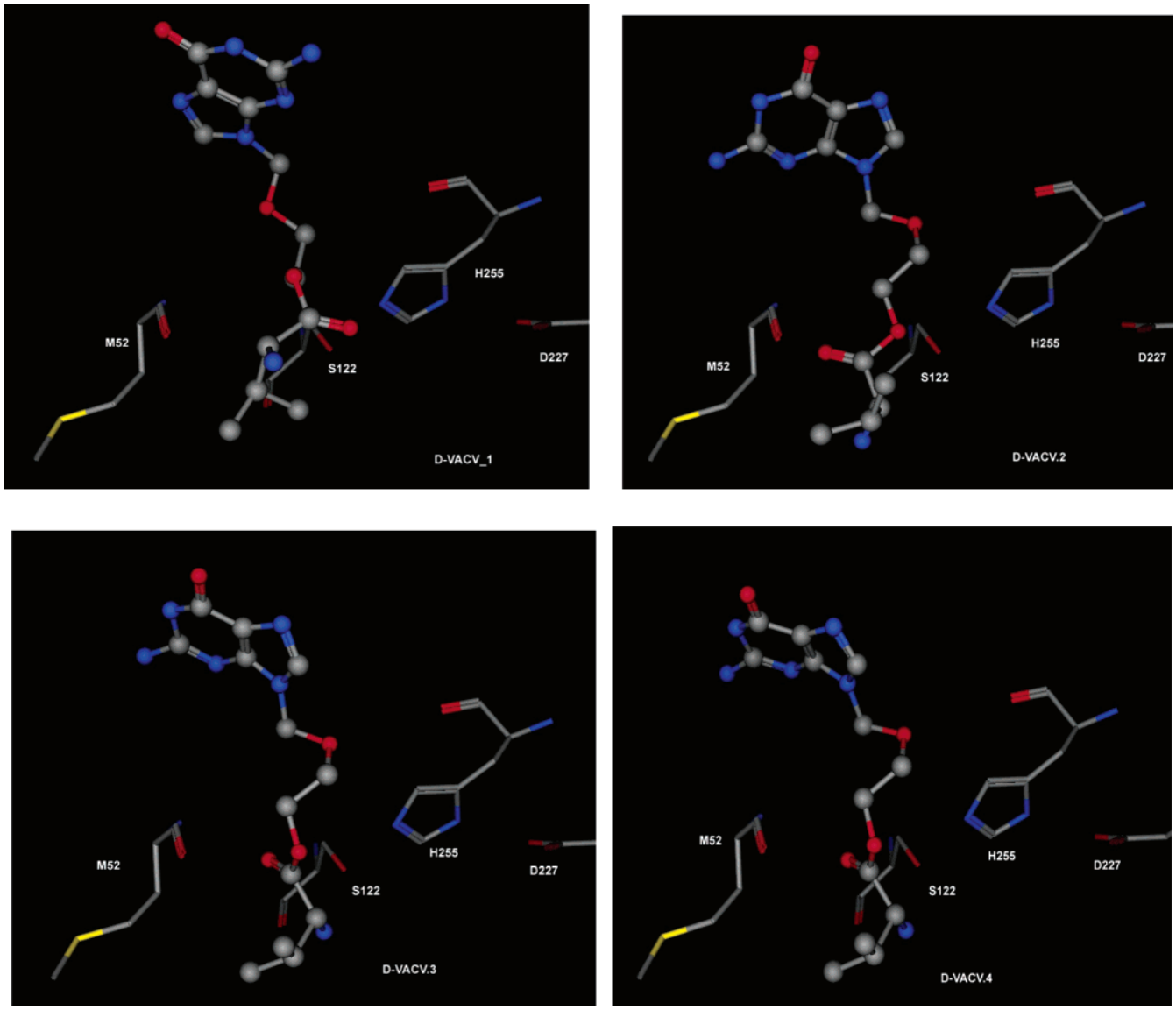

Figure 8. Suggested binding modes of D-VACV to BPHL.

identified (Figure 4A). The residues I158, G161, I162, and L229 of BPHL correspond to residues L139, V142, W143, and V226 of CumD, respectively, which are involved in recognition of the isopropyl group of the ligand bound in CumD (Figure 3), and residues N148, T152, L202, G205, and V230 of BPHL are equivalent to CumD residues involved in the formation of the deeper space of the D-part similar to A129, F133, I199, L202, and V227, respectively.[5] Additionally, M52 and D123 of BPHL are equivalent to S34 and F104, respectively, which form a hydrogen bond with the carbonyl oxygen of isobutyrate in CumD.[5]

Identification of Free Cysteines in BPHL. In order to confirm the number of free cysteines, BPHL was labeled by PCMB and adducts on free cysteines were identified by MALDI-TOF mass spectrometry. Precursor ions of 1024 and 1971 in mass were predicted to contain C208 and C217, respectively, and corresponding modified precursor ions of 1345 and 2292 in mass, respectively, were identified (Figure 5) and further confirmed by tandem mass spectrometry (Figure 6). However, the peptide with C189 was not identifiable either in a control or in a sample because its mass was below the detection range. This suggests that C208 and C217 of BPHL have a free thiol and C189 is presumably in a free form since there is no cysteine remaining to form an intramolecular disulfide bond.

Molecular Docking Study. A molecular docking study was performed with substrates, L-VACV and L-Val-FUdR, and with poor substrates, D-VACV and D-Val-FUdR. All the substrates were able to bind to the identified binding site, and, in general, orientations of most of the conformations of docked substrates were consistent with that of isobutyrate in CumD such that the acyl group resided on the D-part while

(30) Ollis, D.; Cheaf, E.; Cygler, M.; Dijkstra, B.; Frolow, F.; Franken, S. M.; Harel, M.; Remington, S. J.; Silman, I.; Schrag, J.; Sussman, J. L.; Verschueren, K. H. G.; Goldman, A. The α/β hydrolase fold. *Protein Eng.* **1992**, *5*, 197−211.

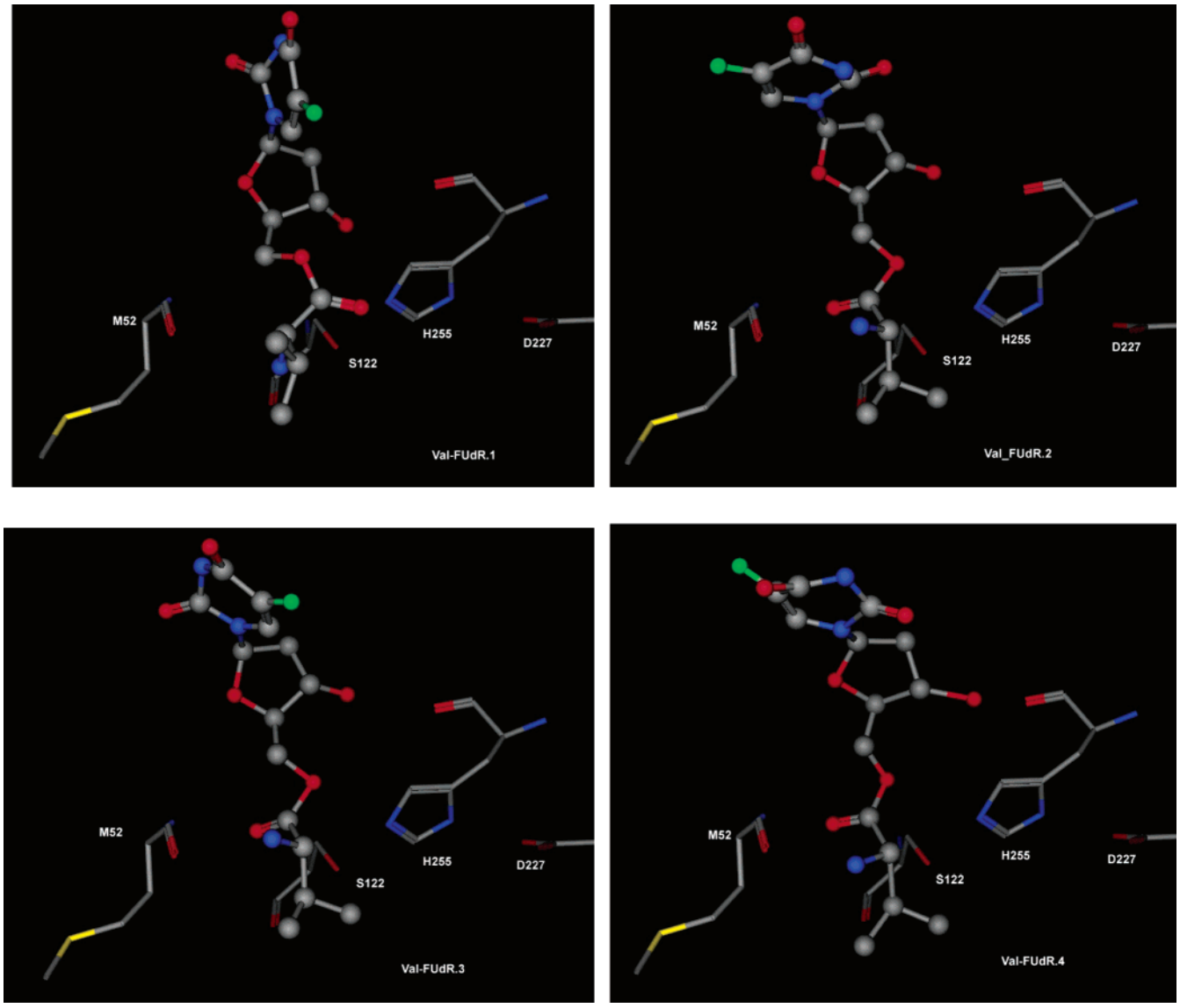

Figure 9. Suggested binding models of L-Val-FUdR to BPHL.

the alcohol leaving group resided on the P-part of the binding site (Figure 4A). The carbonyl carbon of substrates of the best scored docked conformations was positioned close to the active serine (2.64−4.09 Å apart). The residues L53, L179, D59, N256, and L259 were positioned within 4.5 Å of a docked L-VACV and likely form the P-part of the binding pocket (Figure 4A). The P-part was formed by S122 and H255 on one side and a loop from P50 to G56 between β3 and α1 on the other side. The residues L53 and L179 cover the catalytic site forming a tunnel-like path toward the catalytic site (Figure 4B). The presence of two leucines before the catalytic site may serve to limit the access of hydrophilic amino acid acyl groups to the hydrophobic D-part of BPHL, consequently making hydrophilic amino acid esters poor substrates of BPHL, in agreement with the observed substrate specificity.

Several potential binding modes of docked L-VACV, D-VACV, L-Val-FUdR, and D-Val-FUdR were suggested by the docking results (Figures 7−10). The best scored BPHL complex with L-valine esters were energetically favored compared to BPHL complexes with D-valine ester counterparts. The total energy of a BPHL complexed with L-VACV and D-VACV was −455 and −350 kcal/mol, respectively, and that with L-Val-FUdR and D-Val-FUdR was −211 and −187 kcal/mol, respectively. This is consistent with the stereopreference of BPHL for L-valine ester prodrugs relative to the corresponding D-valine ester prodrugs as shown by the 100-fold greater specificity constant (V_{max}/K_m) for L-Val-FUdR than that for D-Val-FUdR (739 and 7.43, respectively),[2] and the significantly lower hydrolysis rate of D-VACV compared to that of L-VACV.[1] However, the relative specificity between valine esters of different parent drugs, L-VACV and L-Val-FUdR, was not reflected in the total energy of the substrate−enzyme complex, and several of the docked structures exhibited a slight nonplanarity of the nucleoside ring, suggesting that further evaluation of docking structures may be needed. In general, the carbonyl oxygen of docked L-VACV and L-Val-FUdR was positioned

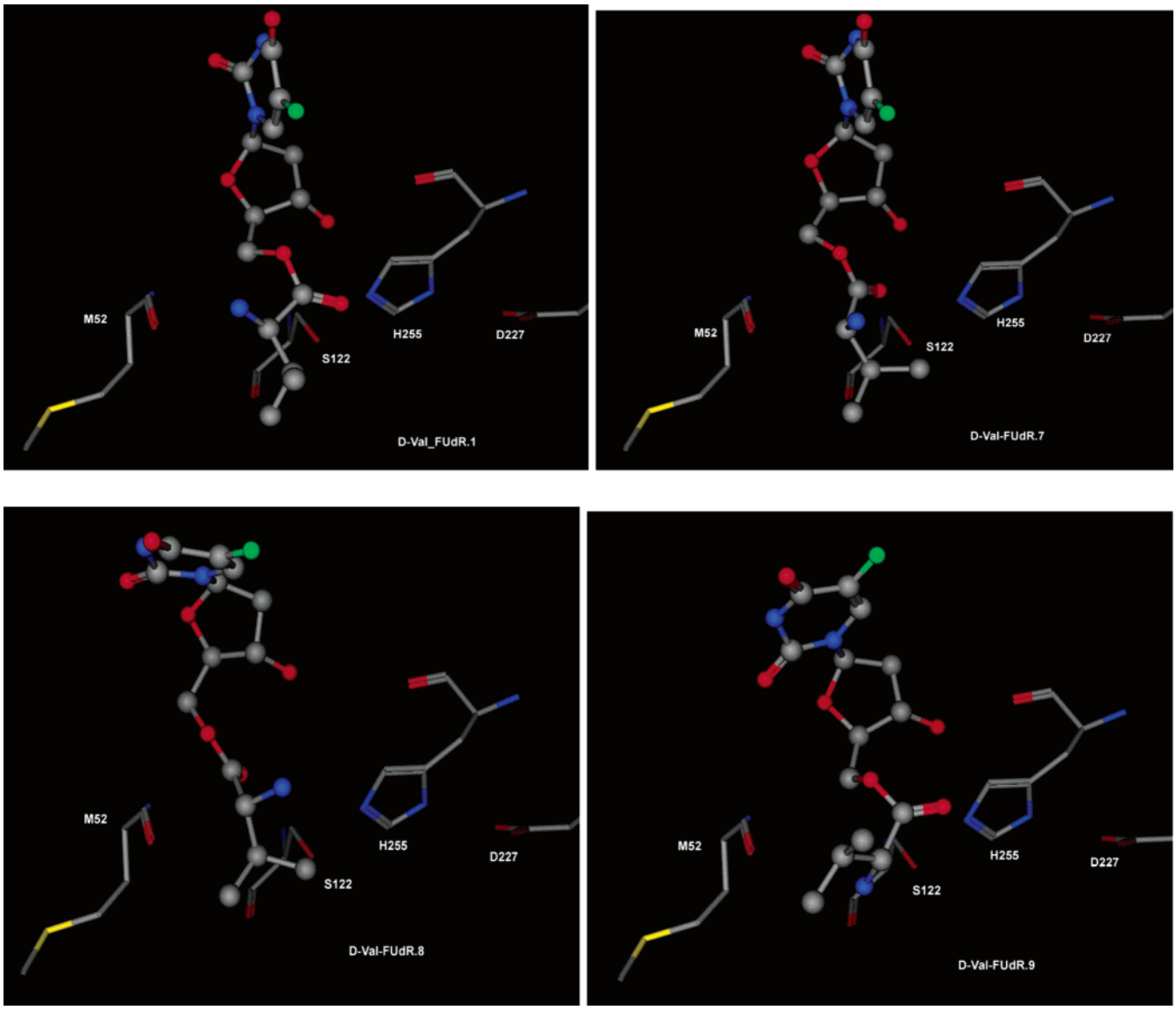

Figure 10. Suggested binding modes of D-Val-FUdR to BPHL.

within a hydrogen-bond distance to the backbone NH group of M52 (Figures 7 and 9), while that of D-VACV and D-Val-FUdR was positioned in the opposite direction (Figures 8 and 10).

Discussion

The α/β-hydrolase fold superfamily is a structurally conserved but functionally diverse protein family.[28,30,31] In general, α/β-hydrolase fold proteins have a well-conserved core domain and a variable lid domain typically inserted between β6 and β7, which plays a role in the diverse substrate specificity of the members.[28] However, the general folding and the catalytic mechanism are very well-conserved despite low sequence identity among the members. Analyses of the tertiary structures of the α/β-hydrolase fold members carboxyl/cholinesterases with only 28% amino acid identity reveal that the positions of about 400 out of 540 residues overlap within several angstroms.[30,32] Similarly the overall rmsd of the Cα positions of CumD (1IUP.A),[5] proline iminopeptidase (1MU0),[20] and chloroperoxidase (1A8S)[21] with about 20−25% sequence identity was 3.9 Å and that of the core domain (based on 218 residues in CumD) was 1.9 Å (data not shown). Moreover, the geometry of the catalytic triad is remarkably consistent among those enzymes, and the binding locations and the general orientations of ligands in each structure are very consistent, such that the carbonyl carbon is positioned toward the oxyanion hole, and the acyl group is oriented in the same direction (toward the D-part in CumD) (Figure 11). Therefore the homology modeling of BPHL was expected to generate a BPHL

(31) Nardini, M.; Dijkstra, B. W. α/β hydrolase fold enzymes: the family keeps growing. *Curr. Opin. Struct. Biol.* **1999**, *9*, 732−737.

(32) Oakeshott, J. G.; Claudianos, C.; Russell, R. J.; Robin, G. C. Carboxyl/cholinesterases: a case study of the evolution of a successful multigene family. *Bioessays* **1999**, *21*, 1031−1042.

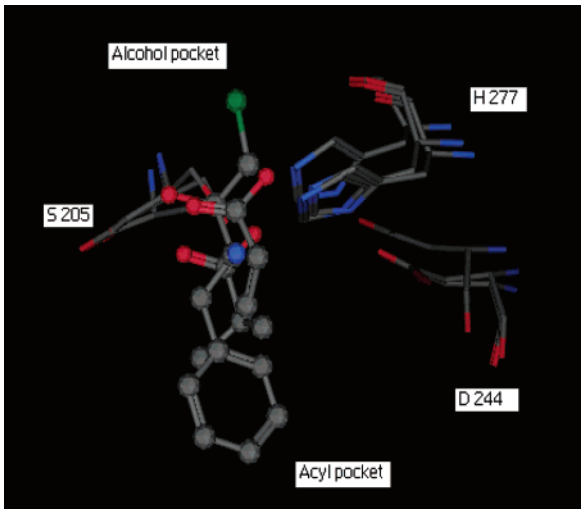

Figure 11. Superposition of the catalytic triad of α/*β*-hydrolase fold enzymes. The catalytic triad composed of serine-histidine-aspartic acid of 1MU0, 1IUP.A, and 1A8S and a ligand bound to each structure, phenylalanyl chloromethylketone, isobutyrate, and isopropionate, respectively, were superposed. The binding pockets for acyl and alcohol group of a substrate based on the orientation of ligands are suggested. Residue numbering is based on 1MU0.

structural model with good overall resemblance to the crystal structure despite the relatively low sequence homology to the template (41% similarity). The experimental confirmation of the presence of cysteines in free form, predicted by the BPHL model, further supports the homology model.

In general, the BPHL model exhibited features that can account for the substrate specificity of BPHL. For example, the model suggests that the substrate preference of BPHL for hydrophobic amino acid ester prodrugs may be attributed to the acyl binding site, formed mainly by hydrophobic amino acids and the presence of hydrophobic L53 and L179 at the entrance to the acyl binding site, preventing hydrophilic amino acyl promoieties from accessing the acyl binding site (Figure 4B). The substrate specificity of BPHL for amino acyl promoieties led us to hypothesize the presence of a charge−charge interaction site for the α-amino group near the active site. In addition, the fact that BPHL hydrolyzes phenylalanyl benzyl ester but not *N*-acetyl phenylalanyl ethyl ester with a blocked α-amino group (Lai, personal communication) suggests that the recognition of the α-amino group of substrate and the charge−charge interaction sites are important for substrate specificity. The examination of the geometry of the putative binding site of BPHL suggested that a backbone carbonyl oxygen of M52 and/or a backbone carbonyl oxygen or the side chain of D123 may function as a charge−charge interaction site for the α-amino group. In addition, in the docking study, some of the energetically favored conformations of docked substrates had the α-amino group positioned within 3.5 Å from the carbonyl oxygen of M52, suggesting a possible hydrogen-bond formation. Interestingly, similar charge−charge interaction between the α-amino group of an irreversible inhibitor and a backbone carbonyl oxygen of G37 of 1MU0 was observed.[20] In the crystal structure of 1MU0, the α-amino group of phenyl chloromethyl ketone, which is covalently bound to the active serine, forms a hydrogen bond with the backbone carbonyl oxygen of G37, which also functions as the second hydrogen donor for the oxyanion hole. M52 of BPHL, which is likely structurally equivalent to G37 of 1MU0, may interact with the α-amino group of a substrate in a similar manner. Further, the well-conserved D123 immediately after the active serine may be important for the substrate specificity for amino acyl promoieties within the mainly hydrophobic acyl-binding site.

The BPHL model suggests that the binding pocket is a deep gorge shape and the space around the catalytic site is limited by adjacent residues and a loop. The less constrained nucleoside parent drug binding site and the limited space around the catalytic site where the ester oxygen of a substrate would be may account for the broad specificity for nucleoside analogues as a parent drug and the significantly lower reactivity of BPHL toward a secondary alcohol than a primary alcohol.[2] Consistent with this is the fact that BPHL did not hydrolyze L-phenylalanyl *tert*-butyl ester (Lai, personal communication).

In some serine hydrolases of the α/*β*-hydrolase fold family, the residue immediately following the active serine is usually involved in the formation of an oxyanion hole, donating one of the two hydrogen bonds to the oxygen of the tetrahedral intermediate to stabilize the oxyanion.[28,30,33] The second group forming the oxyanion hole is usually located on a loop between *β*3 and α1. For example, S32 and G37 of CumD and 1MU0, respectively are hydrogen bonded with the carbonyl oxygen of a ligand.[5,20] Interestingly, CumD and 1MU0 have a conserved sequence motif around the second hydrogen donor such as HG**X**GPG and HG**G**PG, respectively, and the third residue functions as a second hydrogen donor.[5,20] Similarly, BPHL has a well-conserved motif PG-(M/A)LG and M52 is a likely candidate for a second hydrogen donor in BPHL (Figure 1). The molecular docking study also suggested that the carbonyl oxygen of docked L-valine esters was more preferably positioned toward the backbone NH group of M52 (Figures 7, 9) than D-valine esters (Figures 8, 10). The suggested difference in the preferred orientation of the carbonyl oxygen of valine isomers may contribute to the stereoselectivity of BPHL for L-isomer amino acyl ester prodrugs to D-isomer amino acyl ester prodrugs.

The BPHL activity was completely inactivated by PCMB.[1] Combined with the chemical labeling study, the suggested location of free cysteines, which are more than 10 Å away from the active site, suggest that the inhibition of BPHL activity by PCMB may be due to conformational changes upon PCMB binding to the cysteines, which may indirectly block access to the active site rather than by direct blocking of the catalytic residues. Possibly, the binding of PCMB to

(33) Holmquist, M. Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. *Curr. Protein Pept. Sci.* **2000**, *1*, 209−35.

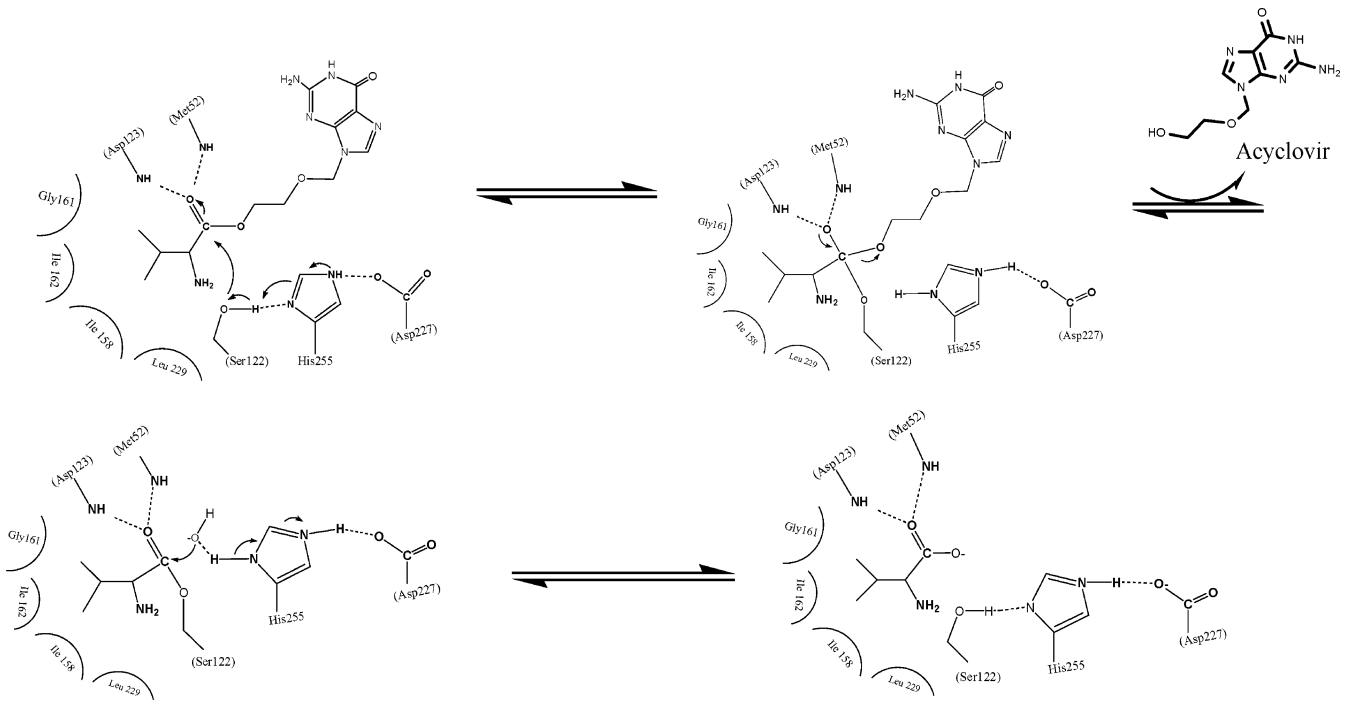

Figure 12. Hypothetical mechanism of VACV hydrolysis by BPHL. The putative mechanism of VACV hydrolysis by BPHL is proposed by an analogy to the classical catalytic mechanism of typical serine hydrolases featuring the "catalytic triad".

C189 which appears to be close to the entrance of the binding pocket may cause the conformational change to block the entrance of the binding site (Figure 2). On the other hand, C208 is completely conserved among eukaryotic homologues, suggesting that it may be structurally important. Experiments with site-directed mutagenesis are in progress to elucidate the mechanism of this inhibition.

In summary, a homology model of BPHL has been developed and evaluated. While the homology model is a low-resolution structure such that insight gathered from the model is considered provisional, it nevertheless can account for most of the experimental results to date on the substrate specificity and provide a basis for mutagenesis studies to further explore the substrate specificity and catalytic mechanism of this novel enzyme. A putative catalytic mechanism of BPHL can be suggested, based on analogy to the mechanism of classical serine esterase hydrolysis and the residues forming the oxyanion hole in the BPHL model (Figure 12). This homology model of BPHL will be very useful in providing further suggestions for mutagenesis studies and study of potential substrates and promoieties for prodrug targeting. Further refinement of the structure, catalytic mechanism, and substrate binding will be based on the ongoing X-ray crystal structure determination.

Abbreviations Used

BPHL, biphenyl hydrolase-like protein; L-VAC, L-valyl ester acyclovir; D-VACV, D-valyl ester acyclovir; L-Val-FUdR, L-valyl ester floxuridine; D-Val-FUdR, D-valyl ester floxuridine; PCMB, *p*-chloromercuribenzoic acid.

Acknowledgment. This research was supported by NIH Grant GM37188. The authors thank John Strahler and Maureen Kochman at the Michigan Proteome Consortium at the University of Michigan for their help on the mass spectrometry. I.K. is a recipient of a University of Michigan Rackham Predoctoral Fellowship and Barbour Scholarship.

MP049959+