In Silico Simulation of DUSP-YIV906 Protein-Ligand Interactions and DUSP3-ERK Protein-Peptide Interactions

by

Shuhan Liu

Bachelor of Science, Huazhong University of Science and Technology, 2018

Submitted to the Graduate Faculty of the

School of Pharmacy in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF PHARMACY

This thesis was presented

by

Shuhan Liu

It was defended on

March 19, 2020

and approved by

Junmei Wang, Associate Professor, Department of Pharmaceutical Sciences

David Koes, Assistant Professor, Department of Computational and Systems Biology

Xiang-Qun Xie, Professor, Department of Pharmaceutical Science

Lirong Wang, Assistant Professor, Department of Pharmaceutical Science

Thesis Advisor: Junmei Wang, Department of Pharmaceutical Sciences

In Silico Simulation of DUSP-YIV906 Protein-Ligand Interactions and DUSP-ERK Protein-Protein Interactions

Shuhan Liu, B.S

University of Pittsburgh, 2020

The dual-specificity phosphatases (DUSPs) are a heterogeneous group of protein enzymes that modulate several critical cellular signaling pathways by dephosphorylating the phosphotyrosine and phosphoserine/phosphothreonine residues within their substrate proteins. One of the best characterized sub-group of DUSPs is the mitogen-activated protein kinase phosphatases (MKPs), which act as the antagonists of associated signaling cascades including the extracellular signal-regulated kinases (ERKs) pathways. Accumulated evidences have highlighted the therapeutic value of DUSPs, as deletion or inhibition of some DUSPs can increase the phosphorylated level of ERKs to cause cancer cell death. In this study, multi-scale molecular modeling simulations were first performed to investigate the mechanism of action of YIV-906, which is an herbal formulation used in cancer treatment targeting DUSP-ERK1/2 pathways. In total, MD simulations and binding free energy calculations were performed for 99 DUSP-ligand complexes. Our results demonstrate that the sulfate moieties and carboxyl moieties of the advantageous ligands, either original herbal chemicals or human metabolites from YIV-906, can occupy the enzymes' catalytic sites, mimicking the endogenous phosphate substrates of DUSPs. With the second aim to improve the accuracy of protein-peptide docking between DUSP3 and a peptide fragment of ERK1/2, a new receptor residue mapping (RR mapping) algorithm was developed to identify hotspots residues on the surface of DUSP3 and improve the peptide docking scoring. By performing all-atom molecular dynamics (MD) simulations with the receptor being

soaked in a water box containing 0.5 moles of capped dipeptide of 20 natural amino acids (AA) plus 3 phosphorylated non-standard AAs, the RR maps with probabilities of AAs interacting with DUSP3's surface residues were obtained. With the interaction probabilities incorporated, the ERK peptide binding models yielded by protein-peptide docking can be re-ranked and generate more accurate predictions. We have demonstrated that multi-scale molecular modeling techniques are able to elucidate molecular mechanisms involving complex molecular systems. Finally, our modeling study provides useful insights into the rational design of high potent anti-cancer drugs targeting DUSPs, and the new RR mapping algorithm is a promising tool that can be universally applied in characterization of protein-protein interactions (PPIs).

Table of Contents

Preface xi

1.0 Introduction 1

1.1 Overview of DUSPs as ERK signaling pathway modulators 1

1.1.1 Physiological and pathological role of DUSP and ERK 1

1.1.2 Structural biology of DUSP vi

1.2 YIV-906 targeting DUSPs as adjuvant in cancer treatment 2

1.3 Molecular simulation techniques in drug discovery and development 5

1.3.1 Molecular docking 5

1.3.1.1 Small molecule-protein docking 5

1.3.1.2 Protein-protein docking 6

1.3.2 Molecular dynamics simulation 7

2.0 Methods 9

2.1 Application of in silico methods in identifying mechanisms of action of YIV-906 9

2.1.1 Structural preparation of ligands and target proteins 9

2.1.2 Molecular docking 10

2.1.3 MD simulation 10

2.1.4 MM/PBSA binding free energy calculation and MM/GBSA free energy decomposition 12

2.2 Development of the RR mapping algorithm to improve the accuracy of protein-protein docking between DUSP3 and ERK peptide 13

2.2.1 Structural preparation of simulation systems 13

2.2.1.1 Generation of conformations for AA analogues and solvate boxes 13

2.2.1.2 Construction of initial systems for protein mapping 13

2.2.2 MD simulations **14**

2.2.3 Data analysis and mapping **15**

2.2.4 Protein-peptide docking **15**

2.2.5 Re-ranking of docking yielded models **16**

3.0 Results and Discussion **18**

3.1 Mechanisms of action of YIV-906 ..... **18**

3.1.1 Unexpected pre-docking results led to the expansion of the compound library to include human metabolites for virtual screening **18**

3.1.2 The high similarity of target proteins allows the implementation of the same docking protocol **20**

3.1.3 Molecular docking yielded ligands of interest **20**

3.1.4 MD simulation and binding free energy prediction **23**

3.1.4.1 RMSD fluctuations of ligands and binding free energy $\Delta GMM/PBSA$ jointly reflect binding stability. **23**

3.1.4.2 Electrostatic energy is a key driving force in DUSP ligand binding. **24**

3.1.4.3 Sulfation and glucuronidation human metabolites of YIV-906 are vital to its DUSP-inhibiting bioactivity. **26**

3.1.4.4 Common ligands identified for DUSP3, DUSP5, and DUSP7. ..... **29**

3.1.4.5 Interactions between identified common ligands and corresponding DUSPs. **31**

3.1.4.6 MM/PBSA is a powerful tool to differentiate compounds with different activities. **34**

3.2 Protein-protein interactions between DUSP and ERK **36**

4.0 Conclusions **41**

Appendix A Supplementary materials for DUSP-YIV906 Protein-Ligand Interactions **43**

Bibliography **62**

List of Tables

Table 1 Basic information of DUSP3, DUSP5, and DUSP7 4

List of Figures

Figure 1 Effect of YIV-906 on the DUSP-ERK1/2 signaling pathway (simplified, schematic) 3

Figure 2 Crystal structures of target proteins in surface representation 4

Figure 3 Establishment of YIV-906 compound library 19

Figure 4 Top-ranked ligands and their docking scores with DUSP3 (A), DUSP5 (B), and DUSP7 (C) 22

Figure 5 Analysis of electrostatic energy ($\Delta Eele + \Delta Gpsol$), van der Waals energy $\Delta Evdw$ and their correlation with binding energy $\Delta GMM/PBSA$ 25

Figure 6 Common substituent groups of advantageous DUSP3 ligands from YIV-906 26

Figure 7 Human metabolites and herbal chemicals from YIV-906 act as phosphate mimic and occupy the catalytic center of DUSP3, which is shown as surface 28

Figure 8 (A) MM-PBSA-reranked ligands of DUSP3, DUSP5 and DUSP7. The axis of this radar is $\Delta GMM/PBSA$ (kcal/mol). 32

Figure 9 Interactions between binding-site residues of DUSP3 and representative ligands. 33

Figure 10 Hotspot residues identification for DUSP3, DUSP5, and DUSP7. 34

Figure 11 The mapping simulation system of DUSP3 mapped with ACE-ALA-NHE 36

Figure 12 Heatmaps for probabilities of each type of AA interacting with each residues of DUSP3 at distance cutoffs of 2, 3, 4, and 5 Å. 38

Figure 13 The correlation between the docking score and RMSD to the crystal structure before and after re-ranking with interaction possibilities 40

Appendix A Supplementary Materials for DUSP-YIV906 Protein-Ligand Interactions

Appendix Figure 1 Sequence alignment for DUSP3, DUSP5, and DUSP7 43

Appendix Figure 2 Validation of the Glide docking protocol by docking the co-crystal ligand, STT, back to the binding pocket of DUSP3 44

Appendix Figure 3 Fluctuational patterns of Root-Mean-Square Deviations along the time course of MD simulations for DUSP3-ligand complexes 48

Appendix Figure 4 RMSD ~ Simulation Time (RST) plots for DUSP5-ligand complexes. 52

Appendix Figure 5 RMSD ~ Simulation Time (RST) plots for DUSP7-ligand complexes 55

Appendix Figure 6 Common substituent groups of advantageous DUSP5 ligands from YIY-906 ... 57

Appendix Figure 7 Common substituent groups of advantageous DUSP7 ligands from YIY-906 ... 58

Appendix Figure 8 Comparison of the binding modes predicted by docking and MD simulations for DUSP3 59

Appendix Figure 9 Comparison of the binding modes predicted by docking and MD simulations for DUSP5 60

Appendix Figure 10 Comparison of the binding modes predicted by docking and MD simulations for DUSP7 61

Preface

I have been enjoying doing molecular modeling in front of my computer at my desk in the Baum Office. When using Pymol to look at my proteins, I always feel the protein structures with helixes and loops are like the roller coasters, well-designed and breathtaking, making me imagine taking a ride and figuring out the related molecular mechanisms at the end.

Doing research, however, takes a lot longer time than taking a roller coaster ride. The first part of this project started off when I was an undergraduate student, and after two years, it turned out to be my first publication during my master's study. Based on that, I took a step further trying to develop a new mapping algorithm to study peptide-protein interactions, which is second major component of this thesis.

I would like to express my gratitude to my undergrad mentor, Dr. Junjun Liu, and my current mentor, Dr. Junmei Wang, for opening the door of molecular modeling to me and giving me generous support whenever I needed. I would also like to thank all professors as my committee members for their valuable suggestions. Thank all my lab fellows, my families and friends who see me through hardships and encourage me all the time.

I am grateful to acknowledge the research support from the National Institutes of Health (R01GM79383 and R21GM097617) and the National Natural Science Foundation of China Grants (21102050). Computational support from the Center for Research Computing of the University of Pittsburgh and Extreme Science and Engineering Discovery Environment (TG-CHE090098) is acknowledged.

Contents in this thesis regarding YIV-906-DUSP interactions come from this publication, Liu *et al.* **New Application of In Silico Methods in Identifying Mechanisms of Action and Key Components of Anti-Cancer Herbal Formulation YIV-906 (PHY906)**, of which the author, I, can adapt the paper for non-commercial exploitation purpose without seeking permissions from the publisher, the Royal Society of Chemistry, according to their **Licence to Publish – Physical Chemistry Chemical Physics**.

1.0 Introduction

1.1 Overview of DUSPs as ERK signaling pathway modulators

1.1.1 Physiological and pathological role of DUSP and ERK

The dual-specificity phosphatases (DUSPs) are a heterogeneous group of protein enzymes that dephosphorylate the phosphotyrosine and phosphoserine/phosphothreonine residues within their substrate proteins[1]. One of the well-established signaling pathway modulated by DUSPs is the extracellular signal-regulated kinase (ERK) pathway[2], where ERK is an important cassette of the mitogen-activated protein kinase (MAPK). Depending on the cell type, subcellular localization of ERK, and the duration and magnitude of stimulus, ERK activation can mediate various cell events such as cellular proliferation, differentiation, apoptosis, and autophagy[3]. As such, abnormal ERK has been implicated in cancer development[4]. Specifically, the ERK pathway is deregulated in approximately 30% of human malignancies[5], probably as a result of the upregulation of several DUSPs in cancer cells. Moreover, prolonged ERK activation has been proved to help stimulate both the apoptosis and autophagy of cancer cells via either intrinsic or extrinsic mechanisms[3, 6-9]. Therefore, inhibition of DUSPs to increase the active phosphorylated ERK (ERK-P) level is of great interest in cancer treatment.

1.1.2 Structural biology of DUSP

In general, the active sites of DUSPs are usually very shallow and highly conserved, consisting of a phosphate-binding loop (P-loop) and a general acid loop (D-loop). Despite the common active site conformation and the same catalytic mechanism, different sub-types of phosphatases in DUSP family may have a different catalytic preference due to their subtle difference of steric and electrostatic characteristics near their catalytic domains. By now, there are 78 DUSP structures of homo sapiens (74 solved by X-ray crystallography and 4 solved by solution nucleic magnetic resonance (NMR)) recorded in the Protein Data Bank (PDB), with 40 DUSP ligands in total. In this study, specifically, DUSP3, DUSP5, and DUSP7 are targets of interest. The reasons why these DUSPs were selected and their structure information are given in the following Section 1.2.

1.2 YIV-906 targeting DUSPs as adjuvant in cancer treatment

YIV-906 (also known as PHY906, KD018) is a standardized herbal formulation inspired by a Traditional Chinese Medicine formula named Huang Qin Tang[10]. With the same ingredients as Huang Qin Tang, the clinical grade YIV-906 is produced following cGMP (current Good Manufacturing Practice) protocol, consisting of spray-dried aqueous extracts from Scutellaria baicalensis Georgi (S), Paeonia lactiflora Pall (P), Glycyrrhiza uralensis Fisch (G), and Ziziphus jujuba Mill (Z)[10]. In both preclinical and clinical studies, YIV-906 has shown great potential to enhance therapeutic outcomes of a broad spectrum of anticancer agents such as Sorafenib[11], Capecitabine[12-14], and Irinotecan[15]. It can also ameliorate the complications

and side effects induced by chemotherapy, especially effective for gastrointestinal symptoms. By now, seven Phase I/II clinical trials of YIV-906 have finished, and no related toxicity has been observed[10].

Although YIV-906 is a promising adjuvant in cancer treatment, the molecular action mechanism of it remains unclear, largely due to its multi-component and multi-target features. It has been reported that YIV-906 improved the therapeutic index of Sorafenib in patients with hepatocellular carcinoma (HCC), as evidenced by the increase of the phosphorylation of extracellular signal-regulated kinases1/2 (ERK1/2), of which the higher level corresponds to the higher susceptibility of HCC to Sorafenib[11]. In this context, YIV-906 works synergistically with Sorafenib by increasing the phosphorylated ERK1/2 (ERK1/2-P) level to trigger apoptosis and

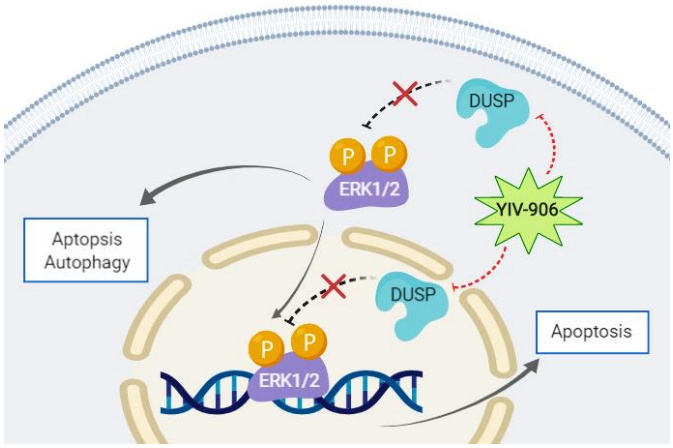


Figure 1 Effect of YIV-906 on the DUSP-ERK1/2 signaling pathway (simplified, schematic)

By inhibiting DUSPs, which are the negative regulators of ERK1/2, YIV-906 can increase the phosphorylated ERK1/2 to promote tumor apoptosis and autophagy.

autophagy in tumors (**Figure 1**). Thus, DUSPs, which show catalytic preference to ERK1/2-P, are potential targets of YIV-906.

Specifically, DUSP3[16, 17], DUSP5[18], and DUSP7[19] have been reported to preferentially dephosphorylate ERK1/2 both *in vitro* and *in vivo*. Among these three DUSPs, DUSP3 is an atypical DUSP, whereas DUSP5 and DUSP7 are both typical DUSPs. Inhibitors of

DUSP3 have been reported to cause cell circle arrest and halt cervix tumor growth without affecting the growth of primary normal keratinocytes[16]. DUSP5 abrogation has been shown to cause ERK hyperactivity and accelerate cell senescence[20]. Similarly, other studies have suggested the potential of DUSP7 inhibitors in cancer therapy. Therefore, we selected DUSP3, DUSP5, and DUSP7 as target proteins based on their normal biological functions and beneficial loss-of-function in cancer. A summary of the PDB ID, subcellular localization, active-site amino acids, and crystal-structure conformation status of target proteins is given in **Table 1**. Schematic crystal structures of DUSP3, DUSP5, and DUSP7 and their active sites are shown in **Figure 2**.

Table 1 Basic information of DUSP3, DUSP5, and DUSP7

Target Protein	PDBID	Subcellular Localization	Active-site Amino Acids	Crystal-Structure Conformation Status
DUSP3	3F81	Nuclear	P-loop: residues 16,23,25,128 D-loop: residues 68,69	With inhibitor
DUSP5	2G6Z	Nuclear & Cytoplasmic	P-loop: residues 263–269 D-loop: residue 232	No inhibitor
DUSP7	4Y2E	Cytoplasmic	P-loop: residues 231–238 D-loop: residues 197–208	No inhibitor

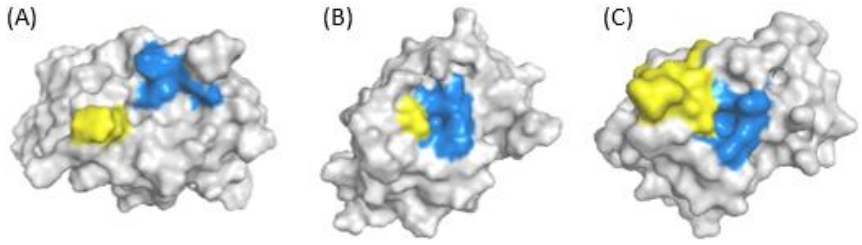


Figure 2 Crystal structures of target proteins in surface representation

(A) DUSP3, (B) DUSP5, (C) DUSP7. For their active sites, P-loops are colored in blue while D-loops in yellow.

1.3 Molecular simulation techniques in drug discovery and development

Molecular modeling has been extensively employed in modern medicinal chemistry for the study of molecular recognition events[21]. The field has grown hand-in-hand with progresses in molecular and structural biology, given increasingly resolved three-dimensional protein structures by X-ray crystallography and nuclear magnetic resonance (NMR)[22, 23]. With the vital structural information of drug targets and the demand to make the best of such information for drug discovery and development, numerous molecular modeling techniques regarding molecular docking and molecular dynamics simulation are developed, while researchers in this field are still actively improving current computational tools and developing new algorithms.

1.3.1 Molecular docking

In most cases, molecular docking refers to predicting the binding conformation of small molecules within the binding site of target proteins[22, 24]. Whereas in a larger context, this technique can also be utilized to predict the binding of ligand peptides or proteins to target proteins. In general, molecular docking is composed of two phases: (1) exploration of a large conformational space and sampling of possible binding modes of the interacting molecules; (2) quantitative prediction of binding energies associated with each binding modes and scoring to identify the correct ones[22, 25, 26].

1.3.1.1 Small molecule-protein docking

Molecular docking is an essential tool in structure-based drug design (SBDD), since it can be performed quickly and conveniently to predict both the binding orientations and the binding

energies of ligands of interest to target proteins[27, 28]. Most docking programs can successfully yield the binding conformations of ligands within the binding pocket of proteins, which can be validated by comparison of re-docked ligand poses with their corresponding poses in the co-crystalized structures[22]. However, these docking programs are not able to reproduce the absolute binding energies of ligand-protein complexes, due to issues including de-solvation and entropic effects[22, 29, 30]. Therefore, molecular docking for small molecules is useful in virtual screening and understanding of intermolecular recognition, narrowing down the scope for further drug discovery. More accurate prediction of biding free energies can be obtained from more sophisticated and computationally expensive molecular dynamics (MD) methods such as molecular mechanics Poison Boltzmann area (MM/PBSA), molecular mechanics generalized Born surface area (MM/GBSA) calculation and thermodynamics integration (TI) calculation.

1.3.1.2 Protein-protein docking

Protein-protein interactions (PPIs) are essential for all cellular processes and dysregulated PPIs underlie a wide range of pathologies[31]. As potential therapeutic targets, however, PPIs have posted challenges to both experimental and computational modeling biologists on understanding the molecular recognition and binding affinities between interacting proteins.

Compared to small molecules, which generally do not interact at large PPI interfaces, peptides are much better candidates in PPI inhibition for two main reasons. One reason is that peptides have more physiochemically favorable natures including large and flexible backbones. The other is the presence of amino acid residues that can interact with residues on the surface of target proteins[32].

PPI docking can guide rapid peptide drug design at an atomic level. Similar to small molecule docking, there are both rigid and flexible docking strategies for PPI. These docking

strategies can also be categorized into local or global docking according to the extent of structural information provided as inputs[31]. The increasing number of protein-peptide complex structures in the Protein Data Bank has promoted the development of docking and refinement algorithms in predicting PPIs. However, major challenges remain despite recent advances in modeling PPIs, including the difficulties in simultaneously consider both the backbone and side-chain flexibilities, ambiguous interpretation of experimental data to be integrated into modeling, and immature scoring methods that might result in lower rank of higher-quality models[31].

1.3.2 Molecular dynamics simulation

Molecular dynamics (MD) simulations capture the behavior of proteins and other biomolecules over time at the atomic level, based on physical principles that govern interatomic interactions[33, 34]. In combination with a wide range of structural biology techniques, such as X-ray crystallography, MD simulations have expanded their impact dramatically in molecular biology and drug discovery. On the other hand, recently introduced computer hardware, especially graphics processing units (GPUs), enables MD simulations to be more accessible and powerful modeling tools with a modest cost[33].

As a strong compensation to wet-lab experimental techniques, MD simulations can not only delineate the motion and position of every atom at any time point, but also can be precisely controlled with different settings of simulation conditions. These simulation conditions include the initial conformation of macromolecules, the ligands bound to them, solvent molecules, temperature, pressure, and so on. By comparing simulation results yielded by different condition settings, one can qualitatively identify the key interactions the ligands make with target macromolecules and quantitatively calculate the binding free energies. For binding free energy

prediction, the alchemical methods such as TI calculation in which a ligand gradually mutate to another one with similar scaffold can offer the most accurate estimates[35], but these methods are computationally expensive and require structure similarities between ligands. In comparison, the MM/PBSA and MM/GBSA methods are cheaper and faster and can deal with more diverse ligand series, since these methods rely on continuum solvent models instead of an explicit representation of water[33].

In this work, we investigated the mechanism of YIV-906 inhibiting DUSPs, using a broad spectrum of molecular modelling techniques, including molecular docking, MD simulations, and binding free energy calculations. To characterize the PPI interface of DUSP3, we also utilized MD simulations to develop a new protein residue mapping algorithm, which in principle can be applied to any protein of interest and be used to improve the accuracy of protein-protein docking.

2.0 Methods

2.1 Application of in silico methods in identifying mechanisms of action of YIV-906

2.1.1 Structural preparation of ligands and target proteins

We acquired 3D structures of known parent chemicals in S, P, G, and Z from the Traditional Chinese Medicine System Pharmacology Database and Analysis Platform (TCMSP http://lsp.nwu.edu.cn/tcmsp.php) and constructed 3D structures of identified human metabolites of YIV-906 with ChemDraw (https://www.perkinelmer.com/category/chemdraw). All ligands were prepared using the LigPrep program implemented in the Schrodinger suite of software (Maestro version 11.2). At target pH 6.8 ± 0.2 which mimicked the pH in the cell cytoplasm and nucleus, possible states of ligands were generated by Epik. Specified chiralities were retained for stereoisomer computation and the maximum number of stereoisomers was set to 32 (as default). After preparation, in total 2502 stereoisomers with possible ionization states were generated for 350 ligands.

Crystal structures of target proteins DUSP3, DUSP5, and DUSP7 were downloaded from Research Collaboratory for Structural Bioinformatics - Protein Data Bank (RCSB – PDB https://www.rcsb.org/). We applied the program Protein Preparation Wizard in the Schrodinger software to prepare target proteins for docking: missing side chains were filled if there were any; restrained minimization (converge heavy atoms RMSD to 0.3 Å) was performed for each target protein to optimize the structure after adding hydrogens and deleting waters. Next, docking grid files were generated using the Receptor Grid Generation module. For each receptor, the center of

the docking site was specified by providing the X, Y, and Z coordinates of the center of the previously reported catalytic sites.

2.1.2 Molecular docking

Flexible ligand docking simulations were performed using the Glide module of the Schrodinger software (Maestro version 11.2).[36] The Glide docking setting is described as follows: (1) using the standard precision version of docking scoring function; (2) applying scaling factor and partial charge cut-off of 0.80 and 0.15, respectively; (3) applying the "Sample nitrogen inversions" and "Sample ring conformations" options for ligand sampling; (4) rewarding intramolecular hydrogen bonds and adding Epik state penalties to docking scores; and (5) writing out at most 10 poses per ligand and performed post-docking minimization.

2.1.3 MD simulation

Large-scale molecular dynamics simulations were performed to study if YIV-906 molecules and their metabolites could bind to DUSPs. The starting protein-ligand conformations for MD simulations came from docking simulations. For each complex, the best docking pose was selected to study the dynamics of the protein-ligand binding. The complex was immersed in a rectangular box with TIP3P water molecules[37] and the shortest distance from any atom of the protein to the edge of the box is at least 12 Å. Proportional Na^+/Cl^- ions were added to make a 0.15 M concentration of the NaCl salt after the system was neutralized with counterions. For the molecular mechanics force field parameters, the partial atomic charges of ligands were derived by using the RESP[38] program to fit the HF/6-31G* electrostatic potentials generated with the

GAUSSIAN16 software package.[39] The other force field parameters came from GAFF[40] in AMBER 18.[41] The residue topologies for ligands were prepared using the Antechamber module.[42] The AMBER FF14SB[43] force fields were used to model proteins.

We performed MD simulations to produce isothermal-isobaric ensembles using the pmemd.cuda program of AMBER 18. The Particle Mesh Ewald (PME) method[44] was used to calculate the full electrostatic energy of a unit cell in a macroscopic lattice of repeating images. All bonds were constrained using the SHAKE algorithm[45] in both the minimization and MD simulation stages. We first performed a series of minimizations with the main chain atoms restrained using the harmonic restraint force constants decreased from 20 to 10, 5, and 1 kcal/mol/$Å^2$. After that, the system was further relaxed by a maximum 10,000-step minimization. The above minimization steps removed any possible steric clashes. The subsequent MD simulation step consists of three phases, namely, the relaxation phase, the equilibrium phase, and the sampling phase. In the relaxation phase, the simulation system was heated up progressively from 50 K to 250 K at steps of 50 K. At each temperature, a 1-nanosecond MD simulation was run without any restraints or constraints. In the following equilibrium phase, the system was equilibrated at 298 K, 1 bar for 15 ns. Finally, a 140-nanosecond MD simulation was performed at 298 K, 1 bar to produce NTP (constant temperature and pressure) ensembles. Due to some intrinsic deficiency of GPU codes, some simulations were restarted if unexpectedly terminated. This led to different total simulation time scales for different protein-ligand complexes. We consider the 120 nanoseconds as the lower-limit time scale for simulation systems to reach the steady states and for us to obtain enough snapshots for analysis. After simulation, all simulation times were over 120 ns. Integration of the equations of motion was conducted at a time step of 1 fs for the relaxation phase and 2 fs for both the equilibrium and sampling phases. In total, 1400 MD frames were recorded. All MD

snapshots were applied to conduct MM/GBSA free energy decomposition and 280 were evenly selected for the MM-PBSA binding free energy calculation.

For extensive sampling, we sampled 20 independent MD trajectories using the last snapshot of the equilibrium phase as the starting conformation. For each MD trajectory, the simulation lasted 20 nanoseconds and 50 MD frames were recorded for the MM/PBSA binding free energy calculations.

2.1.4 MM/PBSA binding free energy calculation and MM/GBSA free energy decomposition

In MM/PBSA methods, the binding free energy is calculated as below:

$$\Delta G_{MM/PBSA} = \Delta H - T\Delta S \approx \Delta E_{ele} + \Delta E_{vdw} + \Delta G_p^{sol} + \Delta G_{np}^{sol} - T\Delta S$$

where ΔE_{ele}, ΔE_{vdw}, ΔG_p^{sol}, ΔG_{np}^{sol}, and $-T\Delta S$ are the change of MM electrostatic energy, the change of MM van der Waals energy, the polar solvation free energy, the nonpolar solvation free energy, and the change of entropy upon ligand binding, respectively.

Key parameters controlling the MM/PBSA and MM/GBSA analyses were the following: external dielectric constant, ~ 80; internal dielectric constant, ~ 1; and the surface tension for estimating the nonpolar solvation energy, ~ 0.054. The Parse radii[46] was used in the MMPB/SA solvation calculation using the Delphi package (http://compbio.clemson.edu/delphi). The entropic term was estimated using a method described elsewhere[47]. Free energy decompositions were performed for all the snapshots collected in the sampling phases of the MD simulations. The interaction energies between the ligand and each residue were calculated with the consideration of solvent effect using an MMGB/SA solvation model developed previously[48].

2.2 Development of the RR mapping algorithm to improve the accuracy of protein-protein docking between DUSP3 and ERK peptide

2.2.1 Structural preparation of simulation systems

2.2.1.1 Generation of conformations for AA analogues and solvate boxes

To mimic the side chains of peptides or proteins, dipeptides were created by adding the acetyl group (ACE) to the N terminal of each type of amino acids (AAs) and the amino group (NHE) to corresponding C terminal (thus the common structure of dipeptides can be expressed as ACE-AA-NHE). In addition to the 20 common AAs, 3 more phosphorylated non-standard AAs were also included, since the protein of interest, DUSP3, catalyzes the dephosphorylation reaction to its substrate ERK-P at its active site. These 3 non-standard AAs are phosphorylated serine, tyrosine, and threonine, respectively. The AMBER FF14SB[43] force fields were used to model dipeptides. For the phosphorylated residues, we used reported parameters in the literature of the partial atomic charges[49] and the phosphate oxygen van der waals[50]. After construction of dipeptides, a 10-ns MD simulation (settings described in the following 2.2.2 section) was performed for each dipeptide to generate structures with different conformations and orientations. From the last 5-ns of the 10-ns simulations, 500 structures of each dipeptide were extracted. Using these dipeptide monomers, cubic dipeptide solvate boxes with grids were generated.

2.2.1.2 Construction of initial systems for protein mapping

A protein-peptide complex structure of DUSP3 and a phosphorylated ERK peptide is available with PDBID: 1J4X. From this crystal structure, the protein DUSP3 was extracted and served as the mapping protein. Next, the mapping protein was positioned into the dipeptide boxes,

where dipeptides originally located within the maximum distance λ_{max} from the mapping protein's surface to its center of mass were removed, and dipeptides originally located farer than the distance of λ_{max} plus 15 Å were removed. Therefore, ball-shaped systems were obtained, in which there were no initial crashes of dipeptide to the mapping protein and sufficient dipeptides to interact with the mapping protein's surface residues. Then these complexes were immersed with TIP3P water molecules[37] and proportional Na^+/Cl^- ions were added to make a 0.15 M concentration of the NaCl salt after the systems were neutralized with counterions. The AMBER FF14SB[43] force fields were used to model proteins. As such, the initial simulation systems consisting of each type of AA (with random conformations), mapping protein DUSP3, water molecules, and ions were constructed. For parallel simulation, 2 initial systems for each type of AA with the mapping proteins were constructed, where both the structures of individual mapping amino acids and their positions related to the protein are different. Therefore, there were 46 systems in total for MD simulations.

2.2.2 MD simulations

All MD simulations in this part were performed using the pmemd.cuda program of AMBER 16.

To generate AA conformations, we first performed minimizations with the heavy atoms restrained using the harmonic restraint force constant of 30 kcal/mol/$Å^2$. In the relaxation phase, the simulation systems were heated up progressively from 0 K to 310 K with the harmonic restraint force constant of 18 kcal/mol/$Å^2$. In the following equilibrium phase, the systems were equilibrated at 310 K for 0.1ns. Finally, a 10-ns MD simulation was performed for each system at 310 K to

produce AA structures with diverse conformations and orientations. From the last 5-ns simulations, 500 structures of each type of dipeptide with different conformations were sampled.

For mapping MD simulations, we adopted the same protocols for minimization, relaxation, and equilibrium as above described. A 100-ns MD simulation was performed for each system at 310 K to predict interactions between AA analogues and the surface residues of the target protein.

2.2.3 Data analysis and mapping

Distances (λ) between the center of mass of each dipeptide in the simulation system and the center of mass of each residue on the surface of DUSP3 were calculated for whole simulation processes. Using cutoffs of 1, 2, 3, 4, and 5 Å, the frequencies of λ falling within these cutoff values were calculated, representing the probabilities of each type of AA interacting with the residues on the surface of DUSP3. This information further indicated the binding affinities of AAs to the target protein. Heatmaps for the receptor residues were obtained.

2.2.4 Protein-peptide docking

The protein-peptide docking program, ZDock[51] (version 3.0.2 for Linux), was used to predict binding modes of the phosphorylated peptide of ERK to the target protein DUSP3. The sequence of the phosphorylated peptide of ERK is DDE(AHP)(TPO)G(PTR)VATR. Since there were no parameters in the ZDcok library file, *uniCHARMM*, for nonstandard residues that were involved, we manually modified the file to include atom names and corresponding charges of these residues. CGenFF[52] (https://cgenff.umaryland.edu/initguess/) was utilized to predict the charges of dipeptides ACE-TPO-NHE, ACE-PTR-NHE, ACE-SEP-NHE, and ACE-AHP-NHE. Then the

charge information for TPO, PTR and SEP were incorporated into the library file for structure preparation. The input structures of the peptide and target protein were extracted from the 1J4X complex structure and were prepared according to ZDock's instruction. Next, 2000 predicted binding modes of the ERK peptide with docking scores were generated.

2.2.5 Re-ranking of docking yielded models

With the interaction probabilities incorporated, the ERK peptide binding models yielded by protein-peptide docking can be re-ranked and generate more accurate predictions. The equations that we used to re-score the docking models are given below:

$$s_{rc} = \sum_i \sum_j P_{rc}^{i.j} \qquad (1)$$

Where s_{rc} is the sum of interaction probabilities $P_{rc}^{i.j}$ at the residue cut-off, rc (rc = 1, 2, 3, 4, and 5 Å), and $P_{rc}^{i.j}$ represents the interaction probability of the protein-residue i to the amino acid j. Note j can also represent a peptide residue which contacts to protein-residue i in the complex poses.

$$S = \left\{\sum_{rc=1,5} \frac{s_{rc}}{rc^2}\right\}^{-1} \qquad (2)$$

Where S is the parameter for re-scoring and is inversely proportional to the interaction probabilities s_{rc}. Therefore, the larger S value of a docking pose, the lower ranking place of the pose is.

$$RS = w * S \qquad (3)$$

Where RS is the re-ranking score and the weighting factor $w = \max\left(RMSD_{poses}\right) / \max\left(S_{poses}\right)$. $RMSD_{pose}$ is the RMSD of a pose of the docking peptide in comparison with its native position (the protein's position is fixed during docking). There are 2000 poses from ZDock result, thus we

can calculate 2000 RMSD values. The $\max(RMSD_{poses})$ is the maximum value of the 2000 values. The max (S_{poses}) is the S value of the pose ranked at 2000, and the min (S_{poses}) is the S value of the top 1 pose.

$$ZSR = v/ZS \quad (4)$$

Where ZSR represents the ZDock ranking, ZS is the docking score given by ZDock, and the weighting factor $v = \max(RMSD_{poses}) * min\ (ZS_{poses})$. The $min\ (ZS_{poses})$ is the docking score of the top 1 pose.

3.0 Results and Discussion

3.1 Mechanisms of action of YIV-906

3.1.1 Unexpected pre-docking results led to the expansion of the compound library to include human metabolites for virtual screening

In order to figure out which herbal ingredient(s), or more specifically, which chemical(s) might play a leading role in modulating DUSPs, we need to firstly establish the ligand library which contains known chemicals in YIV-906 to the greatest extent. A previous study has identified 57 chemicals from YIV-906 aqueous extract and 38 of their metabolites in the plasma of a patient treated with Irinotecan and YIV-906[53]. However, due to the small sample size of this study, there were missing compounds from possible inter-patient ADME (absorption, distribution, metabolism, and excretion) differences. Limitations of detection techniques were also responsible for missing compounds from chemicals' co-precipitation with proteins during the sample preparation process[53]. In addition, a newly-published study on YIV-906 has demonstrated that the chemical profile analysis based on detectable chemicals is not suitable for quality control of herbal formulations and may lead to the false prediction of their biological activities[10]. Since chemical databases provide more comprehensive information, we collected all known parent herbal chemicals of Scutellaria baicalensis Georgi (S), Paeonia lactiflora Pall (P), Glycyrrhiza uralensis Fisch (G), and Ziziphus jujuba Mill (Z) from the Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP)[54]. In total, 500 parent chemicals were included, among which 119, 67, 193, and 121 came from S, P, G, and Z, respectively. In

consideration of the first-pass effect, solubility, and intestinal absorption of drug-like compounds, we performed a preliminary screening according to the following criteria: (1) removing highly hydrophobic molecules, namely molecules with AlogP>5, (2) removing molecules that are composed of only carbon and hydrogen atoms, (3) removing the carbohydrate compounds that will undergo digestion in the gastrointestinal tract. After the preliminary screening, 312 parent chemicals remained in the compound library.

Interestingly, after docking these 312 molecules to DUSPs, the results were not consistent with the previous experimental reports. Top-ranked molecules in our pre-docking simulation were mainly from Glycyrrhiza uralensis Fisch (G), not in agreement with the fact that Scutellaria baicalensis Georgi (S) was most important ingredient of YIV-906 in inhibiting ERK1/2 phosphatases both *in vitro* and *in vivo*[11]. Inspired by the concept of "prodrug" that only if metabolized can the drugs become active, we then added the 38 aforementioned human

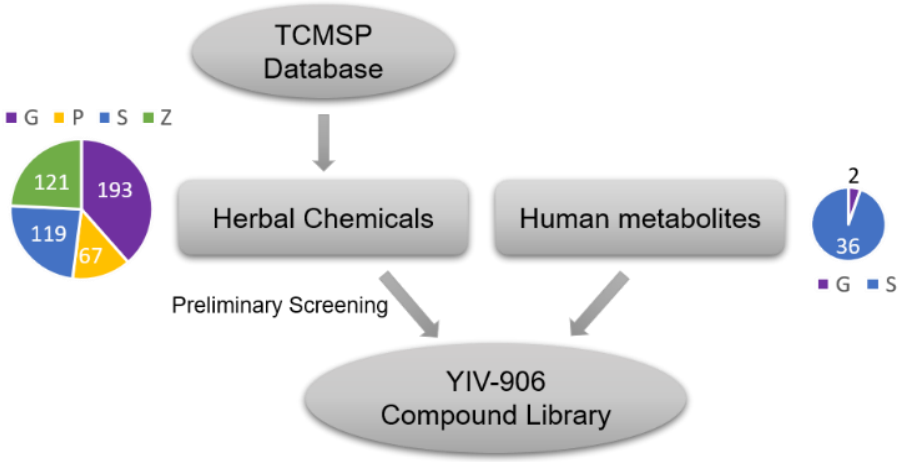


Figure 3 Establishment of YIV-906 compound library

The simplified workflow is demonstrated by the grey flowchart. The counts and origins (G, P, S and Z) of parent herbal chemicals and human metabolites of YIV-906 are shown in pie charts.

metabolites of YIV-906[53] (36 S-derived metabolites and 2 G-derived metabolites) into our virtual screening ligand library (**Figure 3**).

3.1.2 The high similarity of target proteins allows the implementation of the same docking protocol

Currently, there are two crystal structures of DUSP3 reported – one contains an artificial peptide mimicking ERK1/2 (PDB ID: 1J4X[55], referred to in **4.3**), while the other has a co-crystallized inhibitor called STT in its active site (PDB ID: 3F81)[16]. The crystal structure with PDB ID 3F81 was used to establish the docking protocol. Only one crystal structure was reported for DUSP5 and has two sulfate ions located in its active site (PDB ID: 2G6Z)[18]. Similarly, the only known crystal structure of DUSP7 contained a phosphate ion in its active site (PDB ID: 4Y2E)[19]. To test the structure similarity of target proteins, we performed sequence and structural alignment for these three proteins where DUSP3 and DUSP7 were aligned to DUSP5 (**Appendix Figure 1** in the Supplementary Material). The mainchain root-mean-square deviation (RMSD) after least square fitting achieves 0.69 Å for DUSP3 and 0.65 Å for DUSP7, which indicates that the structures of DUSP3, DUSP5, and DUSP7 are similar enough to allow us to adopt the same docking protocol established by using the DUSP3/STT crystal structure.

3.1.3 Molecular docking yielded ligands of interest

First, we docked the reported inhibitor of DUSP3 back to its crystal structure as a validation of our docking protocol. As shown in Appendix Figure 2 in the Supplementary Material, the docked DUSP3 inhibitor overlaps well with the crystallized inhibitor structure with an RMSD of

0.37 Å, indicating our docking protocol is reliable. Then we docked all the parent chemicals and metabolites in our compound library to DUSP3, DUSP5, and DUSP7 with the same protocol. Docked ligands were sorted according to their docking scores, which represented the docking combination free energy. Lower values (more negative values) of docking score indicate higher binding affinity.

In general, top-ranked compounds of DUSP3 are structurally more diverse and have lower binding energy than those of DUSP5 and DUSP7. As demonstrated in Figure 4, several human metabolites stood out, even though the total number of human metabolites was much less than that of parent chemicals (38 metabolites versus 312 herbal chemicals). There are substrate overlaps among DUSP3, DUSP5, and DUSP7, probably due to the conservativeness of their catalytic domains. Common ligands for target proteins include MOL012951, MOL001468, and M18. We selected compounds with docking scores lower than -6.0 kcal/mol for DUSPs for further investigation. Among selected top compounds, most are from S ingredient of YIV-906: 20 S compounds for DUSP3, 6 S compounds for DUSP5, and 13 S compounds for DUSP7. Among these 39 S compounds, only 3 are herbal chemicals, and the rest are all human metabolites. Compounds from ingredient Z of YIV-906 took the second place.

Molecular docking is helpful to narrow down the screening scope. Nevertheless, it only takes the flexibility of ligands into account, utilizes a relatively simple algorithm to achieve computation efficiency, and may produce false positive predictions. We next performed MD simulation for ligands of interest with target proteins, which considers the flexibility of both the

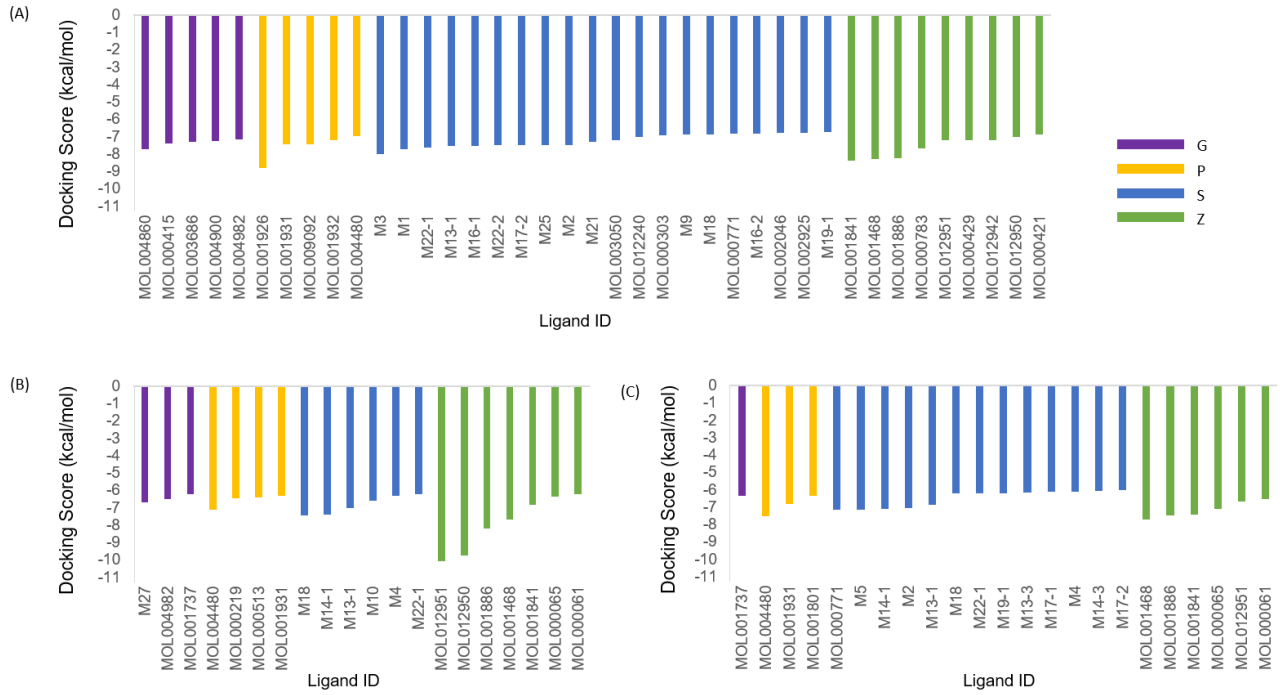


Figure 4 Top-ranked ligands and their docking scores with DUSP3 (A), DUSP5 (B), and DUSP7 (C) Series G, P, S, and Z indicate the origins of ligands. Ligands of which the IDs start with "MOL" are parent chemicals, while IDs starting with only "M" indicate human metabolites of YIV-906.

ligands and receptors. MM/PBSA binding free energy calculations were further conducted to study the ligand-protein interactions.

3.1.4 MD simulation and binding free energy prediction

3.1.4.1 RMSD fluctuations of ligands and binding free energy $\Delta G_{MM/PBSA}$ jointly reflect binding stability.

Stability of docking-yielded binding modes of top YIV-906 ligands with target proteins was assessed by MD simulation. Fluctuations of ligand-protein complex conformation are described by the RMSD ~ Simulation Time (RST) plots which are shown in **Figs. S3**, **S4**, and **S5**. For the receptors, only mainchain atoms were considered for least-square (LS) fitting; while for ligands, all heavy atoms were considered. To measure the dynamics of a ligand in the binding pocket, non-LS fitting RMSDs were calculated for the ligand after the receptor was aligned to the reference structure. Non-LS fitting RMSDs are always equal to or larger than the corresponding LS fitting RMSDs, as the former takes the conformational change, the ligand's translational and rotational movements into account.

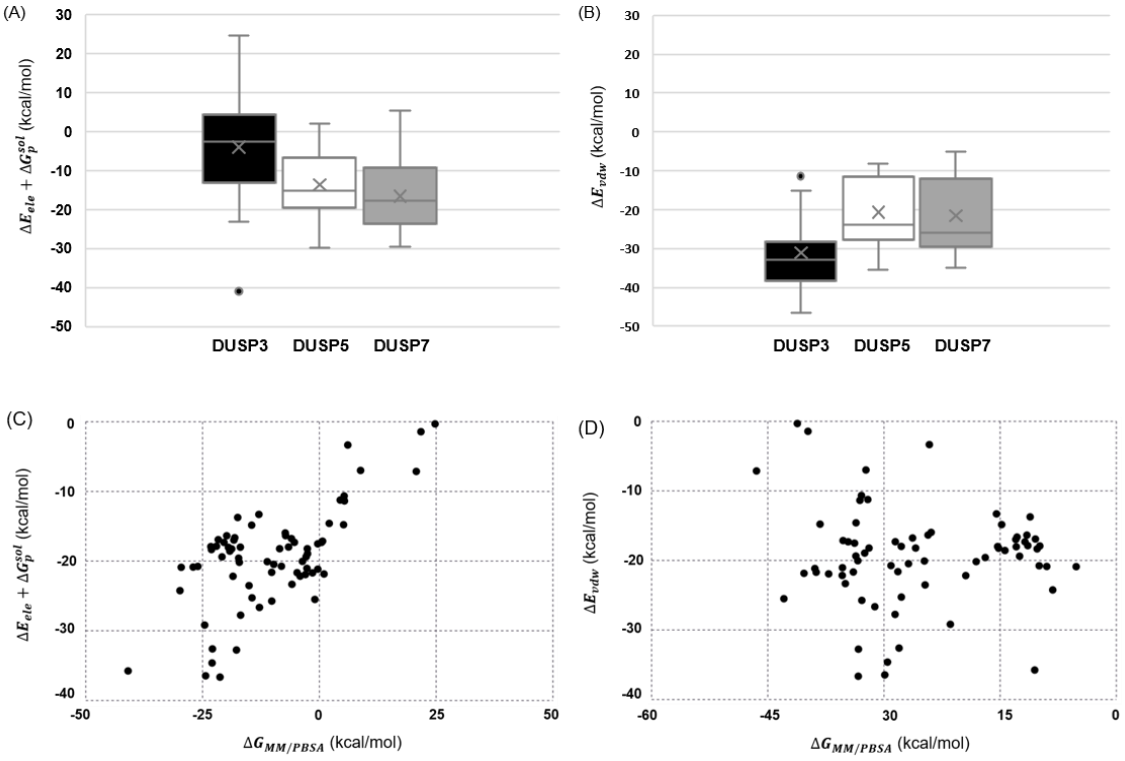
Furthermore, free energy differences between bound and unbound states of solvated proteins and their ligands were calculated using an endpoint free energy method, MM/PBSA. Individual energy terms of binding free energy (in kcal/mol), including the electrostatics, van der Waals, solvation and entropy, are listed in **Tables S1**, **S2**, and **S3** for DUSP3, DUSP5, and DUSP7, respectively.

The conjoining of RMSDs and $\Delta G_{MM/PBSA}$ can be more reliable than when we look at one of the two indicators alone. It is encouraging that our results of RMSDs and $\Delta G_{MM/PBSA}$ are consistent. As shown in **Figs S1-S3** and **Tables S1-S3**, ligands with negative $\boldsymbol{\Delta G_{MM/PBSA}}$ tend to have relatively more narrowed RMSD fluctuations, which indicates they bind stably at DUSPs' binding sites. On the contrary, ligands with positive values of $\boldsymbol{\Delta G_{MM/PBSA}}$ that indicate

unfavourable bound states are also observed as embodying more drastic fluctuations of non-LS fitting RMSDs for ligands. In other words, those ligands had poor binding stability and underwent more conformational change as well as translational and rotational movements during the MD simulations.

3.1.4.2 Electrostatic energy is a key driving force in DUSP ligand binding.

In our MM/PBSA results, we find a significant positive correlation between electrostatic energy ($\Delta \boldsymbol{E_{ele}} + \Delta \boldsymbol{G_p^{sol}}$) and binding energy $\Delta \boldsymbol{G_{MM/PBSA}}$, while no significant correlation is observed between van der Waals energy $\Delta \boldsymbol{E_{vdw}}$ and $\Delta \boldsymbol{G_{MM/PBSA}}$ (**Figure 5**). This indicates the greater importance of electrostatic interaction than van der Waals interaction in DUSP-ligand



	Pearson correlation	Spearman correlation	Kendall correlation
$[\Delta E_{ele} + \Delta G_p^{sol}]$ -- $\Delta G_{MM/PBSA}$	r = 0.6591 p-value = 0.0000	r_s = ρ(rho) = 0.4622 p-value = 0.0000	τ(tau) = 0.3390 p-value = 0.0000
ΔE_{vdw} -- $\Delta G_{MM/PBSA}$	r = -0.0644 p-value = 0.5935	r_s = ρ(rho) = 0.0440 p-value = 0.7115	τ(tau) = 0.0318 p-value = 0.6950

Figure 5 Analysis of electrostatic energy ($\Delta E_{ele} + \Delta G_p^{sol}$), van der Waals energy ΔE_{vdw} and their correlation with binding energy $\Delta G_{MM/PBSA}$

(A) and (B) are boxplots showing the spreads and centers of ($\Delta E_{ele} + \Delta G_p^{sol}$) and ΔE_{vdw} of DUSP-ligand complexes. (C) shows the correlation between ($\Delta E_{ele} + \Delta G_p^{sol}$) and $\Delta G_{MM/PBSA}$. (D) shows the correlation between ΔE_{vdw} and $\Delta G_{MM/PBSA}$. Three types of correlation methods (Pearson, Spearman, and Kendall correlation with different sensitivity to outliers and explanatory power) were applied. Corresponding correlation coefficients r and p-values are listed in the table below the diagrams.

binding. This result is consistent with the previous finding that negatively charged molecules have high affinity to DUSPs, but their hydrophobic entities show high diversity without common structure patterns[15].

3.1.4.3 Sulfation and glucuronidation human metabolites of YIV-906 are vital to its DUSP-inhibiting bioactivity.

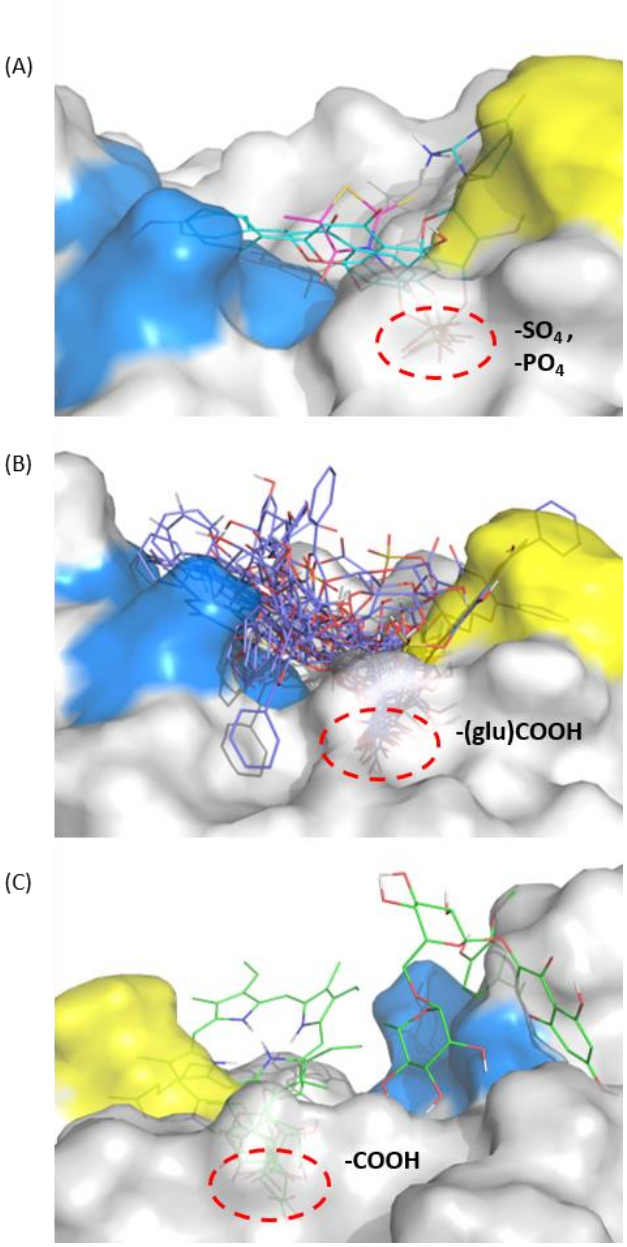


Figure 6 Common substituent groups of advantageous DUSP3 ligands from YIV-906

The protein DUSP3 is represented as surface with the P-loop colored in blue and the D-loops in yellow. Ligand binding modes are demonstrated by representative conformations of MD simulations. The co-crystalized inhibitor of DUSP3 is shown as magenta lines. Common substituent groups (-SO_4, -PO_4, and -COOH) are marked with dashed red circles. (A) The binding of phosphate groups and sulfate groups in the protein's active pocket; (B) The binding of carboxyl groups of glucuronic acids on YIV-906 metabolites in the protein's active site; (C) The binding of carboxyl groups of original YIV-906 chemicals in the protein's active site. The perspective of each ligand-protein complex was slightly adjusted to achieve a clearer view. The surface transparency is set to 60% to show the parts of ligands which stick into the pocket but are concealed by nearby surface.

Ligands with positive values of $\Delta G_{MM/PBSA}$ were considered as false positive docking results. With those false-positive ligands excluded, we performed binding mode visualization for all three target proteins with their advantageous ligands bound in the binding pockets.

As demonstrated in **Figure 6**, we find some common substituents of top-ranked molecules for DUSP3, such as sulfate groups (-SO_4) and carboxyl groups (-COOH), aggregate in the enzymes' catalytic center. Preponderant ligands for DUSP5 and DUSP7 have the same common substituents (**Appendix Figure 6, S7**). These functional groups are all negatively charged as aforementioned. Fluctuations of whole ligands can be observed from their RST plots with the docking poses as the reference structures, but fluctuations of sulfate or carboxyl groups on these ligands are much smaller, indicating the catalytic center is a preferred interaction site for those negatively charged functional groups. Among these substituents, all sulfate groups belong to YIV-906 metabolites generated by Phase II sulfation biotransformation, while carboxyl groups are either from original YIV-906 herbal chemicals or glucuronides generated by the Phase II glucuronidation reaction.

In total, 23 YIV-906 metabolites (22 from the S ingredient and 1 from the G ingredient) stand out, among which 11 are sulfate or glucuronide metabolites of baicalein or baicalin. A recently published study has reported that baicalein or baicalin monotherapy causatively increased ERK-P and caused apoptosis in HCT116 and SW480 colon cancer cells[56]. Thus, we believe that baicalein and baicalin of S ingredient should be two key components of YIV-906 without which the cancer-cell-death related ERK-DUSP pathway is less likely to be modulated. Metabolism to S-source chemicals, including baicalein and baicalin, can endow them with DUSP-inhibiting bioactivity.

The drug binding modes of our results can also be supported by several structural and biophysical studies. The binding location of sulfate moieties of YIV-906 sulfation metabolites overlap well with that of the sulfate inhibitor in the DUSP3 crystal structure (**Figure 6 (A)**), one of the sulfate ions in the DUSP5 crystal structure (**Appendix Figure 6 (A)**) and the phosphate ion

in the DUSP7 crystal structure (**Appendix Figure 7 (A)**). For another crystal structure of DUSP3 (PDB ID: 1J4X)[55], an artificial peptide mimicking a phosphorylated ERK1/2 segment is bound to it. As shown in **Figure 7**, several representative ligands from YIV-906 also overlap well with the crystallized ERK1/2-like peptide. Taken together, it is quite likely that the sulfate moiety and carboxyl moiety act like phosphate mimics to block the P-loop of DUSPs, and hydrophobic moieties of those ligands block the surrounding area of the active sites.

In addition, hydrophobic moieties of advantageous ligands are quite flexible as mentioned, reflecting the fact that the surface residues surrounding the DUSP active sites are structurally very

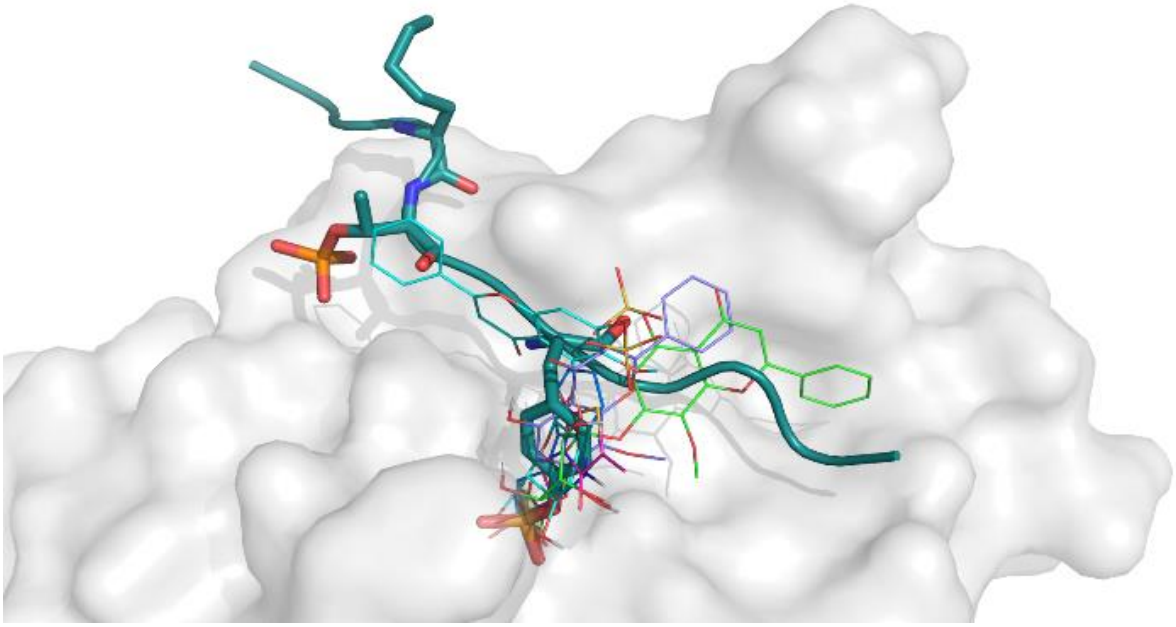


Figure 7 Human metabolites and herbal chemicals from YIV-906 act as phosphate mimic and occupy the catalytic center of DUSP3, which is shown as surface

The artificial peptide that mimics ERK1/2 in the crystal structure of DUSP3 is shown as turquoise ribbon except for the phosphothreonine and phosphotyrosine residues which are shown in sticks and colored by element. Representative ligands are shown in lines and colored by element. Ligands whose carbons are magenta, light purple, green, light blue, and dark blue are STT, M15-1, M18, M19-2, and MOL001841, respectively.

dynamic as indicated by their large B-factors. It is very likely that phosphate-mimic moieties are vital to binding affinity while hydrophobic moieties account for binding selectivity. Glucuronidation metabolites of YIV-906 are observed to bind dominantly to DUSP3 and DUSP7, while sulfation metabolites are in the ascendant when binding to DUSP5.

As mentioned before in **Table 1**, the target proteins DUSP3, DUSP5, and DUSP7 have different subcellular localization. We believe the subcellular localization of different DUSPs is an important issue in terms of how drugs work on them and cause therapeutic effect by directly inhibiting the cytoplasmic DUSPs and/or by inhibiting nuclear DUSPs to modulate gene expression. It is tempting to link the observed DUSPs' ligand preference to their subcellular localization and cellular function to induce apoptosis and autophagy. *In vitro* experiments are needed in the future to promote this topic from molecular level to cellular level and to investigate whether there is a relationship between the DUSPs' ligand preference and their subcellular localization.

Moreover, unlike a mono-component drug which is solely responsible for its therapeutic effect, the chemicals and their metabolites of the YIV-906 formulation work synergistically. There was no significant difference between sulfation and glucuronidation metabolites in their binding free energies. We suggest that the contribution of these metabolites to YIV-906's therapeutic effect is more likely to be related to their relative abundances rather than binding affinities. Future work will probably be needed to uncover the pharmacokinetic profile of the mixture YIV-906.

3.1.4.4 Common ligands identified for DUSP3, DUSP5, and DUSP7.

In many cases, the combination of molecular docking and MM/PBSA rescoring has proven to correctly predict binding poses and binding affinities of a series of ligands[57-62]. MM/PBSA-rescored rank of binding affinities are given in Figure 8 (A). Common ligands of DUSP3, DUSP5,

and DUSP7 are identified as M13-1, M18, M22-1, MOL001468, MOL001841, MOL001886, MOL001931, and MOL012951. Chemical structures of these common ligands are illustrated in Figure 8 (B). Representative positions and orientations of these ligands during MD simulations are given in Appendix Figure 8, S9, and S10 together with the corresponding initial docking-yielded binding modes.

Among these common ligands, M13-1, M18, and M22-1 are metabolites of the S ingredient in YIV-906. The compound MOL001468 is essentially L-malic acid which is derived from the Z ingredient in YIV-906. A crystal structure of the Shp2 DUSP domain in complex with malic acid has been previously published[63], which partially validates our result. Compound MOL001841 is indoleacetic acid (IAA) from the Z ingredient. Very interestingly, IAA has been reported to activate the MAPK pathway in plant[64]. It also causes cell circle arrest and apoptosis in a dose-dependent manner on various human cancer cell lines, including MCF-7, MDA-MB231, HuH-7, and HeLa[65]. Compound MOL001886 and MOL001931 are from the Z and P ingredient, respectively, of which there are no related targets or diseases currently reported. Compound MOL012951 from ingredient Z is the isomer of the secondary messenger, cGMP. What's also worth mentioning is that compound MOL012950 itself is the secondary messenger, cAMP. However, MOL012950 is not a common ligand for DUSPs, and our simulation result indicates that it only binds well with DUSP5. This is probably because near-catalytic-site geometry fitting is also important in addition to the presence of a phosphate group or its mimics. Protein kinases and phosphatases have been reported as secondary messengers' targets[66], whilst the cGMP has also been shown to manipulate the kinase-MLC-phosphatase pathway[67]. Nonetheless, we did not find literature specifically reporting on how cGMP or cAMP targeting DUSPs. It is tempting to speculate that DUSPs are receptors of cGMP and cAMP and have binding-specificity to DUSP5.

3.1.4.5 Interactions between identified common ligands and corresponding DUSPs.

The dynamic interactions between common ligands and DUSP residues were investigated by performing MM/GBSA free energy decomposition. Herein, MOL001841 (indoleacetic acid, IAA) was selected as a representative original compound ligand from plants, and M13-1 was selected as a representative metabolite ligand for DUSP3. As shown in **Figure 9 (A)** and **(C)**, M13-1 interacts with both the P-loop and D-loop of DUSP3, while MOL001841 only interacts with the P-loop of DUSP3. Other common ligands which are not shown here behaved likewise. Original herbal compounds which are generally smaller in size block the P-loops, while larger metabolites of YIV-906 block both the P-loops and D-loops. The sulfate groups or carboxyl groups can form several hydrogen bonds with residues at DUSPs' catalytic centers, which contribute to the most electrostatic binding energy and binding stability (**Figure 9 (B)** and **(D)**). Specifically, as molecules which have both the sulfate moiety and carboxyl moiety in their chemical structures, M13-1 adopt different binding orientations when binding to different DUSPs: with the sulfate moiety binding to the P-loop of DUSP5 and the carboxyl moiety binding to the P-loops of DUSP3 and DUSP7. The binding modes of M13-1 to DUSP3, DUSP5, and DUSP7 are given in **Appendix Figure 11** in the Supplementary Material.

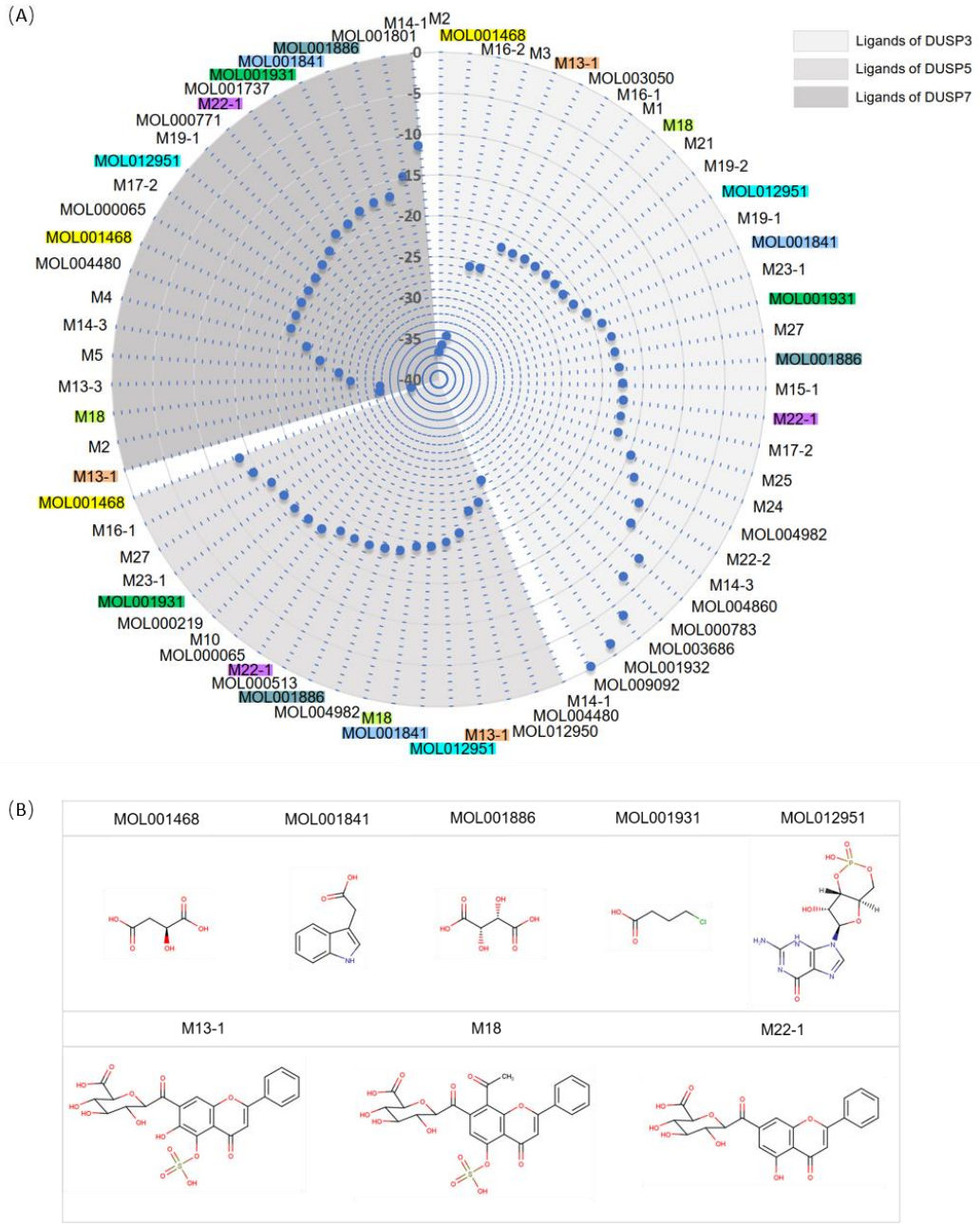


Figure 8 (A) MM-PBSA-reranked ligands of DUSP3, DUSP5 and DUSP7. The axis of this radar is $\Delta G_{MM/PBSA}$ (kcal/mol).

Inner concentric circles indicate lower binding energy. Ligands for each target proteins are ranked clockwise. Three areas with grayscales from shallow to deep indicate the ligands' affiliation to DUSP3, DUSP5, and DUSP7, respectively. Common ligands of these three DUSPs are marked with same colors.

(B) Chemical structures of common ligands of DUSP3, DUSP5, and DUSP7.

For drug discovery, a residue that has high-propensity for ligand binding is important and often referred to as a hotspot residue. Hotspot residues of the three DUSPs in this study are identified (**Figure 10**). All hotspot residues are located on the reported P-loop or D-loop of DUSPs. The binding energy decomposition to P-loop residues was significantly higher than that of D-loop residues, squaring with the fact that the P-loop is responsible for catalysis and the D-loop takes the subsidiary function. The MM/GBSA free energy decomposition results combined

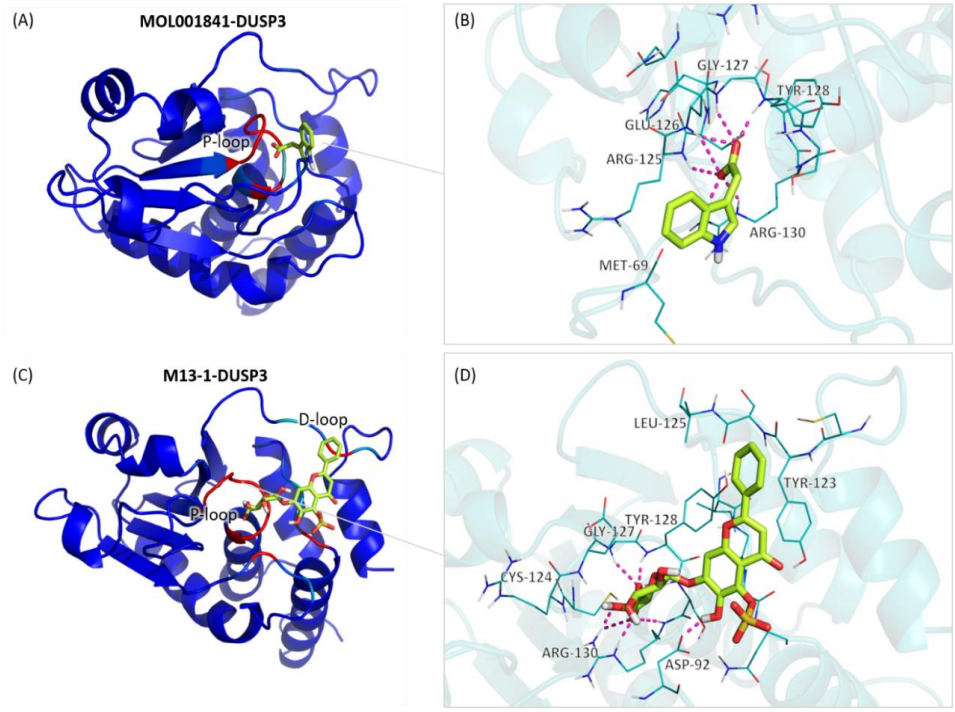


Figure 9 Interactions between binding-site residues of DUSP3 and representative ligands.

(A) Overall view of the original herbal compound MOL001841 at DUSP3's catalytic site. (B) Overall view of the metabolite M13-1 at DUSP3's catalytic site. (C) and (D) are the corresponding amplified views of (A) and (B), respectively. Representative ligands are shown as sticks and colored by element. Hydrogen bonds formed between ligands and DUSP3 are shown in dashed magenta lines. In (C) and (D), correlated residues at catalytic sites are shown in lines. In (A) and (C), DUSP3 is represented as cartoon and residues are colored according to their interaction energies with the ligand. The stronger the ligand-residue interaction is, the more reddish the residue is colored; the weaker the interaction is, the more blueish the residue is colored. In (B) and (D), DUSP3 is represented as cartoon and colored by element.

with the afore discussed DUSPs' ligand preference (in **4.3**) can provide us useful information on developing high potent and selective antagonists of the DUSPs.

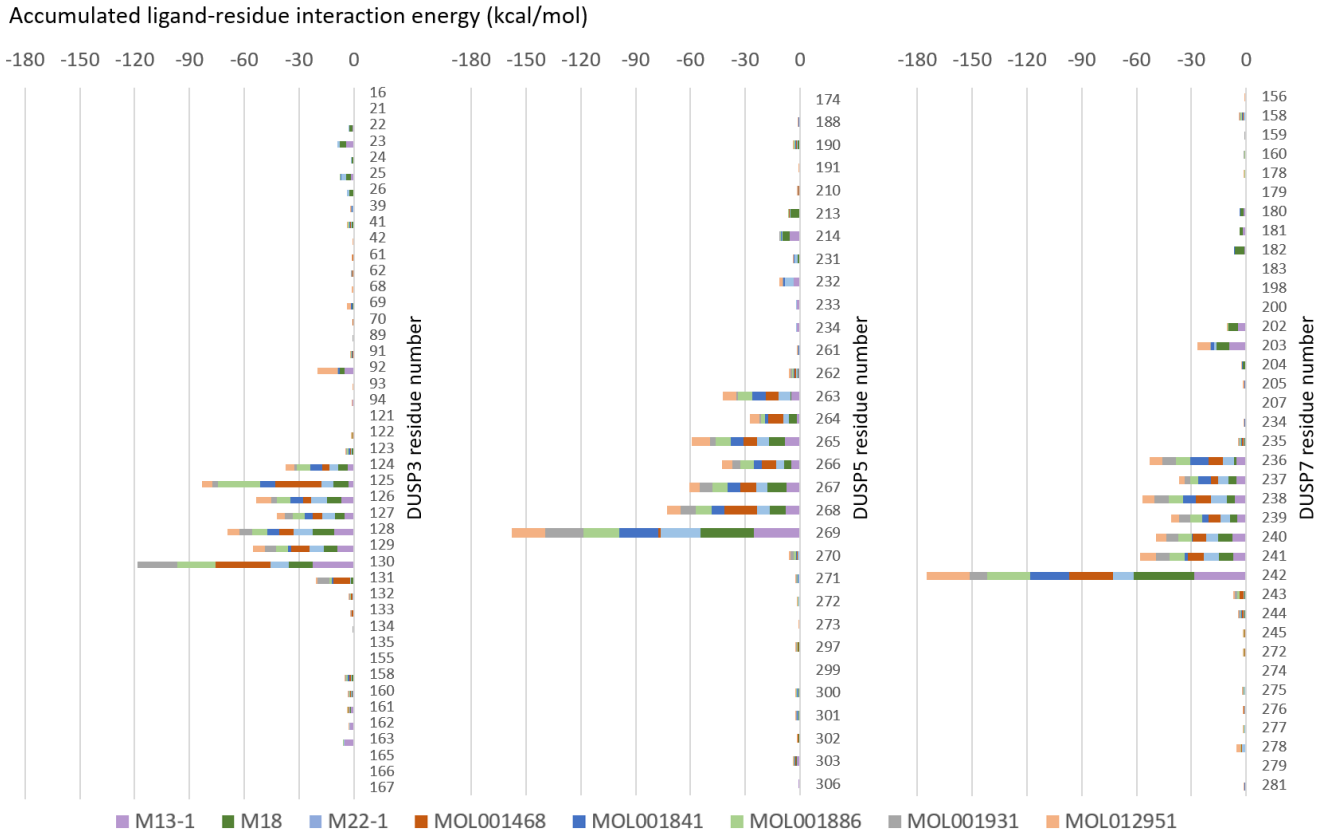


Figure 10 Hotspot residues identification for DUSP3, DUSP5, and DUSP7.

Each fraction with different color on the stacked bars represents the interaction energy of each common ligand to each individual residue of DUSPs and each common ligand has a distinct color.

3.1.4.6 MM/PBSA is a powerful tool to differentiate compounds with different activities.

The MM/PBSA method is a more physical and accurate method compared to the docking scoring function[62, 68-71]. The performance of MM/PBSA, which depends on the setting for desolvation calculations, such as the dielectric constants, is system-dependent. To evaluate how well this technique performs for DUSP proteins, we searched for existing bioassays for DUSPs from Chemical Database of European Molecular Biology Laboratory (CHEMBL Database[72] -

https://www.ebi.ac.uk/chembl/). Since no compounds were reported for DUSP7 and there was only one compound recorded for DUSP5 with no standard values of Ki or IC50, we focused on DUSP3 for which more experimental data are available. We randomly selected 15 bioassay compounds with the experimental $\Delta \boldsymbol{G}$ (calculated from K_i values) ranging from -5.0 to -8.0 kcal/mol for MD simulations and MM/PBSA calculations. The complex structures of those compounds binding to DUSP3 were prepared following the same docking protocol as detailed in **Section 3**.

One of the bioassay compounds, CHEMBL592943, had poor MM/PBSA binding free energy and became an outlier. One explanation of this failed prediction lies in that this compound has a positive charge center and may have a different binding mode than other negatively charged or neutral compounds. Nevertheless, our MM/PBSA method turned out to successfully identify 4 of the top 5 compounds with the lowest K_i values and 3 of the bottom 5 compounds (**Appendix Table 4** in the Supplementary Material). This indicates that the MM/PBSA calculation has the capacity to differentiate active and inactive inhibitors.

With the outlier removed, we calculated the prediction index (PI) and conducted the correlation analysis. The PI between the $(\Delta \boldsymbol{E_{ele}} + \Delta \boldsymbol{G_p^{sol}})$ and experimental $\Delta \boldsymbol{G}$ for these bioassay compounds is 0.7805, which is higher than the PI of 0.6941 between $\Delta \boldsymbol{G_{MM/PBSA}}$ and experimental $\Delta \boldsymbol{G}$. The correlation between the $(\Delta \boldsymbol{E_{ele}} + \Delta \boldsymbol{G_p^{sol}})$ and experimental $\Delta \boldsymbol{G}$ for these bioassay compounds is again better than the correlation between $\Delta \boldsymbol{G_{MM/PBSA}}$ and experimental $\Delta \boldsymbol{G}$. By applying the $(\Delta \boldsymbol{E_{ele}} + \Delta \boldsymbol{G_p^{sol}}) \sim$ experimental $\Delta \boldsymbol{G}$ linear regression equation, the predicted binding energy $\Delta \boldsymbol{G_{MM/PBSA}}$ for each bioassay compound is calibrated (**Appendix Table 4**). Calibrated values of $\Delta \boldsymbol{G_{MM/PBSA}}$ are close to experimental data but exceptions still exist. Possible reasons include the diversity of bioassay compound structures and limitations of the MM/PBSA

method compared to more computationally expensive methods.[58, 68-70, 73] In our study, most chemicals from YIV-906 for MM/PBSA calculations have similar substituting groups and parent skeletons, thus we can expect relatively accurate predictions.

3.2 Protein-protein interactions between DUSP and ERK

An example of the mapping simulation systems is given in Figure 11. Since 2 initial systems for each type of dipeptide with the mapping proteins were constructed, there were 46 systems in total for MD simulations.

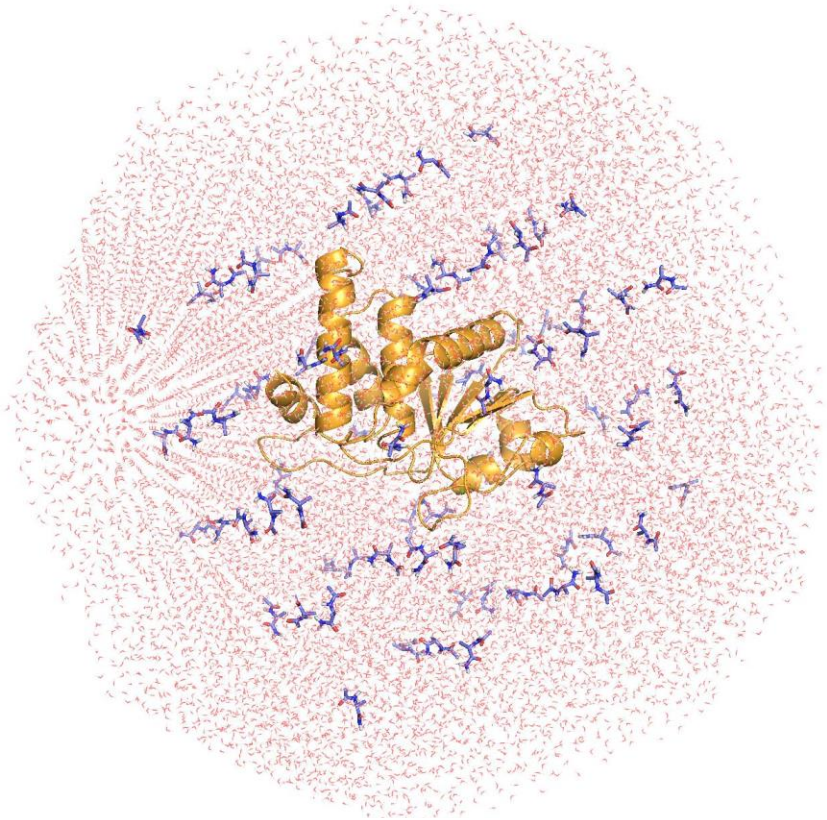
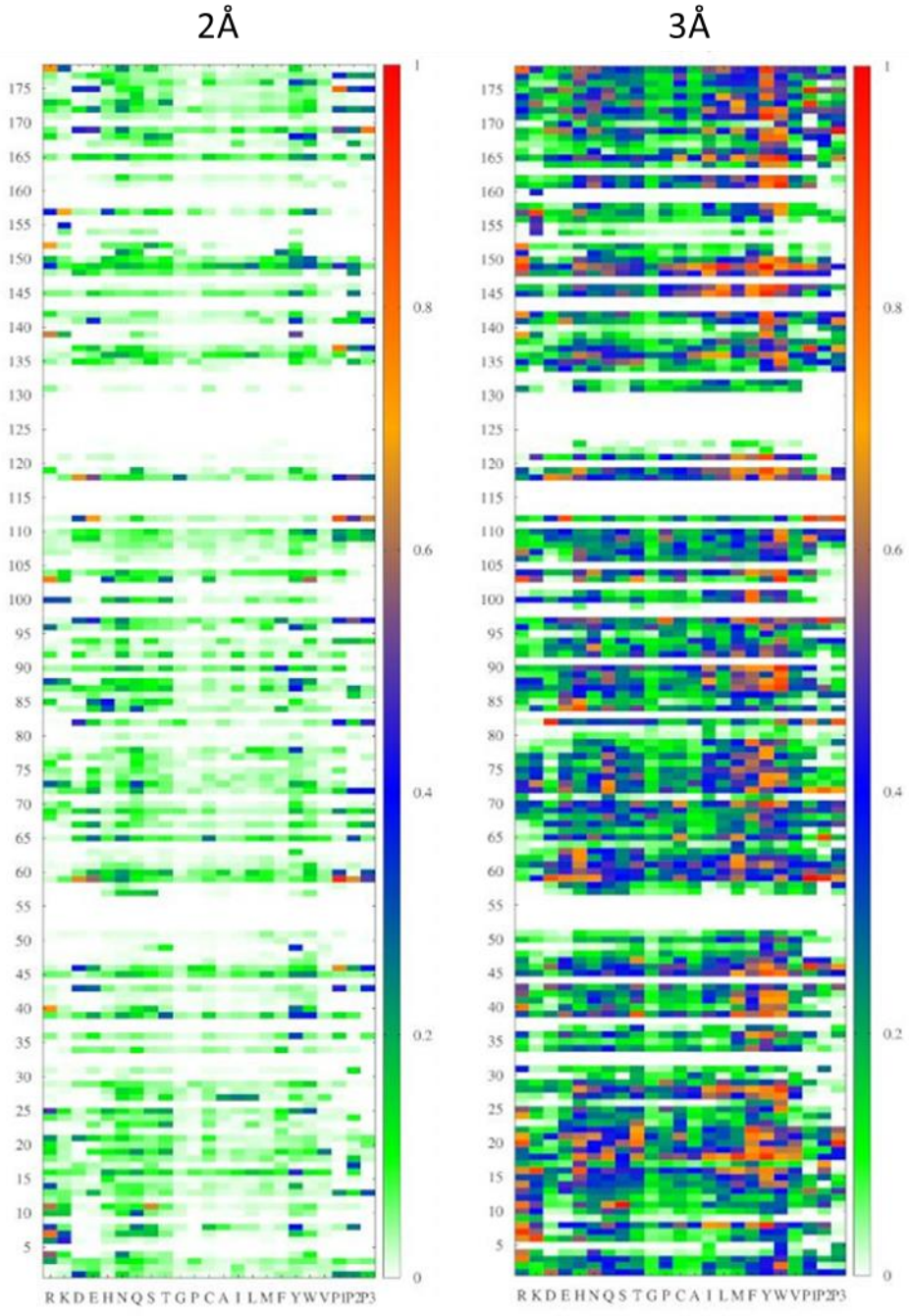


Figure 11 The mapping simulation system of DUSP3 mapped with ACE-ALA-NHE

The protein DUSP3 is shown in cartoon in orange. The dipeptides ACE-ALA-NHE are shown in sticks, while water molecules are shown in lines.

Heatmaps for the receptor residues' affinity to each type of AA were obtained. At the distance cutoff of 1 Å, the probabilities for all dipeptides' AA side chains to interact with residues



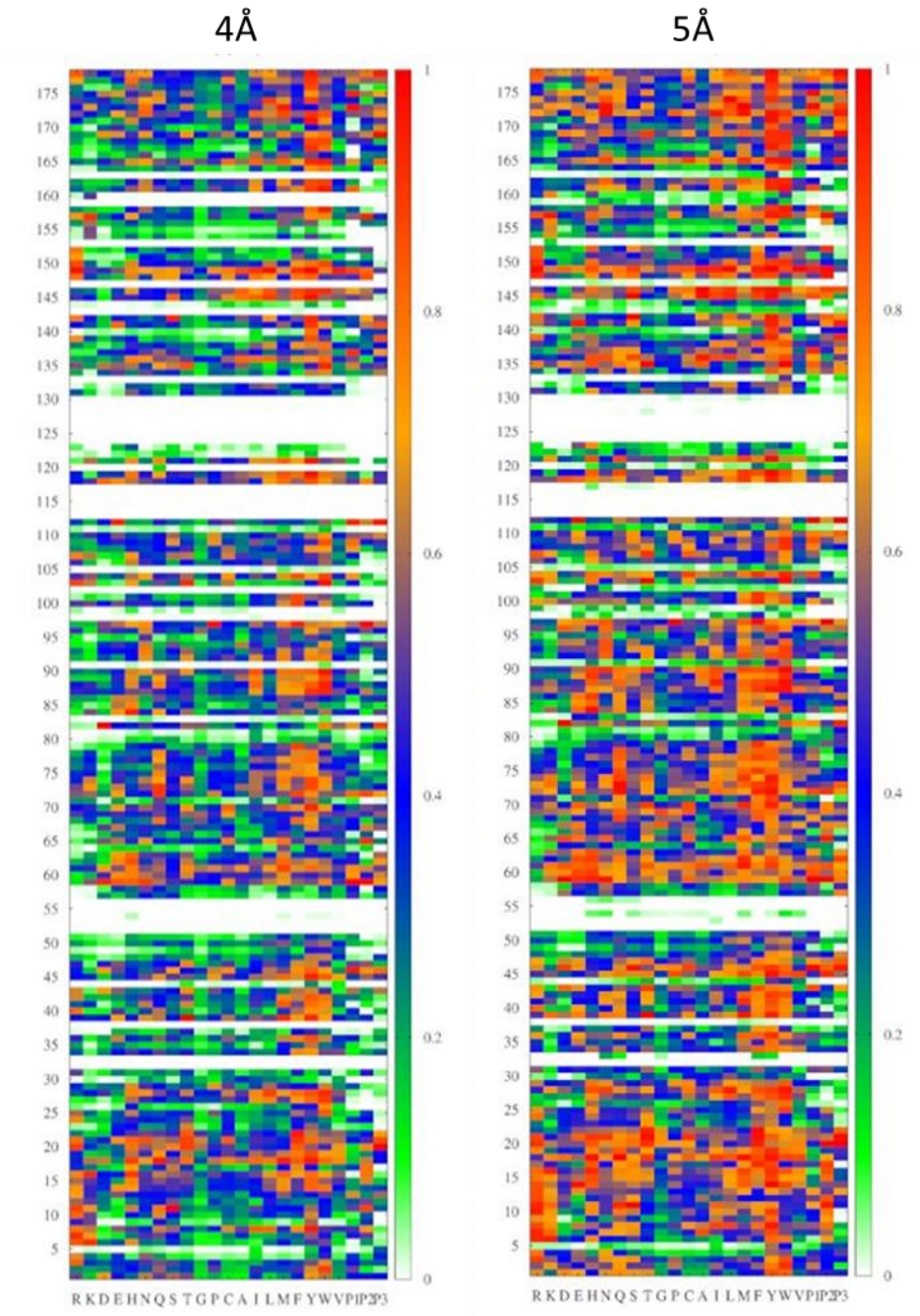


Figure 12 Heatmaps for probabilities of each type of AA interacting with each residues of DUSP3 at distance cutoffs of 2, 3, 4, and 5 Å.

In the sub-figure, on the X axis, the first 20 ACE-AA-NHEs are named as the same single-letter name of the coresponding standard AAs. The last 3, P1, P2, and P3 represednt dipeptide of phoshorylated tyrosine, serine and threonine. The Y axes show the renumbered residue number of DUSP3. The probabilities of different values are colored accroding to the color bar.

of DUSP3 were 0, which is reasonable because of the strong repulsion between residues and

dipeptides at such distance. Despite this, other heatmaps are shown in Figure 12.

Most areas of the heatmaps have colors, indicating that corresponding residues were on the surface of the mapping protein DUSP3 and there were interactions between those residues and AA analogues during simulations. Residues with index from 58 to 62, which form the D-loop of DUSP3, were observed to interact with most AA analogues with relatively higher probabilities. On the other hand, there were "white bands" on the heatmaps, meaning corresponding residues were not located on the shallow protein surface. Specifically, the P-loop residues with residue index from 118 to 125 form the catalytic active site of DUSP3, which was not exposed to the dipeptides as much as residues on the shallow surface but was critical for phosphate binding. It can be seen from the heatmap with the distance cut-off of 2 Å that the three phosphorylated dipeptides interacted with the residue 118 with significantly higher possibilities than other AA analogues, but almost no interactions with other P-loop residues were observed at this cut-off. As the cut-off values increased to 3, 4, and 5 Å, interactions of phosphorylated dipeptide were observed with more P-loop residues. Though the possibilities of such interactions that were key to the catalytic function of DUSP3 were not significantly higher than those of other AA analogues, it was due to the relatively smaller contact area of the binding pocket and the calculated interaction possibilities could still be useful for re-scoring of docking peptides.

Using the ZDock program, 2000 binding modes of the ERK peptide to DUSP3 were generated. As shown in Figure 14, the correlation coefficient R between the docking score from ZDock and the RMSD compared to the peptide's crystal structure is only 0.017, indicating the inaccuracy of ZDock scoring and ranking. After using our equations to re-calculate the scores and re-rank the predicted modes, it is encouraging that the correlation coefficient R increased to 0.217. Though this correlation is still not strong, it can give us several top-ranked binding modes which

are very close to that of the crystal structure. There are many ways to calculate the re-scoring parameter S in the equation (2). Currently we are still exploring the best way to re-scoring, and future work is needed to incorporate the MM/PBSA predicted binding energy into re-scoring.

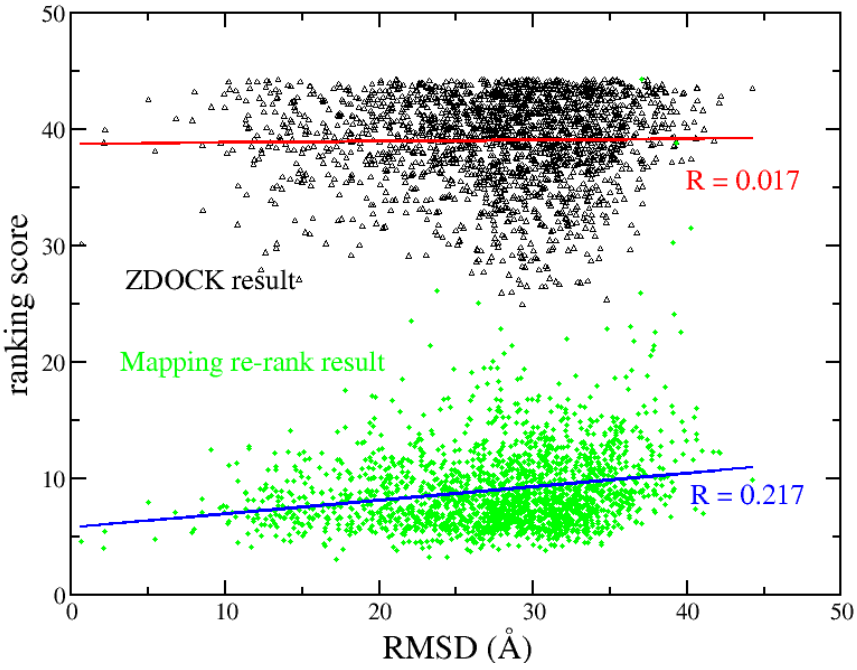


Figure 13 The correlation between the docking score and RMSD to the crystal structure before and after re-ranking with interaction possibilities

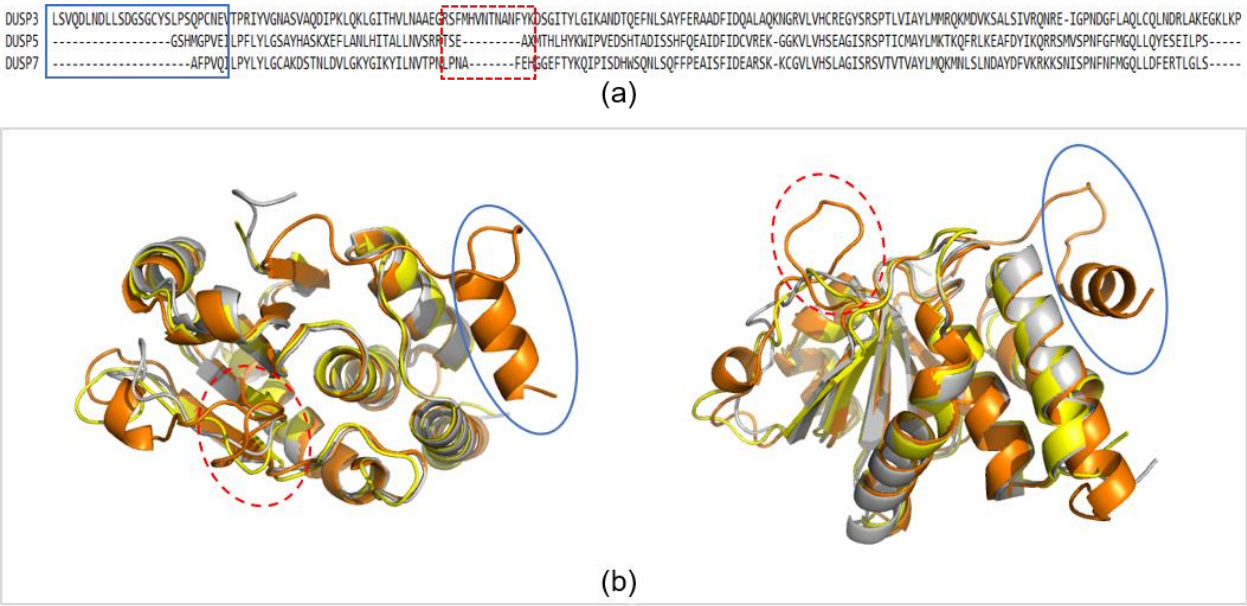
4.0 Conclusions

We applied a set of molecular modeling techniques to decipher the black box of how YIV-906 inhibit DUSPs and applied our newly developed RR algorithm to simulate the interactions between each type of amino acid and each residue of one of the target proteins, DUSP3. It is encouraging that our large-scale molecular modeling study successfully linked the pharmacokinetic metabolism of YIV-906 to its pharmacodynamic effect. We demonstrated that sulfate moieties (produced through the human sulfation metabolism) and carboxyl moieties (either intrinsic or produced through the human glucuronidation metabolism) of the top-ranked ligands from YIV-906 act as the endogenous phosphate mimics. Even though all four ingredients of YIV-906 contribute to the inhibitory effect on DUSPs, the S ingredient plays a dominant role. Specifically, the sulfation and glucuronidation human metabolites of baicalin and baicalein from the S ingredient are identified as key components of the herb formulation without which the cancer-cell-death related ERK-DUSP pathway is less likely to be modulated. Most importantly, the occupancy of the ligands or metabolites at the negative charge centers of the active sites of DUSPs undermines the interaction between DUSPs and ERK1/2 and the resulting elevated phosphorylated ERK1/2 level may trigger apoptosis and autophagy in cancer cells.[74] More efforts should be made to investigate the pharmacokinetics profile and other pharmacological targets of YIV-906 and to explore how different compounds in YIV-906 work synergistically.

To further study the protein-protein interactions between DUSP3 and ERK, the heatmaps describing possibilities of each type of AA interacting with the residues of DUSP3 were obtained with the RR mapping algorithm. By incorporating the interaction probabilities as new scoring arguments, the correlation coefficient between the docking score and the RMSD values compared

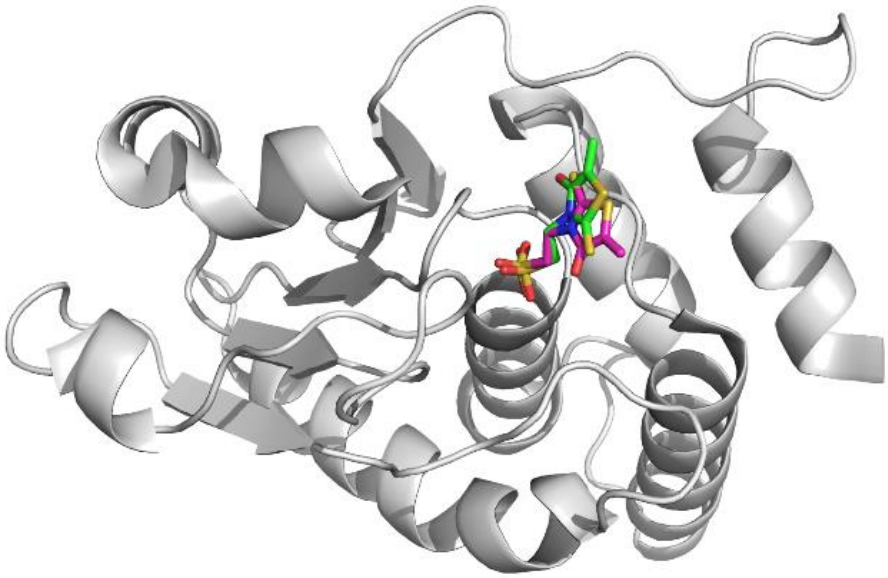
to the peptide's crystal structure has been improved from 0.017 to 0.217. After re-ranking, several predicted binding modes that were close to the crystal structure were ranked on the top. It is also necessary to apply our RR mapping method to more proteins of interest with benchmark or mutagenesis data to validate and optimize our protocol for future use in other protein systems. Nevertheless, our multiscale modeling work well characterized the YIV-906 chemicals binding to DUSPs and provided useful hints on rational design of high potent antagonists, either small molecules or therapeutic peptides, targeting the negative regulators of ERK1/2.

Appendix A Supplementary Figures for DUSP-YIV906 Protein-Ligand Interactions



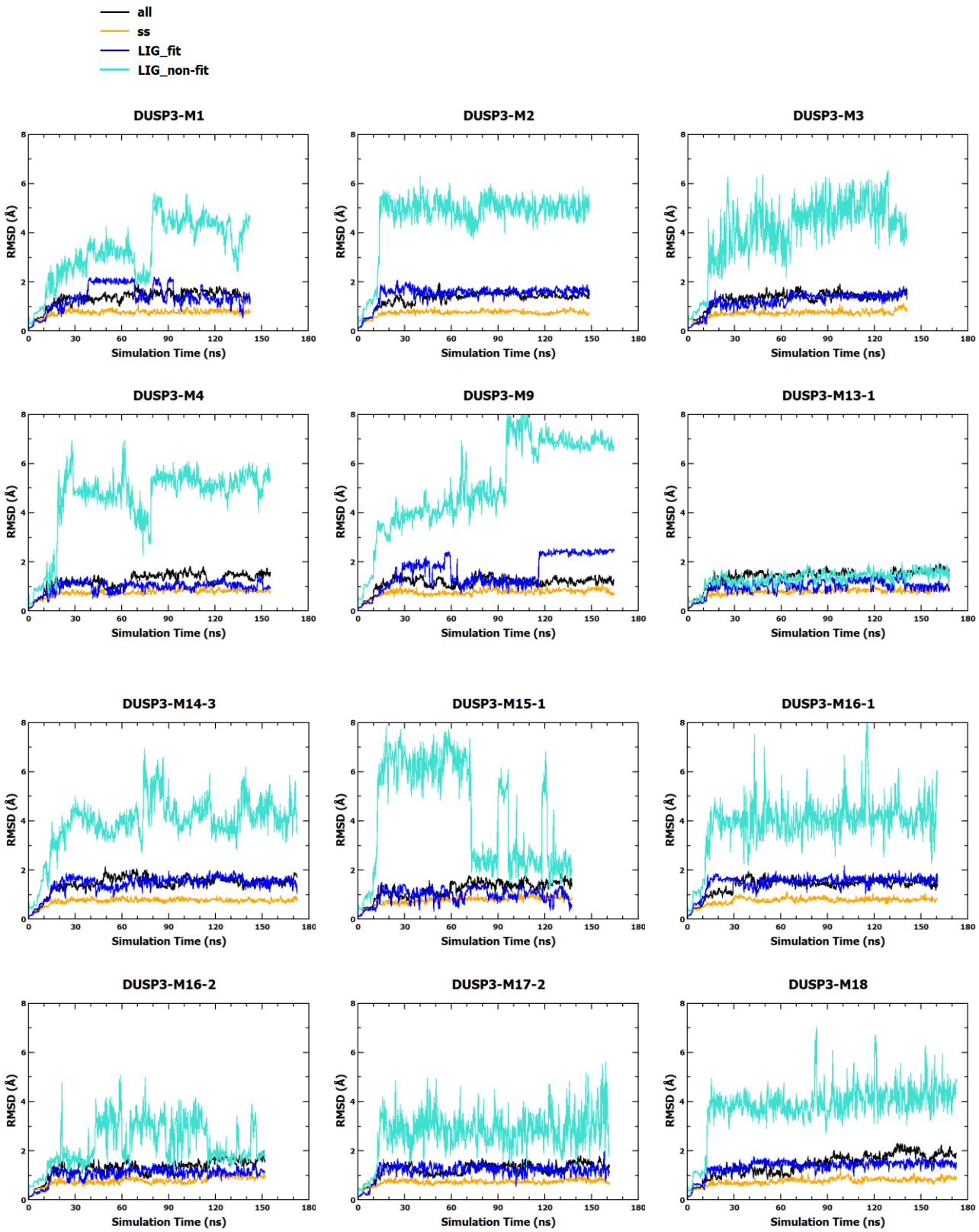
Appendix Figure 1 Sequence alignment for DUSP3, DUSP5, and DUSP7

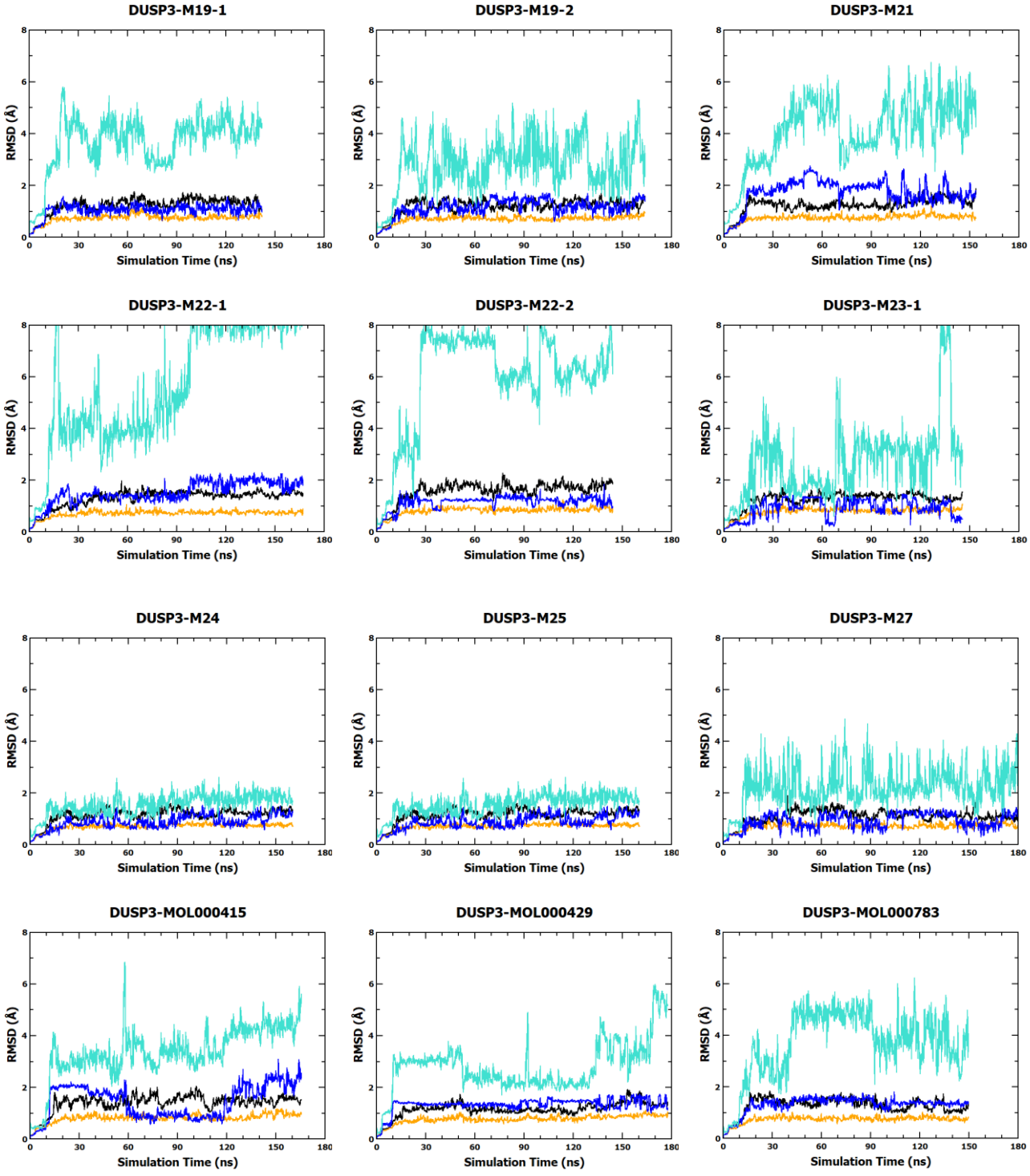
Appendix Figure 1(a) gives sequences of these DUSPs. Appendix Figure 1(b) gives the aligned 3D structures (shown in cartoon) from two different views. DUSP3, DUSP5, and DUSP7 are colored orange, gray, and yellow, respectively. DUSP5 and DUSP7 are quite similar as typical DUSPs. The atypical DUSP3 has a longer loop near the catalytic center's P-loop compared to DUSP5 and DUSP7. This structure difference is marked with the dashed red circle and the sequences are marked with the dashed red rectangle. The extra loop and alpha-helix at the N terminal of DUSP3 is marked with a blue circle. The corresponding sequence difference is marked with a blue rectangle.

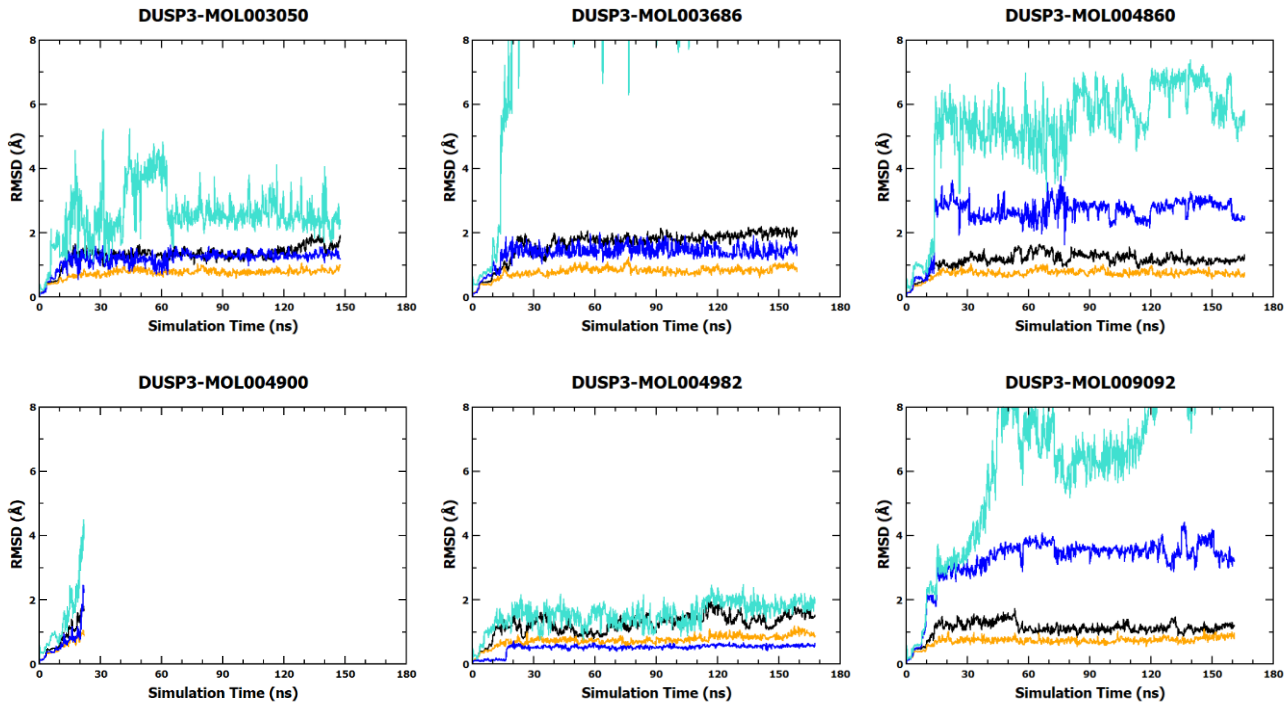


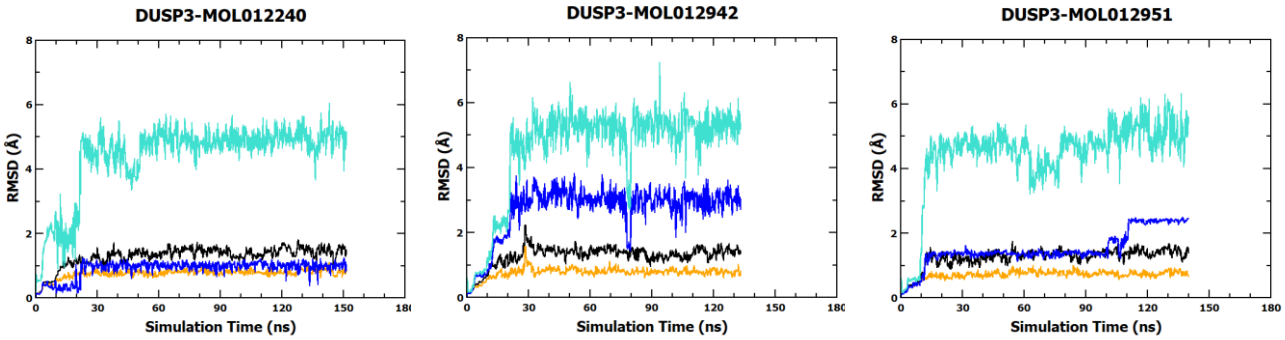
Appendix Figure 2 Validation of the Glide docking protocol by docking the co-crystal ligand, STT, back to the binding pocket of DUSP3

The shown binding mode of redocked STT is the top 1 among redocking results with both the lowest docking energy and the lowest RMSD to STT's crystal structure. DUSP3 is represented as gray cartoon. STT in the crystal structure is shown as sticks and colored by element with the magenta carbon. The docked-back STT is shown as sticks and colored by element with green carbon.


all
ss
LIG_fit
LIG_non-fit
DUSP3-M1
DUSP3-M2
DUSP3-M3
DUSP3-M4
DUSP3-M9
DUSP3-M13-1
DUSP3-M14-3
DUSP3-M15-1
DUSP3-M16-1
DUSP3-M16-2
DUSP3-M17-2
DUSP3-M18
RMSD (Å)
Simulation Time (ns)


DUSP3-M19-1
DUSP3-M19-2
DUSP3-M21
DUSP3-M22-1
DUSP3-M22-2
DUSP3-M23-1
DUSP3-M24
DUSP3-M25
DUSP3-M27
DUSP3-MOL000415
DUSP3-MOL000429
DUSP3-MOL000783
RMSD (Å)
Simulation Time (ns)


DUSP3-MOL003050
DUSP3-MOL003686
DUSP3-MOL004860
DUSP3-MOL004900
DUSP3-MOL004982
DUSP3-MOL009092
RMSD (Å)
Simulation Time (ns)
0
2
4
6
8
0
30
60
90
120
150
180



Appendix Figure 3 Fluctuational patterns of Root-Mean-Square Deviations along the time course of MD simulations for DUSP3-ligand complexes

Black broken lines represent the RMSD of backbone heavy atoms of whole target proteins. Orange broken lines represent the RMSD of backbone heavy atoms of only secondary structures of target proteins. Blue broken lines represent the RMSD of least-square (LS) fitting ligands. Turquoise broken lines represent the RMSD of non-LS fitting ligands. non-LS fitting RMSDs were directly calculated for the ligand after the receptor was aligned to the reference structure.

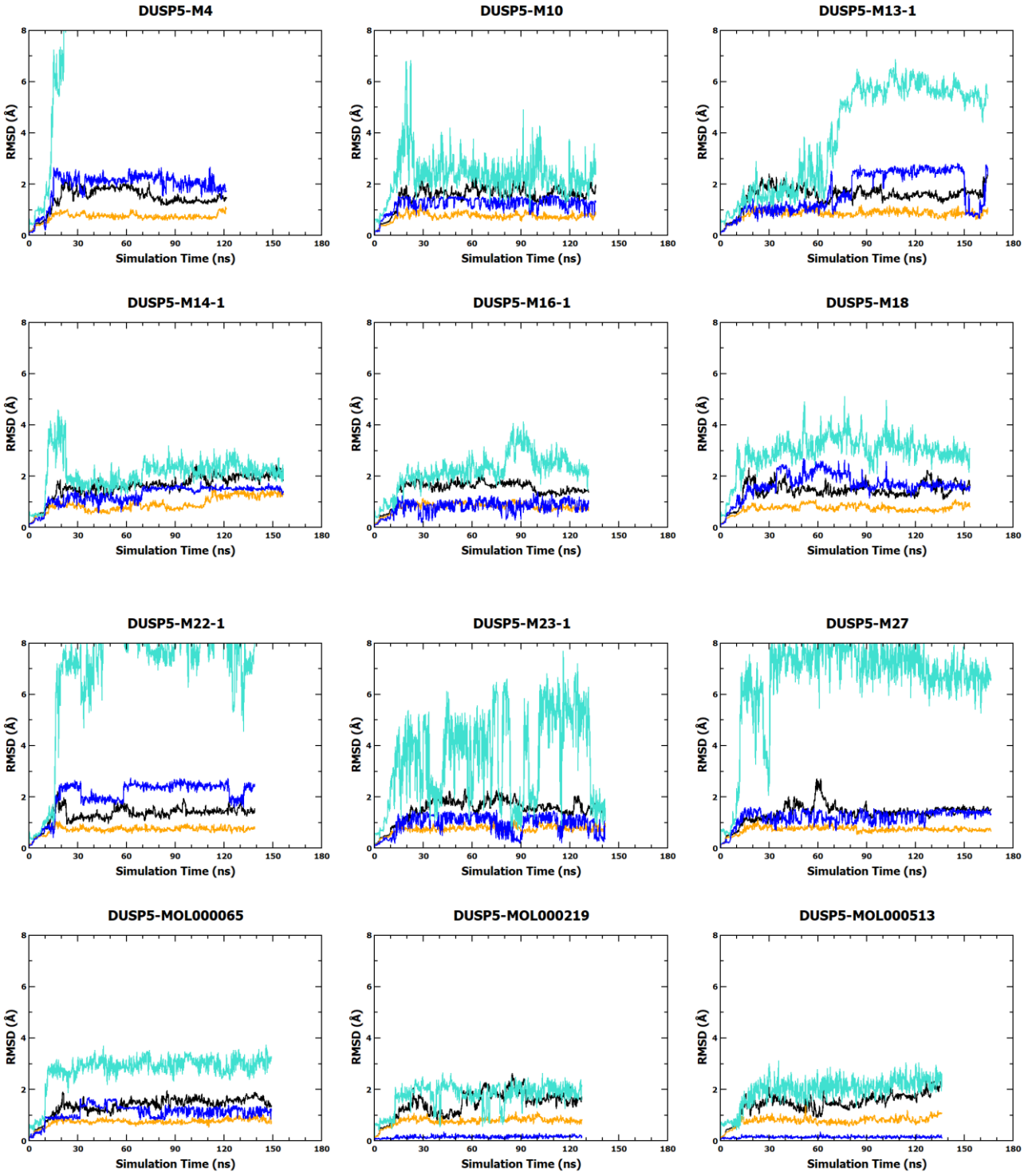
Appendix Table 1 Individual energy terms for DUSP3-ligand binding free energy calculations using the MM-PBSA method (energies in kcal/mol)

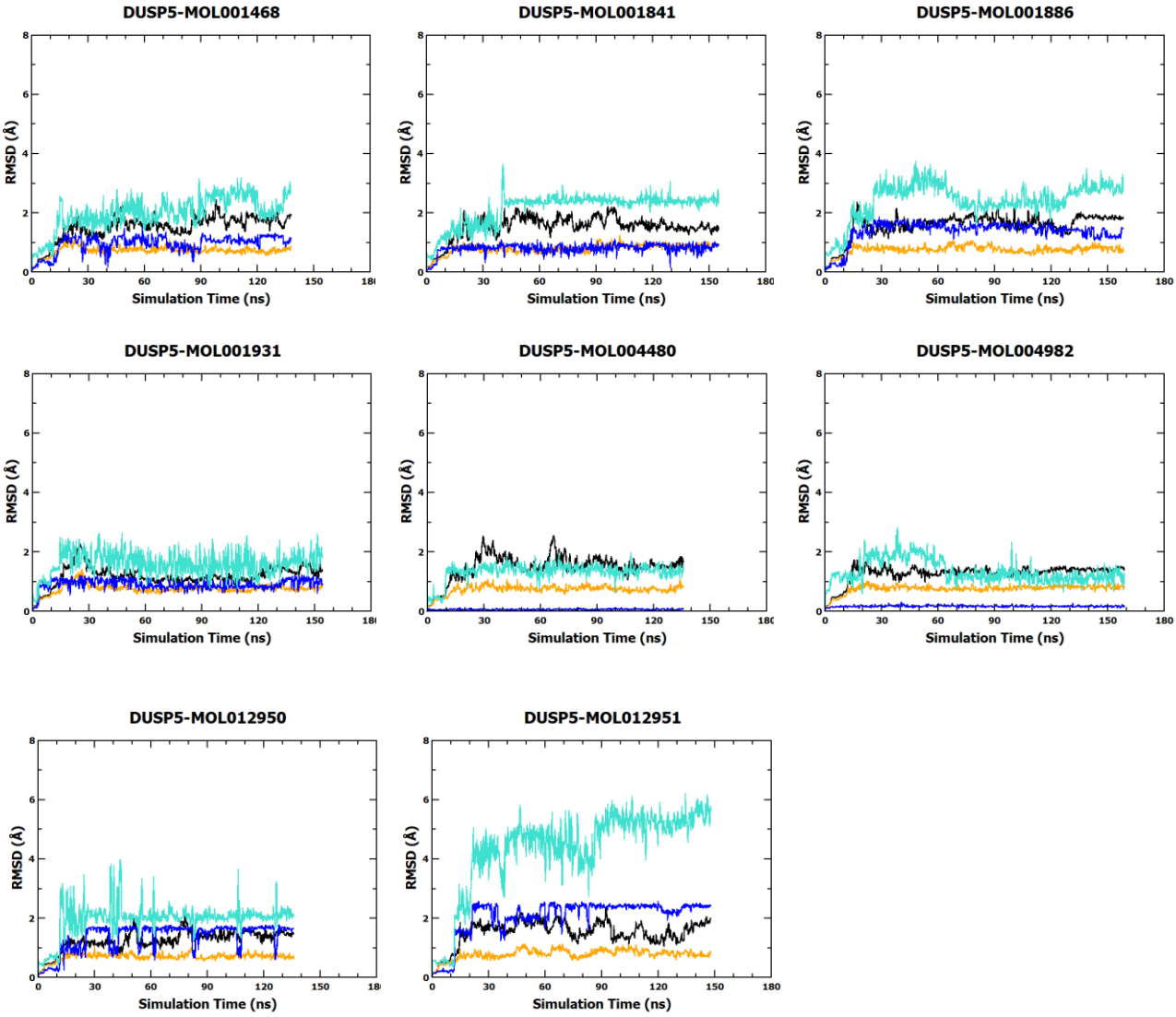
Ligands with a positive value of $\Delta G_{MM/PBSA}$ are shown in red and considered as false positives of docking screening.

LIGAND	ΔE_{vdW}^{gas}	ΔE_{eel}^{gas}	ΔG_{p}^{Sol}	ΔG_{np}^{Sol}	$T\Delta S$	$\Delta G_{MM/PBSA}$
M1	-37.14±0.06	-204.79±0.44	201.80±0.52	-3.25±0.01	-21.44±0.06	-21.93±0.11
M2	-33.32±0.28	-210.73±1.51	189.46±1.63	-2.94±0.01	-20.88±0.04	-36.65±0.33
M3	-32.86±0.10	-196.53±0.72	186.35±0.75	-2.88±0.01	-20.17±0.04	-25.74±0.10

M4	-29.05±0.18	15.56±0.44	10.77±0.48	-2.29±0.02	-17.08±0.01	12.08±0.44
M9	-34.97±0.28	17.19±0.46	12.70±0.24	-2.94±0.02	-18.37±0.06	10.34±0.28
M13-1	-42.92±0.11	-131.10±0.70	130.13±0.49	-3.07±0.00	-21.47±0.02	-25.50±0.27
M14-3	-32.91±0.14	-220.08±0.25	225.39±0.34	-2.49±0.01	-19.43±0.03	-10.66±0.15
M15-1	-33.80±0.04	-143.19±0.42	142.85±0.39	-2.19±0.01	-18.80±0.03	-17.52±0.03
M16-1	-35.39±0.25	-185.69±0.39	181.57±0.75	-3.00±0.02	-20.35±0.04	-22.16±0.18
M16-2	-29.54±0.34	-215.53±0.74	192.57±0.26	-2.77±0.01	-20.64±0.06	-34.62±0.31
M17-2	-28.53±0.17	-173.51±1.04	168.17±1.16	-2.66±0.01	-19.22±0.02	-17.31±0.12
M18	-40.32±0.13	-178.98±1.50	179.97±1.31	-3.07±0.00	-20.54±0.05	-21.85±0.38
M19-1	-38.92±0.16	-205.09±1.50	204.75±1.08	-3.10±0.02	-21.19±0.03	-21.16±0.47
M19-2	-33.96±0.27	-161.23±0.90	156.47±0.96	-2.67±0.01	-19.73±0.00	-21.66±0.11
M21	-38.75±0.23	-202.83±0.89	201.36±0.43	-3.27±0.02	-21.79±0.02	-21.69±0.37
M22-1	-34.62±0.40	-157.37±0.83	157.82±0.50	-2.48±0.01	-19.33±0.05	-17.32±0.53
M22-2	-32.09±0.55	-196.66±0.79	201.12±0.77	-2.77±0.01	-19.16±0.02	-11.24±0.26
M23-1	-32.49±0.11	-133.29±0.32	130.76±0.31	-2.00±0.01	-18.08±0.04	-18.94±0.20
M24	-38.25±0.26	-182.81±0.56	187.99±0.21	-3.42±0.00	-21.72±0.02	-14.78±0.20
M25	-35.28±0.06	-223.38±1.54	224.06±1.32	-3.39±0.01	-20.85±0.04	-17.14±0.40
M27	-31.93±0.11	-125.28±0.23	122.70±0.40	-2.07±0.01	-18.35±0.02	-18.22±0.34
MOL000415	-27.99±0.16	-38.53±0.13	53.77±0.45	-2.74±0.02	-18.03±0.05	2.54±0.37
MOL000429	-12.85±0.23	-103.72±0.28	115.33±0.27	-0.97±0.01	-15.68±0.03	13.47±0.23
MOL000783	-32.33±0.05	-219.76±1.47	228.59±1.06	-3.02±0.00	-19.52±0.01	-6.99±0.47
MOL001468	-10.54±0.14	-365.51±1.26	324.57±1.03	-0.80±0.00	-16.52±0.00	-35.76±0.20
MOL001841	-18.09±0.17	-156.91±0.49	139.78±0.31	-1.21±0.00	-16.26±0.01	-20.17±0.33

MOL001886	-11.43±0.16	-303.55±0.27	281.52±0.53	-0.88±0.00	-16.48±0.00	-17.87±0.37
MOL001926	-31.83±0.25	35.07±0.37	3.94±0.68	-2.94±0.02	-18.34±0.11	22.57±0.67
MOL001931	-15.19±0.06	-162.05±0.74	143.27±0.80	-0.82±0.00	-16.53±0.02	-18.25±0.20
MOL001932	-39.81±0.10	-21.07±0.51	42.81±0.57	-3.17±0.01	-19.81±0.07	-1.43±0.05
MOL003050	-19.44±0.21	-168.83±0.49	150.34±0.43	-1.42±0.01	-17.17±0.03	-22.18±0.16
MOL003686	-24.14±0.14	-41.26±0.52	47.35±0.38	-1.99±0.03	-16.71±0.07	-3.34±0.28
MOL004860	-46.46±0.08	-46.33±0.53	67.07±0.16	-4.21±0.02	-22.80±0.09	-7.13±0.25
MOL004900	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
MOL004982	-15.47±0.29	-167.69±0.28	154.75±0.07	-0.94±0.00	-16.05±0.01	-13.30±0.17
MOL009092	-41.18±0.08	-28.05±0.19	52.82±0.32	-3.57±0.01	-19.67±0.06	-0.31±0.18
MOL012240	-20.73±0.09	-14.85±0.64	30.85±0.37	-1.64±0.00	-15.68±0.03	9.32±0.29
MOL012942	-35.69±0.17	-17.29±0.28	51.82±0.16	-2.92±0.00	-17.83±0.05	13.74±0.11
MOL012951	-28.20±0.30	-175.48±0.64	165.28±0.45	-1.76±0.00	-18.56±0.02	-21.61±0.15


DUSP5-M4
DUSP5-M10
DUSP5-M13-1
DUSP5-M14-1
DUSP5-M16-1
DUSP5-M18
DUSP5-M22-1
DUSP5-M23-1
DUSP5-M27
DUSP5-MOL000065
DUSP5-MOL000219
DUSP5-MOL000513
RMSD (Å)
Simulation Time (ns)



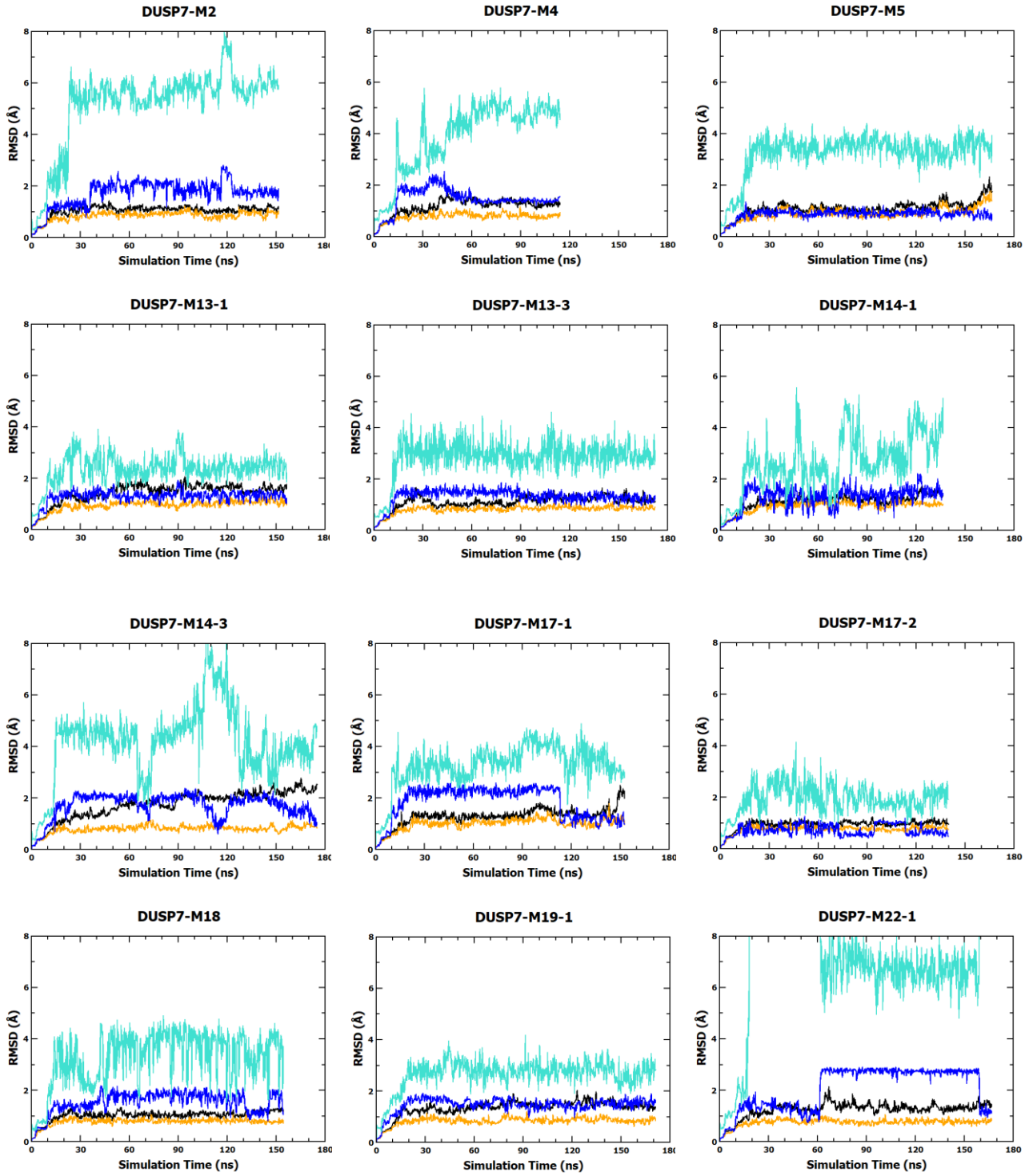
Appendix Figure 4 RMSD ~ Simulation Time (RST) plots for DUSP5-ligand complexes.

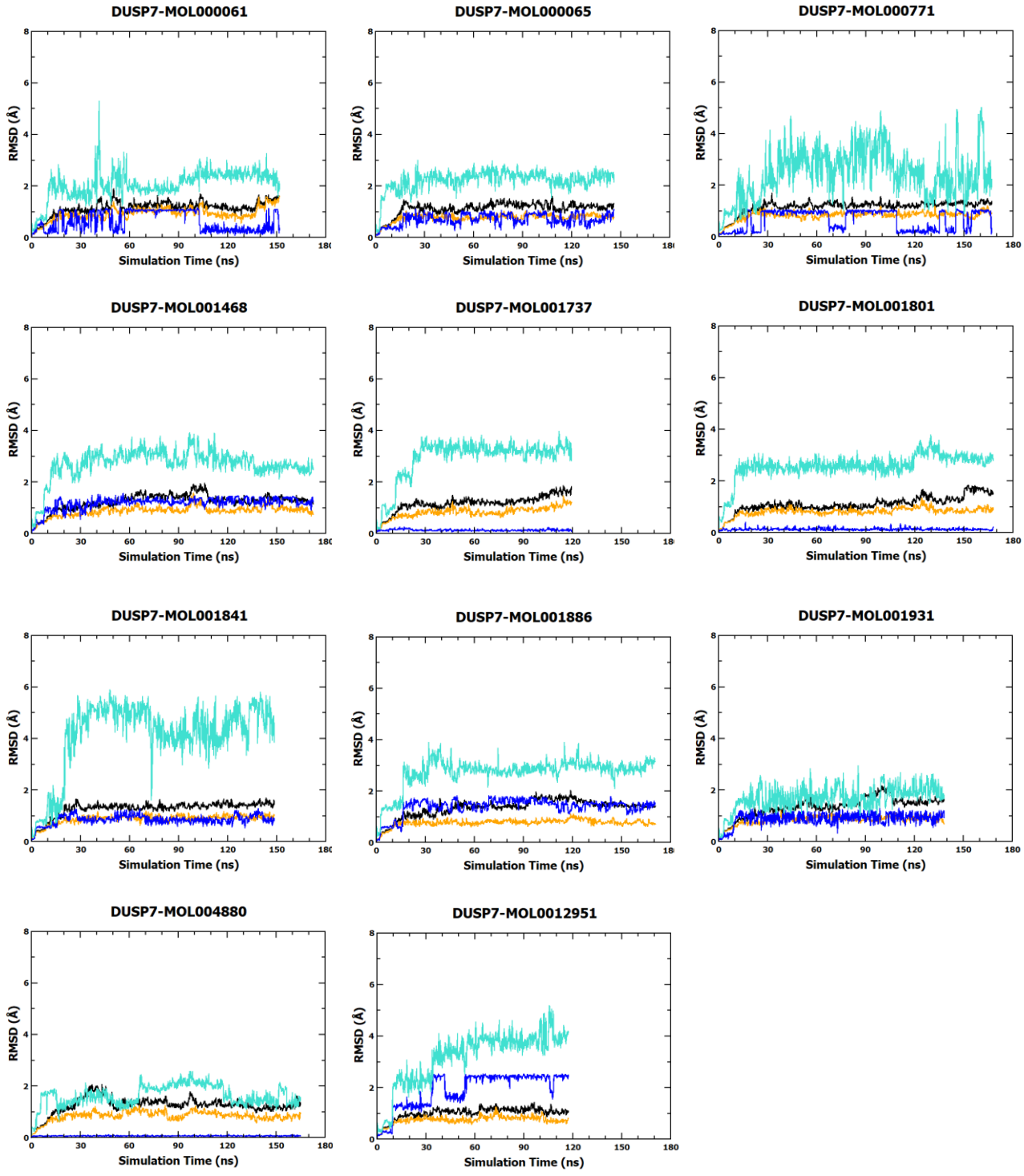
Legend description is the same with Appendix Appendix Figure 2.

Appendix Table 2 Individual energy terms for DUSP5-ligand binding free energy calculations using the MM-PBSA method (energies in kcal/mol)

Ligands with a positive value of $\Delta G_{MM/PBSA}$ are shown in red and considered as false positives of docking screening.

LIGAND	ΔE_{vdW}^{gas}	ΔE_{eel}^{gas}	ΔG_{p}^{Sol}	ΔG_{np}^{Sol}	$T\Delta S$	$\Delta G_{MM/PBSA}$
M4	-22.55±0.06	-102.18±0.54	114.52±0.34	-1.87±0.01	-15.69±0.02	3.61±0.40
M10	-26.31±0.35	-141.49±0.41	135.52±0.31	-1.48±0.01	-16.99±0.02	-16.77±0.27
M13-1	-35.36±0.06	-213.23±1.10	210.57±1.12	-2.32±0.00	-19.29±0.03	-21.04±0.22
M14-1	-31.18±0.19	-293.41±1.09	280.59±1.37	-2.31±0.00	-19.66±0.02	-26.65±0.16
M16-1	-33.61±0.23	-152.90±0.38	155.01±0.29	-2.51±0.01	-19.43±0.04	-14.59±0.25
M18	-33.63±0.09	-243.16±3.08	240.33±2.91	-2.26±0.02	-19.35±0.04	-19.37±0.25
M22-1	-27.76±0.00	-184.87±0.53	178.28±0.60	-1.69±0.00	-18.06±0.06	-17.98±0.33
M23-1	-24.31±0.28	-122.63±0.82	115.41±0.61	-1.25±0.01	-16.43±0.05	-16.35±0.32
M27	-23.91±0.04	-127.68±0.63	120.39±0.77	-1.24±0.01	-16.49±0.02	-15.96±0.23
MOL000065	-9.88±0.20	-214.08±0.96	190.86±0.60	-0.77±0.00	-15.95±0.03	-17.92±0.27
MOL000219	-12.82±0.29	-159.96±0.32	141.84±0.51	-0.69±0.00	-14.99±0.01	-16.64±0.09
MOL000513	-12.92±0.37	-160.52±0.93	141.05±0.66	-0.82±0.00	-15.18±0.03	-18.03±0.23
MOL001468	-11.13±0.32	-218.97±1.93	201.49±1.49	-0.75±0.00	-15.61±0.01	-13.74±0.33
MOL001841	-16.92±0.23	-167.59±0.32	150.31±0.48	-0.91±0.00	-15.53±0.03	-19.58±0.22
MOL001886	-10.23±0.13	-290.37±0.57	267.28±0.66	-0.76±0.00	-15.73±0.01	-18.34±0.17
MOL001931	-11.55±0.04	-162.94±0.40	143.11±0.38	-0.67±0.00	-15.69±0.01	-16.36±0.26
MOL004480	-8.21±0.42	-190.79±0.46	160.96±0.56	-0.24±0.00	-14.04±0.01	-24.24±0.57
MOL004982	-14.38±0.11	-185.51±0.32	166.34±0.29	-0.82±0.00	-15.80±0.01	-18.57±0.13
MOL012950	-24.69±0.29	-177.21±0.07	162.16±0.42	-1.40±0.00	-17.62±0.03	-23.52±0.57
MOL012951	-24.78±0.07	-123.07±0.72	111.92±0.86	-1.29±0.01	-17.14±0.05	-20.08±0.13


DUSP7-M2
DUSP7-M4
DUSP7-M5
DUSP7-M13-1
DUSP7-M13-3
DUSP7-M14-1
DUSP7-M14-3
DUSP7-M17-1
DUSP7-M17-2
DUSP7-M18
DUSP7-M19-1
DUSP7-M22-1
RMSD (Å)
Simulation Time (ns)



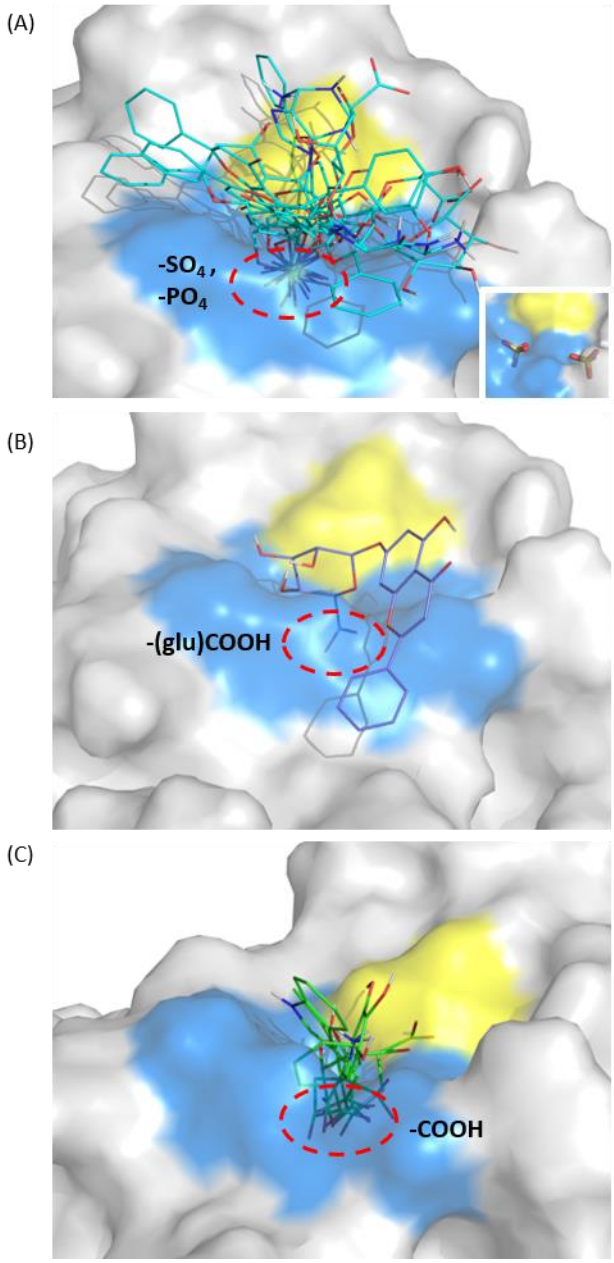
Appendix Figure 5 RMSD ~ Simulation Time (RST) plots for DUSP7-ligand complexes

Legend description is the same as Appendix Appendix Figure 2.

Appendix Table 3. Individual energy terms for DUSP7-ligand binding free energy calculations using the MM-PBSA method (energies in kcal/mol)

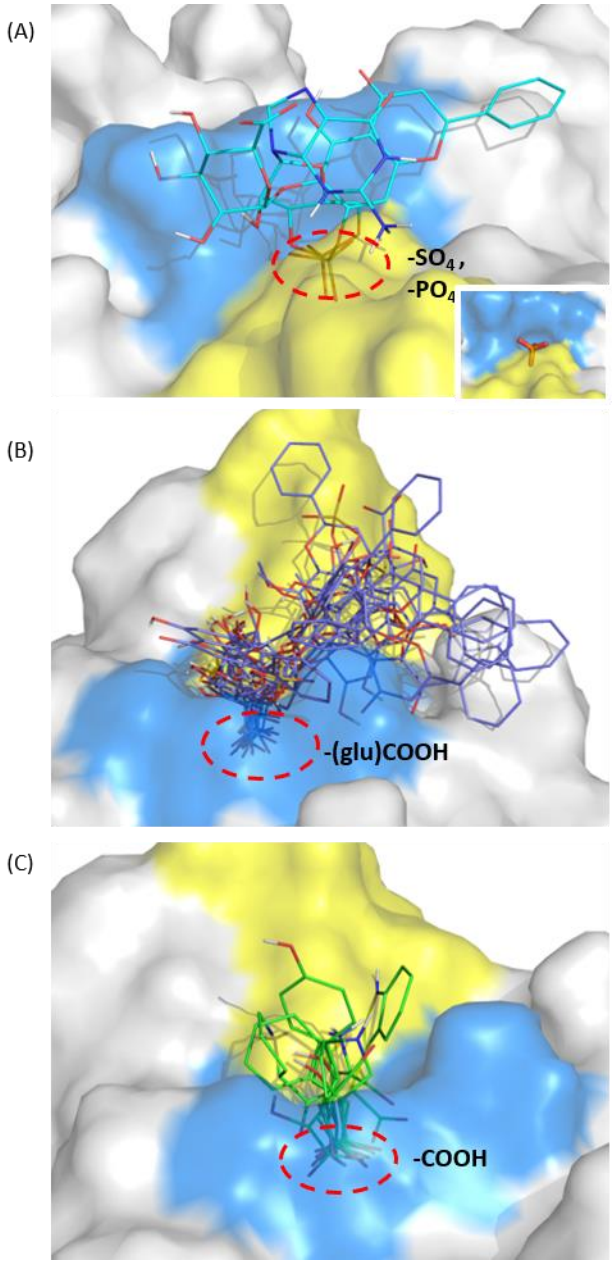
Ligands with a positive value of $\Delta G_{MM/PBSA}$ are shown in red and considered as false positives of docking screening.

LIGAND	ΔE_{vdW}^{gas}	ΔE_{eel}^{gas}	ΔG_{p}^{Sol}	ΔG_{np}^{Sol}	$T\Delta S$	$\Delta G_{MM/PBSA}$
M2	-28.05±0.01	-334.82±0.74	311.93±0.46	-2.57±0.02	-20.91±0.02	-32.60±0.27
M4	-34.99±0.21	-229.22±2.09	223.36±1.39	-2.58±0.01	-20.10±0.05	-23.33±0.53
M5	-28.56±0.44	-243.84±0.83	226.98±0.84	-2.34±0.01	-20.00±0.04	-27.76±0.24
M13-1	-29.89±0.18	-345.54±0.56	321.20±0.57	-2.33±0.02	-20.10±0.05	-36.46±0.01
M13-3	-21.43±0.20	-335.62±0.36	311.08±0.68	-1.62±0.00	-18.41±0.02	-29.18±0.21
M14-1	-33.11±0.11	-251.07±1.06	256.44±1.43	-2.18±0.01	-18.56±0.02	-11.36±0.37
M14-3	-27.75±0.06	-335.77±0.48	321.36±0.47	-1.88±0.01	-18.76±0.20	-25.27±0.30
M17-1	-29.85±0.13	-171.55±0.17	193.87±0.01	-2.38±0.01	-18.09±0.00	8.19±0.32
M17-2	-29.09±0.16	-212.70±0.87	204.59±0.71	-2.24±0.01	-18.70±0.02	-20.74±0.35
M18	-33.30±0.56	-340.89±0.60	323.14±0.70	-2.44±0.03	-20.73±0.09	-32.76±0.49
M19-1	-33.38±0.26	-330.13±0.48	326.51±0.16	-2.69±0.01	-19.66±0.06	-20.04±0.13
M22-1	-25.91±0.32	-211.58±1.35	203.09±1.13	-2.09±0.03	-18.27±0.10	-18.22±0.30
MOL000061	-11.08±0.14	-69.25±0.29	72.11±0.54	-0.71±0.00	-14.84±0.01	5.91±0.31
MOL000065	-9.96±0.20	-278.89±0.89	252.88±0.72	-0.81±0.00	-16.03±0.03	-20.76±0.33
MOL000771	-12.49±0.33	-230.19±0.68	209.35±0.12	-0.92±0.01	-14.86±0.03	-19.39±0.33
MOL001468	-8.98±0.22	-397.56±1.04	370.55±0.69	-0.73±0.00	-15.85±0.02	-20.86±0.22
MOL001737	-15.32±0.19	-218.12±0.20	201.21±0.19	-1.01±0.01	-15.24±0.02	-18.00±0.17
MOL001801	-14.80±0.04	-198.65±0.40	184.16±0.10	-0.82±0.00	-15.26±0.01	-14.84±0.32
MOL001841	-12.98±0.06	-230.24±0.85	211.94±0.45	-1.13±0.00	-15.43±0.03	-16.98±0.43
MOL001886	-10.45±0.29	-376.06±1.16	354.42±1.13	-0.81±0.00	-15.99±0.03	-16.92±0.22
MOL001931	-11.80±0.02	-209.91±0.28	189.50±0.35	-0.79±0.00	-15.68±0.02	-17.32±0.11
MOL004480	-5.18±0.32	-238.60±0.46	209.04±0.17	-0.25±0.00	-14.10±0.01	-20.89±0.12
MOL012951	-26.83±0.17	-185.28±0.63	175.53±0.57	-1.70±0.02	-17.81±0.03	-20.47±0.15



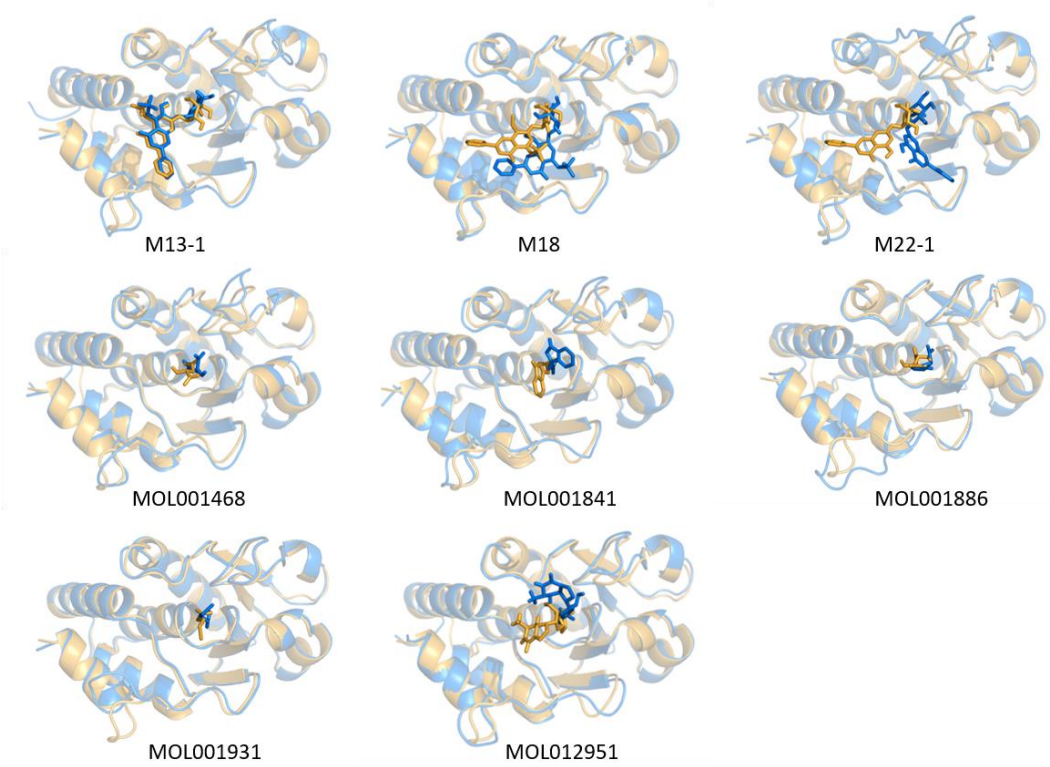
Appendix Figure 6 Common substituent groups of advantageous DUSP5 ligands from YIY-906

The protein DUSP3 is represented as surface with the P-loop colored blue and the D-loops colored yellow. Ligand binding modes are demonstrated using the representative conformations of MD simulations. Specifically, sulfate ions in the crystal structure of DUSP5 are shown in the right corner of Panel (A). Common substituent groups (-SO_4, -PO_4, and -COOH) are marked with dashed red circles. (A) the binding of phosphate groups and sulfate groups in the protein's active pocket; (B) the binding of carboxyl groups of glucuronic acids on YIV-906 metabolites in the protein's active site; (C) the binding of carboxyl groups of original YIV-906 chemicals in the protein's active site. The perspective of each ligand-protein complex was slightly adjusted to achieve a clearer view. The surface transparency was set to 60% to show the parts of ligands which stick into the pocket but are concealed by nearby surface.



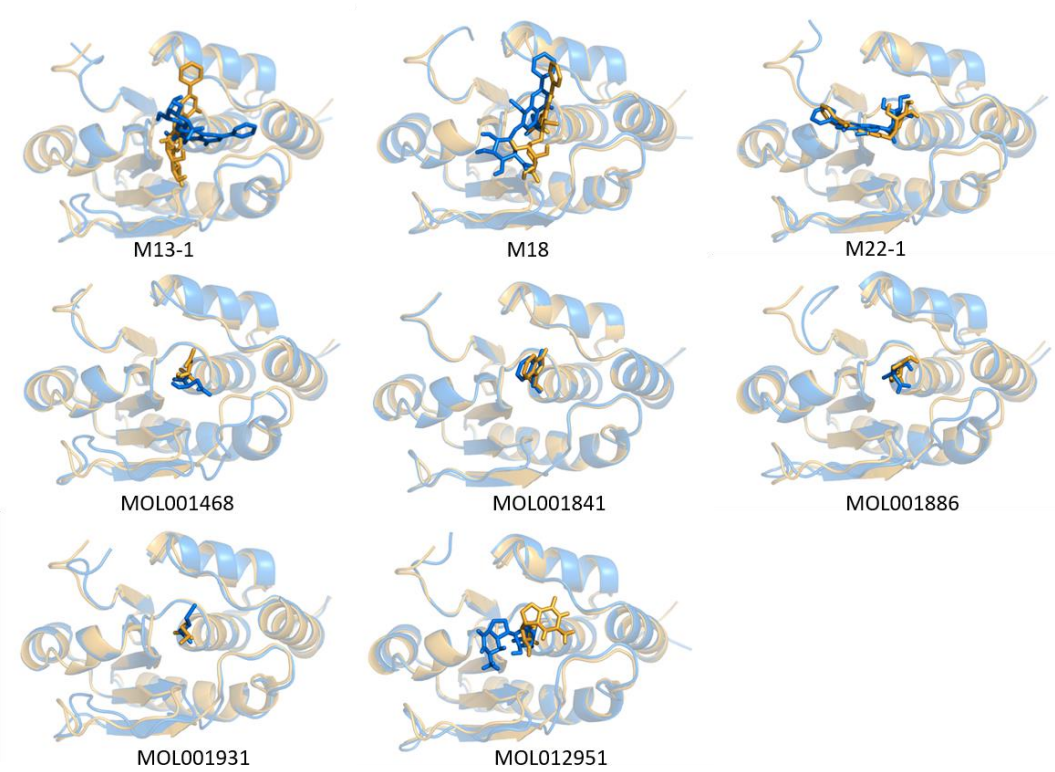
Appendix Figure 7 Common substituent groups of advantageous DUSP7 ligands from YIY-906

Specifically, the phosphate ion in the crystal structure of DUSP7 are shown in the right corner of Panel (A). Other captions are the same as that of Appendix Appendix Figure 6.



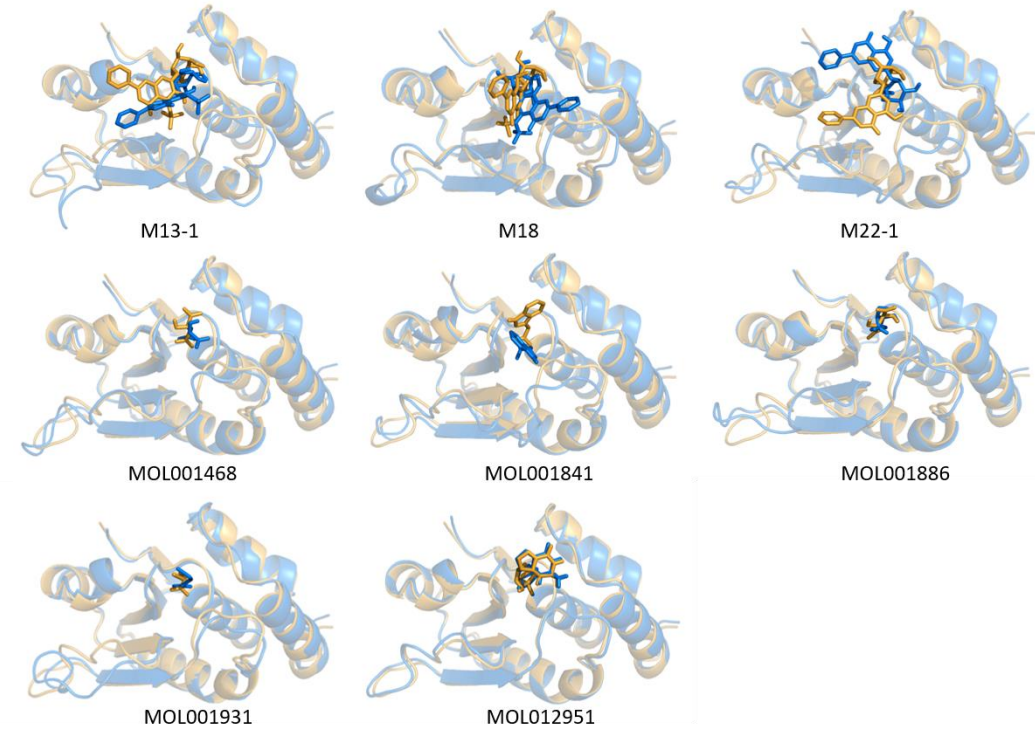
Appendix Figure 8 Comparison of the binding modes predicted by docking and MD simulations for DUSP3

For each DUSP3-ligand complex, the representative MD snapshot, which has the smallest RMSD to the average structure of collected MD frames, was selected and aligned to the crystal structure. The receptor is represented using golden and marine blue cartoons for the docking and MD modes, respectively. The common ligands are shown as sticks using the same color scheme.



Appendix Figure 9 Comparison of the binding modes predicted by docking and MD simulations for DUSP5

For each DUSP5-ligand complex, the representative MD snapshot, which has the smallest RMSD to the average structure of collected MD frames, was selected and aligned to the crystal structure. Other captions are the same as that of Appendix Appendix Figure 8.



Appendix Figure 10 Comparison of the binding modes predicted by docking and MD simulations for DUSP7

For each DUSP7-ligand complex, the representative MD snapshot, which has the smallest RMSD to the average structure of collected MD frames, was selected and aligned to the crystal structure. Other captions are the same as that of Appendix Appendix Figure 8.

Bibliography

1. Patterson, Kate I., et al., *Dual-specificity phosphatases: critical regulators with diverse cellular targets.* Biochemical Journal, 2009. **418**(3): p. 475-489.
2. Kidger, A.M. and S.M. Keyse, *The regulation of oncogenic Ras/ERK signalling by dual-specificity mitogen activated protein kinase phosphatases (MKPs).* Semin Cell Dev Biol, 2016. **50**: p. 125-32.
3. Cagnol, S. and J.C. Chambard, *ERK and cell death: mechanisms of ERK-induced cell death--apoptosis, autophagy and senescence.* FEBS J, 2010. **277**(1): p. 2-21.
4. Samatar, A.A. and P.I. Poulikakos, *Targeting RAS-ERK signalling in cancer: promises and challenges.* Nat Rev Drug Discov, 2014. **13**(12): p. 928-42.
5. Keyse, S.M., *Dual-specificity MAP kinase phosphatases (MKPs) and cancer.* Cancer Metastasis Rev, 2008. **27**(2): p. 253-61.
6. Tewari, R., et al., *Involvement of miltefosine-mediated ERK activation in glioma cell apoptosis through Fas regulation.* J Neurochem, 2008. **107**(3): p. 616-27.
7. Cagnol, S., E. Van Obberghen-Schilling, and J.C. Chambard, *Prolonged activation of ERK1,2 induces FADD-independent caspase 8 activation and cell death.* Apoptosis, 2006. **11**(3): p. 337-46.
8. Cheng, Y., et al., *ERK and JNK mediate TNFalpha-induced p53 activation in apoptotic and autophagic L929 cell death.* Biochem Biophys Res Commun, 2008. **376**(3): p. 483-8.
9. Ogier-Denis, E., et al., *Erk1/2-dependent phosphorylation of Galpha-interacting protein stimulates its GTPase accelerating activity and autophagy in human colon cancer cells.* J Biol Chem, 2000. **275**(50): p. 39090-5.
10. Lam, W., et al., *Mechanism Based Quality Control (MBQC) of Herbal Products: A Case Study YIV-906 (PHY906).* Front Pharmacol, 2018. **9**: p. 1324.
11. Lam, W., et al., *PHY906(KD018), an adjuvant based on a 1800-year-old Chinese medicine, enhanced the anti-tumor activity of Sorafenib by changing the tumor microenvironment.* Sci Rep, 2015. **5**: p. 9384.
12. Saif, M.W., et al., *Phase I study of the botanical formulation PHY906 with capecitabine in advanced pancreatic and other gastrointestinal malignancies.* Phytomedicine, 2010. **17**(3-4): p. 161-9.
13. Saif, M.W., et al., *First-in-human phase II trial of the botanical formulation PHY906 with capecitabine as second-line therapy in patients with advanced pancreatic cancer.* Cancer Chemother Pharmacol, 2014. **73**(2): p. 373-80.
14. Yen, Y., et al., *Phase I/II study of PHY906/capecitabine in advanced hepatocellular carcinoma.* Anticancer Res, 2009. **29**(10): p. 4083-92.
15. Kummar, S., et al., *A phase I study of the chinese herbal medicine PHY906 as a modulator of irinotecan-based chemotherapy in patients with advanced colorectal cancer.* Clin Colorectal Cancer, 2011. **10**(2): p. 85-96.
16. Wu, S., et al., *Multidentate small-molecule inhibitors of vaccinia H1-related (VHR) phosphatase decrease proliferation of cervix cancer cells.* J Med Chem, 2009. **52**(21): p. 6716-23.

17. Pavic, K., G. Duan, and M. Kohn, *VHR/DUSP3 phosphatase: structure, function and regulation.* FEBS J, 2015. **282**(10): p. 1871-90.
18. Jeong, D.G., et al., *Crystal structure of the catalytic domain of human DUSP5, a dual specificity MAP kinase protein phosphatase.* Proteins, 2007. **66**(1): p. 253-8.
19. Lountos, G.T., et al., *Structure of human dual-specificity phosphatase 7, a potential cancer drug target.* Acta Crystallogr F Struct Biol Commun, 2015. **71**(Pt 6): p. 650-6.
20. Kidger, A.M., et al., *Dual-specificity phosphatase 5 controls the localized inhibition, propagation, and transforming potential of ERK signaling.* Proc Natl Acad Sci U S A, 2017. **114**(3): p. E317-E326.
21. Kalyaanamoorthy, S. and Y.P. Chen, *Structure-based drug design to augment hit discovery.* Drug Discov Today, 2011. **16**(17-18): p. 831-9.
22. Ferreira, L.G., et al., *Molecular Docking and Structure-Based Drug Design Strategies.* Molecules, 2015. **20**(7): p. 13384-13421.
23. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
24. Meng, X.Y., et al., *Molecular docking: a powerful approach for structure-based drug discovery.* Curr Comput Aided Drug Des, 2011. **7**(2): p. 146-57.
25. Barradas-Bautista, D., et al., *Structural Prediction of Protein-Protein Interactions by Docking: Application to Biomedical Problems.* Adv Protein Chem Struct Biol, 2018. **110**: p. 203-249.
26. Kapetanovic, I.M., *Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach.* Chem Biol Interact, 2008. **171**(2): p. 165-76.
27. Lopez-Vallejo, F., et al., *Integrating virtual screening and combinatorial chemistry for accelerated drug discovery.* Comb Chem High Throughput Screen, 2011. **14**(6): p. 475-87.
28. Huang, S.Y. and X. Zou, *Advances and challenges in protein-ligand docking.* Int J Mol Sci, 2010. **11**(8): p. 3016-34.
29. Mysinger, M.M. and B.K. Shoichet, *Rapid context-dependent ligand desolvation in molecular docking.* J Chem Inf Model, 2010. **50**(9): p. 1561-73.
30. Ruvinsky, A.M., *Role of binding entropy in the refinement of protein-ligand docking predictions: analysis based on the use of 11 scoring functions.* J Comput Chem, 2007. **28**(8): p. 1364-72.
31. Lee, A.C., et al., *A Comprehensive Review on Current Advances in Peptide Drug Development and Design.* Int J Mol Sci, 2019. **20**(10).
32. Wells, J.A. and C.L. McClendon, *Reaching for high-hanging fruit in drug discovery at protein-protein interfaces.* Nature, 2007. **450**(7172): p. 1001-1009.
33. Hollingsworth, S.A. and R.O. Dror, *Molecular Dynamics Simulation for All.* Neuron, 2018. **99**(6): p. 1129-1143.
34. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules (vol 9, pg 646, 2002).* Nature Structural Biology, 2002. **9**(10): p. 788-788.
35. Chodera, J.D., et al., *Alchemical free energy methods for drug discovery: progress and challenges.* Current Opinion in Structural Biology, 2011. **21**(2): p. 150-160.
36. Friesner, R.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.* J Med Chem, 2004. **47**(7): p. 1739-49.
37. Jorgensen, W.L., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water.* Journal of Chemical Physics, 1983. **79**(2): p. 926-935.

38. Bayly, C.I., et al., *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges- the RESP model.* Journal Of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
39. Frisch, M.J.T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J., *Gaussian 16, Revision B.01*. Gaussian, Inc., Wallingford CT, 2016.
40. Hardcastle, I.R., et al., *N2-substituted O6-cyclohexylmethylguanine derivatives: potent inhibitors of cyclin-dependent kinases 1 and 2.* J Med Chem, 2004. **47**(15): p. 3710-22.
41. Case, D.A., et al., *AMBER 2018.* University of California, San Francisco, 2018.
42. Wang, J., et al., *Automatic atom type and bond type perception in molecular mechanical calculations.* J Mol Graph Model, 2006. **25**(2): p. 247-260.
43. Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* J Chem Theory Comput, 2015. **11**(8): p. 3696-713.
44. Darden, T., et al., *New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations.* Structure, 1999. **7**(3): p. R55-60.
45. Forester, T.R. and W. Smith, *SHAKE, rattle, and roll: Efficient constraint algorithms for linked rigid bodies.* Journal of Computational Chemistry, 1998. **19**(1): p. 102-111.
46. Sitkoff, D., K.A. Sharp, and B. Honig, *Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models.* Journal of Physical Chemistry, 1994. **98**(7): p. 1978-1988.
47. Wang, J.M. and T.J. Hou, *Develop and Test a Solvent Accessible Surface Area-Based Model in Conformational Entropy Calculations.* Journal of Chemical Information and Modeling, 2012. **52**(5): p. 1199-1212.
48. Hawkins, G.D., C.J. Cramer, and D.G. Truhlar, *Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium.* Journal of Physical Chemistry, 1996. **100**(51): p. 19824-19839.
49. Homeyer, N., et al., *AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine.* J Mol Model, 2006. **12**(3): p. 281-9.
50. Steinbrecher, T., J. Latzer, and D.A. Case, *Revised AMBER parameters for bioorganic phosphates.* J Chem Theory Comput, 2012. **8**(11): p. 4405-4412.
51. Pierce, B.G., et al., *ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers.* Bioinformatics, 2014. **30**(12): p. 1771-1773.

52. Vanommeslaeghe, K., et al., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.* J Comput Chem, 2010. **31**(4): p. 671-90.
53. Zhang, W., et al., *Identification of chemicals and their metabolites from PHY906, a Chinese medicine formulation, in the plasma of a patient treated with irinotecan and PHY906 using liquid chromatography/tandem mass spectrometry (LC/MS/MS).* J Chromatogr A, 2010. **1217**(37): p. 5785-93.
54. Ru, J., et al., *TCMSP: a database of systems pharmacology for drug discovery from herbal medicines.* J Cheminform, 2014. **6**: p. 13.
55. Schumacher, M.A., et al., *Structural basis for the recognition of a bisphosphorylated MAP kinase peptide by human VHR protein Phosphatase.* Biochemistry, 2002. **41**(9): p. 3009-17.
56. Jie Dou1, Zhou Wang1, Leon Ma2, Bo Peng2, Ke Mao1, Chengqin Li1, Mengqi Su1, Changlin Zhou1,* and Guangyong Peng2,*, *Baicalein and baicalin inhibit colon cancer using two distinct fashions of apoptosis and senescence.* Oncotarget, 2018. **9**.
57. He, X.M., V. H.; Ji, B.; Xie, X.-Q.; Wang, J. , *Calculate Protein-Ligand Binding Affinities with the Extended Linear Interaction Energy Method: Application on the Cathepsin S Set in the D3R Grand Challenge 3.* Journal of Computer-Aided Molecular Design, 2018: p. Accepted.
58. Genheden, S. and U. Ryde, *The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities.* Expert Opin Drug Discov, 2015. **10**(5): p. 449-61.
59. Wang, J., et al., *Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA.* Journal of the American Chemical Society, 2001. **123**(22): p. 5221-5230.
60. Chen, F., et al., *Assessing the performance of MM/PBSA and MM/GBSA methods. 8. Predicting binding free energies and poses of protein-RNA complexes.* RNA, 2018. **24**(9): p. 1183-1194.
61. Wang, J., et al., *Hierarchical database screenings for HIV-1 reverse transcriptase using a pharmacophore model, rigid docking, solvation docking, and MM-PB/SA.* J Med Chem, 2005. **48**(7): p. 2432-44.
62. Kuhn, B., et al., *Validation and use of the MM-PBSA approach for drug discovery.* J Med Chem, 2005. **48**(12): p. 4040-8.
63. Barr, A.J., et al., *Large-scale structural analysis of the classical human protein tyrosine phosphatome.* Cell, 2009. **136**(2): p. 352-63.
64. Pagnussat, G.C., et al., *Nitric oxide mediates the indole acetic acid induction activation of a mitogen-activated protein kinase cascade involved in adventitious root development.* Plant Physiol, 2004. **135**(1): p. 279-86.
65. Katagi, A., et al., *Inhibitory effect of isoamericanol A from Jatropha curcas seeds on the growth of MCF-7 human breast cancer cell line by G2/M cell cycle arrest.* Heliyon, 2016. **2**(1): p. e00055.
66. Purves D, A.G., Fitzpatrick D, *Second Messenger Targets: Protein Kinases and Phosphatases.* 2001.
67. Aburima, A., et al., *cGMP signaling inhibits platelet shape change through regulation of the RhoA-Rho Kinase-MLC phosphatase signaling pathway.* Journal of Thrombosis and Haemostasis, 2017. **15**(8): p. 1668-1678.

68. Hou, T., et al., *Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations.* J Chem Inf Model, 2011. **51**(1): p. 69-82.
69. Hou, T., et al., *Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking.* J Comput Chem, 2011. **32**(5): p. 866-77.
70. Wang, J.M., T.J. Hou, and X.J. Xu, *Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models.* Current Computer-Aided Drug Design, 2006. **2**(3): p. 287-306.
71. Wang, E., et al., *End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design.* Chem Rev, 2019.
72. Gaulton, A., et al., *ChEMBL: a large-scale bioactivity database for drug discovery.* Nucleic Acids Res, 2012. **40**(Database issue): p. D1100-7.
73. Xu, L., et al., *Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models.* J Phys Chem B, 2013. **117**(28): p. 8408-21.
74. Xu, C., et al., *Activation of the SphK1/ERK/p-ERK pathway promotes autophagy in colon cancer cells.* Oncol Lett, 2018. **15**(6): p. 9719-9724.