UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Applications and method development in the simulation of ligands binding to drug targets

Permalink

https://escholarship.org/uc/item/4rs6v0p5

Author

Sinko, William

Publication Date

2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Applications and Method Development in the Simulation of Ligands Binding to Drug Targets

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Biomedical Sciences

by

William Sinko

Committee In Charge:

Professor J. Andrew McCammon, Chair
Professor Philip E. Bourne
Professor Michael K. Gilson
Professor Alexandra C. Newton
Professor Victor Nizet

2012

Copyright

William Sinko, 2012

All Rights Reserved

The dissertation of William Sinko is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2012

Table of Contents

Signature Page iii

Table of Contents iv

List of Figures vi

List of Tables xi

List of Schemes xii

Acknowledgements xiii

Vita xvi

Abstract of the Dissertation xvii

Chapter 1: Accounting for receptor flexibility and enhanced sampling methods in computer aided drug design 1

- Abstract 1
- Abbreviations 2
- Introduction 2
  - Introduction to receptor flexibility 3
- Induced Fit Docking 7
- Rosetta Ligand 8
- Ensemble Based Screening Methods 8
- Enhanced Sampling: Hardware improvements 11
- Enhanced Sampling: Methodologies 12
  - Umbrella sampling 13
  - Metadynamics 14
  - Accelerated Molecular Dynamics 15
- Accelerated MD in free energy calculations 18
- Concluding remarks 21
- Acknowledgements 22

Chapter 2: Discovery of Small Molecule Inhibitors of the PH domain Leucine-rich repeat Protein Phosphatase PHLPP by Chemical and Virtual Screening 23

- Abstract: 23
- Abbreviations: 24
- Introduction 24
- Results: 28
- Discussion: 46
- Experimental section: 51
  - Experimental *in vitro* screen 52

In vitro inhibition concentration assay 53
Homology modeling 53
Similarity searches and compound library generation 54
Docking 54
Homology model evaluation 55
Cell culture and immunoblotting 55
Inhibitor treatment of cells 56
Apoptosis assay 57
Acknowledgements 57
Supplemental Information 59
Synthesis and purification of compounds 1 and 13 60

Chapter 3: Applying molecular dynamics simulations to identify rarely sampled ligand-bound conformational states of undecaprenyl pyrophosphate synthase (UPPS), an antibacterial target 64
Abstract 64
Abbreviations 65
Introduction 65
Experimental Section 68
Computational Details. 68
Crystallization and structure determination of Apo *E. coli* UPPS. 70
Results 71
Docking of known UPPS inhibitors 71
Identifying inhibitor bound UPPS conformations from MD simulation 74
Breathing Motion of the Catalytic Pocket 80
Conclusions 82
Acknowledgements 86
Supplemental Information 88

Chapter 4: Protecting High Energy Barriers: A New Equation to Regulate Boost Energy in Accelerated Molecular Dynamics Simulations 92
Abstract 92
Introduction 93
Results and discussion 95
Computational Methods 111
Conclusion 114
Acknowledgements 115
Supplemental Information 116

Conclusion 118

References 122

List of Figures

Figure 1.1 Schematic representation of the possible pathways taken in ligand binding. The induced fit pathways and conformational selection pathways are shown via dashed and solid lines. Various methods for binding energy calculations, which account for receptor flexibility are listed in the grey box. Simulation methods for sampling receptor flexibility are listed in the purple box. 5

Figure 1.2 Schematic representation of ensemble based docking. Examples of methods or data used at each step are given inside the dashed grey lined squares. 6

Figure 2.1 Screen of NCI Diversity Set. [a] Scatter plot of the raw data from 25 different plates showing control (DMSO) (●) and background (absence of enzymes) values (◇). The solid lines represent the mean of the control and background data and the broken lines display 3 standard deviations (SD) from the mean of each data set. The z-factor is 0.5 and the signal-over-background ratio is 3.8. [b] Representative time course for the hydrolysis of pNPP by PHLPP2 catalytic domain. pNPP (8 mM) was incubated in 125 μL assay buffer at room temperature, without (◇) or with the phosphatase domain of PHLPP2, in the presence of DMSO (●) or compound 24515 (■). The absorbance of the solution was recorded at 405 nm and the data fit by linear regression. [c] Representative activity data from one 96-well plate. NCI plate 4238 was assayed for PHLPP2 activity as in [b]. Each compound is represented by a black bar, DMSO controls are represented by grey bars (background controls were omitted on this graph). The values obtained for the DMSO controls were averaged and normalized to a relative activity of 1.0 (black line). Compounds that reduced PHLPP2 relative activity to lower than 0.3 (red line) were considered a hit (lane 86). 30

Figure 2.2 Chemical structures of some inhibitors of PHLPP2. Different compounds that have been identified as inhibitors of PHLPP2 are presented along with the experimental IC_{50} determined at 8 mM substrate. The chemical backbones used to determine the structural families are highlighted. 31

Figure 2.3 Virtual screen for inhibitors using modeled structure of PHLPP2 phosphatase domain. [a] Sequence alignment of PP2Cα and PHLPP2. Amino acids in red may be important for coordinating metal ions. * denotes a conserved residue. [b] 3-dimensional representation of model 4 of table c with D806, E989, and D1024 shown coordinating 2 Mn^{2+} ions in the phosphatase active site. [c] Table indicating the functional relevance of modified homology models. Model 4 was chosen for further docking because of its high docking scores and good performance identifying true hits from the Diversity Set. The number of true hits is derived from docking 50 compounds with experimentally derived IC_{50}s for PHLPP2 and determining how many compounds scored better than -7. [d] Structural similarity search to create libraries of compounds similar to known binders of PHLPP2. Selected compounds

were docked with the GLIDE algorithm employing calculations at 3 levels of accuracy, removing lower scoring compounds at each stage.....33
Figure 2.4 PHLPP2 inhibitors from the virtual screening. [a] Table of experimentally confirmed inhibitors of PHLPP2 and their estimated free energy of binding (kcal/mol) as well as IC_{50}. [b] 2-dimensional representations of the compounds which inhibit PHLPP2 from the docking study.....35
Figure 2.5 Model of active site of PHLPP2 docked with substrate or inhibitors. [a] pNPP, [b] a phospho-peptide based on the hydrophobic motif of Akt (HFPQFpSYSAS; where the phosphorylated Ser corresponds to Ser473), [c] compound 134149 [d] compound 13378 [e] compound 117079 and [f] compound 45586 were docked in the PHLPP2 model 4. Ligands are represented in tube format, Mn^{2+} ions are represented as yellow spheres and surface representation of PHLPP2 shows the electrostatic potential where blue represents positive, white neutral and red negative electrostatic potential. Docked substrates or inhibitors are in green, the phosphate groups are highlighted in purple. * denotes the carbonyl group of Gly 745 in all views.....37
Figure 2.6 Cellular potency and specificity of PHLPP inhibitors. [a], [b] Effect of different inhibitors on Akt phosphorylation. HT29 cells in DMEM supplemented with 5% FBS, were treated for 24 hours with different compounds (see table). Blots were probed with antibodies specific to phospho-Akt (S473), phospho-GSK3α/β (S21/9), phospho-FoxO1/3a (T24/32) and actin. [b] Densitometric analysis was performed on blots from 3 separate experiments. The levels of phospho-proteins are normalized to actin. The phosphorylation of Akt at serine 473 (white bars), FoxO 1/3 (light grey bars) and GSK 3α/β (dark grey bars) is calculated relative to control lanes. The graph represents mean values ± SEM from 3 separate experiments. [c] Selectivity towards other phosphatases. pNPP (8 mM) was incubated in 125 μL of the optimal assay buffer for each protein, in the presence of the PP2C domain of PHLPP2 (black bars), PP1 (dark grey bars), PP2B (white bars) or PP2Cα (light grey bars). Compounds (see table) were added. The activity of the enzyme is relative to the DMSO control. The graph represents mean values ± SEM of at least 3 different experiments. * denotes that compound 45586 was tested at 100 μM and not 250 μM. [d], [e] In vitro inhibition curves for compound 117079 [d] and 45586 [e]. The inhibitor, diluted in DMSO, was incubated in assay buffer with 8 mM pNPP, in the presence of 1 μM enzyme. Activity was calculated relative to DMSO alone. The mean activity ± SEM for 5 different experiments are represented (circle:117079, diamond: 45586) and were fit with equation 1.....40
Figure 2.7 Effect of inhibitors on Akt-dependent signaling in cells. COS 7 cells were treated for 35 minutes with compound 117079 [a],[b] or 45586 [c],[d] at different concentrations. The control corresponds to 35 minutes treatment with DMSO alone. The graph represents mean values ± SEM from at least 3 separate experiments. [a], [b] Blots were probed with antibodies specific to phospho-S473, phospho-T308, an antibody specific to phospho-serine or phospho-threonine in a consensus sequence for Akt and actin. [b] The levels of phospho-Akt at Ser 473 (◇), at Thr 308 (○) and phosphorylation of Akt substrates (◆) are normalized to total actin and relative to the level of phosphorylation in the control lanes. [c], [d] Blots were probed with

antibodies specific to phospho-S473, phospho-T308, phospho-GSK3 α/β, phospho-FoxO1/3a and actin. [d] The levels of phospho-Akt at Ser 473 (◇), at Thr 308 (○) and phosphorylation of Akt substrates (GSK 3:□, FoxO:▼) are normalized to total actin and relative to the level of phosphorylation in the control lane. [e], [f] Effect of compound 117079 on the agonist-evoked phosphorylation of Akt. COS 7 cells were incubated in serum-free DMEM for 24 hours. The cells were pre-treated with DMSO (dark symbols) or compound 117079 (50 μM) (open symbols) for 35 minutes. EGF (1 mg mL-1) was then added at various time. Blots were probed with the specified antibodies. [d] The levels of phospho-Akt and phospho-ERK are normalized to actin. The phosphorylation of Akt at S473 (◆ : DMSO, ◇ : 117079), of Akt at T308 (● : DMSO, ○: 117079) and ERK (▲: DMSO, △ : 117079) are relative to the levels of phospho-protein in unstimulated cells treated with DMSO (lane 1).43

Figure 2.8 Effect of inhibitors on etoposide-induced cell death. COS 7 cells in DMEM supplemented with 0.1% FBS were treated for 30 min with DMSO, 117079 or 45586 at 50 μM followed by a 24 hours treatment with DMSO or etoposide. Cells were then submitted to a Trypan blue exclusion assay. The graph shows the etoposide-induced increase in cell death (compared to DMSO) and represents mean values ± SEM of 3 separate experiments.* corresponds to p values < 0.0005 and ** corresponds to p values ≤ $2\ 10^{-5}$. 46

Figure 3.1 (A) Superposition of docked (colored by atom type) and co-crystallized pose (green) of BPH-629 into the 2E98 crystal structure. (B) Docked poses of 29 BPH inhibitors into the 2E98 structure. (C) Docked poses of **1i**, **1j**, **4a**, **4g**, **4j**, **4l** and **4m** into the 2E98 structure. Ligand poses showed a very poor alignment with the substrate, FPP (shown in green). (D) Docked poses of **1i**, **1j**, **4a**, **4g**, **4j**, **4l** and **4m** into the fourth most populated MD-derived structure. All ligand poses showed a very good alignment with the substrate, FPP (shown in green). 72

Figure 3.1 Grid points used to calculate pocket volumes. A) Points spaced 1 Å apart were positioned along a grid that encompassed the UPPS IPP-binding pocket. B) Those grid points near protein atoms were removed, leaving only points within the active site, from which the volume was calculated.....76

Figure 3.2 Volume distribution of the HIP43 UPPS binding pocket 2E98 crystal structure and the apo crystal structure. (A) Volume of the binding pocket along the MD trajectory. The black line shows every data taken every 10 ps, the overlayed gray line is the average over every 100ps. (B) Frequency at which different volumes of the pocket are sampled. The size of the bisphosphonate bound crystal structure (2E98), the newly described apo crystal structure (Apo), and Cluster 4 which bound tetramic acids are represented by labeled dashed lines in both graphs. 77

Figure 3.3 (A) The "apo" crystal structure in green and the bisphosphonate bound crystal structure in blue. (B) The "apo" crystal structure with 1 Å spheres filling the active site pocket. (C) The bisphosphonate bound crystal structure with 1 Å spheres filling the active site pocket. Note the significantly larger pocket size in the bisphosphonate bound structure when compared to the "apo" crystal structure.....78

Figure 3.4 Principal Component Analysises (PCA) (A) Principal Component Analysis (PCA) calculated from the HID43 trajectory. (B) The extreme conformations of the flexible loop (residues 72-82) in HID43 are shown in red and blue. (C) PCA

calculated from the HIP43 trajectory. The green circle indicates the largest conformation, and the blue circle indicates cluster 4. (D) The extreme conformations of the flexible loop (residues 72-82) in HIP43 are shown in red and blue.....80
Figure S3.1 The 2E98 crystal structure, (A) Monomer A is represented in cartoon. bisphosphonate binding sites 1, 2, 3, and 4 are represented in surface in red, gray, white and brown respectively. (B) bisphosphonate binding site 1 interaction map with labeled residues. Residues have light green carbons, BPH-629 has gray carbons.....88
Figure S3.2 RMSD of the Trajectories. HIP43 (blue) simulation and HID43 (red) simulation, both show stability throughout the MD run.....89
Figure S3.3 Scatter plot of pIC_{50} and predicted free energy of binding by GLIDE for BPH compounds. On the left is the 2E98 crystal structure docking results and linear regression line (R=-.52) and on the right is the largest structure from the MD simulation and linear regression line (R=-.79).....89
Figure S3.4 Poses of BPH-629 compared with BPH-641 docked against (A) the 2E98 crystal structure, and (B) the largest MD derived structure. The protein backbone is shown in cartoon representation with a transparent electrostatic surface representation (blue is positive and red negative). The black arrow points to HIS43, the gray arrow points to ARG77, the * denotes the nitrogen in BPH-641, ligands are represented in stick form with standard atom coloring. Note the narrow positively charged bisphosphonate binding region in the 2E98 structure and the unexpected pose generated by GLIDE for BPH-641. 90
Figure S3.5 Root mean square fluctuation (RMSF) calculated for each residue in the MD simulations of HID43 (red) and HIP43 (blue) systems.....91
Figure 4.1 Hypothetical one dimensional potential representing the effect of ΔV^c. In all charts, α_1 =200 and E_1 =-250. The upper and lower solid black lines represent the original potential and the modified potential generated with ΔV^b, respectively. This color scheme is used throughout. A) Effects of different parameters E_2 (dashed colored lines) on the modified potential generated with ΔV^c (solid colored lines). B) Boost levels (ΔV) as V moves away from E_1. For both (A) and (B) α_2=15, and E_2=-100 (red), 0 (blue) and 150 (green). C) Effect of varying α_2 parameter on ΔV^c: α_2=3 (red), 15 (blue), 75 (green) with E_2=0.....98
Figure 4.2 Alanine dipeptide simulation results. A) Psi (Ψ) angle values obtained from cMD and five different aMD simulations. From top to bottom, aMD parameters were set to: $E_2=E_1+15$ and $\alpha_1=5$, $E_2=E_1+20$ and $\alpha_1=5$, $E_2=E_1+25$ and $\alpha_1=5$, $E_2=E_1+25$ and $\alpha_1=2.5$, $E_2=E_1+25$ and $\alpha_1=1.25$. In all simulations, E_1 and α_2 were set to 10 and 5 respectively. Weighted free energy density plots obtained from cMD (B, C, D) and aMD with ΔV^c (E). All values are in kcal/mol.....102
Figure 4.3 Psi (ψ) and Omega (Ω) angle values obtained from aMD simulations with boost potentials ΔV^b and ΔV^c. In all simulations, E_1=10.0 and α_1 was set as shown on the far right. Additional parameters for aMD with ΔV^c were set to $E_2= E_1+15$ and α_2=5. All values are in kcal/mol.....104

Figure 4.4 Decalanine Phi (Φ) - Psi(Ψ) angles distribution obtained from cMD and aMD simulations. For the aMD simulations with ΔV^c parameters were set to E_1=74, E_2= E_1+25, α_2=5, α_1=30 (C) and α_1=15 (D). All values are in kcal/mol. 106
Figure 4.5 Principle Component Analysis obtained from decalanine MD simulations. A) 50 ns of aMD simulation with ΔV^c. Parameters were set to E_1=74, E_2= E_1+25, α_2=5 and α_1=15, same as in Figure 4D. B) 50 ns of cMD simulation and C) 350ns of cMD simulation. Structures 1, 2 and 3 shown in yellow represent relevant populated states in PC subspace sampled by aMD and cMD. 108
Figure 4.6 (A) Relative free energy of binding between $Ac_{2\text{-}L\text{-}}Lys_{\text{-}D\text{-}}Ala_{\text{-}D\text{-}}Ala$ and $Ac_{2\text{-}L\text{-}}Lys_{\text{-}D\text{-}}Ala_{\text{-}D\text{-}}Lac$ to vancomycin calculated from cMD (solid black line) and aMD with ΔV^c (solid red line). Dashed line displays the experimental value, 4.1 kcal/mol. (B) Cumulative free energy curves calculated from simulations of 600ps (left) and 1000ps (right) per λ point. The * show points where there is no overlapping between error bars.....111
Figure S4.1 Decalanine Phi(Φ)-Psi(Ψ) angles distribution obtained from aMD simulations with ΔV^c. Parameters were set to E_1=74, E_2= E_1+35, α_2=5 and α_1=15. All values are in kcal/mol.....116
Figure S4.2 Decalinine root mean square (RMS) deviation from α-helix conformation calculated along: A) 350ns of cMD and 50 ns of aMD simulations, and B) the first 50ns of simulations displayed in (A). Solid lines correspond to MD (black) and aMD with parameters set to E_1=74, α_2=5 (all aMD simulations) E_2= E_1+25 and α_1=30 (cyan); E_2= E_1+25 and α_1=15 (red); E_2= E_1+35 and α_1=15 (green). All values are in kcal/mol. 117

List of Tables

Table 2.1Calculated chemical properties of compounds **13** and **1**. Qikprop was run to calculate the chemical properties that make up the Lipinski rules 48
Table S2.1 NSC IDs of compounds that inhibit PHLPP 2 in vitro grouped in clusters according to their chemical structures. Compounds presenting an IC50 value below 5 μM are written in bold and italic, those with an IC50 value comprised between 5 and 10 μM are in italic and underlined and the compounds only underlined present an IC50 value between 10 and 25 μM. The compounds that were tested in cells are in grey. 59
Table S2.1 The IC50s for the different products as well as the reagents were obtained as described in the Methods section. 62
Table 3.1 UPPS inhibitors. IC_{50} values shown in parentheses. 84
Table 3.2 Data collection and refinement statistics for UPPS apo crystal (3QAS). (Values in parentheses reflect the statistics for the high resolution shell) 86

List of Schemes

Scheme 1.1 Accelerated Molecular Dynamics. Equations used to calculate the boost energy and modified potential energy surface in aMD. A two dimensional representation of the modified potential, V*(r) (dashed lines), and the unmodified potential energy surface, V (thick black line). a was varied as indicated E was always fixed at 60 and is indicated by a thin solid black line. 18
Scheme 2.1 Synthesis of compound 1 61
Scheme 3.1 Condensation reaction of IPP and FPP by UPPS 66

Acknowledgements

I would like to thank Professor Andy McCammon for all the support and guidance during my graduate education. I also must thank Dr. Cesar de Oliveira, Levi Pierce, Dr. Arneh Babakhani, Dr. Jacob Durrant, Dr. Steffen Lindert, Professor Adam van Wynsberghe, Dr. Sarah Williams and the entire McCammon lab for all the help along the way. I also wish to thank Professor Alexandra Newton and Dr. Emma Sierecki for the collaborative work on the PHLPP inhibitors project. Chelsea Buma, Dr. Nadia Fomina, Dr. Adah Almutairi, Dr. Li Xie, Professor Philip E. Bourne, Dr. Yann Gambin and Dr. Edward Lemke were of great help on the PHLPP project. In addition, I would like to thank Professor Eric Oldfield, and members of his lab Dr. Rong Cao, Dr. Wei Zhu, and Dr. Yonghui Zhang at UIUC for the collaborative work on UPPS. I wish to thank Professor Michael Gilson, and Professor Victor Nizet for advice and guidance.

This work was supported by a Juvenile Diabetes Research Foundation grant 3-2008-478, Molecular Biophysics Training Grant GM08326, the National Science Foundation Grant MCB-0506593, NBCR, CTBP, ARCS, Howard Hughes Medical Institute, and National Institutes of Health Grants GM31749 GM067946, and GM065307. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, and Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357. Use of the Life Science Collaborative Access Team Sector 21 was

supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor for the support of this research program (Grant 085P1000817).

Chapter 1 contains text from the article in preparation Sinko W.; Lindert S.; McCammon, J. A. Accounting for Receptor Flexibility, and Enhanced Sampling in Computer Aided Drug Design. (In preparation for *Chem Biol Drug Des*. special edition) I am the first author on this paper.

Chapter 2 is a minimally modified reprint of the material as it appears in Sierecki, E.; Sinko, W.; McCammon, J. A.; Newton, A. C. Discovery of small molecule inhibitors of the PH domain leucine-rich repeat protein phosphatase (PHLPP) by chemical and virtual screening. *J. Med. Chem*. 2010, *53*, 6899–6911. I was a co-first author and co-principal investigator on this paper.

Chapter 3 is a minimally modified reprint of the material as it appears in Sinko, W.; de Oliveira, C.; Williams, S.; Van Wynsberghe, A.; Durrant, J. D.; Cao, R.; Oldfield, E.; McCammon, J. A. Applying molecular dynamics simulations to identify rarely sampled ligand-bound conformational states of undecaprenyl pyrophosphate synthase, an antibacterial target. *Chem Biol Drug Des*. 2011, *77*, 412–420. I was the first author and principal investigator on this paper.

Chapter 4 is a minimally modified reprint of the material as it appears in Sinko, W.; de Oliveira, C. A. F.; Pierce, L. C. T.; McCammon, J. A. Protecting High Energy Barriers: A New Equation to Regulate Boost Energy in Accelerated Molecular Dynamics

Simulations. *J Chem Theory Comput*. 2012, *8*, 17–23. I was a first author and principal investigator on this paper.

Vita

2007 B. S. Exercise Biology University of California, Davis

2012 Ph.D. Biomedical Sciences University of California, San Diego

Publications

1. **Sinko, W.**; de Oliveira, C. A. F.; Pierce, L. C. T.; McCammon, J. A. Protecting High Energy Barriers: A New Equation to Regulate Boost Energy in Accelerated Molecular Dynamics Simulations. *J Chem Theory Comput* **2012**, *8*, 17–23.

2. **Sinko, W.**; Sierecki, E.; de Oliveira, C. A. F.; McCammon, J. A. Guide to Virtual Screening: Application to the Akt Phosphatase PHLPP. *Methods Mol. Biol.* **2012**, *819*, 561–573.

3. **Sinko, W.**; de Oliveira, C.; Williams, S.; Van Wynsberghe, A.; Durrant, J. D.; Cao, R.; Oldfield, E.; McCammon, J. A. Applying molecular dynamics simulations to identify rarely sampled ligand-bound conformational states of undecaprenyl pyrophosphate synthase, an antibacterial target. *Chem Biol Drug Des* **2011**, *77*, 412–420.

4. Sierecki, E.; **Sinko, W.**; McCammon, J. A.; Newton, A. C. Discovery of small molecule inhibitors of the PH domain leucine-rich repeat protein phosphatase (PHLPP) by chemical and virtual screening. *J. Med. Chem.* **2010**, *53*, 6899–6911.

5. Morairty, S. R.; Wisor, J.; Silveira, K.; **Sinko, W.**; Kilduff, T. S. The wake-promoting effects of hypocretin-1 are attenuated in old rats. *Neurobiol. Aging* **2011**, *32*, 1514–1527.

Abstract of the Dissertation

Applications and Method Development in the Simulation of Ligands Binding to Drug Targets

by

William Sinko

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2012

Professor J. Andrew McCammon, Chair

The role of computation in drug discovery has increased considerably in the past few decades. As of 2004 about 50 compounds discovered with computational approaches had entered human clinical trials and some are FDA approved(*1*). Computational methods can be employed to effectively guide medicinal chemists and discover and optimize active compounds(*2*). Here we have focused on two protein targets, which are medically important, to inhibit with novel compounds that we have

discovered virtually. In chapter 2 we describe the discovery of the first inhibitors of the Akt phosphatase PHLPP, (PH domain Leucine-rich repeat Protein Phosphatase). This enzyme opposes the effects of two kinases, Akt and PKC, which play a major role in cell growth and survival making it an attractive drug target in the treatment of diabetes and heart disease(*3*, *4*). We have used a combination of *in silico* and chemical screening tools to find new inhibitor compounds of PHLPP. In chapter 3 we describe drug discovery efforts towards inhibitors of the second target, which is undecaprenyl pyrophosphate synthase, (UPPS) an enzyme necessary for bacterial cell wall synthesis. This recently discovered enzyme is substantially different than the analogous human enzyme making it an attractive target for anti-bacterial drug discovery(*5*). However UPPS is highly flexible target showing great changes in the active site upon substrate, product, or inhibitor binding, thus making it challenging for structure based drug design. Therefore we have characterized these changes using molecular dynamics simulations, and extracted rare conformations important to drug design.

In addition to applying existing methods to new drug targets we have worked on developing a new form of enhanced sampling MD described in chapter 4. This method relies on the same approach as Hamelberg et al. took in accelerated MD simulations but the mathematical equations modifying the potential energy surface increase the recovery of statistics from the simulation. This new boost equation has been applied to statistically rigorous free energy calculations that incorporate solvent, entropy, and dynamics, and are among the most computationally accurate yet costly calculations of ligand-protein binding free energy.

Chapter 1: Accounting for receptor flexibility and enhanced sampling methods in computer aided drug design

Abstract

Protein flexibility plays a major role in biomolecular recognition. In many cases it is not obvious how molecular structure will change upon association with other molecules. In proteins these changes can be major, with large deviations in overall backbone structure, or they can be more subtle as in a side chain rotation. Either way the algorithms that predict the favorability of biomolecular association require relatively accurate predictions of the bound structure to give an accurate assessment of the energy involved in association. Here we review a number of techniques, that have been proposed to accommodate receptor flexibility in the simulation of small molecules binding to protein receptors. We investigate modifications to standard rigid receptor docking algorithms, and also explore enhanced sampling techniques, and the combination of free energy calculations and enhanced sampling techniques. The understanding and allowance for receptor flexibility are helping to make computer simulations of ligand protein binding more accurate. These developments may help improve the efficiency of drug discovery and development. Efficiency will be essential as we begin to see

personalized medicine tailored to individual patients, which means specific drugs are needed for each patient's genetic makeup.

Abbreviations

cytochrome P450s (CYP), root mean square deviation(RMSD), molecular dynamics (MD), Monte Carlo (MC), induced fit docking (IFD), relaxed complex scheme (RCS), graphics processing unit (GPU) , high performance computing (HPC), collective variable (CV), accelerated molecular dynamics (aMD), guided entry of tail anchored proteins (GET), free energy perturbation (FEP), thermodynamic integration (TI), multistate bennett acceptance ratio (MBAR), and molecular mechanics/ generalized born surface area (MM/GBSA)

Introduction

The advent of pharmacogenomics has opened up many new and exciting avenues to optimize drug treatment for a variety of diseases(*6*). While often used to correlate drug efficiency or toxicity with specific mutations, it has great potential in combination with structure based drug discovery as well. Particularly in the realm of computer aided drug design, it is very conceivable that studies on individual mutant receptors can yield "personalized drugs" tailored to target specific mutant protein forms only. These studies ideally will include full receptor and ligand flexibility to find the most optimally suited inhibitors. This review will focus on techniques that are used to account for receptor flexibility in computer aided drug discovery studies.

Introduction to receptor flexibility

Often it is of great interest to know the manner in which different molecules interact. Gibbs free energy can be related to the favorability of one state over another such as state A: a drug bound to its protein target, or state B: unbound. The free energy can be related to the concentration bound and unbound, and affinity can be quantitatively expressed in this way. A large number of methods have been proposed to calculate the free energy of binding in biomolecular recognition, because this property is of great interest to those that develop drugs for biological targets. If one only needs to know the structure of a compound to calculate its binding affinity to a target of interest, it would save a great effort in screening compounds and synthesizing them. In addition, the vast majority of the chemical space possible for drug like molecules has not been probed by chemists, but could in theory be more efficiently explored via computation if rapid and accurate methods are developed. In fact, in terms of drug-sized molecules it is estimated that only a very small percentage of the chemical space available has been synthesized(*7*). Of great relevance and application have been empirical methods like docking, which performs a 3D search of ligand poses in the receptor binding pocket, and then scores them according to specific types of interactions such as: van der Waals forces and electrostatic terms. These methods have proved of great use(*8*) because of their speed and simplicity in setup, but suffer from numerous errors because they don't account well for entropy, water effects, or protein structural changes upon ligand binding(*9*).

Often proteins adopt different conformations upon ligand binding than they do in their unbound form. Significant changes can be seen in the protein structure of some proteins, while minor changes like side chain rotations or subtle rearrangements are often important in terms of binding affinity as well. Some Cytochrome P450s (CYP) including CYP2B4 undergo alpha-helical rearrangements to allow different substrates to bind(*10*). The infamous CYP3A4 isoform adopts major changes in conformation during binding which is believed to be necessary for metabolism of many substrates, which vary greatly in size (1000 daltons variation), geometry, and composition. These changes have been verified by ligand co-crystallography(*11-13*). With all the degrees of freedom and flexibility inherent in protein structure, the changes in receptor conformation are very important in calculating the binding free energy. If the basic geometric shape of the binding site is not described correctly, then it will be quite difficult to predict compound binding correctly using empirical methods like docking. Two theories have been proposed to describe receptor configurational change upon ligand interaction. The first is conformational selection(*14*, *15*) in which all conformations are present in the unbound receptor, but the populations of each configuration change in the bound form, meaning that the average structure is different from bound to unbound. The second theory is induced fit(*16*), in which one configuration is forced into another configuration by the ligand-binding event. See figure 1 for a schematic of these two theories of biomolecular recognition and many methods to account for these conformational changes in protein-ligand interactions.

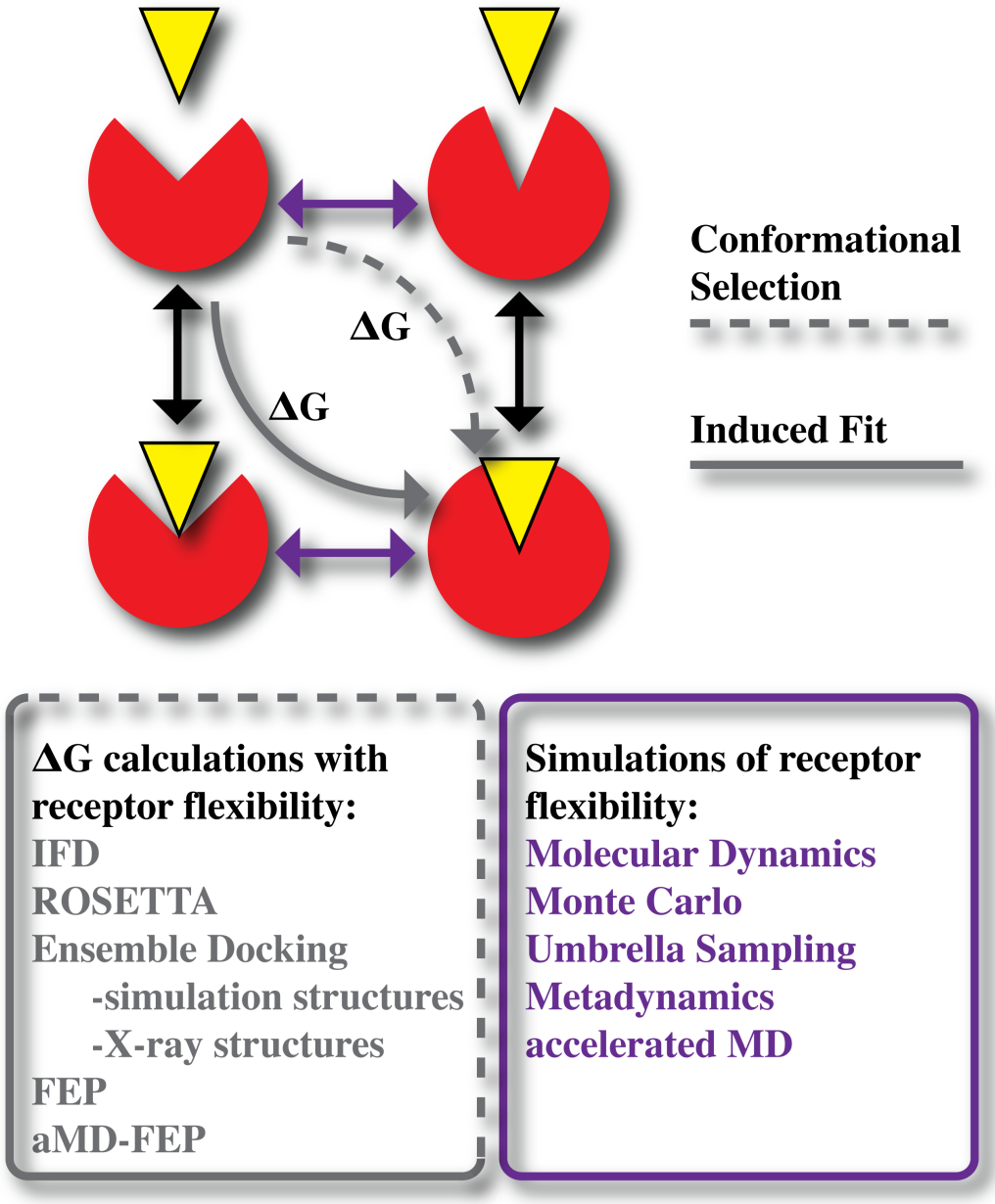


Figure 1.1 Schematic representation of the possible pathways taken in ligand binding. The induced fit pathways and conformational selection pathways are shown via dashed and solid lines. Various methods for binding energy calculations, which account for receptor flexibility are listed in the grey box. Simulation methods for sampling receptor flexibility are listed in the purple box.

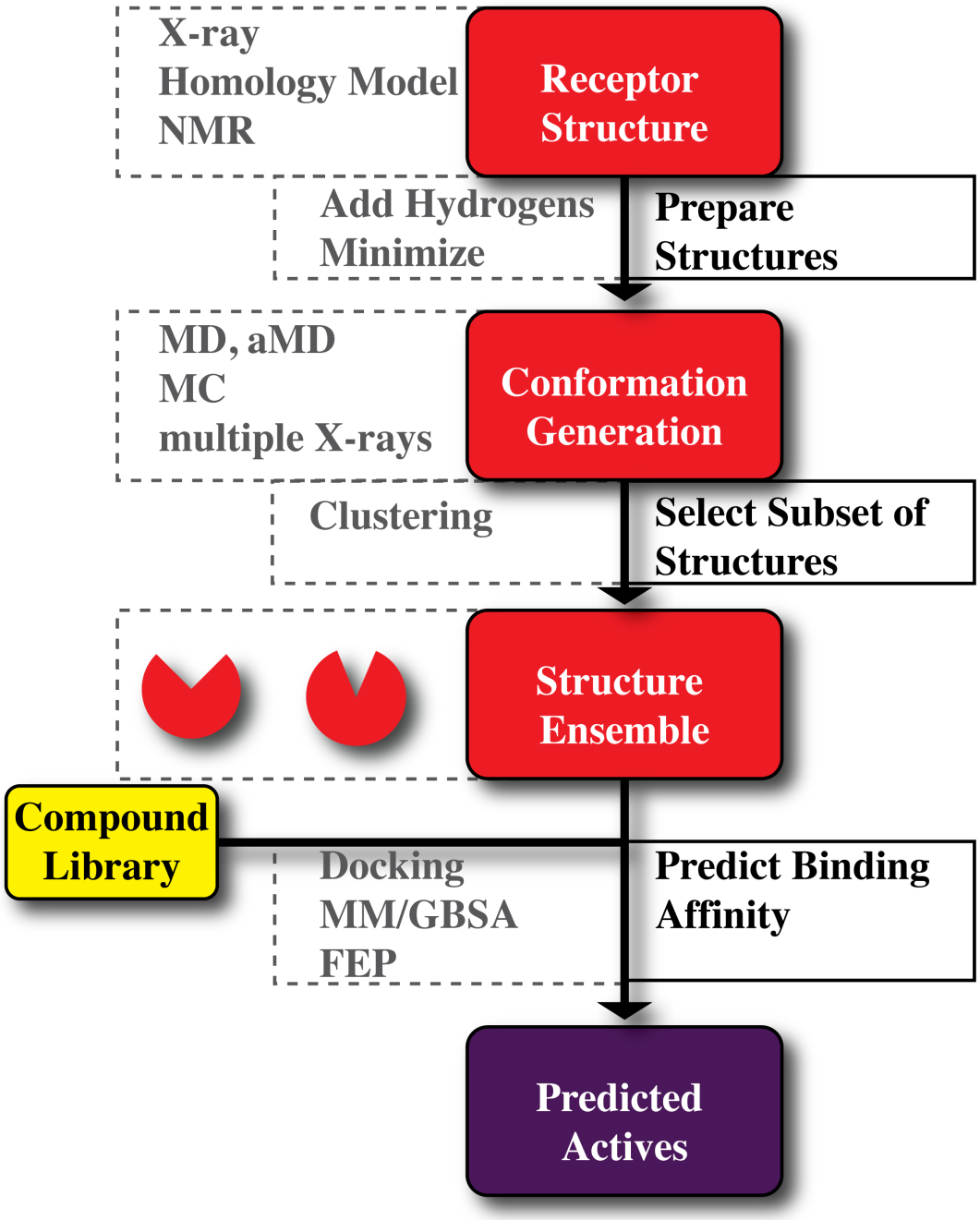


Figure 1.2 Schematic representation of ensemble based docking. Examples of methods or data used at each step are given inside the dashed grey lined squares.

Induced Fit Docking

Docking is a fast method for estimating binding free energy in which first the pose's are generated by performing a three dimensional conformational search of the active site with a small molecule probe, then poses are evaluated based on a scoring function which includes evaluation of steric clashes, hydrophobic interactions, hydrogen bonds and other non-covalent interactions. Traditionally only a single receptor conformation is used to score each pose for computational efficiency, however there are ways in which docking programs try to account for receptor flexibility. By softening the van der Waals potential of the atoms one can allow for small overlaps between the ligand and receptor without large steric penalties. However, this may increase the rate of false positives because more diverse structures are allowed to bind. It also does not allow for larger motions like side chain rotations or protein backbone motions. Allowing for certain side chains to rotate, also can account partially for receptor flexibility. Additionally, one can use induced fit docking where the receptor is allowed to move around the ligand to accommodate it. In this procedure, residue side chains are changed to alanine residues to prevent steric clashes in the initial docking. Then side chain predictions are used to generate possible conformations and the binding site and ligand are energy minimized. This allows for local movements around the protein upon ligand binding, and significant reductions of the root mean square deviation (RMSD) of some flexible protein binding poses as compared to crystal structures(*17*). Nonetheless, it remains a challenge to predict larger scale motions that may lead to different binding conformations.

Rosetta Ligand

An alternate method that accounts for receptor flexibility is RosettaLigand. RosettaLigand is based on the hugely popular modeling software Rosetta(*18*). In contrast to the before aforementioned methods RosettaLigand, does not split the receptor and ligand ensemble generation into separate steps that are combined during docking. Rather it allows for both receptor and ligand structural changes during the docking stage. Also in contrast to the previously mentioned methods, Rosetta uses knowledge-based scoring functions to assess protein conformations while the protein-ligand interactions are modeled employing first-principles. In its original implementation RosettaLigand allowed for ligand flexibility as well as receptor side chain flexibility(*19*). In a subsequent release the developers went beyond just accounting for side chain flexibility in the active site to incorporating full protein backbone and side chain flexibility(*20*). An extensive blind docking benchmark revealed that the performance of RosettaLigand is comparable to that of other commercial software (*21*). In its latest release, RosettaLigand can dock multiple ligands simultaneously, allowing for the redesign of the binding interface during docking, and it has a more user-friendly xml script interface (*22*).

Ensemble Based Screening Methods

Ensemble based screening methods rely on using varied receptor conformations in a docking protocol. The conformations can be determined from crystallographic, NMR structures, Monte Carlo simulation, molecular dynamic simulation, or enhanced sampling methods. See figure 2 for a schematic of ensemble based docking methods. The relaxed complex scheme (RCS)(*23*, *24*) is a type of ensemble docking, which relies on previously

determined conformations from molecular dynamics simulation to perform docking studies against. This ensemble of structures is then used in conjunction with a docking scoring function rather than a single structure, with the idea that an ensemble of low energy higher populated structures will bind a larger variety of compounds, and thus more hits will be obtained from a compound library(*23-25*). While one could just extract structures at equidistant time points, it seems more logical to perform a clustering type analysis, and use a single representative of related structures(*25*) to cover a large conformational space without redundant structures. In the final scoring of compounds, it may be useful to weight the score from each structure based on how often that structure was present in the original simulation, or one can simply take the highest score given. The problem remains though that perhaps very rare structures will be the most important in compound binding. The protein structure can be simulated in the apo state with the hope that without any compounds bound more states will be sampled since the structure will not be biased by ligand-protein interactions. However, in some structures the binding pocket may adopt a structure that does not bind ligands or occlude itself completely without a ligand bound. In this case, it is best to simulate with a ligand bound, however the structures will be biased towards the ligand bound state and may fail to explore a larger variety of potential ligand binding structures.

This method has been successfully used in a number of studies to find compounds for various targets. McCammon and colleagues described a novel binding trench in HIV integrase using RCS methods and docking, which helped inspire the discovery of the FDA approved drug raltegravir(*26*). Other diketo acid HIV integrase inhibitors have been

modeled and designed using RCS(*27*). In addition, inhibitors of essential enzymes for *Trypanosoma brucei* have been found using the RCS method (*28*). The selection of appropriate configurations based on compound class and active site volume have been described for the proposed antibacterial target undecaprenyl pyrophosphate synthase(*29*). New compounds have been described which inhibit neuraminidase in avian influenza(*30*). Although molecular dynamics (MD) structures can enrich active compound predictions, some structures may perform worse than X-ray structures and a broadly applicable protocol for a priori determining the most predictive structures from the simulations has not been determined(*31*). Indeed, ensemble based screening can generate more hits with more chemical diversity than a single structure, but if many crystal structures are available with varied binding site geometries the enrichment may be better for ensembles of crystal structures than for ensembles of simulation structures(*32*). This leads us to the conclusion that ensemble based screening is superior to a single structure screening, but simulation based ensemble screening may be more applicable when extensive crystallographic data is not available. Generally the configurational changes observed with these RCS methods and induced fit methods are smaller changes. The short timescale upon which most molecular simulations are run does not fully map out bio-molecular phase space. Greater sampling of biomolecular phase space as well as a broadly applicable method for determining the most predictive structures a priori would both be big steps forward in ensemble based docking.

Enhanced Sampling: Hardware improvements

Over the course of MD simulations if there is a high energy barrier between two low energy states it is unlikely that the simulation will cross this barrier and describe the second low energy state, unless the simulation is very long. In addition, the free energy surface of proteins is vast and rugged and exploration is slow. The timescales often simulated with MD using today's computers are nanoseconds to microseconds and sometimes even milliseconds(*33*). However, many interesting events occur on the timescales of milliseconds to seconds. In addition, if one wishes to recover a Boltzmann ensemble of structures then the crossing of energy barriers must be performed many times to generate converged statistics. It is of interest then to speed the crossing of high energy barriers and numerous methods have been proposed to perform this.

Conceptually the simplest improvements come from increased speed and efficiency of computer hardware and software. Two recent hardware advances have been graphics processing unit (GPU) accelerated computing and the construction of a special purpose machine Anton for running long MD simulations. GPU computing has shown impressive benchmarks for MD simulation as well as providing a more reasonably priced solution to traditional high performance computing (HPC). Popular simulation packages like NAMD(*34*), and AMBER have released code which runs on GPU's, often with impressive benchmarks. A fine review of GPU-computing in the context of molecular modeling is given Stone et al.(*35*). In addition, special purpose machines like Anton have been built for running fast MD simulations and extending brute force MD simulation into the millisecond timescale(*33*). These extremely long simulations have

been used to show drug binding events to their protein target(*36*). They have also been used to show in full atomic resolution how small proteins fold, and that in many cases a very accurate representation of the folded state can be obtained from MD simulation(*37*). The accuracy of force fields on these millisecond timescales has never truly been tested before and these new hardware developments have been an exciting development in molecular modeling. Protein flexibility on longer timescales is now being accessed much more routinely. In addition to the hardware changes, simulation techniques that help improve sampling have been developed and continue to show promise and efficiency over brute force descriptions of conformational flexibility.

Enhanced Sampling: Methodologies

One can speed sampling in a number of ways by introducing artificial biases into the model upon which the simulation is based. The simplest of artificial biases is raising the temperature, which causes more rapid fluctuations in structure by increasing the average velocity of all atoms. Temperature accelerated replica exchange uses many replicas at varying temperatures to increase sampling, and these replicas exchange with each other based on a Metropolis criterion, finally recovering the canonical ensemble (*38*). In addition, one may use a Hamiltonian modifying technique like accelerated molecular dynamics (aMD) in a replica exchange framework to modify only the potential energy surface at a given temperature (*39*). Other methods like umbrella sampling(*40-42*), and metadynamics(*43*, *44*) have also been used to enhance sampling. However many of these methods suffer from the need to define a reaction coordinate a priori to simulating the system. This is not advantageous if one is looking for new configurations

to bind a drug-like compound to a protein active site without knowledge of what that configuration is, but can be quite useful in determining the free energy change between known conformations.

Umbrella sampling

The calculation of free energy differences from molecular dynamics simulations requires enhanced sampling techniques to probe the energy landscape effectively. Free energy difference calculations are facilitated by the use of a reaction coordinate, a parameter that measures the degree to which the system is near each of the two or more thermodynamic states of interest. One major challenge is the finite simulation time that results in regions close to energy minima being sampled well, whereas regions of higher energy are rarely or never sampled(*45*). Sufficient sampling of the entire space along the reaction coordinate is necessary to derive the free energy difference. Umbrella sampling is one such technique that enforces adequate sampling of high energy regions. In umbrella sampling (or biased MD) (*40-42*), the underlying energy potential is modified in order to allow for an easier transition over the energy barrier. For this an additional energy term, often referred to as bias or bias potential, is applied to the system to ensure efficient sampling along the entire reaction coordinate (*45*). This is done in separate simulations or windows, which overlap. Bias potentials should be chosen to ensure that sampling along the reaction coordinate is as uniform as possible. Umbrella sampling frequently employs harmonic biasing potentials in a series of windows along the reaction coordinate. An alternative to this is the use of a single window in combination with an adaptive biasing potential aimed at matching the free energy profile along the reaction

coordinate as well as possible. This method is referred to as adaptive bias umbrella sampling. The sampling in each individual window can be performed using conventional molecular dynamics (cMD) or employing enhanced sampling techniques such as Hamiltonian replica exchange (*38*, *46*). Careful choice of the reaction coordinate is crucial for correct umbrella sampling results(*47*). The free energy curves are combined using techniques such as the weighted histogram analysis method (WHAM, (*48*, *49*)).

Metadynamics

Metadynamics is another technique designed to accelerate rare events and reconstruct the free energy profile(*44*). Metadynamics (*43*) relies on the introduction of a set of collective variables (CVs) which best describe the process of interest. Within the space of these CVs, a history-dependent potential is build up by dropping Gaussians along the sampled trajectory effectively discouraging the system from revisiting configurations that have already been sampled(*50*). The overall sum of these Gaussians is then used to compute a free energy profile along the CVs. For an excellent example of a metadynamics simulation see Fig. 1 in(*44*). The main difference to umbrella sampling is the non-systematic sampling along the collective variable(s). Important user-defined parameters are the height, width and dropping-frequency of Gaussians. Choosing the right set of CVs is crucial to the success of the simulation and criteria for picking good CVs have been outlined in the literature (*44*, *50*). An instructive example of what happens if the correct CVs are not chosen is given in (*51*). Unfortunately owing to the complexity of large biological systems, determining a set of appropriate collective variables is challenging. However, this has not deterred a large number of applications

using the method. Most notably to mention for this review is work that used metadynamics to characterize small molecule – protein interactions accounting for full receptor flexibility (*52-56*).

Accelerated Molecular Dynamics

Accelerated molecular dynamics (*57*), an extension of hyperdynamics(*58*), is a simulation method that does not require the selection of a reaction coordinate or CVs and is ideally suited for efficiently exploring configurational space. Accelerated MD modifies the potential energy surface based on the difference between a user defined reference energy E and the normal potential energy of the underlying MD force field at each point. Accelerated MD speeds sampling by decreasing the size of energy barriers and smoothing the energetic ruggedness of phase space exploration according to a parameter α. See scheme 1 for the equations used to modify the potential energy surface and a hypothetical 2-dimensional representation of the effect of aMD on a potential energy landscape. Accelerated MD has been used to simulate small benchmark systems like alanine dipeptide and reproduce the Ramachandran plot from MD(*57*). Originally aMD was only applied to torsional angles(*57*) but was subsequently extended to all force field terms including explicit solvent(*59*). These two forms were combined in the dual-boost approach(*60*). Accelerated MD has been used for complex systems and validated with experimental results such as the improved prediction of experimental NMR observables in large protein systems (*61*). Grant et al. suggested that conformational selection is the dominant mechanism for nucleotide binding dependent conformational

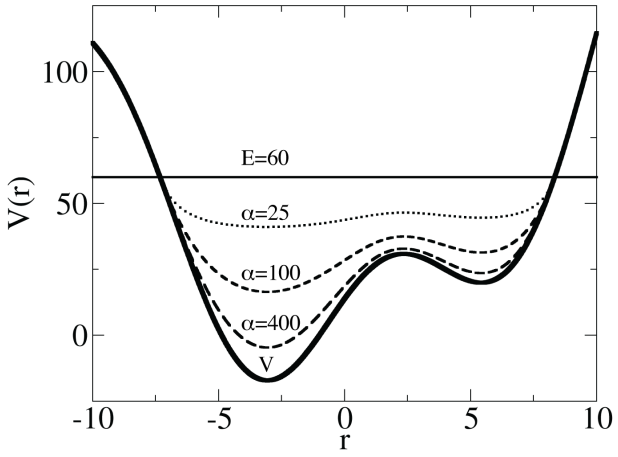
changes in G-proteins, while induced fit plays a smaller role based on data from aMD simulations(*62*).

It is however important to note that recovery of Boltzmann statistics in large simulations remains a challenge because of sampling limitations in of simulating large biomolecules. A non-Boltzmann simulation like aMD requires that each configuration be reweighted in a way that recovers the canonical ensemble(*57*). Often when longer timescales are accessed with aMD the reweighting procedure is subject to statistical error in the estimate of the weighting factor for each point. An excellent review of the statistical issues with reweighting is given by Shen and Hamelberg(*63*). However, since the underlying shape of the energy surface is preserved, although flattened, one can consider the most common configurations from aMD to be likely structures although the calculation of the exact ratios of their populations is still challenging. In simulations with the goal of conformational exploration where the exact free energy of the conformation is not desired, simulations are often described without reweighting, with the risk that a high energy state may have been over-represented. Recently, Wereszczynski et al. explored the conformational space of GET3 (a protein involved in the guided entry of tail anchored proteins into the membrane) with aMD and then performed rigorous potential of mean force calculations to determine the energetics of conformational change in the presence of various bound nucleotides(*64*). Accelerated MD was also recently used to study the dynamics of the important cardiomyocyte calcium binding protein troponin C (*65*). Indeed, enhanced sampling methods can play a critical role in the generation of

structural ensembles with larger conformational changes and these methods are also seeing increased use in combination with rigorous free energy calculations.

$$V^*(r) = \begin{cases} V(r), & V(r) \geq E \\ V(r) + \Delta V(r), & V(r) < E \end{cases}$$

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))}$$



Scheme 1.1 Accelerated Molecular Dynamics. Equations used to calculate the boost energy and modified potential energy surface in aMD. A two dimensional representation of the modified potential, V*(r) (dashed lines), and the unmodified potential energy surface, V (thick black line). a was varied as indicated E was always fixed at 60 and is indicated by a thin solid black line.

Accelerated MD in free energy calculations

Free energy calculations based on computer simulations have been pursued for a few decades now, although they are not as ubiquitous in pharmaceutical settings as quicker methods of binding energy estimation such as docking. Part of the problem is that they are not well automated like docking algorithms. Another major issue is that the computational power required for these calculations is great and a large investment in computing resources is needed before one can predict the affinity of a single ligand let alone a large compound library. We now have the computational power to make predictions of the free energy using methods like free energy perturbations (FEP) and thermodynamic integration (TI). These methods rely on defining a thermodynamic cycle

where alchemical transitions can be used to change between states and then one can calculate a free energy difference between two states. Often the states will be the bound and unbound form of the ligand, but variations allw calculation of other properties such as the free energy of solvation, or the difference in free energy of binding between two ligands(*66*). One of the reasons behind the slow speed of these calculations is that the alchemical change must be made so that there is appropriate phase space overlap between the successive states along the reaction coordinate between the different end states. In thermodynamic integration calculations similar convergence challenges arise. This requires considerable sampling of any conformational changes around the modified portion of the system, as well as accurate convergence of the energy of those conformations, which can lead to large amounts of simulation. Since free energy methods rely on sampling, each simulation will be different and one can generate a simple estimate of error by running simulations in replicate, further increasing the need for computation.

Because rapid generation of low energy configurations is necessary for accurate free energy estimations, it is logical that enhanced sampling methods may be beneficially be applied to these calculations. Accelerated MD has been applied in a number of ways to free energy calculations. Fajer et al. used a replica exchange framework where the difference between replicas was the level of acceleration applied via varied boost parameters(*39*). Since the ground state replica was run on an unadulterated potential energy surface there were no issues with reweighting it as long as the replicas were spaced closely enough that they would exchange rapidly. There are however many ways

to incorporate data from replicas at different levels of acceleration and multistate Bennett acceptance ratio (MBAR) was determined to be roughly 4 times more efficient at recovering data than the ground state alone(*67*).

Oliveira created upside down aMD to overcome some of the issues of reweighting and calculate free energies with a single replica. The simple test system of a butane to butane symmetric transformation was used to show that one can increase the accuracy and speed of convergence by enhancing sampling with aMD in free energy calculations(*68*). This method allows the simulation to populate statistics by running at low to no acceleration in low energy regions, and then jumping energy barriers to rapidly move between different configurations. While quite promising in simple systems, this method proved challenging to parameterize so that high energy barriers were not flattened more than low energy barriers in simulations with many degrees of freedom. To ameliorate this issue, we proposed a boost limiting factor and created windowed aMD. This method efficiently reproduced the free energy surface of alanine dipeptide. In addition the new boost equation was used to calculate the free energy difference between the antibiotic vancomycin and two of its glycopeptide binding partners using TI. The overall results demonstrated more rapid convergance of free energy calculations and easily reweighted statistics, however more parameters were required (*69*).

Since the reweighting statistical error is directly related to the amount of acceleration and the complexity of the system, it seems reasonable to limit the portion of the simulation to which boost is applied. Acceleration of just the dihedral angles was the original way aMD was applied and in biomolecular systems these angle terms provide

much of the restraint in conformational exploration(*57*). This also provided a convenient limitation of the energetic terms to which acceleration was applied since water molecules, which make up the majority of the atoms in explicit solvent simulations, have no dihedral angles. Selectively applied aMD took this idea one step further and limited acceleration only to predefined dihedral angles in alanine dipeptide, and in a free energy simulation of decoupling oseltamivir's binding to neuraminidase. This work demonstrated that reweighting a small subset of dihedral angles helped overcome reweighting issues and that the free energy results converged to the same answer as cMD simulations, but in less time(*70*). Although this provides for better statistical recovery, it requires the prediction of which dihedral angles are important in ligand recognition or protein flexibility, which may be non-trivial.

Concluding remarks

There has been tremendous progress in the field of personalized sequencing of genetic code within the past 20 years. The "$100 genome" is within reach in our generation. The possibility of personalized genetic knowledge of specific diseases has unprecedented potential for targeted drug treatment. This exciting development goes hand in hand with advances in computer aided drug design that can be used to discover novel leads targeting specific mutant receptors. Here we presented recent developments in computational algorithms and hardware used in drug discovery with a focus on using enhanced protein dynamics sampling techniques to aid in the incorporation of full receptor flexibility in structure based drug discovery. In the near future, it should be possible to include the effects of genetic variation in models of drug targets, and speed

the choice of therapies appropriate for individual patients. The effects of genetic variation not only lead to sequence changes, but structural and dynamical changes too. Thus we anticipate that computer-aided drug discovery which accommodates receptor flexibility will be an important component of pharmacogenomics in the near future opening many new and exciting opportunities for combining the two techniques.

Acknowledgements

This work was supported by the Molecular Biophysics Training Grant GM08326 (WS), the National Science Foundation Grant MCB-1020765, NBCR, CTBP, Howard Hughes Medical Institute, and National Institutes of Health Grant GM31749 (JAM).

Chapter 1 contains text from the article in preparation Sinko W.; Lindert S.; McCammon, J. A. Accounting for Receptor Flexibility, and Enhanced Sampling in Computer Aided Drug Design. (In preparation for CBDD special edition)

Chapter 2: Discovery of Small Molecule Inhibitors of the PH domain Leucine-rich repeat Protein Phosphatase PHLPP by Chemical and Virtual Screening

Abstract:

PH domain Leucine-rich repeat Protein Phosphatase (PHLPP) directly dephosphorylates and inactivates Akt and protein kinase C, poising it as a prime target for pharmacological intervention of two major survival pathways. Here we report on the discovery of small molecule inhibitors of the phosphatase activity of PHLPP, a member of the PP2C family of phosphatases for which there are no general pharmacological inhibitors. First, the Diversity Set of the NCI was screened for inhibition of the purified phosphatase domain of PHLPP2 *in vitro*. Second, selected libraries from the open NCI database were docked into a virtual model of the phosphatase domain of PHLPP2, previously trained with our experimental data set, unveiling additional inhibitors. Biochemical and cellular assays resulted in the identification of two structurally-diverse compounds that selectively inhibit PHLPP *in vitro*, increase Akt signaling in cells, and prevent apoptosis. Thus, chemical and virtual screening has resulted in the identification

of small molecules that promote Akt signaling by inhibiting its negative regulator PHLPP.

Abbreviations:

DMEM, Dulbecco's modification of Eagle's medium ; DMSO, dimethyl sulfoxide ; EGF, epidermial growth factor ; ERK, external signal regulated protein ; FBS, fetal bovine serum ; FoxO, Forkhead box O; GSK-3, glycogen synthase kinase-3 ; MAPK, mitogen activated protein kinase; PDK-1, phosphoinositide-dependent kinase-1 ; PH, pleckstrin homology; PHLPP, PH domain Leucine-rich repeat Protein Phosphatase ; PI3K, phosphatidylinositol-3 kinase; PKC, protein kinase C ; pNPP, para-nitrophenyl phosphate ; PPM, protein phosphatase magnesium-dependent ; PMSF, phenylmethyl sulfoxide ; PP1, protein phosphatase 1 ; PP2A, protein phosphatase 2A ; PP2B/calcineurin, protein phosphatase 2B ; PP2C, protein phosphatase 2C ; PPP, phosphoprotein phosphatase ; PTEN, phosphatase and tensin homolog ; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel.

Introduction

Transient phosphorylation of proteins is a fundamental mechanism by which cells integrate and transduce signals. Kinases and phosphatases act in dynamic opposition to control the extent, duration and intensity of signaling and to maintain cellular homeostasis. Dysregulation of the precisely-tuned balance between phosphorylation and dephosphorylation results in pathophysiological states.

The phosphatidylinositol-3 kinase (PI3K)-Akt pathway is one of the major phosphorylation cascades that control cell fate(*71*). Stimulation by growth factors, such

as EGF or insulin, results in phosphorylation of receptor tyrosine kinases and recruitment of effector proteins, notably PI3K, to the receptors. PI3K phosphorylates the lipid phosphatidylinositol-4,5-bisphosphate (PIP_2) to yield phosphatidylinositol-3,4,5-trisphosphate (PIP_3). PIP_3 recruits Akt to the plasma membrane where the protein is phosphorylated by its upstream kinase phosphoinositide-dependent kinase-1 (PDK-1) at the activation loop (Thr308 in Akt1). A subsequent phosphorylation occurs at the hydrophobic motif (Ser473 in Akt1) by a mechanism that depends on the TORC 2 complex(*72*). Once phosphorylated, Akt is released from the membrane and phosphorylates diverse substrates throughout the cell, thus inducing a wide range of physiological effects, notably cell growth, proliferation and survival. In addition, Akt is a master regulator of glucose metabolism, playing a key role in mediating the biological effects of insulin(*73*). The activation of Akt is opposed by 1] lipid phosphatases that dephosphorylate, and thus remove, the lipid second messenger and 2] protein phosphatases that dephosphorylate, and thus inactivate, Akt. Specifically, PTEN dephosphorylates PIP_3 (*74*), to terminate the activation of Akt. Activated Akt is dephosphorylated at the activation loop by okadaic acid sensitive phosphatases(*75*) such as PP2A and at the hydrophobic motif by the recently discovered PH domain Leucine-rich repeat Protein Phosphatase (PHLPP)(*76*, *77*), resulting in inhibition of activity and promotion of apoptosis.

PHLPP was initially discovered as the phosphatase that dephosphorylates and inactivates Akt in cells, but it also dephosphorylates and regulates the levels of protein kinase C (PKC) isozymes(*78*), another important class of kinases that control cell growth and survival. PHLPP is a family of three isoforms: the alternatively-spliced PHLPP1α

and PHLPP1β, and PHLPP2(*79*). The phosphatase domains of the three enzymes are highly similar, with 58% amino acid identity. They belong to the PP2C family of phosphatases, which, in turn, belong to the larger PPM (Protein Phosphatase Magnesium/Manganese dependent) family of serine/threonine protein phosphatases, which require Mn^{2+} or Mg^{2+} for their activity. The primary known function of the PP2C family is to down-regulate stress responses in eukaryotes(*80*, *81*). PP2C phosphatases differ from those in the PPP family (which also require metallic cations for their activity) by their resistance to common serine/threonine phosphatase inhibitors such as okadaic acid and microcystin(*82*). In fact, there are no general inhibitors of the PP2C family available, although cyclic peptide inhibitors for PP2Cδ(*83*) and small molecule inhibitors for PP2Cα, identified by virtual screening(*84*), have been reported. Given the high therapeutic value of inhibitors for protein kinases to target disease(*85*, *86*), discovery of phosphatase inhibitors is likely to have a major impact in future therapeutics.

Because PHLPP dephosphorylates Akt and PKC, positioning it as a suppressor of two major survival pathways, PHLPP inhibition would be particularly relevant therapeutically in diseases where survival pathways are repressed, notably diabetes and heart disease. Indeed, Akt and PKC activities are repressed in both diabetes mellitus and cardiovascular conditions such as myocardium infarction and ischemia-reperfusion (I/R) injury. In diabetes mellitus, the Akt pathway is a therapeutic target for islet transplant and survival as well as in the treatment of associated vascular complications(*4*). Akt activity is important for β-cell growth, survival, and insulin production(*87*, *88*). Studies have demonstrated that transgenic over-expression of Akt in islet β-cells gives rise to larger islets resulting from increases in the number and size of cells(*89*, *90*). This

hypertrophy is combined with an increase in insulin production; mice are also resistant to streptozotocin-induced diabetes. Conversely, over-expression of kinase-dead mutants(*91*) or impaired PDK-1(*92*) in transgenic mice leads to defective insulin production and increased susceptibility to streptozotocin. Activation of Akt by different means has been used to improve transplantation success already(*93*, *94*). In cardiovascular diseases, activation of pro-survival pathways is key to protect the heart from damage because cardiovascular injuries are often linked to myocyte cell loss through apoptosis(*3*, *95*, *96*). Akt has a number of positive effects on I/R-mediated damage of the heart that are mediated by different substrates(*97*, *98*). For example, infarct size is reduced through inhibition of GSK3β and this effect is reversed by the PI3K inhibitors, LY 294002 and wortmannin. In the case of PKC, activation of PKC ε has been established to mediate cardiac protection from cardiac ischemia(*99-103*). Ischemic preconditioning(*103*) and many pharmacological agents(*104*, *105*) including insulin, adenosine A_1/A_2 agonist, bradykinin, natriuretic peptides or erythropoietin achieve their protective effect through activation of Akt and PKC. Thus, inhibition of PHLPP, a repressor of Akt and PKC activity, would provide a novel tool promoting the concomitant activation of the two key survival pathways.

Here we report on the discovery of small molecule inhibitors of PHLPP phosphatase activity. These molecules were identified by medium throughput chemical screening and virtual screening of the NCI repository. We identified molecules that inactivate PHLPP at low micromolar concentrations *in vitro*, increase basal and agonist-evoked Akt phosphorylation in cells, and suppress apoptosis.

Results:

As there is no general inhibitor of PP2C, we started our search for inhibitory small molecules of PHLPP by screening the First Diversity Set of the National Cancer Institute. This set comprises 1990 compounds chosen amongst the 140,000 compounds in the repository to encompass the largest chemical space possible. These molecules were assayed in a 96-well format, at concentrations of 100 μM, using the isolated phosphatase domain of PHLPP2 purified from *E. coli* as the enzyme and pNPP as the substrate. Statistical analysis revealed a z value(*106*) of 0.5 and a signal-over-background ratio of almost 4, indicating the assay was statistically valid (Figure 1a). Dephosphorylation of pNPP results in an increase of the optical density of the solution, thus the slope of the change of O.D over time served as a measure of the activity of the phosphatase (Figure 1b). In one assay, 80 compounds can be tested, as well as 12 controls of uninhibited activity (DMSO) and 4 controls for background (absence of enzyme). Activity in each well was measured and the value normalized to that in the absence of inhibitor; 88 compounds were identified that reduced activity to below 0.3 of the control value (Figure 1c), the criterion chosen to warrant further testing as an inhibitor. Colored compounds, which interfered with the colorimetric assay, were further tested at concentrations of either 10 or 20 μM depending on the intensity of the color. Finally, the IC_{50} values of the 50 most promising compounds were determined; these ranged from 1 to 100 μM, with 10% false positives for which the colorimetric change was unrelated to phosphatase activity. Structural analysis of inhibitory molecules led to the identification of 11 different chemical backbones (Figure 2 and Supp. Figure 1). We then retested other

compounds of the Diversity Set with these backbones and uncovered 49 additional inhibitors.

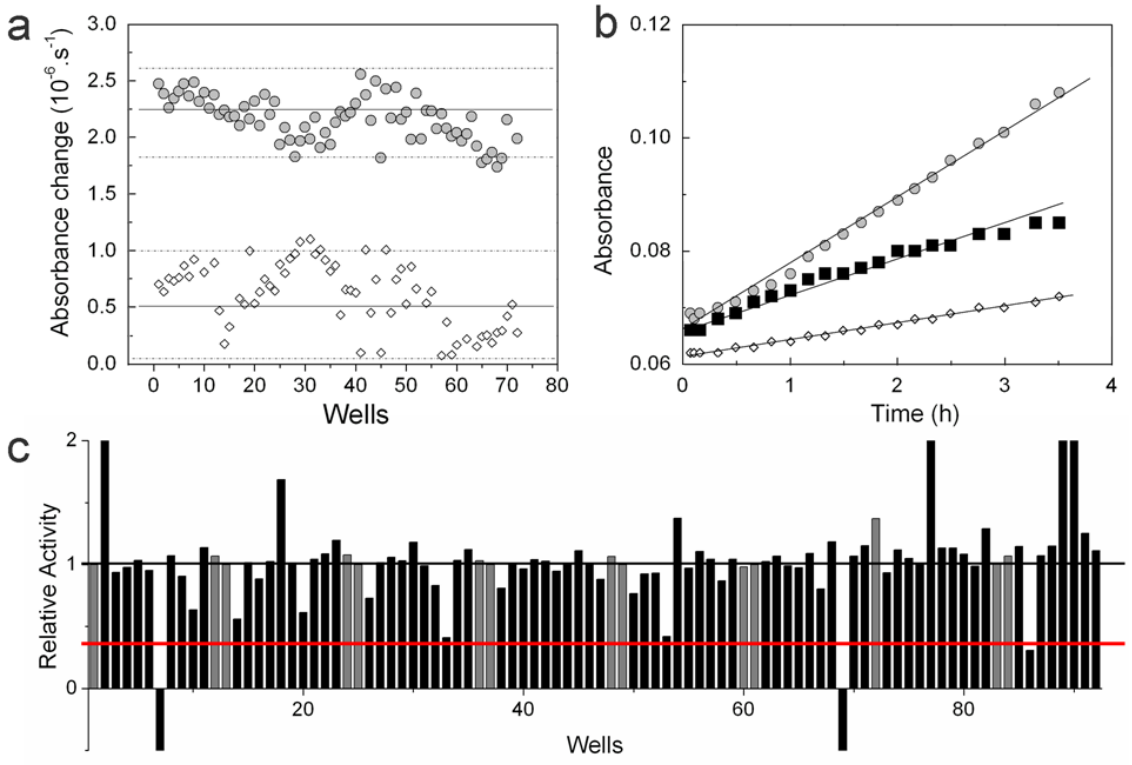


Figure 2.1 Screen of NCI Diversity Set. [a] Scatter plot of the raw data from 25 different plates showing control (DMSO) (●) and background (absence of enzymes) values (◇). The solid lines represent the mean of the control and background data and the broken lines display 3 standard deviations (SD) from the mean of each data set. The z-factor is 0.5 and the signal-over-background ratio is 3.8. [b] Representative time course for the hydrolysis of pNPP by PHLPP2 catalytic domain. pNPP (8 mM) was incubated in 125 µL assay buffer at room temperature, without (◇) or with the phosphatase domain of PHLPP2, in the presence of DMSO (●) or compound 24515 (■). The absorbance of the solution was recorded at 405 nm and the data fit by linear regression. [c] Representative activity data from one 96-well plate. NCI plate 4238 was assayed for PHLPP2 activity as in [b]. Each compound is represented by a black bar, DMSO controls are represented by grey bars (background controls were omitted on this graph). The values obtained for the DMSO controls were averaged and normalized to a relative activity of 1.0 (black line). Compounds that reduced PHLPP2 relative activity to lower than 0.3 (red line) were considered a hit (lane 86).

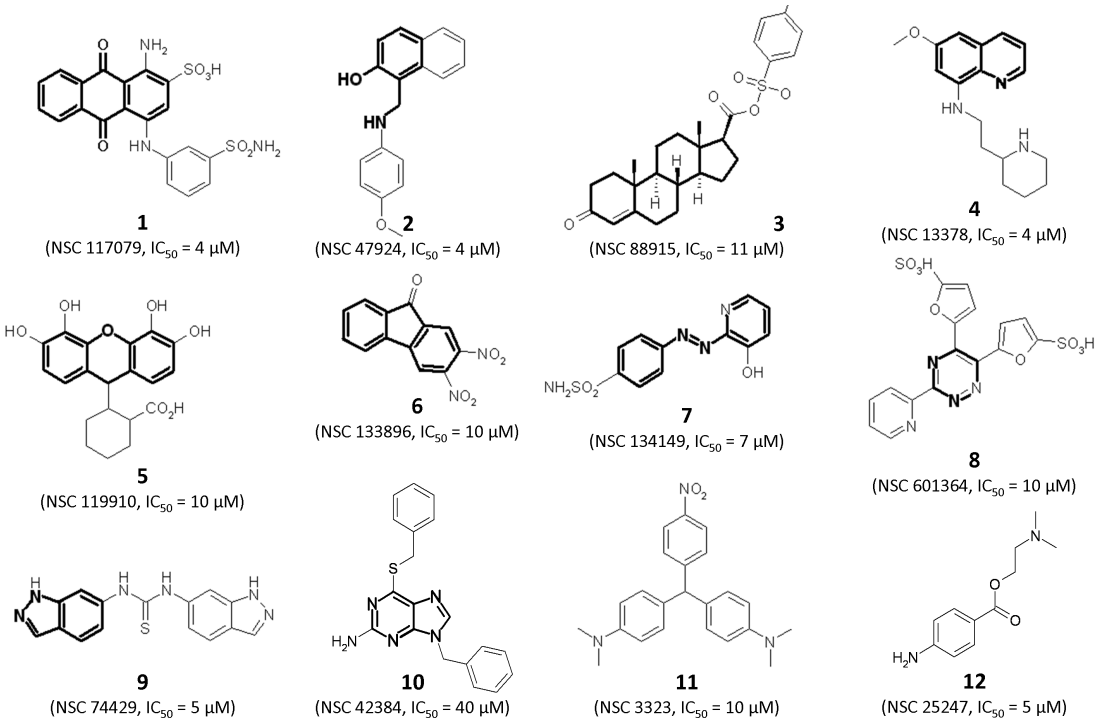


Figure 2.2 Chemical structures of some inhibitors of PHLPP2. Different compounds that have been identified as inhibitors of PHLPP2 are presented along with the experimental IC_{50} determined at 8 mM substrate. The chemical backbones used to determine the structural families are highlighted.

We turned to virtual screening to expand the scope of our investigation. Docking programs have been successfully used to identify novel inhibitory compounds of crystallographically-solved signaling phosphatases(*107-109*), including PP2Cα(*84*). We employed the GLIDE algorithm because it is well-established in virtual drug discovery work and has performed well in comparative docking studies(*110-112*)(*113*).

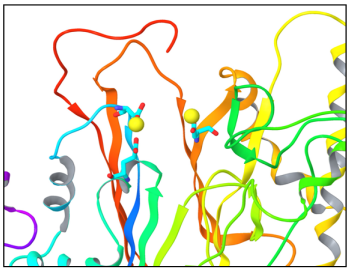
Because of the lack of crystallographic data, a structure for PHLPP2 based on homology modeling was created that was capable of discerning inhibitory binding compounds from non-binding compounds. The first step consisted of designing a model that would correlate best with our experimental data. The amino acid sequence of the PHLPP2 phosphatase domain was aligned with that of the phosphatase domain of

PP2Cα, resulting in an alignment score of 23% (Figure 3a), and a homology model was then produced based on the crystal structure of PP2Cα(*114*). Initial docking results of our control set of known inhibitors using GLIDE did not correlate well with aforementioned inhibitors of PHLPP2. Thus, we concluded that our homology model was not in an optimal conformation or that metal ions or water molecules played a critical role in inhibitor binding. Because many of our inhibitors contain negatively charged moieties, the interaction between these groups and the metal ions are predicted to contribute significantly to the overall binding energy. The crystal structure of PP2Cα coordinates 2 Mn^{2+} ions in the catalytic center. PP2C phosphatases in mycobacteria were found to accommodate a third metallic center in their catalytic core(*115-117*). Previous work has also shown that the PP2Cα active site metal ions are coordinated by 6 different water molecules. Since our structure is a homology model of the phosphatase domain of PP2Cα, it should include 1 to 3 Mn^{2+} ions and coordinated water molecules. We tested this by placing varying numbers of Mn^{2+} ions inside the active site near residues that could coordinate them and relaxed each structure to accommodate the ions. This resulted in a variety of structures, which we tested for the ability to recognize inhibitory compounds. All structures with 1 or more Mn^{2+} ions in the active site recognized inhibitors markedly better than the structure with no Mn^{2+} ions (Figure 3c).

a

```
PP2Cα    ---GAFLDKPKMEKHNAQGQGNGLRYGLSSMQGWRVEMEDAHTAVIGLPS 47
PHLPP2   GNTNLVLEHKTLDIFSHITTLKIDQKPLPTTDSTVTSTFWSHGLAEMAGQ 50
PP2Cα    GLESWSFFAVYDGHAGSQVAKYCCEHLLDHITNNQDFKGSAGAPSVENVK 97
PHLPP2   RNKLCVSALAMDSFAEGVGAVYGMFDGDRNEELPRLLQCTMADVLLEEVQ 100
PP2Cα    NGIRTGFLEIDEHMRVMSEKKHGADRSGSTAVGVLISPQHTY-------- 139
PHLPP2   QSTNDTVFMANTFLVSHRKLGMAGQKLGSSALLCYIRPDTADPASSFSLT 150
PP2Cα    FINCGDSRGLLCRNRK-VHFFTQDHKPSNPLEKERIQNAGGSVMIQR-VN 187
PHLPP2   VANVGTCQAVLCRGGKPVPLSKVFSLEQDPEEAQRVKDQKAIITEDNKVN 200
PP2Cα    GSLAVSRALGDFDYKCVHGKGPTEQLVSPEPEVHDIERSEEDDQFIILAC 237
PHLPP2   GVTCCTRMLG-----CTY----LYPWILPKPHISSTPLTIQD-ELLILGN 240
PP2Cα    DGIWDVMGNEELCDFVRSRLEVTDDLEKVCNEVVDTCLYKGSRDNMSVIL 287
PHLPP2   KALWEHLSYTEAVNAVRHVQDPLAAAKKLCTLAQS----YGCQDNVGAMV 286
PP2Cα    ICFPNAPKVSPEAVKKEAELDKYLECRVEEIIKGVPDLVHVMRT--LASE 335
PHLPP2   VYLN-IGEEGCTCEMNGLTLPGPVGFASTTTIKDAPKPATPSSSSGIASE 335
PP2Cα    NIPSLPPGGELASKRNVIEAVYNRLNPY 363
PHLPP2   FSSEMSTSEVSSEVGSTASDEHN----- 358
```

b



c

Model Name	Number of Mn^{2+} ions	Residues Coordinating Metal Ions	Number of Waters in the Active Site	Number of true hits
1	3	D806, E989, D1024	0	15
2	2	D806, D1024	6	13
3	2	D806, D1024	0	12
4	2	D806, E989, D1024	0	11
5	1	E989, D1024	0	8
6	0	----------	0	0

d

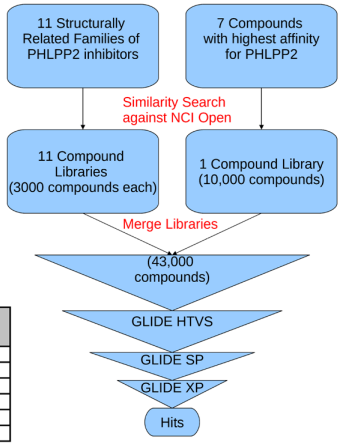


Figure 2.3 Virtual screen for inhibitors using modeled structure of PHLPP2 phosphatase domain. [a] Sequence alignment of PP2Cα and PHLPP2. Amino acids in red may be important for coordinating metal ions. * denotes a conserved residue. [b] 3-dimensional representation of model 4 of table c with D806, E989, and D1024 shown coordinating 2 Mn^{2+} ions in the phosphatase active site. [c] Table indicating the functional relevance of modified homology models. Model 4 was chosen for further docking because of its high docking scores and good performance identifying true hits from the Diversity Set. The number of true hits is derived from docking 50 compounds with experimentally derived IC_{50}s for PHLPP2 and determining how many compounds scored better than -7. [d] Structural similarity search to create libraries of compounds similar to known binders of PHLPP2. Selected compounds were docked with the GLIDE algorithm employing calculations at 3 levels of accuracy, removing lower scoring compounds at each stage.

Next, the entire Diversity Set was docked against our model. This served as a means to test the model for its ability to discriminate true inhibitors from a decoy set of ligands with no experimental activity. The docking protocol was modified so that only the top 4% of ligands were given final docking scores, as would be the case during virtual screening. From these studies, we determined that the model with 2 Mn^{2+} ions in the active site coordinated by D806, E989, and D1024 was most capable of discriminating true binders from decoys. In addition, this model had the highest range of G-scores for true hits (Figure 3c, model 4 and Figure 3b). Addition of water molecules did not improve detection of true inhibitors, although it is likely that they contribute to the coordination of ions in the active site. 40 new compounds were found to dock with G-

scores better than -7 kcal/mol, in addition to some of the previously characterized inhibitors. These new virtual hits were tested experimentally and 14 of these new compounds were determined to have IC_{50} values below 100 μM. Rarely do docking studies serve as a means to identify false negatives in a chemical screen but, in this case, combining chemical testing and virtual testing prevented us from missing 14 inhibitors of PHLPP. Model 4 was chosen for further studies because of its ability to distinguish hits from decoys and value in identifying 14 false negatives in the chemical screen.

Armed with a substantial dataset of inhibitory molecules, we hypothesized that finding similar structures and docking them might enlarge our pool of known binders and increase our hit rate over random virtual screening of the NCI repository. As previously mentioned, 11 structurally-related compound families were identified from *in vitro* screening; these were used as the references for similarity searches performed on the NCI Open compound library (Figure 3d). In addition, 7 of the highest affinity compounds were also used as reference compounds for similarity searches. A total of 43,000 compounds were identified from these similarity searches and docked to model 4. 80 compounds amongst the top ranked structurally-similar compounds were tested experimentally, at concentrations of 50 μM, using the same protocol as described for the original screen. These 80 compounds were selected based on good docking scores, structural diversity, and availability from the NCI. 23 compounds reduced the relative activity of the PHLPP2 phosphatase domain to below 0.5 of control and were considered hits. Of these, 20 compounds had an IC_{50} below 100 μM, with 15 of these having an IC_{50} value below 50 μM (Figure 4). Thus, we discovered a number of new, experimentally-verified low-μM inhibitors by integrating chemical data into our virtual screening effort.

a

N°	NSC ID	GLIDE G-SCORE (kcal/mol)	PHLPP2 IC_{50}
13	45586	-11.07	4
14	79714	-11.56	10
15	45618	-12.57	15
16	51532	-12.07	20
17	134146	-11.31	30
18	65541	-12.24	32
19	75913	-11.03	33
20	134145	-12.09	34
21	12509	-10.27	35
22	47754	-11.68	39
23	366218	-10.81	41
24	11404	-10.04	45
25	47760	-13.9	49
26	114428	-12.16	49
27	79729	-12.28	52
28	51536	-11.3	58
29	45619	-11.32	71
30	27236	-10.36	75
31	12129	-11.46	80
32	326196	-10.5	83

b

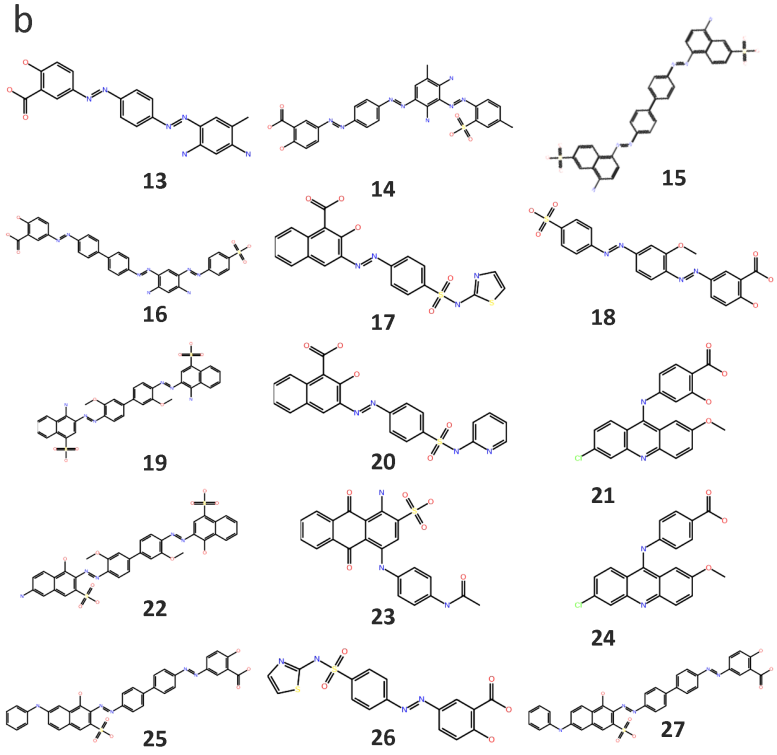


Figure 2.4 PHLPP2 inhibitors from the virtual screening. [a] Table of experimentally confirmed inhibitors of PHLPP2 and their estimated free energy of binding (kcal/mol) as well as IC_{50}. [b] 2-dimensional representations of the compounds which inhibit PHLPP2 from the docking study.

We next undertook a kinetic analysis of select compounds to determine their mechanism of inhibition. Because the chemical and virtual screen focused on the isolated phosphatase domain, we expected inhibitors to be primarily active-site directed rather than allosteric modulators. Determination of the rate of substrate dephosphorylation in the presence of increasing concentrations of the inhibitors revealed three types of inhibition: competitive, uncompetitive and non-competitive (data not shown). We docked pNPP (Figure 5a) and a phosphorylated decapeptide based on the hydrophobic motif sequence of Akt (Figure 5b; HFPQF^{p}SYSAS) into the active site of our best homology model, in the same manner as described for the inhibitors, to determine which substrate binding sites our inhibitor compounds could be blocking. Competitive inhibitors (e.g. **7**

(NCS 134149); Figure 5c and 5e) were predicted to effectively block the binding site of pNPP, as expected for a competitive inhibitor. In contrast, uncompetitive inhibitors (e.g. **4** (NCS 13378); Figure 5d) and most of the compounds determined from our virtual screen (e.g. **13** (NCS 45586); Figure 5f) were predicted to bind the hydrophobic cleft near the active site and interact with one of the Mn^{2+} ions. Non-competitive inhibitors (e.g. **2** (NCS 47924)) tended to dock poorly into our model, as expected if they bind sites distal to the substrate-binding cavity. Note that pNPP is a small molecule which, although it binds the active site and is effectively dephosphorylated, does not recreate the complex interactions of PHLPP with hydrophobic motifs and large peptides. Therefore, the type of inhibition we observe towards pNPP may not necessarily hold for peptides or full-length proteins. Importantly, we identified a number of inhibitors predicted to dock well in the active site and with kinetic parameters (competitive or uncompetitive inhibition) consistent with such docking.

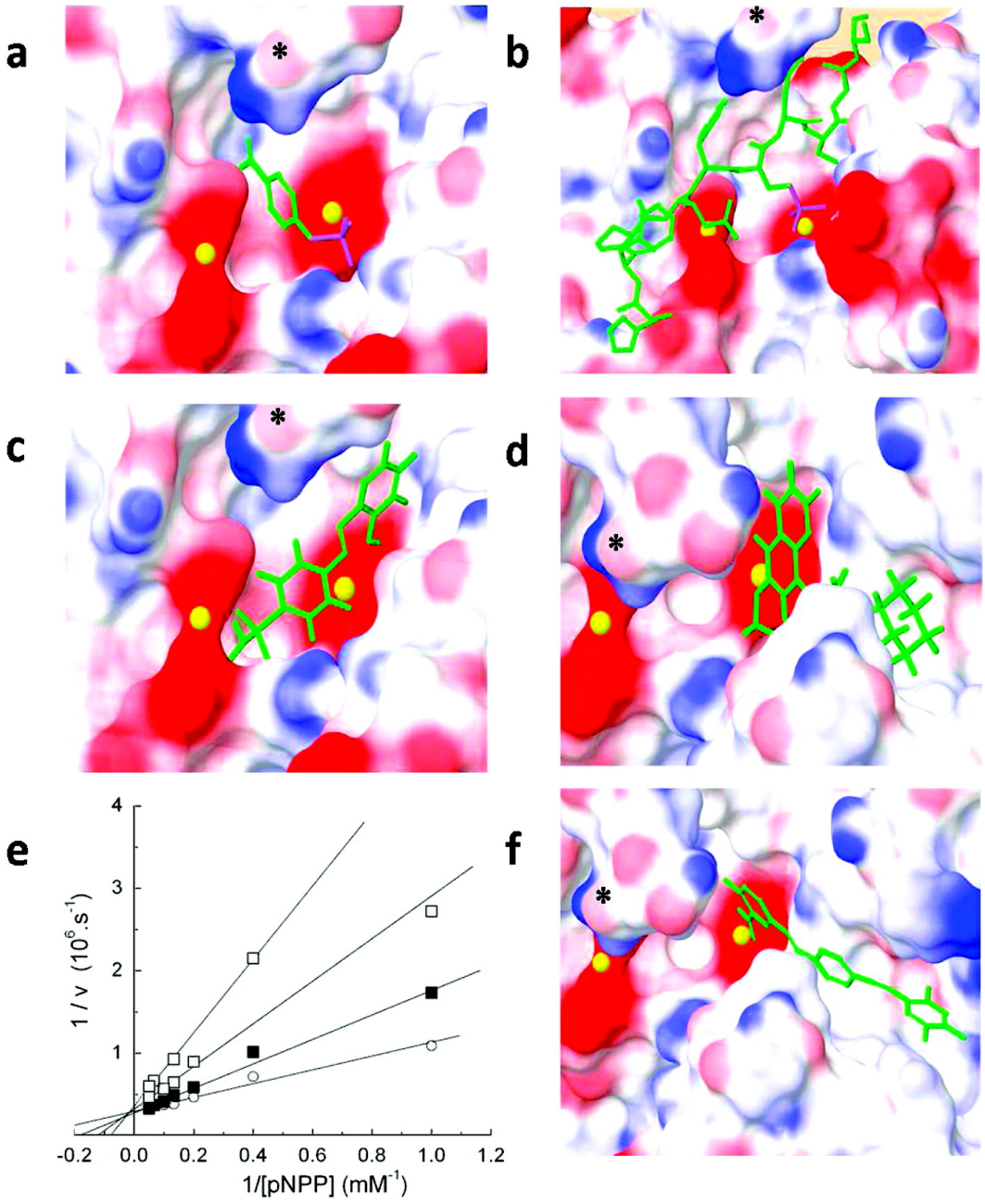
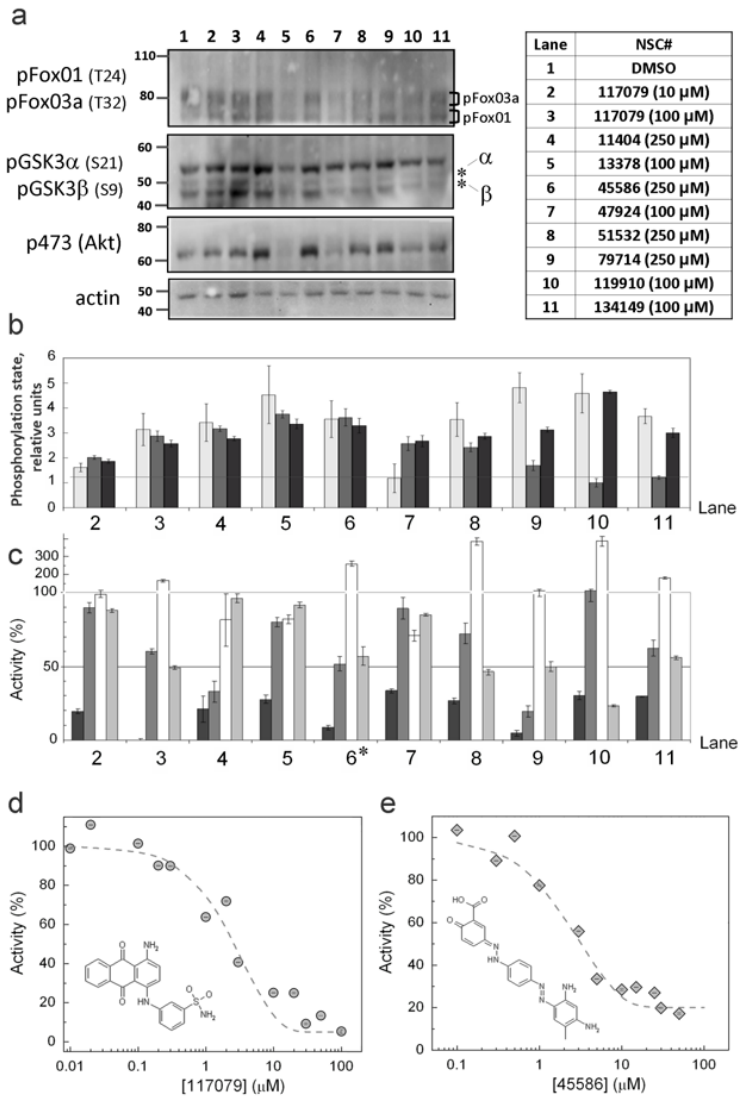


Figure 2.5 Model of active site of PHLPP2 docked with substrate or inhibitors. [a] pNPP, [b] a phospho-peptide based on the hydrophobic motif of Akt (HFPQFpSYSAS; where the phosphorylated Ser corresponds to Ser473), [c] compound 134149 [d] compound 13378 [e] compound 117079 and [f] compound 45586 were docked in the PHLPP2 model 4. Ligands are represented in tube format, Mn^{2+} ions are represented as yellow spheres and surface representation of PHLPP2 shows the electrostatic potential where blue represents positive, white neutral and red negative electrostatic potential. Docked substrates or inhibitors are in green, the phosphate groups are highlighted in purple. * denotes the carbonyl group of Gly 745 in all views.

We next tested whether the six most-promising compounds 1] inhibited PHLPP in cells, and 2] were selective for PHLPP compared with other phosphatases *in vitro*. To investigate PHLPP inhibition in cells, HT29 cells (colon cancer cell line) were treated for 24 hours with compounds at concentrations of either 100 μM or 250 μM, and the effect on Akt was assessed by examining the phosphorylation state of Akt on Ser 473 (the direct target of PHLPP) and, in addition, the phosphorylation state of two downstream targets of Akt, FoxO1 and GSK3 (Figures 6a and 6b). We chose to use HT29 cells for this study because the protein levels of PHLPP are not regulated by the level of Akt activity, as occurs in other cell lines via a recently described negative feedback loop(*118*, *119*). All compounds except **2** (Figures 6a and b, lane 7) caused an increase in the phosphorylation of Akt on Ser 473, with maximal increases of 4-fold caused by several of the compounds (Figure 6b). Of these, 4 compounds (**1** (NCS 117079), **24** (NCS 11404), **4** and **13**) caused corresponding increases in the phosphorylation of the downstream substrates GSK3 α/β on Ser21/9 and FoxO1/3α on Thr 24/32. We have previously shown that knockdown of either PHLPP1 or PHLPP2 increases the phosphorylation of FoxO1 on Thr 24 and GSK3β on Ser9(*77*). Some compounds selectively increased the phosphorylation of the downstream substrates, but not Akt (e.g. lane 7), and others caused an increase in the phosphorylation of Akt but only one of the downstream substrates (e.g. lane 8). Compound **4** (lane 5) induced cells to detach from culture dishes, reflecting toxicity of the compound.

In parallel with the cell study above, we tested the *in vitro* selectivity of the inhibitors by measuring their effect on the activity of the phosphatase domain of related and unrelated phosphatases. Figure 6c shows the effect of these inhibitors on the *in vitro*

activity of the phosphatase domain of PHLPP2 (black bars), PP1 (dark grey bars), PP2B (white), and PP2Cα (light grey bars). A number of compounds proved to be selective for the PP2C domain of PHLPP2 over the other phosphatases tested, including the related family member, PP2Cα. We should point out that, amongst the 54 inhibitors for PHLPP2 tested against PHLPP1, none was specific; at best, IC_{50}s were 5-fold different (data not shown), not unexpected given the high sequence homology of the phosphatase domains of the two isoforms. The most selective molecule for the PHLPP phosphatase domain was compound **1**: a concentration of 10 μM resulted in 80% inhibition of PHLPP2, with no significant effect on the activity of the other phosphatases. A 10-fold higher concentration resulted in approximately 50% inhibition of PP1 and PP2Cα, indicating that the selectivity for PHLPP was over an order of magnitude. Importantly, compound **1** increased Akt phosphorylation and activity in cells. Compound **13** was also noteworthy: it selectively inhibited PHLPP2 compared to the other phosphatases tested (Figure 6c, lane 6) and was one of the compounds that induced a robust increase in the activity of Akt (Figure 6b, lane 6). Thus, compounds **1** and **13** were chosen for further studies. Their IC_{50} values for inhibition of pNPP dephosphorylation were 5.45 ± 0.05 μM and 3.70 ± 0.06 μM, respectively (Figures 6d and 6e).



Lane	NSC#
1	DMSO
2	117079 (10 μM)
3	117079 (100 μM)
4	11404 (250 μM)
5	13378 (100 μM)
6	45586 (250 μM)
7	47924 (100 μM)
8	51532 (250 μM)
9	79714 (250 μM)
10	119910 (100 μM)
11	134149 (100 μM)

Figure 2.6 Cellular potency and specificity of PHLPP inhibitors. [a], [b] Effect of different inhibitors on Akt phosphorylation. HT29 cells in DMEM supplemented with 5% FBS, were treated for 24 hours with different compounds (see table). Blots were probed with antibodies specific to phospho-Akt (S473), phospho-GSK3α/β (S21/9), phospho-FoxO1/3a (T24/32) and actin. [b] Densitometric analysis was performed on blots from 3 separate experiments. The levels of phospho-proteins are normalized to actin. The phosphorylation of Akt at serine 473 (white bars), FoxO 1/3 (light grey bars) and GSK 3α/β (dark grey bars) is calculated relative to control lanes. The graph represents mean values ± SEM from 3 separate experiments. [c] Selectivity towards other phosphatases. pNPP (8 mM) was incubated in 125 μL of the optimal assay buffer for each protein, in the presence of the PP2C domain of PHLPP2 (black bars), PP1 (dark grey bars), PP2B (white bars) or PP2Cα (light grey bars). Compounds (see table) were added. The activity of the enzyme is relative to the DMSO control. The graph represents mean values ± SEM of at least 3 different experiments. * denotes that compound 45586 was tested at 100 μM and not 250 μM. [d], [e] In vitro inhibition curves for compound 117079 [d] and 45586 [e]. The inhibitor, diluted in DMSO, was incubated in assay buffer with 8 mM pNPP, in the presence of 1 μM enzyme. Activity was calculated relative to DMSO alone. The mean activity ± SEM for 5 different experiments are represented (circle:117079, diamond: 45586) and were fit with equation 1.

The inhibitory potency of compounds **1** and **13** on PHLPP activity in cells was determined next. To discriminate between specific effects of the compounds on PHLPP activity vs. nonspecific effects, we took advantage of the finding that PHLPP specifically and directly dephosphorylates Ser473 of Akt and does not dephosphorylate Thr308(*76*). For these experiments, we examined the effect of the compounds on Akt phosphorylation in serum-starved cells, in case PHLPP suppression is more dominant when Akt phosphorylation is maximally-suppressed.

COS 7 cells, serum-starved for 24 hours, were treated with increasing concentrations of the inhibitors for 35 minutes and the phosphorylation of Akt on Ser473 and Thr308 was determined; we also examined the activity of Akt by probing for the phosphorylation of downstream substrates with antibodies that recognize phosphorylated Akt substrates (Figures 7a-d). Treatment of cells with compound **1** resulted in an approximately 6-fold increase in the phosphorylation of Ser473 and a 4-fold increase in the phosphorylation of downstream substrates, with half-maximal increases caused by 29.1 ± 0.3 μM and 36.4 ± 0.3 μM inhibitor for Ser473 and substrate phosphorylation, respectively. Importantly, at the concentration of compound causing a half-maximal increase in phosphorylation of Ser473, there was little effect of compound **1** on the phosphorylation of Thr308 (Figure 7b, arrow). Higher concentrations did cause an increase in phosphorylation of Thr308; half-maximal increase in Thr308 phosphorylation was observed for 132 ± 1 μM inhibitor. Thus, the compound selectively increased Akt phosphorylation on Ser473 and this resulted in a commensurate increase in the activity of Akt. A similar increase in the phosphorylation on Ser473 and substrate phosphorylation was observed with compound **13**. Specifically, this compound also caused an

approximately 4-fold increase in the basal phosphorylation of Akt, with half-maximal increase caused by 70.6 ± 0.6 μM compound. The phosphorylation of the downstream substrates FoxO 1/3α and GSK3 β increased with comparable kinetics. For this compound, the sensitivity of phosphorylation of Akt on Thr308 was not significantly different from that of Ser473 (Figure 7d, lower panel). These data reveal that compounds **1** and **13** selectively inhibit the activity of PHLPP towards Akt in cells, with IC_{50} values of approximately 30 μM and 70 μM, respectively. Compound **1** has higher selectivity towards PHLPP as assessed by the uncoupling of phosphorylation at Ser473 and Thr308. At concentrations above 100 μM, this compound loses specificity as evidenced by the increase in Akt phosphorylation at both Ser473 and Thr308. Compound **13** was considerably less effective at modulating Ser473 phosphorylation in cells grown in serum (data not shown). In contrast, compound **1** increased Akt phosphorylation on Ser473 by 2-fold with comparable kinetics in the presence of serum. This acute treatment of cells with inhibitors did not change the levels of PKC isozymes (data not shown), as expected because PHLPP provides chronic control of PKC levels(*78*) (see Discussion).

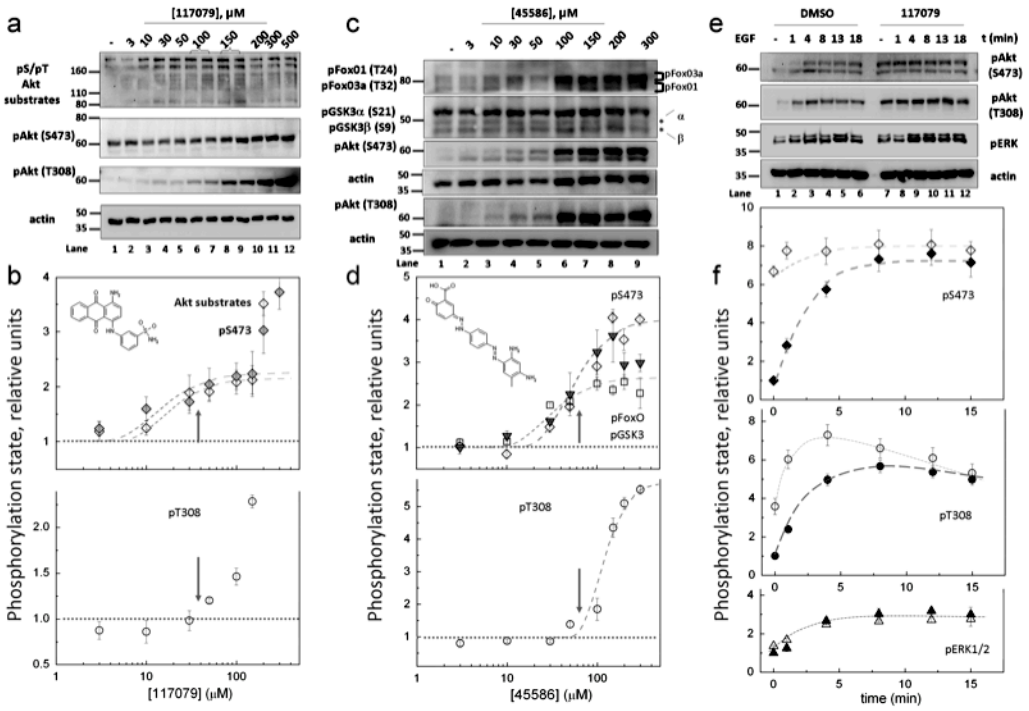


Figure 2.7 Effect of inhibitors on Akt-dependent signaling in cells. COS 7 cells were treated for 35 minutes with compound 117079 [a],[b] or 45586 [c],[d] at different concentrations. The control corresponds to 35 minutes treatment with DMSO alone. The graph represents mean values ± SEM from at least 3 separate experiments. [a], [b] Blots were probed with antibodies specific to phospho-S473, phospho-T308, an antibody specific to phospho-serine or phospho-threonine in a consensus sequence for Akt and actin. [b] The levels of phospho-Akt at Ser 473 (◇), at Thr 308 (○) and phosphorylation of Akt substrates (⬥) are normalized to total actin and relative to the level of phosphorylation in the control lanes. [c], [d] Blots were probed with antibodies specific to phospho-S473, phospho-T308, phospho-GSK3 α/β, phospho-FoxO1/3a and actin. [d] The levels of phospho-Akt at Ser 473 (◇), at Thr 308 (○) and phosphorylation of Akt substrates (GSK 3:□, FoxO:▼) are normalized to total actin and relative to the level of phosphorylation in the control lane. [e], [f] Effect of compound 117079 on the agonist-evoked phosphorylation of Akt. COS 7 cells were incubated in serum-free DMEM for 24 hours. The cells were pre-treated with DMSO (dark symbols) or compound 117079 (50 µM) (open symbols) for 35 minutes. EGF (1 mg mL-1) was then added at various time. Blots were probed with the specified antibodies. [d] The levels of phospho-Akt and phospho-ERK are normalized to actin. The phosphorylation of Akt at S473 (⬥ : DMSO, ◇ : 117079), of Akt at T308 (● : DMSO, ○: 117079) and ERK (▲: DMSO, △ : 117079) are relative to the levels of phospho-protein in unstimulated cells treated with DMSO (lane 1).

PHLPP controls the basal phosphorylation state of Akt as well as the amplitude of the agonist-evoked increase in phosphorylation of Akt(*77*). We therefore tested the effect of the inhibitors on agonist-evoked phosphorylation of Akt by pre-treating serum-starved COS-7 cells with or without 50 μM of **1** and then stimulating with EGF (Figure 7e-f, open symbols (DMSO) and dark symbols (inhibitor)). As in previous experiments, the basal phosphorylation at Ser473 was significantly higher in cells treated with **1** compared with DMSO (Figure 7e, compare lanes 1 and 7). In cells treated with DMSO, addition of EGF caused an approximately 7-fold increase in the phosphorylation of Akt on Ser473 that peaked after 8 min (Figure 7f, top panel). In contrast, EGF had a smaller effect on the already elevated phosphorylation of Akt on Ser473 in cells treated with **1** (Figure 7f, top panel). Phosphorylation at Thr308 was slightly elevated under basal conditions in cells treated with the inhibitor compared to control cells (Figure 7e, lanes 1 and 7). EGF treatment resulted in an approximately 6-fold increase in p308 phosphorylation for both control and treated cells, which peaked earlier in inhibitor-treated cells (Figure 7f, middle panel). Thus, the magnitude of the increase in p308 and p473 phosphorylation was comparable in inhibitor vs DMSO-treated cells, but the rate of phosphorylation on p308 was significantly faster in inhibitor-treated cells and, most strikingly, the basal phosphorylation on Ser473 was highly elevated in inhibitor-treated cells. To discern whether this coupled phosphorylation of p473 and p308 resulted from off target effects of the inhibitor or reflected the stabilization of phosphate on T308 when Ser473 is phosphorylated(*77*), we examined the EGF-dependent phosphorylation of ERK 1/2: the kinetics and magnitude of the EGF-stimulated increase in ERK phosphorylation were the same for control cells and cells treated with the inhibitor (Figure 7f, bottom panel).

Because a major function of activated Akt is to promote cell survival, a function enhanced by loss of PHLPP(*76*), we asked whether treatment of cells with compounds **1** or **13** suppressed etoposide-induced apoptosis. COS 7 cells were pre-treated with DMSO, **1** or **13** (at 50 μM) for 30 minutes, then treated with DMSO or etoposide for 24 hours (Figure 8). Etoposide treatment of control cells resulted in a (3.4 ± 0.4)-fold increase in apoptotic cells, as assessed by Trypan blue exclusion. Pre-treatment of cells with compound **1** reduced the magnitude of this increase by approximately 30%, to only (2.6 ± 0.3)-fold, and pre-treatment with compound **13** essentially abolished the etoposide-induced increase in apoptotic cells. Note that the basal level of apoptotic cells was comparable in control cells and cells treated with compound **13** (approximately 3.4%), but elevated in cells treated with compound **1** (approximately 12%). These data reveal that the PHLPP inhibitors protect cells against etoposide-induced apoptosis.

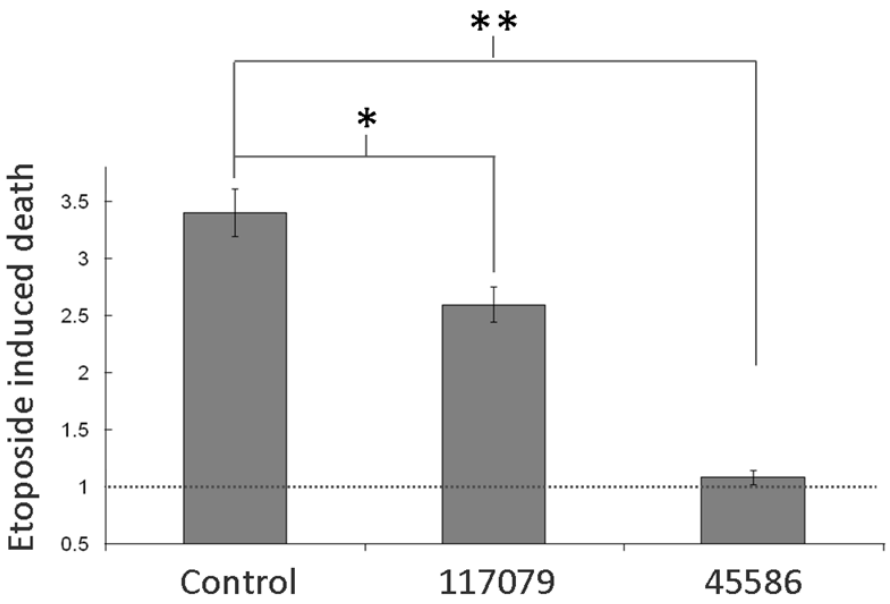


Figure 2.8 Effect of inhibitors on etoposide-induced cell death. COS 7 cells in DMEM supplemented with 0.1% FBS were treated for 30 min with DMSO, 117079 or 45586 at 50 μM followed by a 24 hours treatment with DMSO or etoposide. Cells were then submitted to a Trypan blue exclusion assay. The graph shows the etoposide-induced increase in cell death (compared to DMSO) and represents mean values ± SEM of 3 separate experiments.* corresponds to p values < 0.0005 and ** corresponds to p values ≤ $2\ 10^{-5}$.

Discussion:

By combining experimental and computational methods, we have identified the first set of inhibitors of the phosphatase PHLPP, a member of the PP2C family of phosphatases that has hitherto remained refractory to identification of general inhibitors. Specifically, we have identified small molecules that selectively inhibit PHLPP and show that treatment of cells with these inhibitors increases both the basal and agonist-evoked phosphorylation of Akt. Most relevant for therapeutic goals, these inhibitors selectively suppress cellular apoptosis.

We have specifically identified two molecules, with chemically-distinct backbones, that display selectivity for PHLPP both *in vitro* and in cells. Compound **1** (1-amino-9,10-dioxo-4-(3-sulfamoylanilino)anthracene-2-sulfonic acid, sodium salt) possesses an anthracene core whereas compound **13** (1,3-[[4-(2,4-diamino-5-methylphenyl)diazenylphenyl]hydrazinylidene]-6-oxocyclohexa-1,4-diene-1-carboxylic acid) has aromatic groups linked by 2 diazene bonds. They inhibit PHLPP2 activity *in vitro* with IC_{50} values of 5.45 ± 0.05 μM and 3.70 ± 0.06 μM, respectively. Kinetic analyses reveal that both molecules are non-competitive inhibitors with respect to pNPP. Both molecules are selective for PHLPP compared to other phosphatases tested: neither molecule affected the activity of PP2B at the highest concentration tested (100 μM), and inhibited PP1 and PP2Cα with IC_{50} values of approximately 100 μM (Figure 6c).

Both compound **1** and **13** show the potential for therapeutic development. Quikprop from the Schrödinger Suite was run to estimate properties that are potentially important to compound solubility, permeability, and drug development(*120*). The Lipinski rules indicate that a potential drug compound should not contain more than 5 H-bond donors, 10 H-bond acceptors, a LogP (predicted octanol/water partition coefficient) greater than 5, or a molecular weight greater than 500 daltons(*121*) (Table I). There are no Lipinski violations for **13** and **1** contains one violation from extra H-bond acceptors. Virtual docking of **13** (Figure 5f) shows multiple interactions between the aromatic cycles of the compounds and residues composing the hydrophobic cleft as well as coordination of one Mn^{2+} by the acid moiety. Compound **1** was discovered by chemical screening and does not perform well in the virtual docking so little information can be

gained this way. Note that both compounds are a dark color and both tend to precipitate in the cell culture medium at high concentration (> 300 μM).

Table 2.1Calculated chemical properties of compounds **13** and **1**. Qikprop was run to calculate the chemical properties that make up the Lipinski rules

Compound	mol. weight (daltons)	donor H-bonds	acceptor H-bonds	LogP
13	390	4	7.75	2.27
1	473	4	12	0.097

Cellular studies with compound **1** revealed that, at concentrations below 100 μM, it selectively inhibited the PHLPP-catalyzed dephosphorylation of Akt on Ser473 with little effect on the dephosphorylation on Thr308, a site that is not recognized by PHLPP. Indeed the IC_{50} value for inhibition of Ser473 dephosphorylation (29.1 ± 0.3 μM) was considerably lower than that for Thr308 dephosphorylation (132 ± 1 μM). At concentrations above 100 μM, the phosphorylation of Thr308 increased. This could result from off-target effects at higher concentrations, possibly by modulation of other phosphatases, or could reflect the stabilization of the phosphorylation on Thr308 by phosphorylation on Ser473(*77*). Interestingly, elevation of the phosphorylation of Ser473 alone, and not Thr308, resulted in an accompanying increase in the phosphorylation of downstream substrates of Akt, including FoxO1/3 and GSK3 α/β. These data reveal that phosphorylation on only Ser473 activates cellular Akt sufficiently to mediate downstream signaling. Compound **13** was also an effective inhibitor of Akt dephosphorylation, but displayed less selectively towards inhibiting the dephosphorylation of Ser473 compared

to Thr308. Thus, both compounds are effective inhibitors of Akt dephosphorylation, with compound **1** displaying almost one order of magnitude selectivity for Ser473 compared to Thr308.

Akt plays a key role in controlling the balance between cell survival and cell death. Disruption of this balance results in severe pathological states. Most notably, in heart disease and diabetes, the balance is pushed towards cell death. Therapies are thus targeted towards enhancing cell survival. The ability of our compounds to selectively activate Akt, by inhibiting its negative regulator PHLPP, poises them as prime therapeutic molecules for diseases in which Akt activity is repressed. Indeed, both molecules effectively suppressed etoposide-induced apoptosis, with compound **13** abolishing etoposide-induced apoptosis under the conditions of our assays. The more dramatic effect of **13** may reflect its ability to modulate the phosphorylation of both Ser473 and Thr308. Thus, these molecules hold promise as the basis for novel therapeutics for diseases in which cell survival pathways are suppressed.

PHLPP also controls PKC isozymes: particularly in the case of the conventional isozymes, phosphorylation of the hydrophobic motif is constitutive and protects PKC from degradation[a]. Dephosphorylation at this site triggers the down-regulation of the enzyme. Thus, PHLPP controls the amplitude of the PKC signal by controlling the levels of PKC. Genetic depletion of PHLPP results in elevated levels of PKC because phosphate on the hydrophobic motif is favored. However, PKC has a relatively long half-life (greater than 24 hours), so chronic inhibition of PHLPP is required to modulate PKC levels. For this reason, the acute treatment of cells with inhibitors of PHLPP in this study did not modulate PKC levels.

In addition to uncovering new molecules to activate Akt signaling in cells, this work underscores both the value of computational work in drug discovery campaigns and the value of using chemical data to increase the efficiency of computational work. The efficiency of the experimental work was largely increased by the use and refinement of a virtual model in this drug development campaign. The primary screen of the Diversity Set identified 50 inhibitors out of 1990 tested, a random search yielding 2.5% hits. Overall, we identified 95 inhibitors of PHLPP in this set, representing 5% of the tested compounds. When the same set was first submitted to a virtual screen, 14 inhibitors were found out of 36 tested, increasing the experimental hit ratio to 16%. Finally, when the virtual screen was combined with a structural component, the experimental hit ratio reached 25%, a 10-fold increase, highlighting the value of computational methods for drug discovery.

Given the lack of structural information available for PHLPP, this work had to branch out from traditional docking studies. Although crystal structures are commonly preferred for docking studies, we have demonstrated that homology models, when carefully constructed and correlated to experimental results, can provide success in the search for new inhibitory ligands. The construction of our model included the modeling of metal ions and water molecules in the active site, and some changes in conformations. As demonstrated by these data, placing metal ions in the active site can be extremely important in molecular modeling studies of this nature. Water molecules can also play a large role in inhibitor binding, but it is difficult to predict how the water molecules will be coordinated in the active site and which water molecules may be displaced by different inhibitors without crystallographic information. This may be the reason we did not

observe an improvement in performance when adding water molecules to our homology structure. In addition to identifying new inhibitors by virtual docking, our docking studies also revealed the mechanism of binding of the various inhibitions (competitive, uncompetitive, or non-competitive) with results validated by biochemical studies.

Compounds that inhibit PP2C phosphatases have been fairly refractory to identification, with few published studies(*83*, *84*). Here, we have identified a number of small molecules that not only inhibit this family of phosphatases, but also selectively inhibit PHLPP compared to other phosphatases, including the highly-related phosphatase PP2Cα. The combination of computational and chemical work allowed us to identify a variety of structurally-distinct inhibitors for a phosphatase target without the need for a massive high-throughput chemical screen. It is noteworthy that these tests were performed without the use of robotics or highly-automated methods, and the virtual screening was performed on a common desktop computer. Thus, collaboration between chemical and virtual screening provides an extraordinarily effective approach to drug discovery. Further refinement of these compounds to tune them to higher affinity and more specific inhibitors offers great therapeutic potential. Our identification of these new inhibitors for a PP2C family member is particularly relevant because these compounds could be potential therapeutics given the strategic position of PHLPP in cell survival pathways.

Experimental section:

The Diversity Set and compounds identified by virtual screen were obtained from the Drug Synthesis and Chemistry Branch, Developmental Therapeutics Program,

Division of Cancer Treatment and Diagnosis of the National Cancer Institute (http://dtp.nci.nih.gov/branches/dscb/repo_open.html). The compounds were used as provided in the *in vitro* assay. Purity of the compounds used to treat cells was verified by LC/MS using a Thermo LCQdeca mass spectrometer coupled with a Michrom BioResource HPLC at the UCSD Chemistry & Biochemistry Mass Spectroscopy Facility. Negative ion mode electrospray ionization was used. Purity was found to be >90% for compounds **24** and **4**, 80% for compound **2**, 60% for compound **7** and 55.5% for compound **13**. See Supporting Information for compound **1**.

The following phosphatases were purchased: PP1 (New England Biolabs), PP2B/calcineurin (Enzo). PP2Cα was purified from *E.coli* as previously described(*121*).

The following polyclonal antibodies were purchased from Cell Signaling: phospho-specific to phosphorylated Akt at Ser473 (S473) (#9271), phospho-specific to phosphorylated Akt at Thr308 (T308) (#9275), phospho-specific to phosphorylated Ser/Thr Akt substrate (#9611), phospho-specific to phosphorylated GSK 3 α/β at Ser 21 and Ser 9 respectively (S21/9) (#9331), phospho-specific to phosphorylated FoxO1/3a at Thr 24 and Thr 32 respectively (T24/32) (#9464), phospho-specific to phosphorylated p44/42 MAPK (Erk 1/2) at Thr 202 and Tyr 204 (T202/Y204) (#9101), antibody against p44/42 MAPK (Erk 1/2) (#9102). Monoclonal antibody against actin was purchased from Sigma-Aldrich (#A4700).

Experimental *in vitro* screen

In each well of a 96-well plate, 125 μL of a reaction mixture (100 mM tricine pH 7.5, 100 mM NaCl, 4 mM DTT, 0.02 mg.mL^{-1} BSA) containing 8 mM pNPP as the

substrate, 1 μM enzyme and 100 μM compound were added. Reactions occurred at 23°C. The optical density was monitored over time at 405 nm using an Emax Precision microplate reader (Molecular Devices). The absorbance was plotted against the time and the slope was calculated. Background was averaged from 4 different reactions in the absence of enzyme and subtracted. 8 different controls (containing 5 μL DMSO) were averaged and used to calculate the relative activity.

***In vitro* inhibition concentration assay**

The reactions occurred in the same conditions as described above except that the inhibitor was added at 7 different concentrations and DMSO served as a control. The relative activity was set at 100% for DMSO. The data were then fit to the Equation 1 :

(Equation 1) $$y = 100e^{\left(\frac{-C}{C_0}\right)}$$

The IC_{50} value is defined by $-C_0 \ln(1/2)$.

Homology modeling

The PP2C domain sequence of PHLPP2 (residues 745-1102) was used to make a homology model with the program MODELLER using the PP2C domain of PP2Cα as the reference structure(*87*, *88*). The two sequences were aligned using ClustalW. Next a model of PHLPP2 was created from the reference structure using MODELLER with default parameters. Further refinement of the model was performed by placing varying amounts of Mn^{2+} ions or water molecules in the active site and then relaxing the structure with Macromodel from the Schrödinger Suite[49]. The OPLS_2005 force field was used with 500 iterations of the gradient method.

Similarity searches and compound library generation

Accelrys software was used to search the NCI open repository, using PHLPP2 inhibitors determined previously in this study as reference compounds. Groups of inhibitors (families of compounds, or top inhibitors) were submitted as the reference compounds using the "Find Similar Molecules by Fingerprints" protocol provided with Accelrys Discovery Studio. Long range functional class fingerprint description 6 keys (FCFP 6 keys) were used with a Tanimoto distance coefficient to compute a similarity score. Top scoring compounds were selected for virtual screening.

Docking

The GLIDE virtual screening application in Schrödinger Molecular Modeling Suite was used to screen compounds using 3 levels of docking precision. A modified version of the Chemscore function is employed by GLIDE to assign a score to each ligand in all poses. Glide HTVS was run on all compounds to perform a complete conformational and positional search of 3-dimensional space in the active site. For the compounds that scored in the highest 20%, GLIDE SP (standard precision) was run. Similarily for the top 20% high scoring GLIDE SP compounds GLIDE XP (extra precision) was run too. Flexible docking was allowed in all stages and default parameters from the Virtual Screening Workflow were used in all docking studies, in addition to the aforementioned modifications to the percent of compounds entering each stage. All final scores and poses came from GLIDE XP. The grids were generated for each model in the phosphatase active site with the XYZ coordinates (-27.54, 26.44, 46.38). An inner box which must contain the center of each ligand docked was 14 angstroms in each direction,

and the outer box in which all parts of the ligand must bind was 44 angstroms in each direction. The relatively large box allowed for a variety of docking poses and accepted large compounds. No other constraints were placed on the grids.

Homology model evaluation

Docking of the inhibitors found in the chemical screen into each modified homology model helped determine the best model to use for further docking studies. In addition, the experimentally validated binding compounds with a G-score below -7 were considered to be virtual hits in this study (Figure 2c). The docking protocol was modified slightly in the evaluation of models so that all known inhibitors were permitted to proceed through the 3 stages of GLIDE. Also the entire diversity set was docked into the models with Mn^{2+} in the active site, these studies were performed with only the top 4% of compounds reaching the final GLIDE XP stage and receiving scores as described in the docking protocol. Models were evaluated based on the number of experimentally-confirmed inhibitors receiving a G-score of better than -7.

Cell culture and immunoblotting

Cells are maintained in DMEM supplemented with 5% FBS and 1% penicillin/streptomycin, at 37°C in 5% CO_2.The medium was aspirated before addition of lysis buffer (50 mM Tris pH 7.5, 1 mM sodium pyrophosphate, 20 mM NaF, 2 mM EDTA, 1% Triton, 200 μM benzamidine, 40 μg mL^{-1} leupeptin, 1 mM PMSF and 1 μM microcystin). Whole cell lysates were analyzed on 7.5% SDS-PAGE gels and Western blotting using the indicated antibodies. Chemiluminescent signals were imaged by an

Alpha Innotech MultiImage III and densitometric analysis was performed using AlphaView software (version 2.0.0.9).

Inhibitor treatment of cells

HT29 cells were plated in 12-well plates and maintained in 5% FBS in DMEM. 4 μL of the different compounds diluted in DMSO were added to fresh medium in each well. Cells were lysed in 200 μL lysis buffer, on ice, 24 hours after addition of the inhibitors. For cellular IC_{50} studies, COS 7 cells were plated in 24-well plates. The cells were starved for serum 24 hours prior to the assay by placing them in serum-free DMEM. The inhibitor (**1** or **13**) was added at various concentrations (4-400 μM or 3-300 μM, respectively) to each well and the cells were allowed to sit for 35 minutes in the incubator before the plate was placed on ice. Cells were lysed in 100 μL lysis buffer and lysates were analyzed as described above. DMSO was used as a control. The relative activity was set to 1 for DMSO. The data were then fit to Equation 2.

(Equation 2) $$y = A\left(1 - e^{\left(-\frac{C}{C_0}\right)}\right) + 1$$

For time course studies, COS 7 cells were plated in a 24-wells plate. The cells were starved for serum 24 hours prior to the assay by placing them in serum-free DMEM. Cells were pre-treated with DMSO or **1** (50 μM) for 35 minutes. EGF was then added at different time to a final concentration of 1 μg mL^{-1}. Between the different time points, cells were placed back in the incubator. The control was obtained in the absence of EGF. The plate was then placed on ice and the cells were lysed in 100 μL lysis buffer.

Apoptosis assay

COS 7 cells were plated in a 6-well plates. When cells reached 80% confluency, medium was aspirated and replaced with 0.1% FBS in DMEM. DMSO, **1** or **13** were added to each well to a yield a final concentration of 50 μM. Following 30 minutes incubation at 37°C, DMSO or etoposide (50 μM) was added. After 24 hours, cells were detached from the plate using trypsin, centrifuged at 100 x g for 5 min, 4 °C, and stained with Trypan blue. Cells were manually counted and the percentage of dead cells was calculated.

Acknowledgements

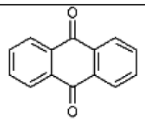
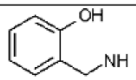
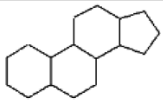
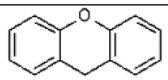
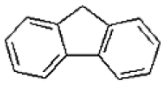
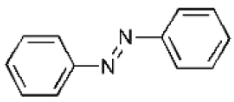
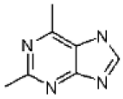
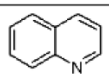
We wish to thank members of the Newton lab for critical reading of the manuscript, Chelsea Buma for help with the *in vitro* screen, Dr Arneh Babakhani for guidance with the *in silico* screen, Drs Nadia Fomina and Adah Almutairi for help with the synthesis, Drs Li Xie and Philip E. Bourne for help with the homology modeling, and Drs Yann Gambin and Edward Lemke for help with data analysis. This work was supported by a Juvenile Diabetes Research Foundation grant 3-2008-478 (ES), Molecular Biophysics Training Grant GM08326 (WS), the National Science Foundation Grant MCB-0506593, NBCR, CTBP, and Howard Hughes Medical Institute (JAM), and National Institutes of Health Grants GM31749 (JAM) and GM067946 (ACN).

Chapter 2 is a minimally modified reprint of the material as it appears in Sierecki, E.; Sinko, W.; McCammon, J. A.; Newton, A. C. Discovery of small molecule inhibitors of the PH domain leucine-rich repeat protein phosphatase (PHLPP) by chemical and

virtual screening. *J. Med. Chem.* 2010, *53*, 6899–6911. I was a co-first author and co-principal investigator on this paper.

Supplemental Information

Table S2.1 NSC IDs of compounds that inhibit PHLPP 2 in vitro grouped in clusters according to their chemical structures. Compounds presenting an IC50 value below 5 μM are written in bold and italic, those with an IC50 value comprised between 5 and 10 μM are in italic and underlined and the compounds only underlined present an IC50 value between 10 and 25 μM. The compounds that were tested in cells are in grey.

Cluster	Chemical determinant	NCS #
A		***117079***, *125908*, 117274, 115448, 69343, 170008, 23173, *125910*, 23128, 13987, 255980, 48949, 16211, 117949, 634663, 366218
C		***119910***, ***47924***, *109248*, 114609, 47949, 67608, 88852, ***103779***, 375985, 120917, 88850, 11404
D		88915, 93352, 75541, *72254*, 48010, 109131, 121137, 88135
E		119913, *119886*, 31076, 114885
F		***133896***, 23173, 375985, 175650, 49692
G		*134149*, *77393*, 125908, 47716, 65537, 37031, 67608, 45208, 134196, 361672, *338517*, 354961, 75140, 134148, 4895 *45586*, *79714*, 45618, 51532, 65541, 75913, 134145, 47754, 47760, 114428, 79729, 51536, 45619, 27236, 12129, 326196, 134146
H		***348401***, 77393, 42384, 109131, 154829, 75140
I		*348401*, *74429*, ***207895***, 26699, 60043
R		***48961***, *24479*, *270156*, *601364*, *327704*, 327705, 41663, 361672, 11241, 106505, 130813, 321518, 634663
Q		***131734***, ***13378***, *130796*, *67091*, 639174, 116708, 210627, 47106, 117198, 354961, 375985, 1013, 130813, 48949,103779, 136469, 134244, 12509, 11404
Other	none	***25247***, *37173*, *3323*, 13161, 99550, 80126, 117241, 105444, 72938, 103019, 117438, 360215

Synthesis and purification of compounds 1 and 13

Purification of 13.

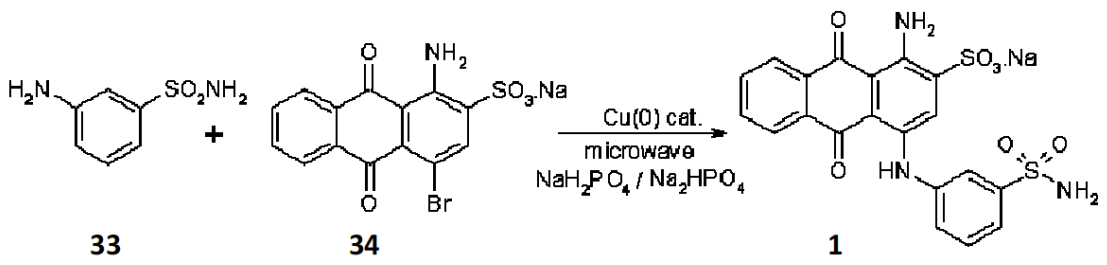
A sample was purified by HPLC, on a C18 column (4.6x250 mm) using a mix of H2O+0.001% TFA and CH3CN + 0.001% TFA. The flow was set at 1 ml/min, the eluent was maintained at 20% CH3CN for 5 min then a gradient was applied for 55 min to 80% CH3 CN. 7 peaks were isolated. The main product which corresponds to the right molecular formula was isolated in >95% purity. We verified that the results for the purified compound were identical to those obtained with the mix.

Purification of 1.

A sample was purified by HPLC, on a C18 column (4.6x250 mm) using a mix of H2O+0.001% TFA and CH3CN + 0.001% TFA. The flow was set at 1 ml/min, the eluent was maintained at 10% CH3CN for 5 min then a gradient was applied for 45 min to 95% CH3 CN. Using color as criterium, we identified at least 8 products, none of which showed the right mass, according to MS.

Synthesis of 1.

3-aminobenzenesulfonamide 33 (97%, Alfa-Aesar) and bromaminic acid, sodium salt 34 (90%, Aldrich) were used without further purification. The elementary copper (99.9% pure) is powdered (170-400 mesh) and was purchased from Alfa-Aesar.



Scheme 2.1 Synthesis of compound 1

Compound 1 was synthesized using a modified Ullmann coupling(*122*). 3-aminobenzenesulfonamide 33 (69.1 mg, 0.4 mmol), solubilized in 1 mL of aqueous Na2HPO4 (0.2 M, pH 9.5) with 5% v/v DMSO, was added to bromaminic acid 34 (81.2 mg, 0.2 mmol), in a 5-mL microwave reaction vial equipped with a magnetic stirring bar. To the mix was added 3 mL Na H2PO4 (0.12 M, pH 4.3), 1mL Na2HPO4 (0.2 M, pH 9.5) and a catalytic amount of finely powdered elemental copper. The vial was placed inside the microwave reactor where the reaction mix was stirred for 30 seconds before the temperature was ramped up to 120°C. This temperature was maintained for 5 minutes before being cooled to r.t by a flow of air. The reaction was filtered and 200 mL H2O was added to the filtrate. This aqueous phase was washed 3 times with 200 mL CHCl3 then reduced to a volume of ~ 10 mL by rotary evaporation. The different products were 33 34 1 S 3 separated by flash chromatography on a column RP18 reverse phase using a H2O/ MeOH eluent. The flow was set at 40 mL/min. Specifically, the column was equilibrated with a 5% MeOH eluent. This eluent was maintained for 5 min then the percentage of methanol was gradually increased to 50% over 25 min then to 100% over the following 10 minutes.

Different products were isolated from this chromatography: the first fraction (F1)

is orange and is eluted from 10% MeOH followed by a red product (F2) eluted from 20% MeOH. The resolution is not as good for the following products: a purple fraction (F3), eluting from 30% MeOH; a blue compound (F4) at 35% MeOH with a light pink product (F5) following. The different fractions were reduced by rotary evaporation and lyophilized. The major product of the reaction is F2, which corresponds to the hydroxy derivative.

LC/MS analyses were obtained using a Thermo LCQdeca mass spectrometer coupledwith a Michrom BioResource HPLC. HPLC was performed on a Majic C-18 column(1.0mmx150mm) at 50 μL/min. The solvent used was a mix of solvent A (water with0.01% TFA) and B (acetonitrile with 0.01% TFA). To detect the product, a gradient from 10% B to 95% B was run in 12 minutes, followed by a hold at 95% B for 6 minutes, then back to 10%B in 2 minutes and hold again at 10% B for 6 minutes. Negative ion mode electrospray ionization was used. According to these analyses, both F4 and F5 have the right mass. F5 was obtained with a poor 8% yield and F4 was obtained in such a small quantity that determination of concentration was not possible.

Table S2.1 The IC50s for the different products as well as the reagents were obtained as described in the Methods section.

Compound	IC_{50} (μM)	Compound	IC_{50} (μM)
1	5.45 ± 0.05	F2	22.6 ±0.2
33	>300	F3	41.4 ± 0.2
34	5.22 ± 0.08	F4	n.d
F1	74.3 ± 0.5	F5	4.15 ± 0.09

From these data and the LC-MS data, F5 was identified as the sodium salt of 1. The bromo derivative 34 inhibits PHLPP with the same efficiency as the product, underlining again the potential of this molecular scaffold to design PHLPP inhibitors.

Chapter 3: Applying molecular dynamics simulations to identify rarely sampled ligand-bound conformational states of undecaprenyl pyrophosphate synthase (UPPS), an antibacterial target

Abstract

Undecaprenyl pyrophosphate synthase (UPPS) is a *cis*-prenyltransferase enzyme, which is required for cell-wall biosynthesis in bacteria. UPPS is an attractive target for anti-microbial therapy. We performed long molecular dynamics (MD) simulations and docking studies on UPPS to investigate its dynamic behavior and the influence of protein flexibility on the design of UPPS inhibitors. We also describe the first x-ray crystallographic structure of *E. coli* apo-UPPS. The MD simulations indicate that UPPS is a highly flexible protein, with mobile binding pockets in the active site. By carrying out docking studies with experimentally validated UPPS inhibitors using high and low populated conformational states extracted from the MD simulations, we show that

structurally dissimilar compounds can bind preferentially to different and rarely sampled conformational states. By performing structural analyses on the newly obtained apo-UPPS and other crystal structures previously published, we show that the changes observed during the MD simulation are very similar to those seen in the crystal structures obtained in the presence or absense of ligands. We believe that this is the first time that a rare "expanded pocket" state, a key to drug design and verified by crystallography, has been extracted from an MD simulation.

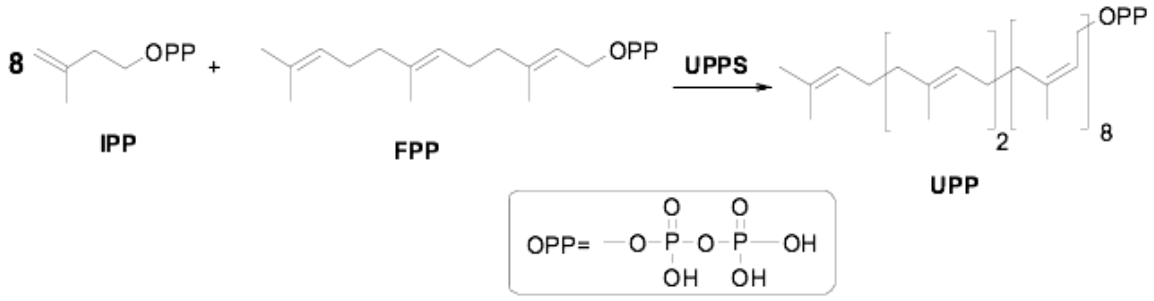
Abbreviations

Undecaprenyl Pyrophosphate Synthase (UPPS), Molecular Dynamics (MD), isopentenyl diphosphate (IPP), farnesyl pyrophosphate (FPP), high throughput screens (HTS), bisphosphonate (BPH), S-*thiolo*-farnesyl diphosphate (FSPP), principle components (PC), Principal Component Analysis (PCA), root mean square deviation (RMSD), root mean square fluctuation (RMSF)

Introduction

The evolutionary pressure of antibiotics has created drug resistant strains in most species, creating the need for new drugs that inhibit targets not previously exploited by clinicians(*123*, *124*). Inhibition of enzymes involved in isoprenoid biosynthesis is a promising new target area since many of these enzymes are not present in humans. For example, the enzymes that comprise the non-mevalonate pathway leading to the biosynthesis of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate are not present in humans, which use the mevalonate pathway[3](*5*). The enzymes involved in the

later stages of isoprenoid biosynthesis, such as undecaprenyl diphosphate synthase (UPPS), are also targets of interest in drug discovery. UPPS is a *cis*-prenyl transferase and is responsible for the condensation of eight molecules of IPP with farnesyl diphosphate to form the C_{55} species, undecaprenyl diphosphate (scheme 1).



Scheme 3.1 Condensation reaction of IPP and FPP by UPPS

Following the condensation, undecaprenyl diphosphate is converted to Lipid I, Lipid II and lastly, to bacterial cell wall peptidoglycans. Many antibiotics (e.g. the beta-lactams penicillin and methicillin; cephalosporins, glycopeptides such as vancomycin and teicoplanin) inhibit cell wall biosynthesis. UPPS is an attractive target in e.g. *S. aureus* since there is only modest homology (34% identity, 54% similarity, E-value= $8e^{-38}$) to the human *cis*-prenyl transferase, dehydrodolicol diphosphate synthase.

Numerous X-ray crystal structures of UPPS have been solved with the farnesyl pyrophosphate (FPP) and IPP substrates, or inhibitors bound(*125-129*). Comparison of these crystal structures reveals that UPPS is a highly flexible protein system, showing very mobile binding pockets in the active site region. The catalytic site is composed of four binding pockets(*129*) (Figure S1A) and it is responsible for accommodating the large substrates, FPP and IPP, and for transferring the phosphate groups and metal ions

between the two substrates during the condensation reaction(*128*). However, the available crystallographic data has proven not to be sufficient for a successful structure based drug design campaign. There have been several reports of high throughput screens (HTS) against UPPS. In one study, several hits were claimed, but the structures were not disclosed(*130*). In a second experimental screen, no hits were reported,(*131*) but in a third screen workers at Novartis reported several promising leads; tetramic and tetronic acids as well as dihydropyridin-2-ones(*132*, *133*). On the *in silico* or virtual HTS front, based on a *Helicobacter pylori* crystal structure, Kuo *et al*.(*134*) reported two hits (with IC_{50} values in the ~70-500 μM range) against *H. pylori* UPPS and *E. coli* UPPS, with some selectivity against the *H. pylori* protein(*134*). In earlier work, we have characterized 29 bisphosphonate compounds and 5 co-crystal structures(*129*). These bisphosphonates and the Novartis Peukart *et al*.(*132*) substituted tetramic acid and dihydoropyridin-2-one-3-carboxamides are shown in chart 1.

In this work, we use experimentally validated compounds (chart 1) to correlate our *in silico* data with known experimental results. In addition, long molecular dynamics (MD) simulations of the *apo* form of the enzyme, starting from an inhibited conformation, and docking studies were performed on UPPS to investigate its dynamic behavior and the influence of protein flexibility on the design of UPPS inhibitors. We were able to identify "*active conformational states"* of the apo form of the UPPS that recognize different classes of known inhibitor molecules, which may be extremely useful in virtual screening efforts. We used a new descriptor, active site volume, to find rare pocket conformations. In addition, we have crystallized *E. coli* UPPS with the flexible loop to investigate further the large pocket size fluctuation in our MD simulations.

Experimental Section

Computational Details.

The crystal structure of the UPPS enzyme in complex with the bisphosphonate, BPH-629 (PDB ID 2E98), was used to build the models for the MD simulations(*129*). All bisphosphonate ligands were removed from the active sites of monomer, and the protein system was simulated in the apo state. Chain B of the UPPS homodimer lacks crystallographic information for residues 73-82, which were modeled using MODLOOP(*135*, *136*). The protonation states of the residues were determined through the use of the PROPKA program(*137*, *138*), with special attention paid to His43. Residue His43 is well positioned to hydrogen bond the diphosphate groups of the bound ligands, and it is thought to not only play an important role in the binding process of the natural substrates, but also to be of key importance to the catalytic activity of the enzyme(*139*). Owing to the importance of this residue, the effect of different protonation states of His43 on the dynamics of UPPS was investigated through two MD simulations; one where HIS43 is singly protonated (HID43), and the other where HIS43 is doubly protonated (HIP43).

In both simulations TIP3P water molecules were used as the solvent model in a truncated octahedron. Water molecules and counter ions, Cl^-, were added to solvate the structure and neutralize the total charge of the protein using the AMBER program xLeap. Simulations were performed using the sander.MPI module of AMBER 10, the AMBER ff99SB forcefield, and Particle Mesh Ewald to describe the electrostatics interactions(*140*). Temperature control was achieved with the weak-coupling algorithm,

and pressure control was accomplished via isotropic position scaling(*141*). Energy minimization of the solvated system was performed with an initial 1500 steps of steepest descent, followed by 500 steps of conjugate gradient minimization. To bring the system to the correct density and allow proper re-orientation of the water molecules, an MD simulation of 100 ps was performed in the NPT ensemble where the protein was fixed and only the water molecules were allowed to move freely. Following this, the entire system was heated from 0 K to 300 K over 500 ps of MD simulation in the NVT ensemble. To ensure complete equilibration of the system at 300 K, 200 ps of MD simulation was performed. All analyses were carried out on an additional MD simulation of 85 ns, in which the NVT ensemble was applied. All simulations were stable as shown by RMSD (root mean squared deviation) plots (Figure S2). The setup, equilibration and production protocols were applied to both HID and HIP systems.

To calculate the volume of the active site of UPPS, frames were extracted from the MD simulations every 10 ps and aligned. POVME software was then used to define a volume that encompassed the active sites by taking into account carefully positioned spheres of a 10 Å radii, respectively centered in the active site region manually(*142*). Points spaced 1 Å apart were positioned along a grid within this defined volume (Figure 2). The same positioning of grid points was used for all structures so volumes are comparable. For each frame, a hydrogen atom was positioned at each grid point, and wherever van der Waals clashes with protein atoms occurred, the point was removed. Small, isolated clusters of points were likewise removed. The remaining points were judged to be contained within the IPP active site (Figure 2). As the points were originally spaced 1 Å apart, each point effectively corresponded to a region of space 1 $Å^3$ in

volume, allowing the volume of the binding pocket to be easily calculated. The pocket volume of the holo (PDB ID: 2E98, 2E99(*129*) 1X06, 1X07, 1X08, 1X09(*128*)) and the apo crystal structures were calculated as described above. Crystallographic waters and co-crystallized ligands were removed. While other crystal structures of *E. coli* UPPS have been solved, they lack some residues in the active site and therefore their volume could not be accurately calculated.

Crystallization and structure determination of Apo *E. coli* UPPS.

The *E. coli* UPPS enzymes were expressed and purified as described previously(*143*). Apo UPPS crystal was obtained as described previously with modifications(*129*). Briefly, 1 μl UPPS protein was mixed with 1 μl of mother liquor (5% PEG4K), then equilibrated with 500 μl of mother liquor at room temperature by using hanging drop evaporation. X-ray diffraction data were collected at LS-CAT (Life Science Colloborative Access Team) 21-ID-F at the Advanced Photon Source of Argonne National Laboratory. Diffraction data were processed and scaled by using the program HKL2000(*144*). The statistics for data collection are included in table 1. For structure determination, a molecular replacement calculation was carried out by using a model prepared from the BPH-629 (PDB ID 2E98) with ligands and solvent removed. The 2Fo-Fc difference Fourier map showed clear electron densities for most amino acid residues, including those in the flexible loop in chain A. Iterative rounds of refinement using Refmac(*145*, *146*) and rebuilding of loop residues using Coot(*147*) were then carried out. Further refinement was performed in Refmac5 with the TLS parameters

generated by the TLSMD server(*148*). The resulting final structure has an R-factor of ~18.6% (R_{free} ~22.4%). The refinement statistics are included in table 1.

Results

Docking of known UPPS inhibitors

In earlier work(*129*), we reported the x-ray crystallographic structures of five bisphosphonate compounds bound to *E. coli* UPPS. We found up to four distinct binding sites for BPH-629 (PDB 1D code 2E98). However, only Site-1 (Figure 1A) was occupied in all of the five bisphosphonate/UPPS structures reported, with ligand interactions involving ASP-26, ASN-28, ARG-39, HIS-43, ARG-51, ARG-77, PHE-89, ARG-102 and HIS-103 being common to all structures. On average, there were ~14 bisphosphonate-protein interactions in Site-1, but only ~9 for Sites 2-4, to be compared with 15 interactions (in Site-1) for the substrate analog, S-*thiolo*-farnesyl diphosphate (FSPP). This suggests that Site-1 might represent the strongest binding site for bisphosphonate UPPS inhibitors. To test this hypothesis further, we have now carried out a computational docking investigation using the Glide program(*110*, *111*, *149*).

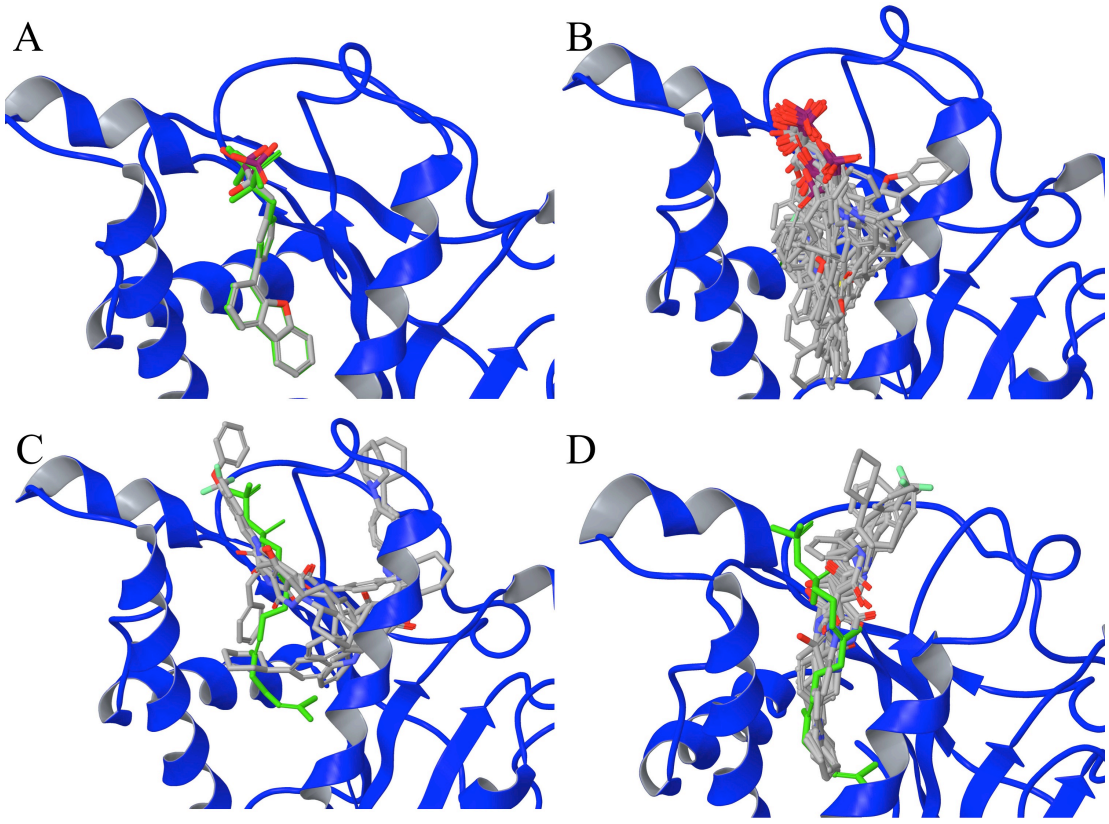


Figure 3.1 (A) Superposition of docked (colored by atom type) and co-crystallized pose (green) of BPH-629 into the 2E98 crystal structure. (B) Docked poses of 29 BPH inhibitors into the 2E98 structure. (C) Docked poses of **1i**, **1j**, **4a**, **4g**, **4j**, **4l** and **4m** into the 2E98 structure. Ligand poses showed a very poor alignment with the substrate, FPP (shown in green). (D) Docked poses of **1i**, **1j**, **4a**, **4g**, **4j**, **4l** and **4m** into the fourth most populated MD-derived structure. All ligand poses showed a very good alignment with the substrate, FPP (shown in green).

The co-crystallized structure of UPPS with BPH-629 (PDB ID 2E98), after removal of ligands, was subjected to numerous docking calculations with the program GLIDE at the XP level. Four BPH-629 ligands are present in the active site of Chain A of the UPPS in the crystal structure, identifying four distinct binding sites. As shown in Figure 1, the docking calculation accurately reproduced the crystallographic poses for BPH-629 in the UPPS active site. Although four binding sites were identified in the UPPS crystal structure, the poses of all bisphosphonate (BPH) ligands generated by

GLIDE were located primarily in the first binding site (Figure 1B). These docking results are also in very good agreement with the experimental observation that, in the crystal structure, ligands bind more tightly to the first binding site(*129*). To further validate our model, the 29 BPH ligands (Figure 1), with known potency against UPPS, were docked into the structure and the estimated free energy of binding (Glide XP score) of each ligand compared with their respective pIC50s. When these docking results are compared with the enzyme inhibition pIC_{50} results ($pIC_{50} = -\log_{10} IC_{50}$ [M]), we find a correlation coefficient of 0.5 (Figure S3). The correlation coefficient is promising but not strong. This could be due in part to many of the estimations inherent in docking software, or artifacts from crystallographic conditions including the high concentration of ligand in the medium (5mM). The structural changes induced by the 3 other inhibitors occupying the active site, which were removed for docking, may also affect the correlation.

In recent work, Peukart *et al*. described a class of potent and selective UPPS inhibitors(*132*). The scaffold from a known UPPS inhibitory compound was modified to create a small library of substituted tetramic acid and dihydropyridin-2-one-3-carboxamides. These compounds, inspired by the binding mode of the FPP, posses two hydrogen bond acceptors and a hydrophobic group, which are important interaction sites. The proposed inhibitors showed sub-micromolar UPPS inhibition and antibacterial activity against Gram-positive bacteria. To investigate the binding mode of this class of inhibitors, inhibitors **1i**, **1j**, **4a**, **4g**, **4j**, **4l** and **4m** (Figure 1) were docked into the crystal structure 2E98. It is worth mentioning that these inhibitors bear no structural similarity with the BPH compounds, and they were designed to adopt a binding mode similar to the natural substrate, FPP. As displayed in Figure 1C, docking of these inhibitors into 2E98

generated unexpectly poor results with unreasonable binding poses. The molecules were distributed over all four binding sites with none of the pose showing a good alignment with the natural substrate, FPP. This raised the question: is the protein is in a different conformation when it binds these non-bisphosphonate inhibitors?

Identifying inhibitor bound UPPS conformations from MD simulation

To investigate the dynamical behavior of UPPS, we performed clustering analysis of the MD trajectories of the HIP43 and HID43 systems. Each trajectory was fit to the alpha carbons of all UPPS subunit A residues, with the exception of the C and N-termini and the flexible segment containing residues 73-80. Clustering was performed by employing the Gromos method with RMS differences of a selection of active site residues (residues 23-51, 67-93, 96, 110, 141-145, 194, 204, 221-222) within Gromacs version 3.3.1(*150*). To choose the appropriate cutoff radius, several cutoffs were investigated; resulting in a final cutoff of 1.8 Å for the HID43 simulation and a cutoff of 2.2 Å for the HIP43 simulation. The first 5 clusters represent more than 90% of the entire trajectory. Docking of the 29 BPH ligands into the 5 representative MD cluster structures did not show any improvement over the results from the crystallographic structure. The clustered structures demonstrate that the crystallographic conformation is not highly populated in the apo enzyme trajectory. It is worth noting that the crystal structure 2E98 contains four BPH-629 molecules on its active site due to the high inhibitor concentration used in the experiment(*129*). Thus, in this case, it is expected that induced fit effects are playing an important role in the binding process, promoting large conformational changes in the active site of the enzyme, expanding the volume from the unbound state. The

active site of 2E98 is very large in size when compared with the most populated structures extracted from the MD simulations (1032 $Å^3$ as compared to an average of 332 $Å^3$ volume in the MD simulations). Clustering analysis showed that all representative structures from the MD simulation displayed active sites with significantly decreased volume when compared to the crystal structure. We believed that this highly open conformation could be favoring the binding of BPH-containing ligands. To test this hypothesis, we used POVME software to calculate the volume of the pocket throughout the MD trajectory and numerous crystal structures. Figure 3A displays the volume of the HIP simulation UPPS binding pocket calculated for each selected frame along the timecourse of the simulation, as well as a time averaged size along the simulation, and the size of selected crystal structures. As we can see, conformational states that show pocket volume close to the one observed in the *holo* crystal structure (2E98) are rarely sampled in the simulation of the *apo* form of the UPPS. The crystal structure, 2E99, is bound to another bisphosphonate BPH-608(*129*), and also has a large pocket size, 873 $Å^3$. Although protein coordinates from the crystal structure, 2E98, were used to build our UPPS model, when simulated in the absence of ligands, the pocket volume reduces significantly after the initial equilibration steps and never returns to its initial value during the production phase. Interestingly, at approximately 12 ns, the pocket widens furthest and reaches within 100 $Å^3$ of the size observed in the crystal structure (Figure 3A). In order to evaluate the influence of the protein pocket size on the docking results, the frame with largest pocket was selected and used for further calculations. Docking of the BPH compounds to the selected structure showed a very good agreement with the experimental IC_{50} values, with a correlation coefficient of 0.8 between the docking score

and pIC_{50} (Figure S3). This represents a significant improvement when compared to the results obtained from the crystal structure 2E98 (correlation coefficient of 0.5). We partially attribute this improvement to the opening of the bisphosphonate binding region in site 1. In the 2E98 crystal structure the positively charged residues HIS43 and ARG77 are in close proximity. This region widens in the largest MD derived structure, minimizing the scoring penalty for ligands with a positively charged N near the positively charged region of the protein such as BPH-641 and BPH-642 (the two lowest scoring ligands when docked against 2E98). This repulsive force is strong enough that both BPH-641 and BPH-642 take on unexpected poses with poor scores (Figure S1 and S4). The wider bisphosphonate binding region is able to accommodate a larger variety of bisphoshonate containing ligands.

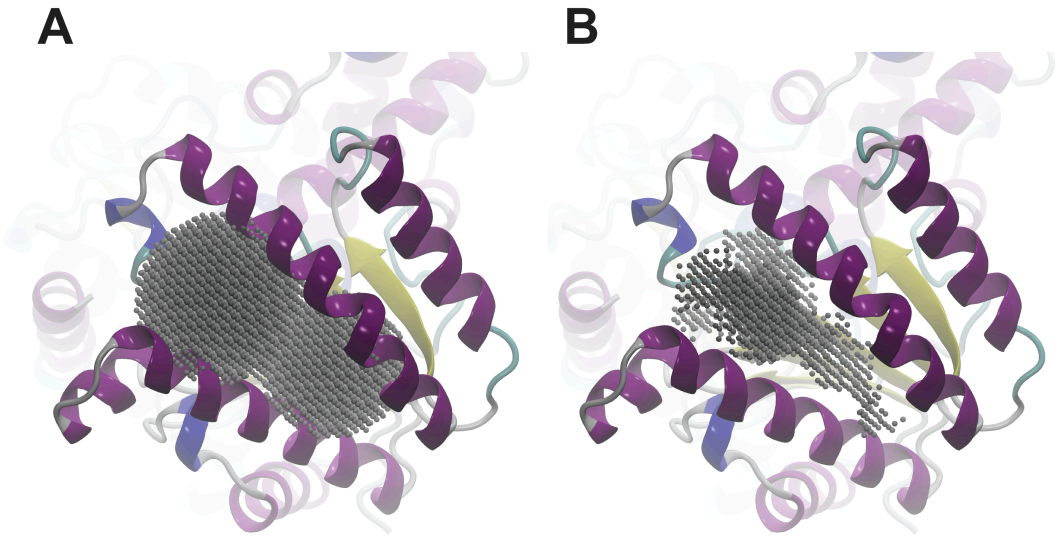


Figure 3.1 Grid points used to calculate pocket volumes. A) Points spaced 1 Å apart were positioned along a grid that encompassed the UPPS IPP-binding pocket. B) Those grid points near protein atoms were removed, leaving only points within the active site, from which the volume was calculated.

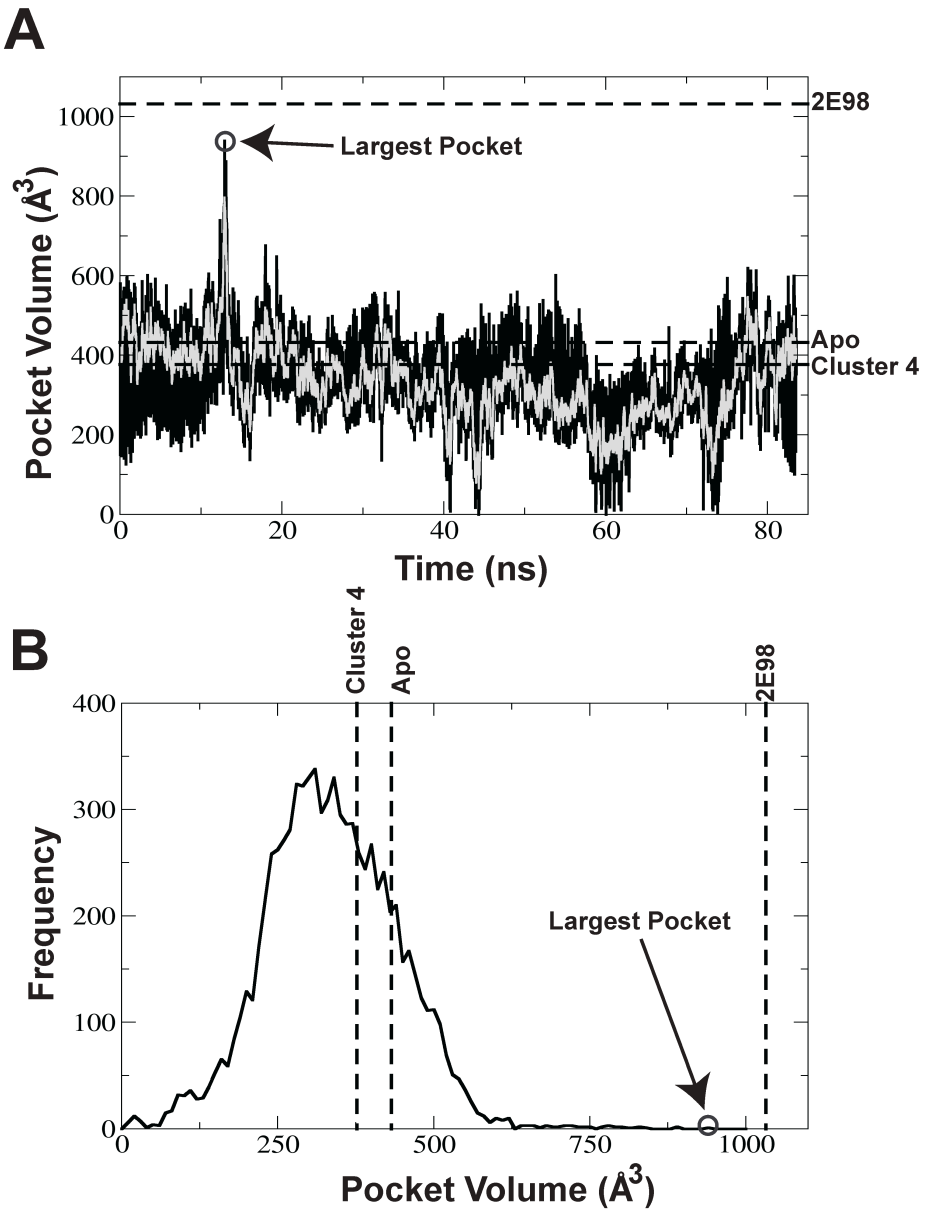


Figure 3.2 Volume distribution of the HIP43 UPPS binding pocket 2E98 crystal structure and the apo crystal structure. (A) Volume of the binding pocket along the MD trajectory. The black line shows every data taken every 10 ps, the overlayed gray line is the average over every 100ps. (B) Frequency at which different volumes of the pocket are sampled. The size of the bisphosphonate bound crystal structure (2E98), the newly described apo crystal structure (Apo), and Cluster 4 which bound tetramic acids are represented by labeled dashed lines in both graphs.

To help confirm the nature of this pocket closure on ligand removal (or expansion on ligand binding), we crystallized the *E. coli* UPPS in the ligand free form. Full

crystallographic data and structure refinement details are given in table 1 and a comparison of this structure (PDB ID code 3QAS) with that of the bisphosphonate-containing species (PDB ID code 2E98) are shown in Figure 4. Crystal structures of the ligand bound and apo form of UPPS reveals that a major pocket closure upon ligand removal, from 1032 $Å^3$ to 432 $Å^3$, is in good agreement with the time-average pocket volume (332 $Å^3$) of the apo form obtained from the MD simulation.

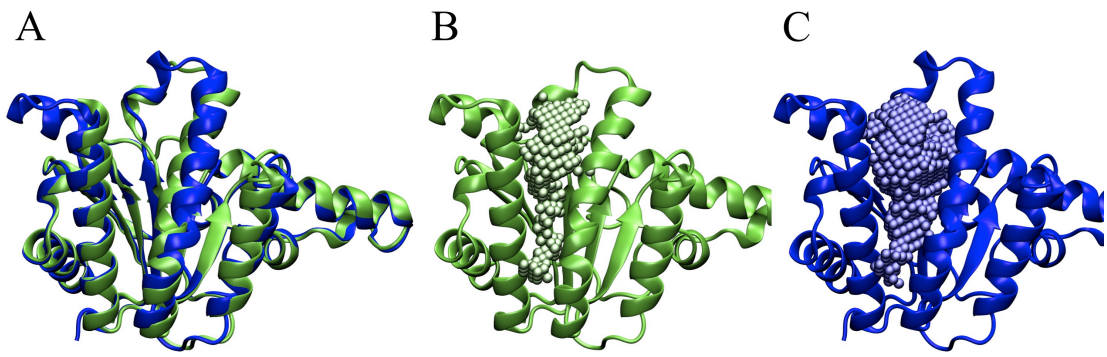


Figure 3.3 (A) The "apo" crystal structure in green and the bisphosphonate bound crystal structure in blue. (B) The "apo" crystal structure with 1 Å spheres filling the active site pocket. (C) The bisphosphonate bound crystal structure with 1 Å spheres filling the active site pocket. Note the significantly larger pocket size in the bisphosphonate bound structure when compared to the "apo" crystal structure.

In order to further investigate the binding mode of tetramic acid and dihydropyridin-2-one-3-carboxamide inhibitors, molecules **1i**, **1j**, **4a**, **4g**, **4j**, **4l** and **4m** (Chart 1) were docked into the five most representative MD structures. Interestingly, only docking into the most representative member of the fourth cluster generated ligand poses very similar to the one observed for the FPP natural substrate (Figure 1D). Unlike the results obtained from the crystal structure 2E98, all docked tetramic acid and dihydropyridin-2-one-3-carboxamide inhibitors reproduced the binding mode of the substrate FPP very well. (Figure 1D). Since our bisphosphonates bind to structures with

open active sites, we wanted to know if these compounds, which were designed using a pharmacophore hypothesis of FPP binding, bind to the same size pockets as FPP bound or *apo* crystal structures. Four crystal structures are described by Guo *et. al*, which have IPP or FSPP, an FPP analogue, in the active site[8]. These structures had active site volumes ranging from 295 $Å^3$ to 330 $Å^3$, which is similar to the calculated volume of the fourth cluster, 377 $Å^3$. This indicates that tetramic acids and dihydropyridin inhibitors, as well as the natural substrates, bind to more closed forms of the enzyme, similar to the apo state, while bisphosphonates bind to an open form.

Since the conformational states that bind tetramic acid and dihydropyridine inihibitors are lower populated, and those that bind bisphosphonates are rarely sampled in our MD simulations, our results suggest that a population shift mechanism(*151*) may be playing an important role in changing the equilibrium towards other conformations upon inhibitor binding. Furthermore, it appears that the active site of the apo structure opens and expands considerably upon binding of bisphosphonate ligands, shifting the population of UPPS enzymes to a markedly different structure. To further investigate this population shift we plotted the principle components (PC) of our HIP trajectory (Figure 5C) and highlighted the BPH binding structure (green), and tetramic acid and dihydropyridine binding structure (blue). PC analyses break the complex motion of molecular dynamics into just a few variables. The two eigenvectors shown are the principle components of motion that account for the most motion. In the event of inhibitor binding, the PC results suggest a shift away from the center of the most highly sampled area of the *apo* MD simulation. Therefore, our results indicate that structurally

diverse inhibitors recognize a specific set of conformational states of the receptor, which can vary significantly between families of ligands.

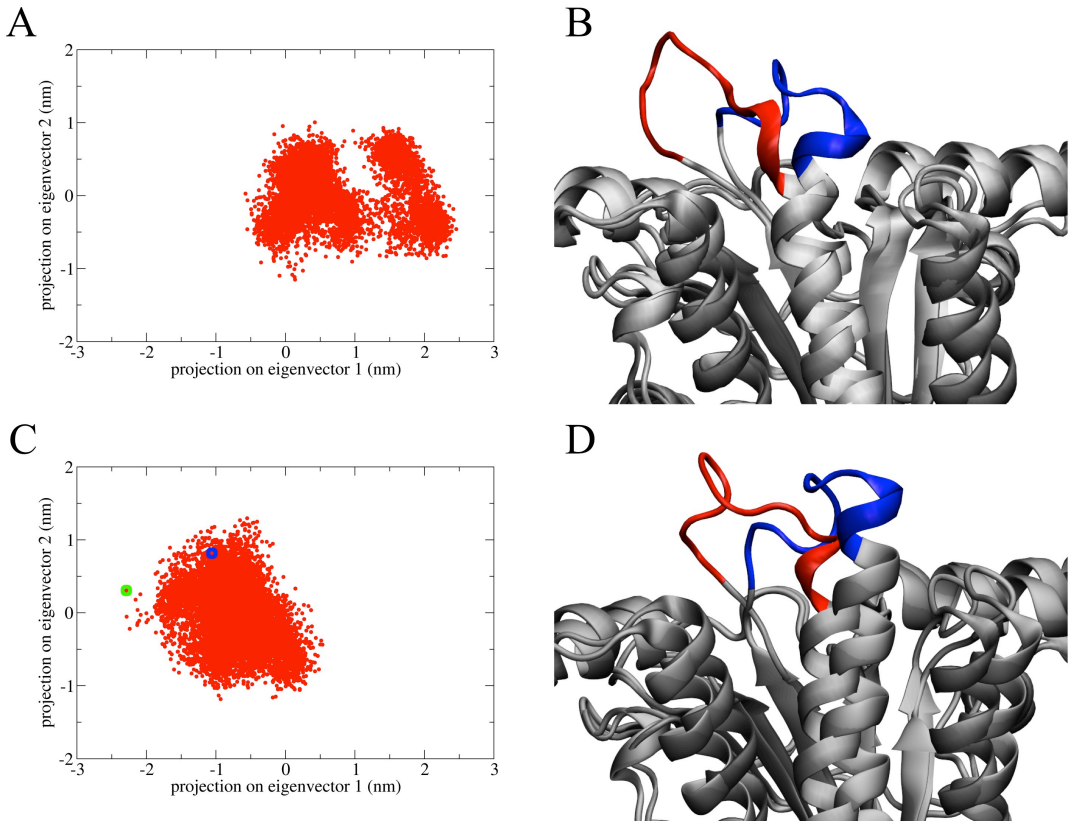


Figure 3.4 Principal Component Analysises (PCA) (A) Principal Component Analysis (PCA) calculated from the HID43 trajectory. (B) The extreme conformations of the flexible loop (residues 72-82) in HID43 are shown in red and blue. (C) PCA calculated from the HIP43 trajectory. The green circle indicates the largest conformation, and the blue circle indicates cluster 4. (D) The extreme conformations of the flexible loop (residues 72-82) in HIP43 are shown in red and blue.

Breathing Motion of the Catalytic Pocket

The most dynamic region of UPPS is the loop comprising residues 72-82. Besides the terminals, the root mean square fluctuation (RMSF) of this loop region is by far the dominant feature in Figure S5 for both the HID43 and HIP43 simulations. The RMSF plots also reveal that the HIP43 simulations show a slightly more flexible protein,

especially the helices as defined by residues 80-100 and 120-140. One of the highly populated conformational states of HID43 shows ARG77 deep inside the UPPS active site. Similar conformations were not observed in the HIP43 simulation. This result suggests that differences in electrostatic interactions, originating from different protonation states of active site HIS43, may affect the loop dynamics and the motion of the catalytically important residues on the flexible loop. It has been proposed that ARG77 plays an important role in the catalytic mechanism by helping the transfer of the pyrophosphate group from one substrate to another in the active site[5]. Kinetic studies have shown a 1000 fold decrease in activity when ARG77 is mutated to ALA77(*128*). Our MD trajectories revealed that ARG77 samples more extensively regions between the first and second sites in the HID43 than the HIP43 simulation. This behavior can be attributed to the difference in electrostatic forces originating from the doubly protonated HIP43, which prevents this movement. The flexible loop in the HID43 simulation shows substantially more movement into and out of the pocket. The effect of different protonation states of HIS43 on the dynamic behavior of UPPS can also be seen in Figure 5. The projection of the trajectories onto the first two eigenvectors calculated from Principal Component Analysis (PCA) reveals that each system clearly samples different regions of the conformational space. This result may have catalytic implications since it suggests that ARG77 is only able to successfully transfer the pyrophosphate group between the two sites when HIS43 is singly protonated. Additionally, the imidazole group of HIS43 interacts directly with the pyrophosphate, and experimental results have shown that a 1000 fold decrease in catalytic activity is observed when mutated to ALA43. These results support the hypothesis proposed by Chang *et. al* who suggests the

initial binding of FPP is encouraged by the protonated imidazole group followed by proton donation from HIS43 to FPP(*127*), inducing changes in the dynamic behavior of the flexible loop.

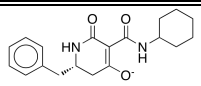
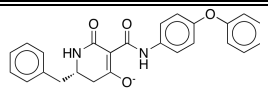
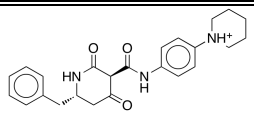
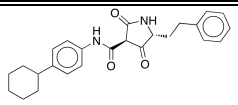
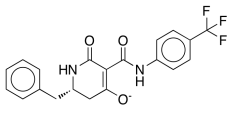
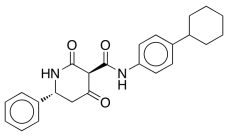
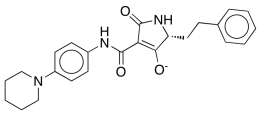
Conclusions

In this work, we performed long MD simulations, crystallography, docking studies and pocket volume analysis on UPPS to investigate its dynamic behavior and the influence of protein flexibility on the design of UPPS inhibitors. We combined virtual screening procedures with MD simulations in order to incorporate protein dynamics into the drug design efforts(*23-25*, *152*). Our MD simulations showed that UPPS is a highly flexible system, displaying very mobile active site pockets. Our results suggest that different classes of inhibitor molecules may recognize different "*active conformational states,"* some of which can be low populated in the *apo* form of the enzyme. We identified rare conformations of UPPS with expanded pocket volumes from an MD simulation that primarily sampled much smaller pockets. Moreover, we observed that poses of co-crystallized bisphosphonate inhibitors can only be reproduced on rarely sampled states with expanded pockets, that exhibit conformations similar to bisphosphonate UPPS co-crystal structures. Conversely, other inhibitor classes that mimic FPP binding require a conformation that is less expanded. Proper identification of the conformational state that specific inhibitors bind to is highly important for structure based drug design in UPPS. We also described the first complete *apo* E. coli UPPS structure with an intact active site. We observed an excellent agreement between the calculated average pocket volume from the *apo* enzyme simulation and the newly solved

crystal structure. Since the successful design of new UPPS-inhibitors is complicated by the intrinsic dynamic behavior of the receptor, the characterization and identification of these low populated, but extremely relevant, *active conformational states* is of key importance in virtual screening efforts.

Table 3.1 UPPS inhibitors. IC_{50} values shown in parentheses.

Substituted tetramic acid and dihydropyridin-2-one-3-carboxamides[a]

			
4m (0.07 μM)	4g (0.11 μM)	4j (0.11 μM)	1i (0.12 μM)
			
4l (0.14 μM)	4a (0.2 μM)	1j (0.8 μM)	

BPH[b]

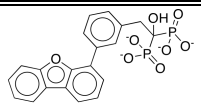
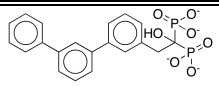
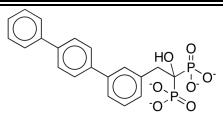
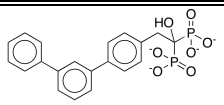
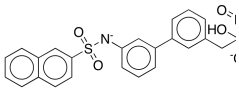
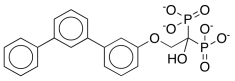
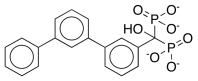
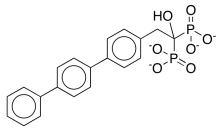
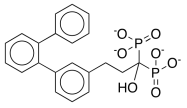
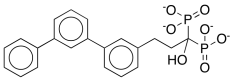
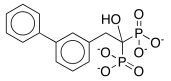
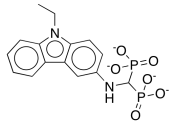
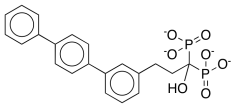
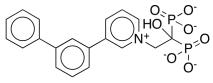
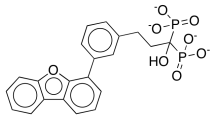
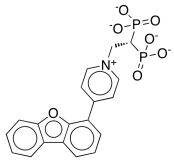
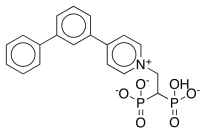
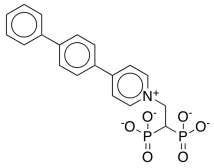
			
BPH-629 (0.59 μM)	BPH-608 (0.69 μM)	BPH-628 (0.91 μM)	BPH-625 (1.0 μM)
			
BPH-675 (1.1 μM)	BPH-676 (1.3 μM)	BPH-640 (1.8 μM)	BPH-626 (2.0 μM)
			
BPH-621 (3.1 μM)	BPH-622 (5.5 μM)	BPH-294 (6.8 μM)	BPH-228 (7.8 μM)
			
BPH-618 (10.0 μM)	BPH-364 (10.0 μM)	BPH-646 (12.0 μM)	BPH-642 (13.0 μM)

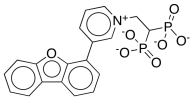
Table 3.1 continued



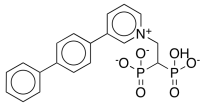
BPH-633 (21.0 μM)



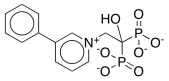
BPH-632 (28.0 μM)



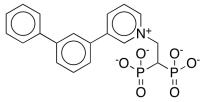
BPH-641 (34.0 μM)



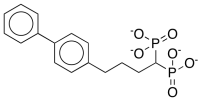
BPH-620 (35.0 μM)



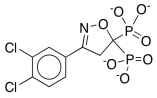
BPH-300 (48.0 μM)



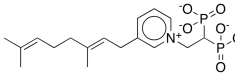
BPH-614 (68.0 μM)



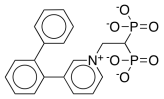
BPH-674 (69.0 μM)



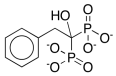
BPH-663 (87.0 μM)



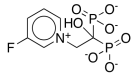
BPH-601 (89.0 μM)



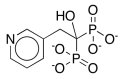
BPH-619 (120.0 μM)



BPH-11 (200.0 μM)



BPH-461 (490.0 μM)



BPH-2 (660.0 μM)

[a] Peukert 2008, (*S. pneumoniae* UPPS tested)

[b] Guo 2007, (*E. Coli* UPPS tested)

Table 3.2 Data collection and refinement statistics for UPPS apo crystal (3QAS). (Values in parentheses reflect the statistics for the high resolution shell)

Data collection	
Space group	$P2_12_12_1$
Unit cell dimension (Å)	
a, *b*, *c* (Å)	62.633, 68.762, 111.826
X-ray source	LS-CAS 21-ID-F
Wavelength (Å)	0.9787
Resolution (Å)	50-1.70 (1.73-1.70)
No. of reflection observed	385, 942
Unique	53,420 (2,600)
Completeness (%)	97.5 (99.7)
R-merge	0.051 (0.400)
$I/\sigma I$	51.2 (3.7)
Multiplicity	7.2 (5.8)
Refinement statistics	
Resolution range (Å)	50.0–1.70
R-work/*R*-free (%)	18.6/22.4
RMSD	
Bond lengths	0.026
Bond angles	2.115
No. of Protein atoms	3,453
B average ($Å^2$) of protein	28.77
Ramachandran plot statistics	
Residues in preferred regions (%)	95.3
Residues in allowed regions (%)	4.4
Residues in generously allowed regions (%)	0.3

Acknowledgements

This work was supported by the Molecular Biophysics Training Grant GM08326 (WS), the National Science Foundation Grant MCB-0506593, NBCR, CTBP, and Howard Hughes Medical Institute (JAM), National Institutes of Health Grant GM31749

(JAM) and GM065307 (EO). Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357. Use of the Life Science Collaborative Access Team Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor for the support of this research program (Grant 085P1000817).

Chapter 3 is a minimally modified reprint of the material as it appears in Sinko, W.; de Oliveira, C.; Williams, S.; Van Wynsberghe, A.; Durrant, J. D.; Cao, R.; Oldfield, E.; McCammon, J. A. Applying molecular dynamics simulations to identify rarely sampled ligand-bound conformational states of undecaprenyl pyrophosphate synthase, an antibacterial target. *Chem Biol Drug Des* 2011, *77*, 412–420. I was the first author an principal investigator on this paper.

Supplemental Information

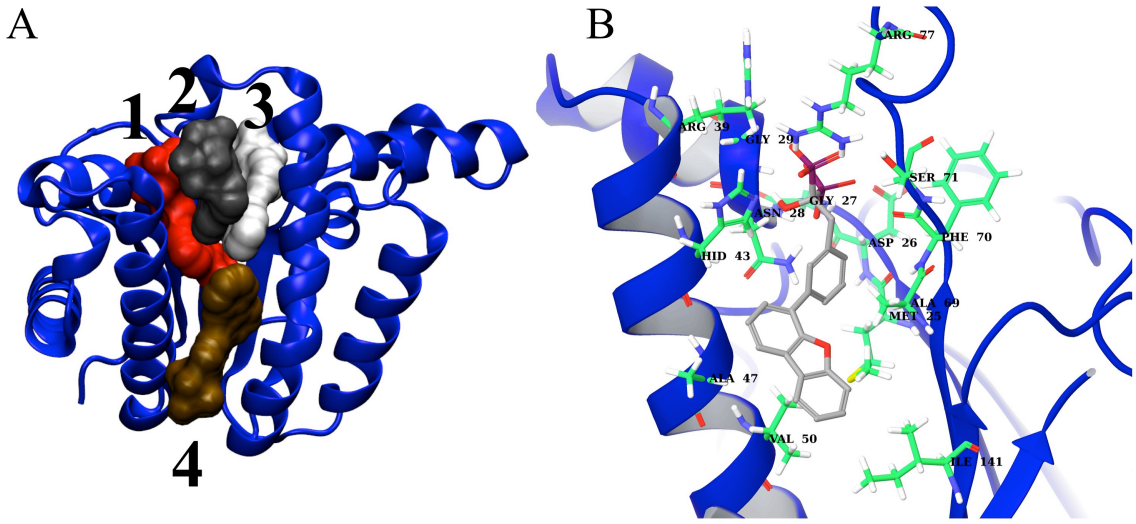


Figure S3.1 The 2E98 crystal structure, (A) Monomer A is represented in cartoon. bisphosphonate binding sites 1, 2, 3, and 4 are represented in surface in red, gray, white and brown respectively. (B) bisphosphonate binding site 1 interaction map with labeled residues. Residues have light green carbons, BPH-629 has gray carbons.

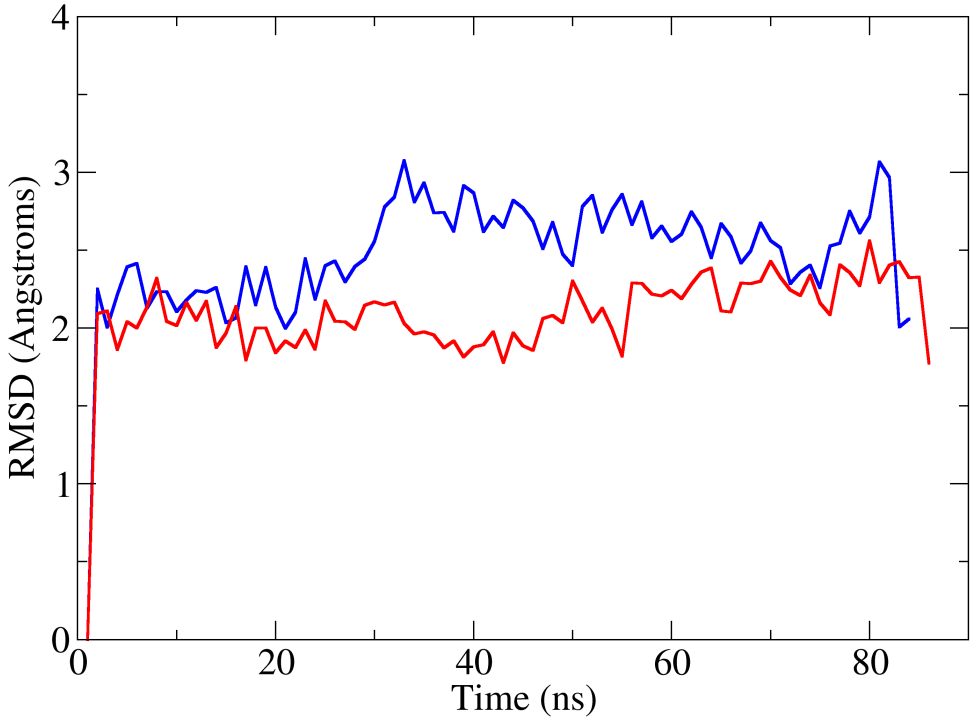


Figure S3.2 RMSD of the Trajectories. HIP43 (blue) simulation and HID43 (red) simulation, both show stability throughout the MD run.

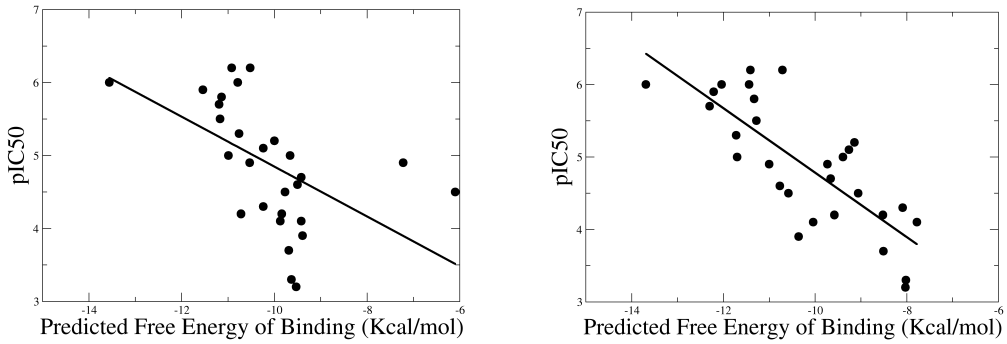


Figure S3.3 Scatter plot of pIC_{50} and predicted free energy of binding by GLIDE for BPH compounds. On the left is the 2E98 crystal structure docking results and linear regression line (R=-.52) and on the right is the largest structure from the MD simulation and linear regression line (R=-.79).

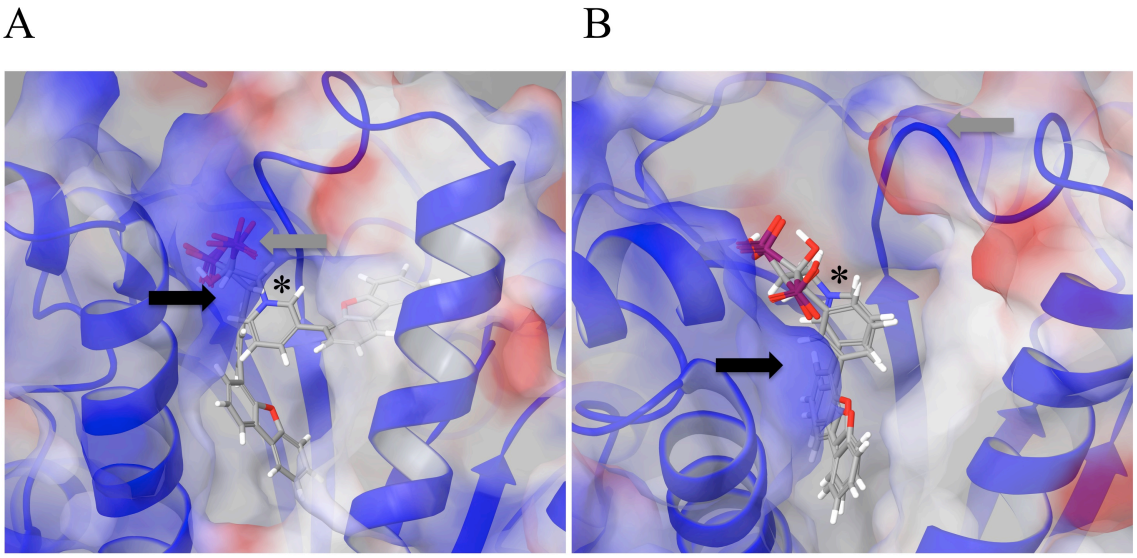


Figure S3.4 Poses of BPH-629 compared with BPH-641 docked against (A) the 2E98 crystal structure, and (B) the largest MD derived structure. The protein backbone is shown in cartoon representation with a transparent electrostatic surface representation (blue is positive and red negative). The black arrow points to HIS43, the gray arrow points to ARG77, the * denotes the nitrogen in BPH-641, ligands are represented in stick form with standard atom coloring. Note the narrow positively charged bisphosphonate binding region in the 2E98 structure and the unexpected pose generated by GLIDE for BPH-641.

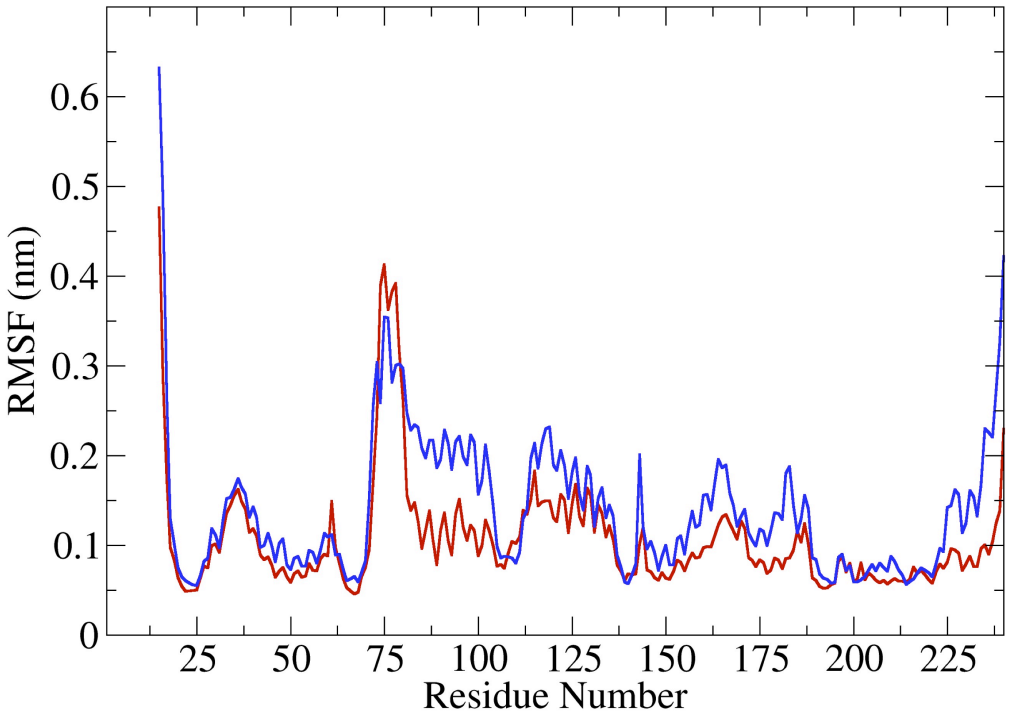


Figure S3.5 Root mean square fluctuation (RMSF) calculated for each residue in the MD simulations of HID43 (red) and HIP43 (blue) systems.

Chapter 4: Protecting High Energy Barriers: A New Equation to Regulate Boost Energy in Accelerated Molecular Dynamics Simulations

Abstract

Molecular dynamics (MD) is one of the most common tools in computational chemistry. Recently our group has employed accelerated molecular dynamics (aMD) to improve the conformational sampling over conventional molecular dynamics techniques. In the original aMD implementation, sampling is greatly improved by raising energy wells below a pre-defined energy level. Recently, our group presented an alternative aMD implementation where simulations are accelerated by lowering energy barriers of the potential energy surface. When coupled with thermodynamic integration simulations, this implementation showed very promising results. However, when applied to large systems, such as proteins, the simulation tends to be biased towards high energy regions of the potential landscape. The reason for this behavior lies in the boost equation used since the highest energy barriers are dramatically more affected than the lower ones. To address this issue, in this work, we present a new boost equation that prevents over-sampling of unfavorable high energy conformational states. The new boost potential not only provides better recovery of statistics throughout the simulation but also enhanced sampling of statistically relevant regions in explicit solvent MD simulations.

Introduction

Molecular Dynamics simulation (MD) is one of the most common tools used by computational chemists to study the dynamic behavior of biomolecules(*153*, *154*) However, conventional MD techniques (cMD) are still limited to relatively short timescales, which hinder observation of conformational transitions that are essential for protein function(*57*, *153*). Most of these transitions occur on timescales of milliseconds to seconds or longer and often involve the rare crossing of high energy barriers. In an effort to extend the time scale of all-atom molecular dynamics simulations of biomolecules, our group recently proposed an enhanced sampling technique called accelerated molecular dynamics (aMD). This method, which is based on the Hyperdynamics technique introduced by Voter *et al*.(*58*), has been shown to increase conformational sampling of biomolecules over cMD(*57*). Recently, our group has been successfully using aMD in a wide range of applications and biological systems(*57*, *59-61*, *155-157*).

Two major implementations of the boost equation for aMD have been proposed. In the original implementation, the boost potential is defined according to equation 1(*57*, *61*).

Equation 1

$$\Delta V^a = \frac{(E_1 - V(r))^2}{(\alpha_1 + E_1 - V(r))}$$

A continuous non-negative boost potential function, $\Delta V^a(r)$, is added the original potential surface, V(r), such that regions around the energy minima are raised and those near high barriers or saddle points are left unaffected. Thus, whenever V(r) is below a chosen threshold boost energy, E_1, the simulation is performed on the modified potential $V^*(r)=V(r)+\Delta V^a$, otherwise sampling is performed on the original potential $V^*(r)=V(r)$. The parameter α_1 modulates roughness and the depth of the energy minima on the modified surface.

To recover the correct canonical ensemble, each frame of the simulation must be re-weighted using the Boltzmann factor $e^{\beta \Delta V[r]}$. Since the lowest energy wells may be associated with the largest boost values, the reweighting can have a detrimental effect on the statistics(*63*, *68*).

To address this issue, a second implementation was introduced in which energy barriers are modified, instead of energy minima(*68*).

Equation 2

$$\Delta V^b = \frac{(V(r) - E_1)^2}{(\alpha_1 + V(r) - E_1)}$$

A continuous negative boost potential function, $\Delta V^b(r)$, is added the original potential surface, V(r) such that regions around the energy barriers are lowered and those near the minima are left unaffected. Thus, whenever V(r) is above the boost energy, E_1, the simulation is performed on the modified potential $V^*(r)=V(r)-\Delta V^b$, otherwise sampling is performed on the original potential $V^*(r)=V(r)$.

This implementation improves the statistical reweighting problem by allowing much of the simulation to remain in the original potential surface, which in this case needs no reweighting. However, application of ΔV^b tends to over sample high energy regions of the potential landscape. As can be seen in equation 2, the boost potential is proportional to the difference V-E and as a consequence regions of the potential surface displaying large V (or high-energy regions) are affected significantly more than regions with relatively low energy barriers. When applied to large systems, such as proteins, the simulation tends to be biased towards high energy regions of the potential landscape. In small systems, application of ΔV^b revealed promising results when combined with free energy calculations, such as thermodynamic integration (TI)(*68*).

In this work, we describe a new boost potential (equation 3) in an attempt to combine the strengths of the two previous implementations.

Results and discussion

A possible way to overcome the sampling issues associated with ΔV^b is to modify the boost potential equation so that its magnitude reduces significantly at large values of V-E.

New Equation ΔV^c:

Equation 3

$$\Delta V^c = \frac{(V(r) - E_1)^2}{(\alpha_1 + V(r) - E_1)\left(1 + e^{\frac{-(E_2 - V(r))}{\alpha_2}}\right)}$$

We defined a second energy level (E_2) in order to return the modified potential surface back to the original one whenever the potential energy of the system is larger than

E_2. This boost equation is shown above as ΔV^c (equation 3). The second energy level allows the user to define a window of acceleration between E_1 and E_2. To regulate the return to the original potential upon crossing E_2, a second parameter alpha is required (α_2). The term in the large brackets of the denominator is responsible for bringing the boost to zero when the potential energy V(r) is higher than E_2. Thus, when V(r) is higher than E_2, $\left(1+e^{\frac{-(E_2-V(r))}{\alpha_2}}\right)$ tends to a very large positive number, and as a result, the modified potential converges to the original one, V(r). On the other hand, when V(r) is lower than E_2, the term $\left(1+e^{\frac{-(E_2-V(r))}{\alpha_2}}\right)$ tends to 1, which results in $\Delta V^c = \Delta V^b$ or equation 2.

We explored the new boost equation by creating a hypothetical one-dimensional potential using the analytical equation below.

Equation 4

$$v(r) = -200 + 50 * \left(\cos(x * \pi) - 1 - \frac{x^2}{x + (1-x) * \sqrt{\frac{3-x}{4}}} \right)$$

Figure 1 displays the effect of boost energy E (E_1 and E_2) and α (α_1 and α_2) on equations 2 and 3. The upper solid black line represents the unmodified potential V(r), while the lower solid black line represents modified potential V(r)* generated after application of the equation 2, ΔV^b. Boost energies E are shown as dashed lines. The solid colored lines represent different modified potentials, V(r)*, generated by ΔV^c with different sets of parameters. Figure 1A shows that high energy barriers can be selectively

protected by setting different values of E_2. It is worth noting that the modified potential generated by ΔV^c, follows closely along ΔV^b until the difference between E_2 and E_1 is similar to the difference between V and E_1 (Figure 1A and B).

Like in the original implementation, the degree of acceleration is controlled by the parameter α_1 and E_1. Parameter α_2 controls how strongly energy barriers higher than E_2 are protected. For instance, when V is higher than E_2, in the limit $\alpha_2 \to \infty$, the term $\left(1 + e^{\frac{-(E_2 - V(r))}{\alpha_2}}\right) \to 2$, and ΔV^c converges to ½ΔV^b, and as a result large energy barrier are not effectively protected anymore. On the other hand, when $\alpha_2 \to 0$, the term $\left(1 + e^{\frac{-(E_2 - V(r))}{\alpha_2}}\right) \to \infty$ and $\Delta V^c \to 0$; thus, keeping all energy regions, where V(r) is higher than E_2, unchanged. Figure 1C displays the effects of α_2 on $V^*(r)$.

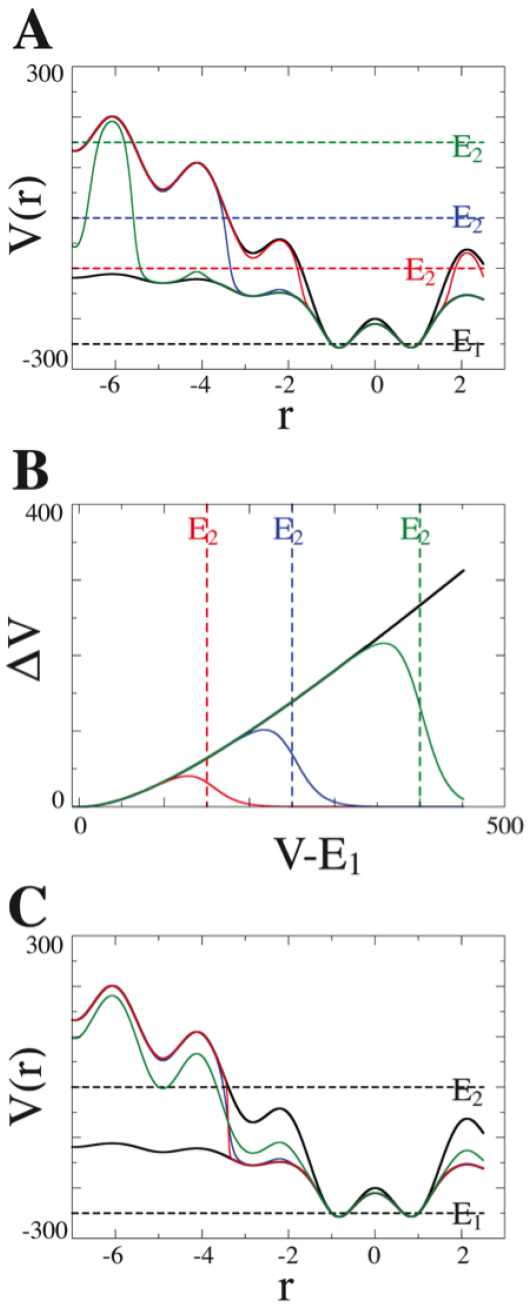


Figure 4.1 Hypothetical one dimensional potential representing the effect of ΔV^c. In all charts, α_1 =200 and E_1 =-250. The upper and lower solid black lines represent the original potential and the modified potential generated with ΔV^b, respectively. This color scheme is used throughout. A) Effects of different parameters E_2 (dashed colored lines) on the modified potential generated with ΔV^c (solid colored lines). B) Boost levels (ΔV) as V moves away from E_1. For both (A) and (B) α_2 =15, and E_2 =-100 (red), 0 (blue) and 150 (green). C) Effect of varying α_2 parameter on ΔV^c: α_2 =3 (red), 15 (blue), 75 (green) with E_2 =0.

Although this new implementation introduces two new parameters, E_2 and α_2 are easily estimated. Initial guesses for α_2 are based on the hypothetical one-dimensional potential shown in Figure 1. To keep the underlying shape of the original potential surface and effectively protect energy barriers higher than E_2, α_2 is recommended to be proportional to the difference ~ (E_2-E_1). More specifically, we estimate α_2 to be between 20-60% of the difference (E_2-E_1). Energy levels E_1 and E_2 are estimated from short cMD simulations. Since ΔV^c is only effectively applied to the system whenever the potential V(r) is higher than E_1, it is important to set E_1 not higher than the average potential energy of system, <V(r)>, in order to guarantee a minimum degree of acceleration. In this work, V(r) and <V(r)> correspond to the total and average dihedral energy, respectively. E_2 is simply defined as $E_2=E_1+\Delta E$, where ΔE is the highest energy barrier allowed to be crossed. Selection of optimum boost parameters is bound to be system dependent. For this reason, short aMD runs are strongly recommended to fine tune parameters α_1 and E_1. Failure in obtaining suitable parameters may lead to two possible scenarios: i) none or extremely low acceleration is effectively applied to the system. In this case, aMD and cMD will likely generate very similar trajectories; ii) extremely high acceleration is applied to system, which results in serious structural and energetic instabilities.

Unless otherwise stated, all simulations were performed applying the boost potential ΔV^c to the dihedral terms of the potential energy function. Enhanced sampling techniques, such as aMD, based on the dihedral energy contributions have been successfully used to effectively enhance conformational sampling of biomolecules(*156-161*). The approach presented in this work can be easily extended to the nonbonded energy terms via the dual boost method(*60*). To investigate the use of the new boost

equation ΔV^c, we first compared our aMD simulations results of fully solvated alanine dipeptide to cMD protocols. Alanine dipeptide has been extensively studied as a model system to evaluate free energy and conformational change in biomolecular simulations(*162-168*). Figure 2A displays the time evolution of the Psi (Ψ) angle during the cMD and five aMD simulations of 10ns. As can be clearly seen, the number of Ψ transitions dramatically increases as we modify boost parameters E_2 and α_2. Figures 2B-E show the free-energy density plots obtained from cMD simulations of 10ns, 100ns, 1 μs and an aMD simulation of 10ns. The free-energy density plots were calculated from the population of states sampled on each simulation. To recover the corrected canonical ensemble each frame of the aMD trajectory was Boltzmann weighted by its respective boost factor. Figure 2B reveals that the conformational sampling obtained from 10ns of cMD is mainly restricted to α-helical ($\Phi < 0°$ and $-60° < \Psi < 0°$) and β-strand regions ($\Phi < 0°$ and $120° < \Psi < 180°$), with the α-helical region displaying the most populated states. A significant increase in conformational sampling is evident when the cMD is extended to 100ns (Figure 2C). The most pronounced change can be seen in the left-handed α-helix region (Φ~50 and Ψ~50), which is now well sampled and is not observed in the cMD of 10ns. It is also noted a dramatic increase in the number of transitions between the α-helical and β-strand regions. To provide some insights concerning the time scale accessed by our aMD runs, we further extended the cMD simulation to 1 μs. Comparison of figures 2D and 2E clearly shows that there is a good agreement between the regions sampled by our short aMD of 10ns and the cMD of 100ns and 1 μs. For the

alanine dipeptide system, these results suggest that aMD simulations with ΔV^c can accelerate conformational sampling by at least 10-100 fold.

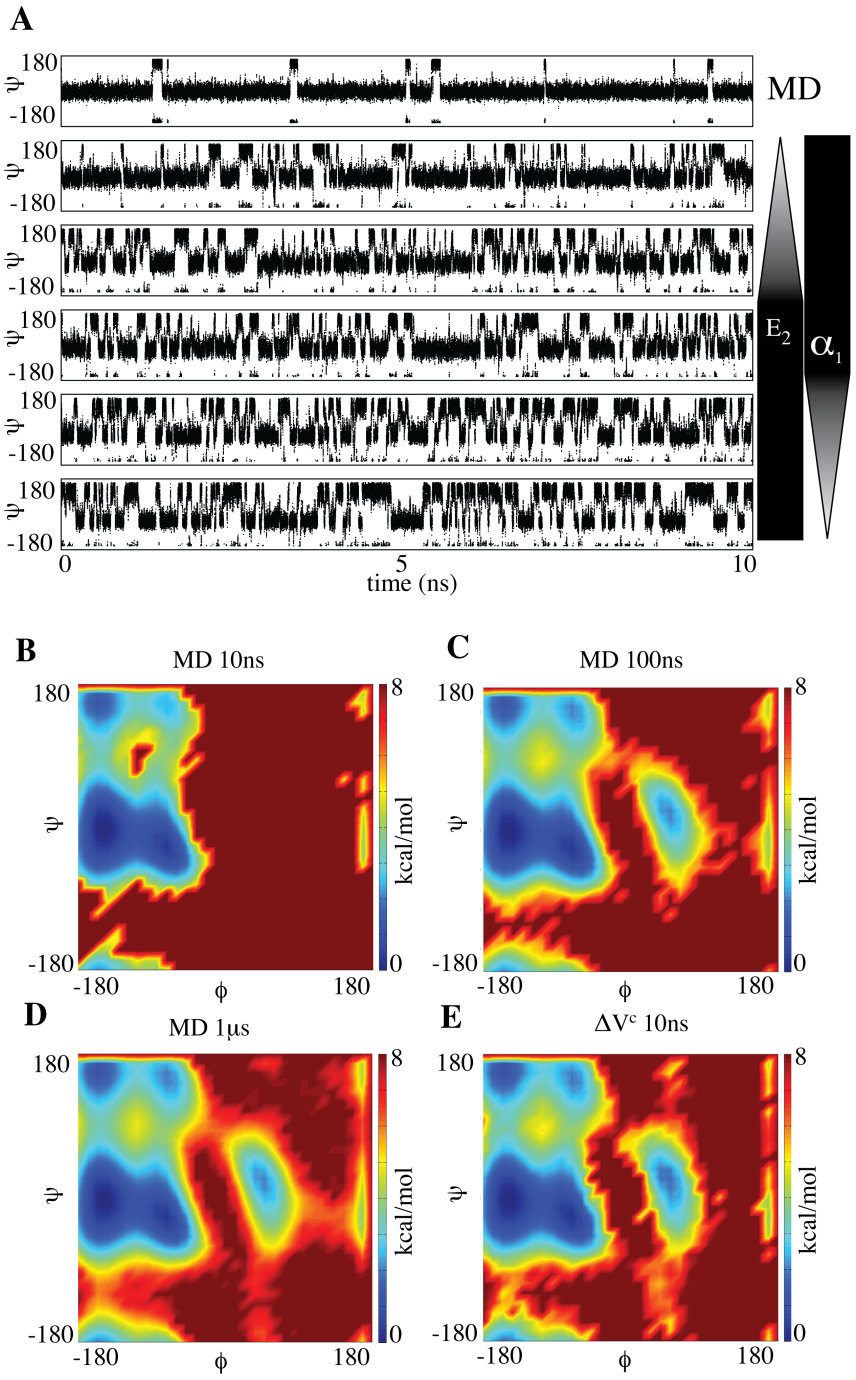


Figure 4.2 Alanine dipeptide simulation results. A) Psi (Ψ) angle values obtained from cMD and five different aMD simulations. From top to bottom, aMD parameters were set to: $E_2=E_1+15$ and $\alpha_1=5$, $E_2=E_1+20$ and $\alpha_1=5$, $E_2=E_1+25$ and $\alpha_1=5$, $E_2= E_1+25$ and $\alpha_1=2.5$, $E_2=E_1+25$ and $\alpha_1=1.25$. In all simulations, E_1 and α_2 were set to 10 and 5 respectively. Weighted free energy density plots obtained from cMD (B, C, D) and aMD with ΔV^c (E). All values are in kcal/mol.

While boosting through energy barriers is important for sampling, limiting the boost to reduce the population of thermodynamically unfavorable states is equally important. To illustrate the advantage of ΔV^c and its boost limiting capabilities over ΔV^b, we analyzed and compared the Psi (Ψ) and Omega (Ω) angle transitions (cis/trans isomerisation) obtained from the alanine dipeptide simulations in both implementations. As seen in Figure 3, as the degree of acceleration is increased (by reducing the value of parameter α_1), ΔV^b increases dramatically not only Ψ but also Ω dihedral transitions. Conversely, ΔV^c promotes a very similar increase in Ψ dihedral transitions without affect the Ω dihedral angles. This result confirms the capability of ΔV^c to accelerate conformational transitions by selectively crossing energy barriers lower than the pre-defined energy level. It is worth mentioning that ΔV^b notably under samples the normally preferred region $-50 > \Psi > 50$ at high acceleration conditions.

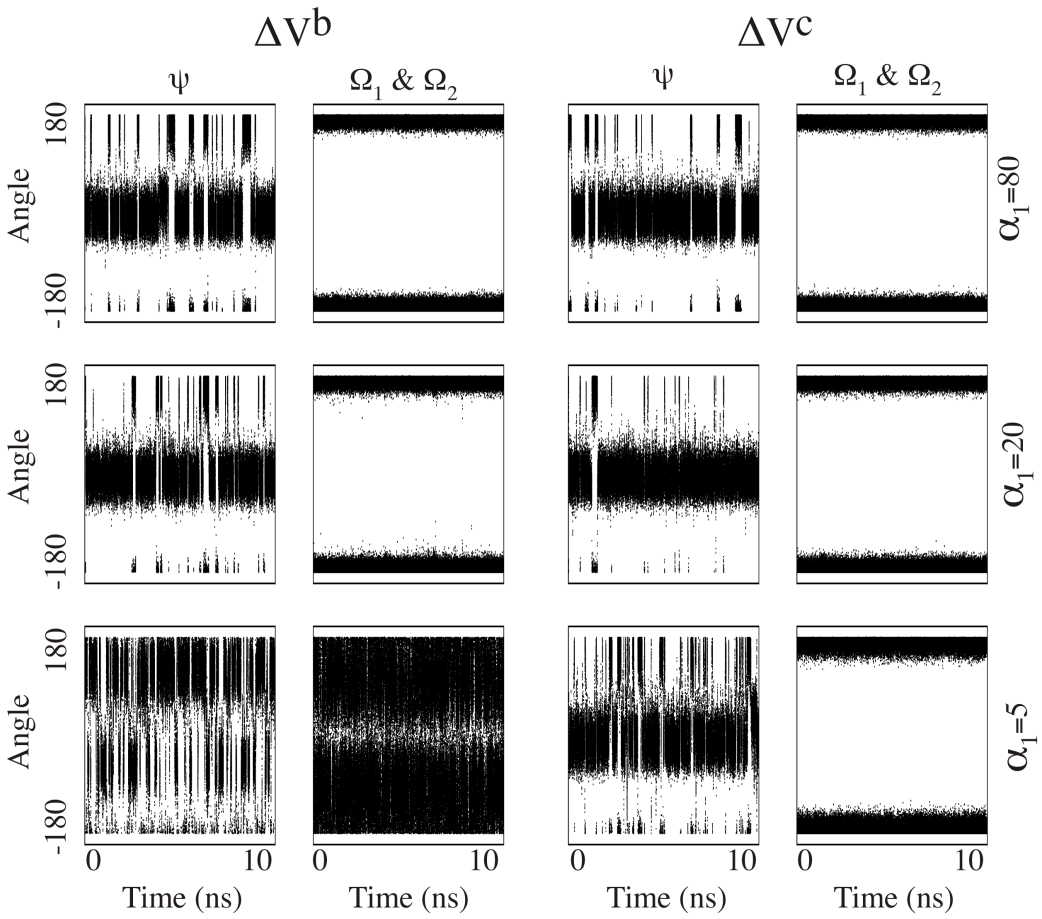


Figure 4.3 Psi (ψ) and Omega (Ω) angle values obtained from aMD simulations with boost potentials ΔV^b and ΔV^c. In all simulations, E_1=10.0 and α_1 was set as shown on the far right. Additional parameters for aMD with ΔV^c were set to $E_2 = E_1$+15 and α_2=5. All values are in kcal/mol.

To evaluate the applicability of equation ΔV^c to biomolecules, we also performed aMD studies on a more complex model system, decalanine(*169*). Figure 4 displays the distribution of eight Phi-Psi (Ψ,Φ) angles monitored along two cMD simulations of 50 and 350ns, and two independent aMD of 50ns. All simulations started from a fully solvated and extended conformation. As expected, there is a substantial improvement in conformational sampling when the cMD simulation is extended from 50ns to 350ns (Figure 4A and B). Similar results are obtained for Phi-Psi angles 4-8 when we compare

aMD with both cMD simulations (Figure 4A, B and C). Interestingly, opposite behavior is observed for Phi-Psi angles 1-3 (Figure 4C). We attribute this result to the low degree of acceleration used on the aMD simulation. Even though application of ΔV^c enhances conformational transitions of decalanine, the small boost used in this simulation, as a test case, may not generate the 7-fold acceleration expected from Figure 4C and 4B. To investigate this issue and further explore the capability of ΔV^c, we carried out two extra aMD simulations of 50ns in which we a) increased the acceleration by reducing the α_1 value by a factor of 2 (result is shown in Figure 4D) and b) increased the degree of acceleration by raising the energy level E_2, ($E_2 = E_1 + 35$ kcal/mol),in addition to reducing α_1 by a factor of 2 (Figure S1). As expected, the different aMD simulations of 50ns each (Figure 4C, D, and S1) cover different regions of the Phi-Psi sub-space, with the more accelerated ones (Figure 4D) showing better agreement with the cMD simulation of 350 ns (Figure 4B). These results also agree with the fact that, by lowering energy barriers, aMD increases the rate of escape from minimum wells and thus generates more diverse trajectories for complex systems with multidimensional energy landscapes such as decalanine. Figure S1 displays the Phi-Psi angle distributions obtained with the highest degree of acceleration tested. It is worth noting that there is better agreement with the conformational sampling obtained from the 350ns of cMD, as a result of the longer time scale accessed by this aMD simulation.

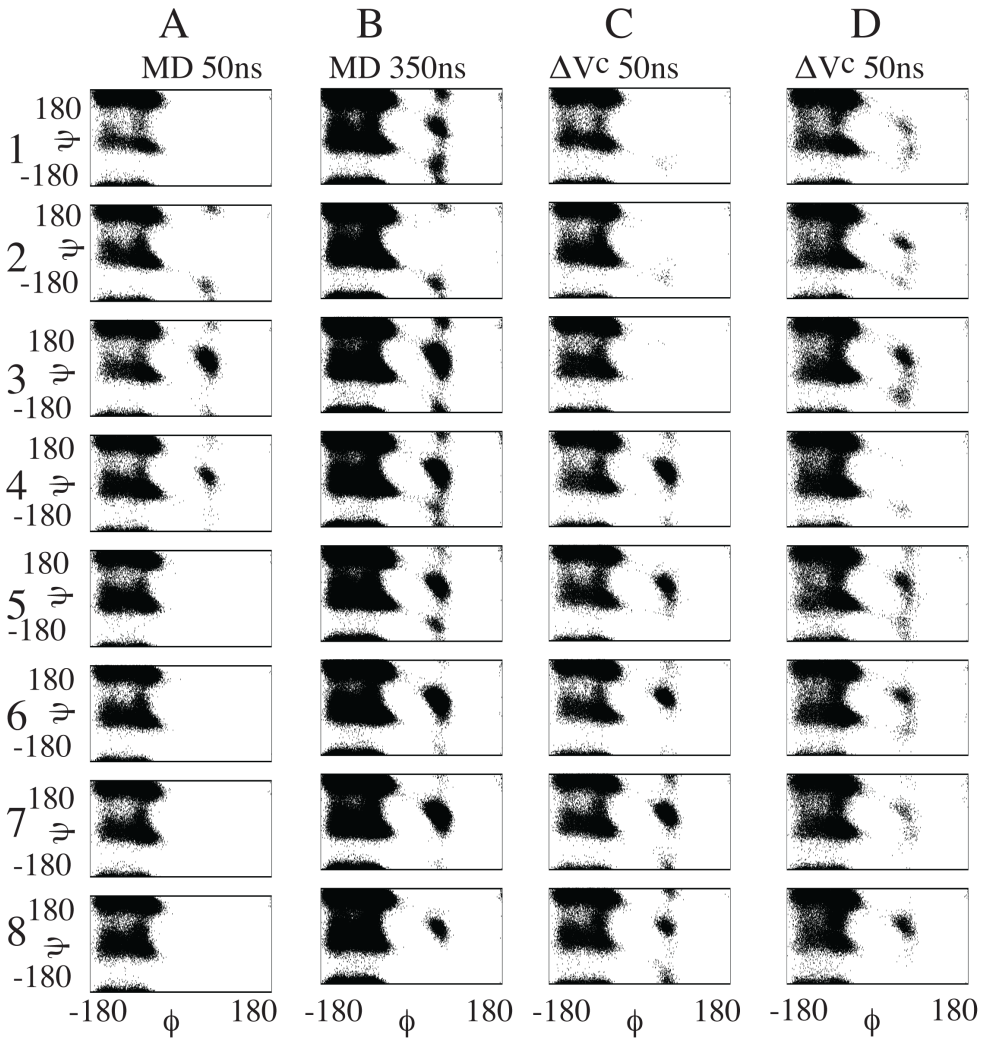


Figure 4.4 Decalanine Phi (Φ) - Psi(Ψ) angles distribution obtained from cMD and aMD simulations. For the aMD simulations with ΔV^c parameters were set to E_1=74, E_2= E_1+25, α_2=5, α_1=30 (C) and α_1=15 (D). All values are in kcal/mol.

Decalanine can adopt numerous secondary structures making it a challenging test case for enhanced sampling methods(*169*). Principal component analysis (PCA) shows that our ΔV^c aMD simulation explores energy wells that are not adequately sampled by 350ns of cMD simulation (figure 5A, B, C). One of these regions represents the state in

which decalanine adopts an α-helical conformation, energetically the most stable configuration(*169*). This folding event is evident in the aMD simulations, but not in the cMD simulations despite the latter being 7-fold longer (Figure 2S).

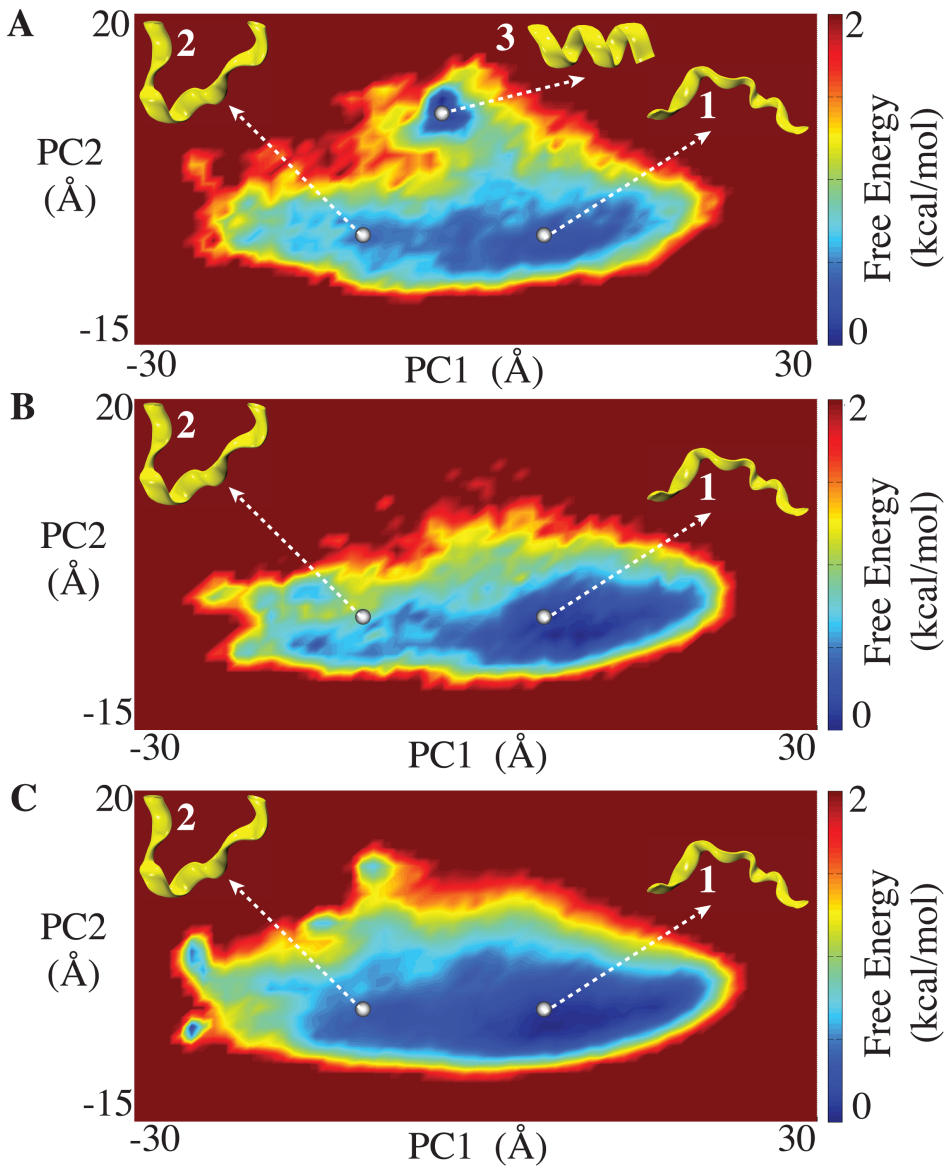


Figure 4.5 Principle Component Analysis obtained from decalanine MD simulations. A) 50 ns of aMD simulation with ΔV^c. Parameters were set to E_1=74, E_2= E_1+25, α_2=5 and α_1=15, same as in Figure 4D. B) 50 ns of cMD simulation and C) 350ns of cMD simulation. Structures 1, 2 and 3 shown in yellow represent relevant populated states in PC subspace sampled by aMD and cMD.

Free energy calculations are useful in the optimization of compounds for biological targets and host systems(*170*). However, these calculations usually require a

computationally expensive ensemble generation either from Monte Carlo calculations or MD simulations(*171*, *172*). As previously shown, coupling of aMD methods with free energy calculations, such as thermodynamic integration (TI), revealed promising results when applied to simple model systems(*68*). To further extend the applicability of aMD-based approaches to free energy calculations, in this work, we modified our original implementation by incorporating the boost equation ΔV^c into the TI simulations. As a test case, we calculated the relative free energy difference between $Ac_{2\text{-}L}Lys_{\text{-}D\text{-}}Ala_{\text{-}D\text{-}}Ala$ and $Ac_{2\text{-}L}Lys_{\text{-}D\text{-}}Ala_{\text{-}D\text{-}}Lac$ bound to vancomycin. This mutation, Ala to Lac, confers to bacteria a resistance against vancomycin(*173*). The experimental change in binding free energy has been determined to be 4.1 kcal/mol, which corresponds to approximately a thousand fold decrease in affinity from $_D$Ala to $_D$Lac (*174*).

Figure 6A compares the relative free energy of binding ($\Delta\Delta G$) calculated from TI simulations using cMD and aMD with ΔV^c. To calculate the final free energy values, we divided the trajectories in blocks of 200ps, with the last block representing the production phase; e.g., a TI simulation of 800ps corresponds to an equilibration phase of 600ps (three blocks) followed by a collecting data phase of 200ps, and a TI simulation of 1000ps corresponds to an equilibration phase of 800ps (four blocks) followed by the collecting data phase of 200ps. Thus, the points displayed in Figure 6A reveals how the calculated $\Delta\Delta G$ changes as a function of equilibration time.

It is worth mentioning that application of ΔV^c notably improves the convergence of $\Delta\Delta G$ when compared to standard cMD TI simulation. In addition, the final free energy value obtained with ΔV^c (4.3 +/-0.3 kcal/mol) shows very good agreement with the experimental value of 4.1 kcal/mol(*174*), while final free energy value from TI with cMD

is 5.3 +/- 0.3 kcal/mol. Since the same force field and simulation conditions were applied to both TI simulations, we attribute this difference solely on the conformational sampling enhancement provided by the ΔV^c. Moreover, the error associated with each point suggests that the faster convergence towards the final free energy value is statistically relevant. Interestingly, the cumulative free energy values (Figure 6B) demonstrate that the TI simulations coupled with cMD are indeed converging towards the ones coupled with aMD as we increase simulation time. Hence, inaccuracies in the final value are likely to be primarily due to the lack of convergence on lambda points. These results indicate that ΔV^c can effectively enhance conformational sampling when coupled with TI simulations and hence shorten the equilibration period required for accurate free energy calculation.

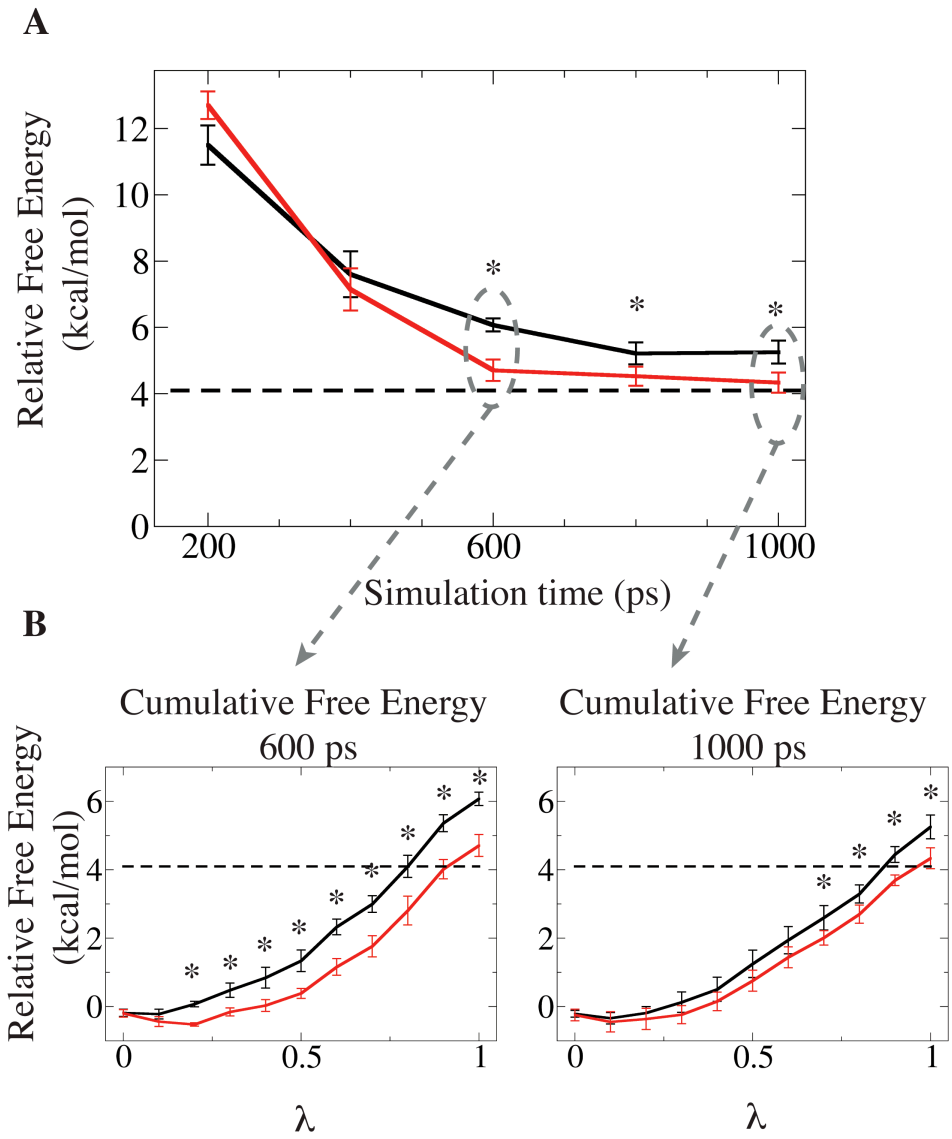


Figure 4.6 (A) Relative free energy of binding between $Ac_{2\text{-}L}Lys_{\text{-}D}Ala_{\text{-}D}Ala$ and $Ac_{2\text{-}L}Lys_{\text{-}D}Ala_{\text{-}D}Lac$ to vancomycin calculated from cMD (solid black line) and aMD with ΔV^c (solid red line). Dashed line displays the experimental value, 4.1 kcal/mol. (B) Cumulative free energy curves calculated from simulations of 600ps (left) and 1000ps (right) per λ point. The * show points where there is no overlapping between error bars.

Computational Methods

ΔV^c was implemented into the AMBER10 code(*140*) as previously reported(*68*).

Equation 5

$$V^* = V - \Delta V^c \quad \Rightarrow \quad \Delta V^c = \begin{cases} \dfrac{(V(r) - E_1)^2}{(\alpha_1 + V(r) - E_1)\left(1 + e^{\frac{-(E_2 - V(r))}{\alpha_2}}\right)} & V > E_1 \\ 0 & V \le E_1 \end{cases}$$

All cMD, aMD and TI simulations were performed using a modified version of the sander module of the AMBER10 package(*140*). TIP3P water molecules were used to solvate both the alanine dipeptide and decalanine systems(*175*). A buffer region of 12 Å was used in all systems. To eliminate any steric clashes, 100 steps of conjugant gradient minimization were performed on all systems. To bring the systems to the right density, we carried out cMD simulations of 50ps in which NPT ensemble was applied. Then, long cMD and aMD simulations were performed in which the NVT ensemble was applied. All bonds involving hydrogen atoms were constrained using the SHAKE algorithm(*176*). The temperature and pressure were controlled using weak coupling to external temperature and pressure baths(*141*). Electrostatic interactions were computed via PME (particle mesh Ewald summation) with a cutoff of 8.0 angstroms. All simulations were performed at temperature of 300k. In all accelerated simulations, the boost potential was based on the total dihedral energy. Principal Components Analysis was performed using ptraj module of AMBER10 package. All cMD simulations were projected onto the PC subspace obtained from the aMD simulation displayed at Figure 4D. Alignment of the trajectory was performed on backbone atoms of decalanine.

To study the use of the new boost equation on thermodynamic integration calculations, we calculated the relative difference in the free energy of binding of $Ac_{2\text{-}L\text{-}}Lys_{\text{-}D\text{-}}Ala_{\text{-}D\text{-}}Ala$ and $Ac_{2\text{-}L\text{-}}Lys_{\text{-}D\text{-}}Ala_{\text{-}D\text{-}}Lac$ to a vancomycin dimer, starting from the crystal structure of $Ac_{2\text{-}L\text{-}}Lys_{\text{-}D\text{-}}Ala_{\text{-}D\text{-}}Ala$ bound to vancomycin (PDB ID: 1FVM). The glycopeptides and vancomycin were parameterized using Antechamber. The system was solvated in a cubic box of TIP3P water molecules, with a buffer region of 10 angstroms(*140*)Owing to the strong correlation between glycopeptides binding affinity and vancomycin dimerization(*177*), we simulated the "back to back" homodimer of vancomycin, as present in the x-ray crystal structure. Both ligands were included in the model and modified alchemically.

TI simulations were performed with nine equally spaced λ parameters (λ = 0.1 to 0.9) in solution and in the vancomycin receptor. In all transformations, electrostatic and van der Waals contributions were decoupled and computed separately. More specifically, in this work, the alchemical transformation of $_{D}$Ala to $_{D}$Lac was carried out in three steps: i) removal of partial charges of NH group from $_{D}$Ala; ii) transformation of van der Waals parameters of NH group to O (oxygen) atom and iii) partial charge creation on O (oxygen) atom. Softcore potentials were used for step ii(*178*, *179*). The $\Delta V/\Delta\lambda$ values were calculated over a production period of 200ps along with five equilibration periods 0, 200, 400, 600, 800 ps. The final free energy values were averaged over three independent (with reassigned initial atomic velocities) cMD or aMD simulations. As previously shown, in order to recover the correct canonical ensemble, $\Delta V/\Delta\lambda$ values collected from aMD runs were reweighed by their respective boost factor $e^{\beta\Delta V[r]}$(*57*, *68*).

Error bars were calculated using $\sigma_{\langle A \rangle} \approx \sigma / \sqrt{M}$, where M is the number of independent simulations, and $\sigma_{\langle A \rangle}$ is the standard deviation of the average value **A** obtained from M independent data values (M=3 in all cases). Analysis of the trajectories was performed using ptraj(*140*).

Our aMD parameters were estimated based on the average dihedral energy term obtained from short cMD simulations. For all alanine dipeptide aMD simulations parameter E_1 was set to 10 kcal/mol. Parameter α_2 was set to 5 kcal/mol, which correspond to 0.2*(E_2-E_1). In Figure 2A, from top to bottom, aMD simulations used the following parameters: E_2=E_1+15 and α_1=5, E_2=E_1+20 and α_1=5, E_2=E_1+25 and α_1=5, E_2= E_1+25 and α_1=2.5, and E_2=E_1+25 and α_1=1.25. In Figure 2B the aMD simulation used the parameters: E_2= E_1+15 α_1=5. In Figure 3, E_2= E_1+15 (for ΔV^c) and α_1 was varied as indicated in the far right column.

Boost parameters for Decalanine simulations were: E_1=74, E_2= E_1+25, α_1=30, and α_2=5. Boost parameters for the vancomycin-glycopeptides simulations were E_1=211, E_2=E_1+25, α_1=30 and α_2=15.

Conclusion

In this work, we introduced a new boost equation, ΔV^c, for aMD simulations aiming to overcome sampling issues introduced by ΔV^b. Since energy barriers located above a pre-defined energy level can now be protected, the new boost equation ΔV^c provided much better control over high energy regions of conformational landscape when compared to ΔV^b. We used two model systems, alanine dipeptide and decalanine, to study the applicability and efficiency of ΔV^c in enhancing conformational sampling. In

both cases, the new boost potential not only provides better recovery of statistics throughout the simulation but also enhanced sampling of statistically relevant regions in explicit solvent MD simulations. When coupled with Thermodynamic Integration, our results indicate that ΔV^c can effectively enhance conformational sampling and accelerate convergence for a more accurate free energy calculation.

Acknowledgements

This work was supported by the Molecular Biophysics Training Grant GM08326 (WS), the National Science Foundation Grant MCB-0506593, NBCR, CTBP, and Howard Hughes Medical Institute (JAM), and National Institutes of Health Grant GM31749 (JAM).

Chapter 4 is a minimally modified reprint of the material as it appears in Sinko, W.; de Oliveira, C. A. F.; Pierce, L. C. T.; McCammon, J. A. Protecting High Energy Barriers: A New Equation to Regulate Boost Energy in Accelerated Molecular Dynamics Simulations. *J Chem Theory Comput* 2012, *8*, 17–23. I was a co-first author and co-principal investigator on this paper.

Supplemental Information

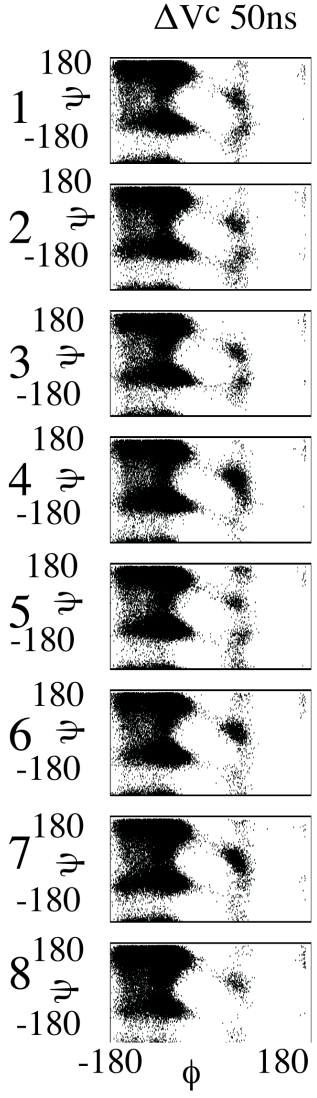


Figure S4.1 Decalanine Phi(Φ)-Psi(Ψ) angles distribution obtained from aMD simulations with ΔV^c. Parameters were set to E_1=74, E_2= E_1+35, α_2=5 and α_1=15. All values are in kcal/mol.

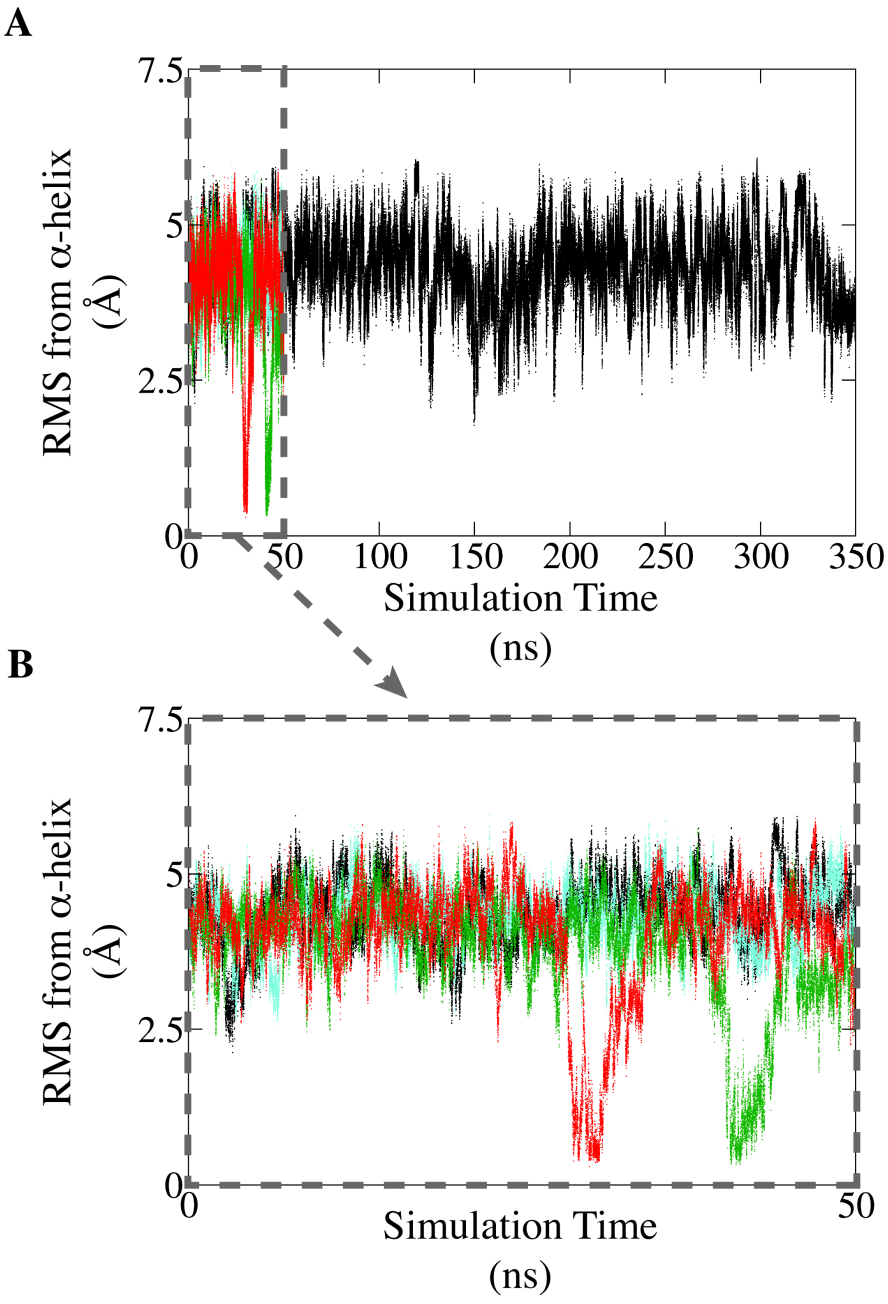


Figure S4.2 Decalinine root mean square (RMS) deviation from α-helix conformation calculated along: A) 350ns of cMD and 50 ns of aMD simulations, and B) the first 50ns of simulations displayed in (A). Solid lines correspond to MD (black) and aMD with parameters set to E_1=74, α_2=5 (all aMD simulations) E_2= E_1+25 and α_1=30 (cyan); E_2= E_1+25 and α_1=15 (red); E_2= E_1+35 and α_1=15 (green). All values are in kcal/mol.

Conclusion

In this work we demonstrated that synergy between computational and experimental observations can be used to rapidly and efficiently identify new modulators of potential drug targets. In the case of PHLPP, both the experimental observations and computation were used to determine diverse active compounds, that were narrowed down by their selectivity over other phosphatases and their cellular effectiveness. We identified two compounds that show a downstream effect on the PI3K-Akt pathway and exhibit the desired cellular effect, preventing cell death. Continuing work in the area of drug design will be to determine compounds of higher affinity for the PHLPP target. This work would benefit greatly from the crystallization of the PHLPP phosphatase domain, although our homology model has shown great predictive abilities increasing the chances of finding an inhibitor molecule from the inhibitor assay by 10-fold when paired with similarity screening. However, we do not know if we have the structural level of detail to begin making subtle modifications of the compound to improve affinity via direct interactions with the phosphatase. Since we have many compounds with structural diversity, we may be able to make a pharmacophoric model or a quantitative structure-activity relationship, QSAR model, which are methods to predict compound affinity based solely on the compound structure.

In the case of UPPS we again correlated our work with experimental details about the system. Target flexibility is highly important in biomolecular recognition and drug design. Given the crystallographic changes in structure observed in UPPS we carefully selected structures to use in docking studies, validating our models before making predictions. We applied new active-site volume estimating tools to select an individual frame of the trajectory that was most predictive in terms of correlating docking scores with experimental pIC_{50}s for the bisphosphonate class of UPPS inhibitors. These docking studies and careful selection of receptor structure continue to be key in the virtual identification of inhibitors of UPPS. Our mostly unpublished work docking compounds into UPPS and FPPS combined with the evaluation of their affinities via spectrophotometric and radiometric assays has directly led to the discovery of 9 compounds with varied scaffolds and R-groups down into the nM range. The computational work on FPPS with Jacob Durrant and Steffen Lindert and our UPPS work has inspired the synthesis and evaluation of nearly 100 verified inhibitors of UPPS some of which are efficacious in the low nM range and effectively kill *E. coli* and *S. aureus* cells at ~250ng/ml. Some of these compounds are now being tested in mice models of infection. Through a structure based approach we have engineered high affinity inhibitors of UPPS, and continue to improve their cellular and physiochemical properties. Bacterial drug candidates tend to have poor physiochemical properties, and many of the initial hit compounds were afflicted by this problem. In recent work we have engineered the charged properties out of some of the molecules as well as reducing the

hydrophobicitiy of the molecules significantly. We have still retained nM affinity in some cases. Our computational work has provided structure-based guidance, prediction of ligand protein complexes in a flexible target, and direct prediction of new inhibitors. The lessons learned early on about the target flexibility and the importance of picking the best conformation to make structure based predictions have been carried on as a theme in this work. The success we have had with UPPS inhibitors has led to simulation and structure based drug design for *M. Tuberculosis* decaprenyl diphosphate synthase, cis-FPPS, and rv3378c (a TB virulence factor). Many of the known UPPS inhibitors also work on these enzymes necessary for TB survival or virulence due to their related structures, so these are excellent drug discovery projects to be done in parallel.

Due to the importance of target and ligand flexibility in molecular recognition we were compelled to produce a new form of accelerated MD in which we could better recover statistics from our simulations. We demonstrated that this new method of aMD could sample simple systems on longer timescales than cMD. With a single replica this aMD method could recover the free energy surface of alanine dipeptide in 1-10% of the simulation time needed to produce a similar free energy surface using cMD, because the time needed to cross energy barriers was much less. Given these positive results we used the method to calculate the free energy difference between vancomycin binding the native glycopeptides substrate and a mutated glycopeptide with 1000 fold less affinity. This calculation helped demonstrate that we could couple our enhanced sampling method with free energy calculations, and achieve more rapid convergence of the calculated free energy value. This method is now being tested in a replica exchange framework.

Accelerated MD is becoming more popular and will be officially released using the original boost equation in the AMBER 12 MD package.

The work described here is of general interest to those involved in structure based drug design because it highlights the importance of receptor flexibility in the calculation of biomolecular recognition. Additionally we have stressed the importance of mixing computational methods with experimental methods to produce a working model whose predictions can be verified at the bench. Many of the inhibitors discovered will be of use as reagents in the laboratory, and may be starting points for the design of drugs. Finally, we addressed the issue of incomplete sampling in MD simulations and moved accelerated MD methods forward in a way that is both theoretically sound, and computationally efficient.

References

1. Jorgensen, W. L. (2004) The many roles of computation in drug discovery, *Science 303*, 1813–1818.

2. Zoete, V., Grosdidier, A., and Michielin, O. (2009) Docking, virtual high throughput screening and in silico fragment-based drug design, *Journal of Cellular and Molecular Medicine*, *13*, 238–248.

3. Armstrong, S. C. (2004) Protein kinase activation and myocardial ischemia/reperfusion injury., *Cardiovasc. Res. 61*, 427–436.

4. Zdychová, J., and Komers, R. (2005) Emerging role of Akt kinase/protein kinase B signaling in pathophysiology of diabetes and its complications., *Physiol Res 54*, 1–16.

5. Rohdich, F., Bacher, A., and Eisenreich, W. (2005) Isoprenoid biosynthetic pathways as anti-infective drug targets., *Biochem. Soc. Trans.*, *33*, 785–791.

6. Wang, L. (2010) Pharmacogenomics: a systems approach., *Wiley Interdiscip Rev Syst Biol Med 2*, 3–22.

7. (2004) New horizons in chemical space., *Nat Rev Drug Discov 3*, 375.

8. McInnes, C. (2007) Virtual screening strategies in drug discovery., *Curr Opin Chem Biol 11*, 494–502.

9. Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications., *Nat Rev Drug Discov 3*, 935–949.

10. Gay, S. C., Sun, L., Maekawa, K., Halpert, J. R., and Stout, C. D. (2009) Crystal structures of cytochrome P450 2B4 in complex with the inhibitor 1-biphenyl-4-methyl-1H-imidazole: ligand-induced structural response through alpha-helical repositioning., *Biochemistry 48*, 4762–4771.

11. Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution., *J. Biol. Chem. 279*, 38091–38094.

12. Ekroos, M., and Sjögren, T. (2006) Structural basis for ligand promiscuity in cytochrome P450 3A4., *P Natl Acad Sci 103*, 13682–13687.

13. Sevrioukova, I. F., and Poulos, T. L. (2010) Structure and mechanism of the complex between cytochrome P4503A4 and ritonavir., *P Natl Acad Sci 107*, 18422–18427.

14. Monod, J., Wyman, J., and Changeux, J. P. (1965) On the nature of allosteric transitions: a plausible model., *J Mol Biol 12*, 88–118.

15. Ma, B., Kumar, S., Tsai, C. J., and Nussinov, R. (1999) Folding funnels and binding mechanisms., *Protein Eng. 12*, 713–720.

16. Koshland, D. E. (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis., *P Natl Acad Sci 44*, 98–104.

17. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., and Farid, R. (2006) Novel procedure for modeling ligand/receptor induced fit effects., *J. Med. Chem. 49*, 534–553.

18. Simons, K. T., Kooperberg, C., Huang, E., and Baker, D. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions, *J Mol Biol 268*, 209–225.

19. Meiler, J., and Baker, D. (2006) ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility, *Proteins 65*, 538–548.

20. Davis, I. W., and Baker, D. (2009) RosettaLigand docking with full ligand and receptor flexibility, *J Mol Biol 385*, 381–392.

21. Davis, I. W., Raha, K., Head, M. S., and Baker, D. (2009) Blind docking of pharmaceutically relevant compounds using RosettaLigand, *Protein Sci 18*, 1998–2002.

22. Lemmon, G., and Meiler, J. (2012) Rosetta Ligand docking with flexible XML protocols, *Methods Mol. Biol. 819*, 143–155.

23. Lin, J.-H., Perryman, A. L., Schames, J. R., and McCammon, J. A. (2002) Computational drug design accommodating receptor flexibility: the relaxed complex scheme., *J Am Chem Soc 124*, 5632–5633.

24. Lin, J.-H., Perryman, A. L., Schames, J. R., and McCammon, J. A. (2003) The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme., *Biopolymers 68*, 47–62.

25. Amaro, R. E., Baron, R., and McCammon, J. A. (2008) An improved relaxed complex scheme for receptor flexibility in computer-aided drug design., *J. Comput. Aided Mol. Des. 22*, 693–705.

26. Schames, J. R., Henchman, R. H., Siegel, J. S., Sotriffer, C. A., Ni, H., and McCammon, J. A. (2004) Discovery of a novel binding trench in HIV integrase., *J. Med. Chem. 47*, 1879–1881.

27. Di Santo, R., Costi, R., Roux, A., Artico, M., Lavecchia, A., Marinelli, L., Novellino, E., Palmisano, L., Andreotti, M., Amici, R., Galluzzo, C. M., Nencioni, L., Palamara, A. T., Pommier, Y., and Marchand, C. (2006) Novel bifunctional quinolonyl diketo acid derivatives as HIV-1 integrase inhibitors: design, synthesis, biological activities, and mechanism of action., *J. Med. Chem. 49*, 1939–1945.

28. Durrant, J. D., Urbaniak, M. D., Ferguson, M. A. J., and McCammon, J. A. (2010) Computer-aided identification of Trypanosoma brucei uridine diphosphate galactose 4'-epimerase inhibitors: toward the development of novel therapies for African sleeping sickness., *J. Med. Chem. 53*, 5025–5032.

29. Sinko, W., de Oliveira, C., Williams, S., Van Wynsberghe, A., Durrant, J. D., Cao, R., Oldfield, E., and McCammon, J. A. (2011) Applying molecular dynamics simulations to identify rarely sampled ligand-bound conformational states of undecaprenyl pyrophosphate synthase, an antibacterial target., *Chem Biol Drug Des 77*, 412–420.

30. Cheng, L. S., Amaro, R. E., Xu, D., Li, W. W., Arzberger, P. W., and McCammon, J. A. (2008) Ensemble-based virtual screening reveals potential

novel antiviral compounds for avian influenza neuraminidase., *J. Med. Chem. 51*, 3878–3894.

31. Nichols, S. E., Baron, R., Ivetac, A., and McCammon, J. A. (2011) Predictive power of molecular dynamics receptor structures in virtual screening., *J Chem Inf Model 51*, 1439–1446.

32. Osguthorpe, D. J., Sherman, W., and Hagler, A. T. (2012) Exploring Protein Flexibility: Incorporating Structural Ensembles From Crystal Structures and Simulation into Virtual Screening Protocols., *J Phys Chem B*.

33. Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., Young, C., Batson, B., Bowers, K. J., Chao, J. C., Eastwood, M. P., Gagliardo, J., Grossman, J. P., Ho, C. R., Ierardi, D. J., Kolossvary, I., Klepeis, J. L., Layman, T., Mcleavey, C., Moraes, M. A., Mueller, R., Priest, E. C., Shan, Y., Spengler, J., Theobald, M., Towles, B., and Wang, S. C. (2008) Anton, a special-purpose machine for molecular dynamics simulation, *Commun ACM 51*, 91–97.

34. Stone, J. E., Phillips, J. C., Freddolino, P. L., Hardy, D. J., Trabuco, L. G., and Schulten, K. (2007) Accelerating molecular modeling applications with graphics processors., *J Comput Chem 28*, 2618–2640.

35. Stone, J. E., Hardy, D. J., Ufimtsev, I. S., and Schulten, K. (2010) GPU-accelerated molecular modeling coming of age., *J. Mol. Graph. Model. 29*, 116–125.

36. Shan, Y., Kim, E. T., Eastwood, M. P., Dror, R. O., Seeliger, M. A., and Shaw, D. E. (2011) How does a drug molecule find its target binding site?, *J Am Chem Soc 133*, 9181–9183.

37. Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. (2011) How fast-folding proteins fold., *Science 334*, 517–520.

38. Sugita, Y., and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding, *Chem Phys Lett 314*, 141–151.

39. Fajer, M., Hamelberg, D., and McCammon, J. A. (2008) Replica-Exchange Accelerated Molecular Dynamics (REXAMD) Applied to Thermodynamic Integration., *J Chem Theory Comput 4*, 1565–1569.

40. Torrie, G. M., and Valleau, J. P. (1974) Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid, *Chem Phys Lett 28*, 578–581.

41. Torrie, G. M., and Valleau, J. P. (1977) Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling, *J Comput Phys 23*, 187–199.

42. Patey, G. N., and Valleau, J. P. (1975) Monte-Carlo Method for Obtaining Interionic Potential of Mean Force in Ionic Solution, *J Chem Phys 63*, 2334–2339.

43. Laio, A., and Parrinello, M. (2002) Escaping free-energy minima, *P Natl Acad Sci 99*, 12562–12566.

44. Laio, A., and Gervasio, F. L. (2008) Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science, *Rep Prog Phys 71*.

45. Kastner, J. (2011) Umbrella sampling, *WIRES Comput Mol Sci 1*, 932–942.

46. Sugita, Y., Kitao, A., and Okamoto, Y. (2000) Multidimensional replica-exchange method for free-energy calculations, *J Chem Phys 113*, 6042–6051.

47. Rosta, E., Woodcock, H. L., Brooks, B. R., and Hummer, G. (2009) Artificial Reaction Coordinate "Tunneling" in Free-Energy Calculations: The Catalytic Reaction of RNase H, *J Comput Chem 30*, 1634–1641.

48. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. (1992) The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method, *J Comput Chem 13*, 1011–1021.

49. Souaille, M., and Roux, B. (2001) Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations, *Comput Phys Commun 135*, 40–57.

50. Barducci, A., Bonomi, M., and Parrinello, M. (2011) Metadynamics, *WIRES Comput Mol Sci 1*, 826–843.

51. Branduardi, D., Gervasio, F. L., Cavalli, A., Recanatini, M., and Parrinello, M. (2005) The role of the peripheral anionic site and cation-pi interactions in the ligand penetration of the human AChE gorge, *J Am Chem Soc 127*, 9147–9155.

52. Gervasio, F. L., Laio, A., and Parrinello, M. (2005) Flexible docking in solution using metadynamics, *J Am Chem Soc 127*, 2600–2607.

53. Limongelli, V., Bonomi, M., Marinelli, L., Gervasio, F. L., Cavalli, A., Novellino, E., and Parrinello, M. (2010) Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition, *P Natl Acad Sci 107*, 5411–5416.

54. Provasi, D., Bortolato, A., and Filizola, M. (2009) Exploring molecular mechanisms of ligand recognition by opioid receptors with metadynamics, *Biochemistry 48*, 10020–10029.

55. Masetti, M., Cavalli, A., Recanatini, M., and Gervasio, F. L. (2009) Exploring complex protein-ligand recognition mechanisms with coarse metadynamics, *J Phys Chem B 113*, 4807–4816.

56. Pietrucci, F., Marinelli, F., Carloni, P., and Laio, A. (2009) Substrate binding mechanism of HIV-1 protease from explicit-solvent atomistic simulations, *J Am Chem Soc 131*, 11811–11818.

57. Hamelberg, D., Mongan, J., and McCammon, J. A. (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules, *J Chem Phys*, American Institute of Physics *120*, 11919–11929.

58. Voter, A. F. (1997) Hyperdynamics: Accelerated molecular dynamics of infrequent events, *Phys Rev Lett 78*, 3908–3911.

59. de Oliveira, C. A., Hamelberg, D., and McCammon, J. A. (2006) On the application of accelerated molecular dynamics to liquid water simulations, *J Phys Chem B 110*, 22695–22701.

60. Hamelberg, D., de Oliveira, C. A. F., and McCammon, J. A. (2007) Sampling of slow diffusive conformational transitions with accelerated molecular dynamics., *J Chem Phys 127*, 155102.

61. Markwick, P. R., Cervantes, C. F., Abel, B. L., Komives, E. A., Blackledge, M., and McCammon, J. A. (2010) Enhanced conformational space sampling improves the prediction of chemical shifts in proteins, *J Am Chem Soc 132*, 1220–1221.

62. Grant, B. J., McCammon, J. A., and Gorfe, A. A. (2010) Conformational selection in G-proteins: lessons from Ras and Rho., *Biophys J 99*, L87–9.

63. Shen, T., and Hamelberg, D. (2008) A statistical analysis of the precision of reweighting-based simulations, *J Chem Phys 129*, 034103.

64. Wereszczynski, J., and McCammon, J. A. (2012) Nucleotide-dependent mechanism of Get3 as elucidated from free energy calculations., *Proc Natl Acad Sci U S A*.

65. Lindert, S., Kekenes-Huskey, P. M., Huber, G., Pierce, L., and McCammon, J. A. (2012) Dynamics and Calcium Association to the N-Terminal Regulatory Domain of Human Cardiac Troponin C: A Multiscale Computational Study., *J Phys Chem B*.

66. Simonson, T., Archontis, G., and Karplus, M. (2002) Free energy simulations come of age: protein-ligand recognition, *Accounts of Chemical Research 35*, 430–437.

67. Fajer, M., Swift, R. V., and McCammon, J. A. (2009) Using multistate free energy techniques to improve the efficiency of replica exchange accelerated molecular dynamics., *J Comput Chem 30*, 1719–1725.

68. de Oliveira, C. A., Hamelberg, D., and McCammon, J. A. (2008) Coupling Accelerated Molecular Dynamics Methods with Thermodynamic Integration Simulations, *J Chem Theory Comput 4*, 1516–1525.

69. Sinko, W., de Oliveira, C. A. F., Pierce, L. C. T., and McCammon, J. A. (2012) Protecting High Energy Barriers: A New Equation to Regulate Boost Energy in Accelerated Molecular Dynamics Simulations, *J Chem Theory Comput*, *8*, 17–23.

70. Wereszczynski, J., and McCammon, J. A. (2012) Using Selectively Applied Accelerated Molecular Dynamics to Enhance Free Energy Calculations, *J Chem Theory Comput, 6*, 3285–3292.

71. Cantley, L. C. (2002) The Phosphoinositide 3-Kinase Pathway, *Science 296*, 1655–1657.

72. Newton, A. C. (2009) Lipid activation of protein kinases, *J. Lipid Res. 50*, S266–271.

73. Manning, B. D., and Cantley, L. C. (2007) AKT/PKB Signaling: Navigating Downstream, *Cell 129*, 1261–1274.

74. Maehama, T., and Dixon, J. E. (1999) PTEN: a tumour suppressor that functions as a phospholipid phosphatase, *Trends in cell biology 9*, 125–128.

75. Trotman, L. C., Alimonti, A., Scaglioni, P. P., Koutcher, J. A., Cordon-Cardo, C., and Pandolfi, P. P. (2006) Identification of a tumour suppressor network opposing nuclear Akt function, *Nature 441*, 523–527.

76. Gao, T., Furnari, F., and Newton, A. C. (2005) PHLPP: a phosphatase that directly dephosphorylates Akt, promotes apoptosis, and suppresses tumor growth, *Molecular cell*, *18*, 13–24.

77. Brognard, J., Sierecki, E., Gao, T., and Newton, A. C. (2007) PHLPP and a second isoform, PHLPP2, differentially attenuate the amplitude of Akt signaling by regulating distinct Akt isoforms, *Molecular cell*, *25*, 917–931.

78. Gao, T., Brognard, J., and Newton, A. C. (2008) The phosphatase PHLPP controls the cellular levels of protein kinase C, *J. Biol. Chem.*, *283*, 6300–6311.

79. Brognard, J., and Newton, A. C. (2008) PHLiPPing the switch on Akt and protein kinase C signaling, *Trends in endocrinology and metabolism: TEM*, *19*, 223–230.

80. Schweighofer, A., Hirt, H., and Meskiene, I. (2004) Plant PP2C phosphatases: emerging functions in stress signaling, *Trends in Plant Science 9*, 236–243.

81. Lammers, T., and Lavi, S. (2007) Role of Type 2C Protein Phosphatases in Growth Regulation and in Cellular Stress Signaling, *Critical Reviews in Biochemistry and Molecular Biology 42*, 437–461.

82. Mumby, M. C., and Walter, G. (1993) Protein serine/threonine phosphatases: structure, regulation, and functions in cell growth, *Physiol. Rev 73*, 673–699.

83. Yamaguchi, H., Durell, S. R., Feng, H., Bai, Y., Anderson, C. W., and Appella, E. (2006) Development of a Substrate-Based Cyclic Phosphopeptide Inhibitor of Protein Phosphatase 2CÎ´, Wip1â€, *Biochemistry*, *45*, 13193–13202.

84. Rogers, J. P., Beuscher, A. E. T., Flajolet, M., McAvoy, T., Nairn, A. C., Olson, A. J., and Greengard, P. (2006) Discovery of protein phosphatase 2C inhibitors by virtual screening, *J. Med. Chem.*, *49*, 1658–1667.

85. Bain, J., Plater, L., Elliott, M., Shpiro, N. C., Hastie, J., McLauchlan, H., Klevernic, I., Arthur, J. S. C., Alessi, D. R., and Cohen, P. (2007) The selectivity of protein kinase inhibitors: a further update, *Biochemical Journal 408*, 297–315.

86. Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., Chan, K. W., Ciceri, P., Davis, M. I., Edeen, P. T., Faraoni, R.,

Floyd, M., Hunt, J. P., Lockhart, D. J., Milanov, Z. V., Morrison, M. J., Pallares, G., Patel, H. K., Pritchard, S., Wodicka, L. M., and Zarrinkar, P. P. (2008) A quantitative analysis of kinase inhibitor selectivity, *Nat Biotech*, Nature Publishing Group *26*, 127–132.

87. Elghazi, L., and Bernal-Mizrachi, E. (2009) Akt and PTEN: [beta]-cell mass and pancreas plasticity, *Trends in Endocrinology & Metabolism 20*, 243–251.

88. Elghazi, L., Rachdi, L., Weiss, A. J., Cras-Méneur, C., and Bernal-Mizrachi, E. (2007) Regulation of beta-cell mass and function by the Akt/protein kinase B signalling pathway, *Diabetes, Obesity and Metabolism 9*, 147–157.

89. Tuttle, R. L., Gill, N. S., Pugh, W., Lee, J.-P., Koeberlein, B., Furth, E. E., Polonsky, K. S., Naji, A., and Birnbaum, M. J. (2001) Regulation of pancreatic [beta]-cell growth and survival by the serine/threonine protein kinase Akt1/PKB[alpha], *Nat Med 7*, 1133–1137.

90. Bernal-Mizrachi, E., Wen, W., Stahlhut, S., Welling, C. M., and Permutt, M. A. (2001) Islet beta cell expression of constitutively active Akt1/PKBalpha induces striking hypertrophy, hyperplasia, and hyperinsulinemia, *The Journal of Clinical Investigation*, The American Society for Clinical Investigation *108*, 1631–1638.

91. Bernal-Mizrachi, E., Fatrai, S., Johnson, J. D., Ohsugi, M., Otani, K., Han, Z., Polonsky, K. S., and Permutt, M. A. (2004) Defective insulin secretion and increased susceptibility to experimental diabetes are induced by reduced Akt activity in pancreatic islet Î² cells, *The Journal of Clinical Investigation*, The American Society for Clinical Investigation *114*, 928–936.

92. Bayascas, J. R., Wullschleger, S., Sakamoto, K., Garcia-Martinez, J. M., Clacher, C., Komander, D., van Aalten, D. M. F., Boini, K. M., Lang, F., Lipina, C., Logie, L., Sutherland, C., Chudek, J. A., van Diepen, J. A., Voshol, P. J., Lucocq, J. M., and Alessi, D. R. (2008) Mutation of the PDK1 PH Domain Inhibits Protein Kinase B/Akt, Leading to Small Size and Insulin Resistance, *Mol. Cell. Biol. 28*, 3258–3272.

93. Garcia-Ocana, A., Takane, K. K., Reddy, V. T., Lopez-Talavera, J.-C., Vasavada, R. C., and Stewart, A. F. (2003) Adenovirus-mediated Hepatocyte Growth Factor Expression in Mouse Islets Improves Pancreatic Islet Transplant Performance and Reduces Beta Cell Death, *Journal of Biological Chemistry 278*, 343–351.

94. Contreras, J. L., Smyth, C. A., Bilbao, G., Young, C. J., Anthony Thompson, J., and Eckhoff, D. E. Simvastatin induces activation of the serine-threonine protein kinase AKT and increases survival of isolated human pancreatic islets. [Miscellaneous Article], Transplantation October 27, 2002;74(8):1063-1069.

95. Hausenloy, D. J., Tsang, A., and Yellon, D. M. (2005) The Reperfusion Injury Salvage Kinase Pathway: A Common Target for Both Ischemic Preconditioning and Postconditioning, *Trends in Cardiovascular Medicine 15*, 69–75.

96. Hausenloy, D. J., and Yellon, D. M. (2006) Survival kinases in ischemic preconditioning and postconditioning, *Cardiovasc. Res. 70*, 240–253.

97. Mullonkal, C. J., and Toledo-Pereyra, L. H. (2007) Akt in Ischemia and Reperfusion, *Journal of Investigative Surgery 20*, 195–203.

98. Zhu, M., Feng, J., Lucchinetti, E., Fischer, G., Xu, L., Pedrazzini, T., Schaub, M. C., and Zaugg, M. (2006) Ischemic postconditioning protects remodeled myocardium via the PI3K-PKB/Akt reperfusion injury salvage kinase pathway, *Cardiovasc. Res. 72*, 152–162.

99. Dorn, G. W., Souroujon, M. C., Liron, T., Chen, C.-H., Gray, M. O., Zhou, H. Z., Csukai, M., Wu, G., Lorenz, J. N., and Mochly-Rosen, D. (1999) Sustained in vivo cardiac protection by a rationally designed peptide that causes É› protein kinase C translocation, *Proc Natl Acad Sci U S A 96*, 12798–12803.

100. Churchill, E. N., and Mochly-rosen, D. (2007) The roles of PKC delta and epsilon isoenzymes in the regulation of myocardial ischaemia/reperfusion injury, *Biochem. Soc. Trans. 035*, 1040–1042.

101. Budas, G. R., Churchill, E. N., and Mochly-Rosen, D. (2007) Cardioprotective mechanisms of PKC isozyme-selective activators and inhibitors in the treatment of ischemia-reperfusion injury, *Pharmacological Research 55*, 523–536.

102. Churchill, E., Budas, G., Vallentin, A., Koyanagi, T., and Mochly-Rosen, D. (2008) PKC Isozymes in Chronic Cardiac Disease: Possible Therapeutic Targets?, *Annual Review of Pharmacology and Toxicology 48*, 569–599.

103. Downey, J., Davis, A., and Cohen, M. (2007) Signaling pathways in ischemic preconditioning, *Heart Failure Reviews 12*, 181–188.

104. Xue-Jiao, Z., Zi-Bo, X., An-Li, T., Hong, M., Yue-Dong, M., Jing-Guo, W., and Yu-Gang, D. (2009) Rosiglitazone -induced myocardial protection against ischaemia/reperfusion injury is mediated through a PI3K/Akt-dependent pathway, *Clinical and Experimental Pharmacology and Physiology 9999*.

105. Przyklenk, K., Maynard, M., and Whittaker, P. (2006) Reduction of infarct size with D-myo-inositol trisphosphate: role of PI3-kinase and mitochondrial KATP channels, *Am J Physiol Heart Circ Physiol 290*, H830–836.

106. Zhang, J.-H., Chung, T. D. Y., and Oldenburg, K. R. (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays, *J Biomol Screen 4*, 67–73.

107. Park, H., Bahn, Y. J., Jung, S.-K., Jeong, D. G., Lee, S.-H., Seo, I., Yoon, T.-S., Kim, S. J., and Ryu, S. E. (2008) Discovery of Novel Cdc25 Phosphatase Inhibitors with Micromolar Activity Based on the Structure-Based Virtual Screening, *J. Med. Chem.*, *51*, 5533–5541.

108. Vidal, D., Blobel, J., Pérez, Y., Thormann, M., and Pons, M. (2007) Structure-based discovery of new small molecule inhibitors of low molecular weight protein tyrosine phosphatase, *European Journal of Medicinal Chemistry 42*, 1102–1108.

109. Yu, W.-M., Guvench, O., MacKerell, A. D., and Qu, C.-K. (2008) Identification of Small Molecular Weight Inhibitors of Src Homology 2 Domain-Containing Tyrosine Phosphatase 2 (SHP-2) via in Silico Database Screening Combined with Experimental Assay, *J. Med. Chem.*, *51*, 7396–7404.

110. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem.*, *47*, 1739–1749.

111. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, *J. Med. Chem.*, *47*, 1750–1759.

112. Kontoyianni, M., McClellan, L. M., and Sokol, G. S. (2003) Evaluation of Docking Performance: Comparative Data on Docking Algorithms, *J. Med. Chem.*, *47*, 558–565.

113. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, *J. Med. Chem.*, *47*, 1739–1749.

114. Das, A. K., Helps, N. R., Cohen, P. T., and Barford, D. (1996) Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 A resolution, *The EMBO journal*, *15*, 6798–6809.

115. Wehenkel, A., Bellinzoni, M., Schaeffer, F., Villarino, A., and Alzari, P. M. (2007) Structural and Binding Studies of the Three-metal Center in Two Mycobacterial PPM Ser/Thr Protein Phosphatases, *J Mol Biol 374*, 890–898.

116. Pullen, K. E., Ng, H.-L., Sung, P.-Y., Good, M. C., Smith, S. M., and Alber, T. (2004) An Alternate Conformation and a Third Metal in PstP/Ppp, the M. tuberculosis PP2C-Family Ser/Thr Protein Phosphatase, *Structure 12*, 1947–1954.

117. Schlicker, C., Fokina, O., Kloft, N., Grüne, T., Becker, S., Sheldrick, G. M., and Forchhammer, K. (2008) Structural Analysis of the PP2C Phosphatase tPphA from Thermosynechococcus elongatus: A Flexible Flap Subdomain Controls Access to the Catalytic Site, *J Mol Biol 376*, 570–581.

118. Li, X., Liu, J., and Gao, T. (2009) {beta}-TrCP-Mediated Ubiquitination and Degradation of PHLPP1 Are Negatively Regulated by Akt, *Mol. Cell. Biol. 29*, 6192–6205.

119. Warfel, N., Niederst, M., and Newton, A. , *submitted*.

120. Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Advanced Drug Delivery Reviews 46*, 3–26.

121. Fjeld, C. C., and Denu, J. M. (1999) Kinetic Analysis of Human Serine/Threonine Protein Phosphatase 2CÎ±, *Journal of Biological Chemistry 274*, 20336–20343.

122. Baqi, Y., and Müller, C. E. (2007) Rapid and efficient microwave-assisted copper(0)-catalyzed ullmann coupling reaction: general access to anilinoanthraquinone derivatives., *Org. Lett. 9*, 1271–1274.

123. Fischbach, M. A., and Walsh, C. T. (2009) Antibiotics for emerging pathogens., *Science 325*, 1089–1093.

124. Chambers, H. F., and Deleo, F. R. (2009) Waves of resistance: Staphylococcus aureus in the antibiotic era., *Nat. Rev. Microbiol.*, *7*, 629–641.

125. Fujihashi, M., Zhang, Y. W., Higuchi, Y., Li, X. Y., Koyama, T., and Miki, K. (2001) Crystal structure of cis-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase, *P Natl Acad Sci*, *98*, 4337–4342.

126. Ko, T. P., Chen, Y. K., Robinson, H., Tsai, P. C., Gao, Y. G., Chen, A. P., Wang, A. H., and Liang, P. H. (2001) Mechanism of product chain length determination

and the role of a flexible loop in Escherichia coli undecaprenyl-pyrophosphate synthase catalysis, *J. Biol. Chem.*, *276*, 47474–47482.

127. Chang, S. Y., Ko, T. P., Liang, P. H., and Wang, A. H. (2003) Catalytic mechanism revealed by the crystal structure of undecaprenyl pyrophosphate synthase in complex with sulfate, magnesium, and triton, *J. Biol. Chem.*, *278*, 29298–29307.

128. Guo, R. T., Ko, T. P., Chen, A. P., Kuo, C. J., Wang, A. H., and Liang, P. H. (2005) Crystal structures of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl pyrophosphate, and farnesyl thiopyrophosphate: roles of the metal ion and conserved residues in catalysis, *J. Biol. Chem.*, *280*, 20762–20774.

129. Guo, R. T., Cao, R., Liang, P. H., Ko, T. P., Chang, T. H., Hudock, M. P., Jeng, W. Y., Chen, C. K., Zhang, Y., Song, Y., Kuo, C. J., Yin, F., Oldfield, E., and Wang, A. H. (2007) Bisphosphonates target multiple sites in both cis- and trans-prenyltransferases, *P Natl Acad Sci*, *104*, 10022–10027.

130. Li, H., Huang, J., Jiang, X., Seefeld, M., McQueney, M., and Macarron, R. (2003) The effect of triton concentration on the activity of undecaprenyl pyrophosphate synthase inhibitors, *Journal of Biomolecular Screening : the official journal of the Society for Biomolecular Screening*, *8*, 712–715.

131. Payne, D. J., Gwynn, M. N., Holmes, D. J., and Pompliano, D. L. (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery., *Nat Rev Drug Discov*, *6*, 29–40.

132. Peukert, S., Sun, Y., Zhang, R., Hurley, B., Sabio, M., Shen, X., Gray, C., Dzink-Fox, J., Tao, J., Cebula, R., and Wattanasin, S. (2008) Design and structure-activity relationships of potent and selective inhibitors of undecaprenyl pyrophosphate synthase (UPPS): tetramic, tetronic acids and dihydropyridin-2-ones., *Bioorg. Med. Chem. Lett.*, *18*, 1840–1844.

133. Lee, L. V., Granda, B., Dean, K., Tao, J., Liu, E., Zhang, R., Peukert, S., Wattanasin, S., Xie, X., Ryder, N. S., Tommasi, R., and Deng, G. (2010) Biophysical investigation of the mode of inhibition of tetramic acids, the allosteric inhibitors of undecaprenyl pyrophosphate synthase., *Biochemistry 49*, 5366–5376.

134. Kuo, C. J., Guo, R. T., Lu, I. L., Liu, H. G., Wu, S. Y., Ko, T. P., Wang, A. H., and Liang, P. H. (2008) Structure-based inhibitors exhibit differential activities against Helicobacter pylori and Escherichia coli undecaprenyl pyrophosphate synthases, *Journal of biomedicine & biotechnology*, *2008*, 841312.

135. Fiser, A., Do, R. K., and Sali, A. (2000) Modeling of loops in protein structures., *Protein Sci 9*, 1753–1773.

136. Fiser, A., and Sali, A. (2003) ModLoop: automated modeling of loops in protein structures., *Bioinformatics 19*, 2500–2501.

137. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations., *Nucleic Acids Res. 32*, W665–7.

138. Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pKa values., *Proteins 61*, 704–721.

139. Chang, S. Y., Chen, Y. K., Wang, A. H., and Liang, P. H. (2003) Identification of the active conformation and the importance of length of the flexible loop 72-83 in regulating the conformational change of undecaprenyl pyrophosphate synthase, *Biochemistry*, *42*, 14452–14459.

140. D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossvai, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2008) AMBER 10, *University of California, San Francisco*.

141. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *J Chem Phys, 81*, 3684–3690.

142. Durrant, J. D., de Oliveira, C. A. F., and McCammon, J. A. (2011) POVME: an algorithm for measuring binding-pocket volumes., *J. Mol. Graph. Model. 29*, 773–776.

143. Pan, J. J., Chiou, S. T., and Liang, P. H. (2000) Product distribution and pre-steady-state kinetic analysis of Escherichia coli undecaprenyl pyrophosphate synthase reaction., *Biochemistry 39*, 10936–10942.

144. Otwinowski, Z. (1997) Processing of X-ray diffraction data, *Methods in enzymology*.

145. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method., *Acta Crystallogr. D Biol. Crystallogr. 53*, 240–255.

146. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite., *Acta Crystallogr. D Biol. Crystallogr. 59*, 1131–1137.

147. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics., *Acta Crystallogr. D Biol. Crystallogr. 60*, 2126–2132.

148. Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion., *Acta Crystallogr. D Biol. Crystallogr. 62*, 439–450.

149. GLIDE 4.5, Schrödinger, New York, NY.

150. van der Spoel D, E, L., B, H., van Buuren AR, E, A., PJ, M., DP, T., ALTM, S., KA, F., van Drunen R, and HJC, B. Gromacs User Manual Version 3.3.

151. Volkman, B. F., Lipson, D., Wemmer, D. E., and Kern, D. (2001) Two-state allosteric behavior in a single-domain signaling protein., *Science 291*, 2429–2433.

152. Amaro, R. E., and Li, W. W. (2009) Emerging Methods for Ensemble-Based Virtual Screening, *Current topics in medicinal chemistry*, 10 (1), 3-13.

153. Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules, *Nat Struct Biol*, *9*, 646–652.

154. Adcock, S. A., and McCammon, J. A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins, *Chem Rev 106*, 1589–1615.

155. De Oliveira, C. A. F., Hamelberg, D., and McCammon, J. A. (2007) Estimating kinetic rates from accelerated molecular dynamics simulations: alanine dipeptide in explicit solvent as a case study, *J Chem Phys 127*, 175105.

156. Hamelberg, D., and McCammon, J. A. (2005) Fast peptidyl cis-trans isomerization within the flexible Gly-rich flaps of HIV-1 protease, *J Am Chem Soc 127*, 13778–13779.

157. Hamelberg, D., Shen, T., and Andrew McCammon, J. (2005) Relating kinetic rates and local energetic roughness by accelerated molecular-dynamics simulations, *J Chem Phys 122*, 241103.

158. Hamelberg, D., Shen, T., and McCammon, J. A. (2005) Phosphorylation Effects on cis/trans Isomerization and the Backbone Conformation of Serine,àíProline Motifs:,Äâ Accelerated Molecular Dynamics Analysis, *J Am Chem Soc*, *127*, 1969–1974.

159. Markwick, P. R. L., Bouvignies, G., and Blackledge, M. (2007) Exploring Multiple Timescale Motions in Protein GB3 Using Accelerated Molecular Dynamics and NMR Spectroscopy, *J Am Chem Soc*, *129*, 4724–4730.

160. Yang, W., Li, H., Min, D., and Liu, Y. (2007) Essential energy space random walk via energy space metadynamics method to accelerate molecular dynamics simulations, *J Chem Phys 127*.

161. Yang, W., and Zheng, L. Q. (2008) Essential energy space random walks to accelerate molecular dynamics simulations: Convergence improvements via an adaptive-length self-healing strategy, *J Chem Phys 129*.

162. Yonezawa, Y., Fukuda, I., Kamiya, N., Shimoyama, H., and Nakamura, H. (2011) Free Energy Landscapes of Alanine Dipeptide in Explicit Water Reproduced by the Force-Switching Wolf Method, *J Chem Theory Comput 7*, 1484–1493.

163. Ng, K. M., Solayappan, M., and Poh, K. L. (2011) Global energy minimization of alanine dipeptide via barrier function methods, *Comput Biol Chem 35*, 19–23.

164. Ferguson, A. F. A. L., Panagiotopoulos, A. Z., Debenedetti, P. G., and Kevrekidis, I. G. (2011) Integrating diffusion maps with umbrella sampling: Application to alanine dipeptide, *J Chem Phys 134*.

165. Cruz, V., Ramos, J., and Martinez-Salazar, J. (2011) Water-Mediated Conformations of the Alanine Dipeptide as Revealed by Distributed Umbrella Sampling Simulations, Quantum Mechanics Based Calculations, and Experimental Data, *J Phys Chem B 115*, 4880–4886.

166. Vondrasek, J., and Vymetal, J. (2010) Metadynamics As a Tool for Mapping the Conformational and Free-Energy Space of Peptides - The Alanine Dipeptide Case Study, *J Phys Chem B 114*, 5632–5642.

167. Ishizuka, R., Huber, G. A., and McCammon, J. A. (2010) Solvation Effect on the Conformations of Alanine Dipeptide: Integral Equation Approach, *J Phys Chem Lett 1*, 2279–2283.

168. Adams, J. P., and Smith, D. A. (1993) Amber and Opls Studies of the Alanine Dipeptide Using the Gb Sa Solvation Method, *Abstr Pap Am Chem S 206*, 42–COMP.

169. Hénin, J., Fiorin, G., Chipot, C., and Klein, M. L. (2009) Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables, *J Chem Theory Comput, 6*, 35–47.

170. Michel, J., Foloppe, N., and Essex, J. W. Rigorous Free Energy Calculations in Structure-Based Drug Design, *Molecular Informatics*, *29*, 570–578.

171. Gilson, M. K., Moghaddam, S., and Inoue, Y. (2009) Host-Guest Complexes with Protein-Ligand-like Affinities: Computational Analysis and Design, *J Am Chem Soc 131*, 4012–4021.

172. Kollman, P. (1993) Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena, *Chem Rev 93*, 2395–2417.

173. Bugg, T. D. H., Wright, G. D., Dutka-Malen, S., Arthur, M., Courvalin, P., and Walsh, C. T. (1991) Molecular basis for vancomycin resistance in Enterococcus faecium BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA, *Biochemistry*, *30*, 10408–10415.

174. McComas, C. C., Crowley, B. M., and Boger, D. L. (2003) Partitioning the loss in vancomycin binding affinity for D-Ala-D-Lac into lost H-bond and repulsive lone pair contributions, *J Am Chem Soc 125*, 9314–9315.

175. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J Chem Phys, 79*, 926–935.

176. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J Comput Phys 23*, 327–341.

177. Mackay, J. P., Gerhard, U., Beauregard, D. A., Williams, D. H., Westwell, M. S., and Searle, M. S. (1994) Glycopeptide Antibiotic Activity and the Possible Role of Dimerization: A Model for Biological Signaling, *J Am Chem Soc*, *116*, 4581–4590.

178. Zacharias, M., Straatsma, T. P., and McCammon, J. A. (1994) Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration, *J Chem Phys*, *100*, 9025–9031.

179. Beutler, T. C., Mark, A. E., van Schaik, R. C., Gerber, P. R., and van Gunsteren, W. F. (1994) Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations, *Chem Phys Lett 222*, 529–539.