JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Identification of Sumoylation Activating Enzyme 1 Inhibitors by Structure-Based Virtual Screening

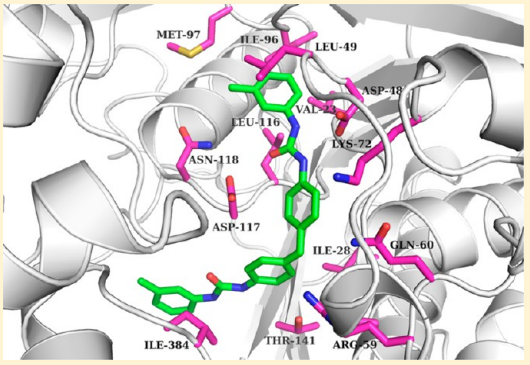
Ashutosh Kumar,[†] Akihiro Ito,[‡] Mikako Hirohama,[‡] Minoru Yoshida,[‡] and Kam Y. J. Zhang[†,*]

[†]Zhang Initiative Research Unit, Advanced Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

[‡]Chemical Genetics Laboratory/Chemical Genomics Research Group, Advanced Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Supporting Information

ABSTRACT: SUMO activating enzyme 1 (SUMO E1) is responsible for the activation of SUMO in the first step of the sumoylation cascade. SUMO E1 is linked to many human diseases including cancer, thus making it a potential therapeutic target. There are few reported SUMO E1 inhibitors including several natural products. To identify small molecule inhibitors of SUMO E1 with better drug-like properties for potential therapeutic studies, we have used structure-based virtual screening to identify hits from the Maybridge small molecule library for biological assay. Our virtual screening protocol involves fast docking of the entire small molecule library with rigid protein and ligands followed by redocking of top hits using a method that incorporates both ligand and protein flexibility. Subsequently, the top-ranking compounds were prioritized using the molecular dynamics simulation-based binding free energy calculation. Out of 24 compounds that were acquired and tested using in vitro sumoylation assay, four of them showed more than 85% inhibition of sumoylation with the most active compound showing an IC_{50} of 14.4 μM. A similarity search with the most active compound in the ZINC database has identified three more compounds with improved potency. These compounds share a common phenyl urea scaffold and have been confirmed to inhibit SUMO E1 by in vitro SUMO-1 thioester bond formation assay. Our study suggests that these phenyl urea compounds could be used as a starting point for the development of novel therapeutic agents.



INTRODUCTION

Protein sumoylation is a reversible post-translational modification of substrate proteins affecting diverse cellular processes including DNA replication and repair, chromosome packing and dynamics, genome integrity, nuclear transport, signal transduction, and cell proliferation.[1−6] In sumoylation, small ubiquitin like modifier (SUMO) protein is covalently attached to the ϵ-amino group of lysine residues in specific substrate proteins via an enzymatic cascade that requires sequential action of a set of enzymes that includes an activating enzyme E1, a conjugating enzyme E2, and a ligase E3 (Figure 1). SUMO proteins are expressed as precursors, and they are processed by SUMO specific proteases (SENPs) exposing two C-terminal glycines to generate mature SUMO.[5,7−10] The conjugation of SUMO protein to SUMO activating enzyme 1 (SUMO E1) is the first step in the sumoylation pathway and involves the adenylation of SUMO C-terminus to form SUMO-AMP intermediate, which is followed by the transfer of SUMO to the catalytic residue Cys173 on SUMO E1.[11−13] In the next step, SUMO protein is transferred from SUMO E1 to SUMO E2 (also known as Ubc9), again with the formation of thioester linkage between SUMO protein C-terminal glycine and the active site Cys93 in Ubc9. Finally, Ubc9 catalyzes the attachment of SUMO protein to lysine residue of substrate proteins. The efficiency of this step is increased by SUMO E3 ligase protein which associates both with Ubc9 and substrate protein to stimulate protein sumoylation.[2,14] As sumoylation is a reversible process, therefore the SUMO protein is recycled from its substrate conjugated form by the action of SENPs.[5,7−10]

Sumoylation is an important pathway for the normal functioning of the cell. However, over the past few years several studies suggest the role of sumoylation proteins in pathogenesis of human diseases like cancer, neurodegenerative diseases, and heart diseases.[15,16] Specifically, several studies indicate the involvement of sumoylation in cancer development and tumorogenesis.[15−17] All sumoylation proteins have been associated with cancer as elevated levels of SUMO E1,[18] Ubc9,[19,20] SUMO E3,[21] and SENPs[22−24] have been observed in various cancers. Also, sumoylation regulates the activity of important tumor suppressor proteins including p53, retinoblastoma protein (pRB), p63, p73, and murine double minute 2 (MDM2).[15−17]

Among all sumoylation proteins, SUMO E1 is the first enzyme of sumoylation cascade and is the only SUMO

Received: December 20, 2012
Published: April 1, 2013

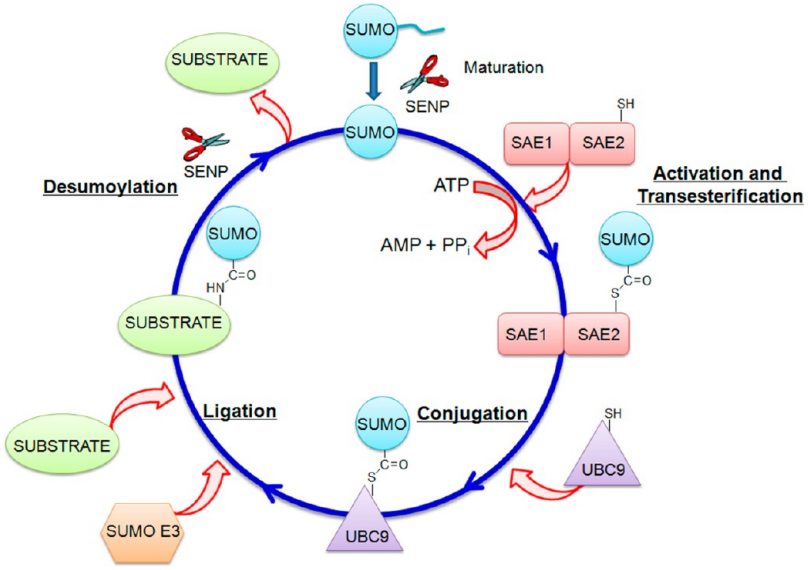


Figure 1. Schematic representation of the sumoylation pathway.

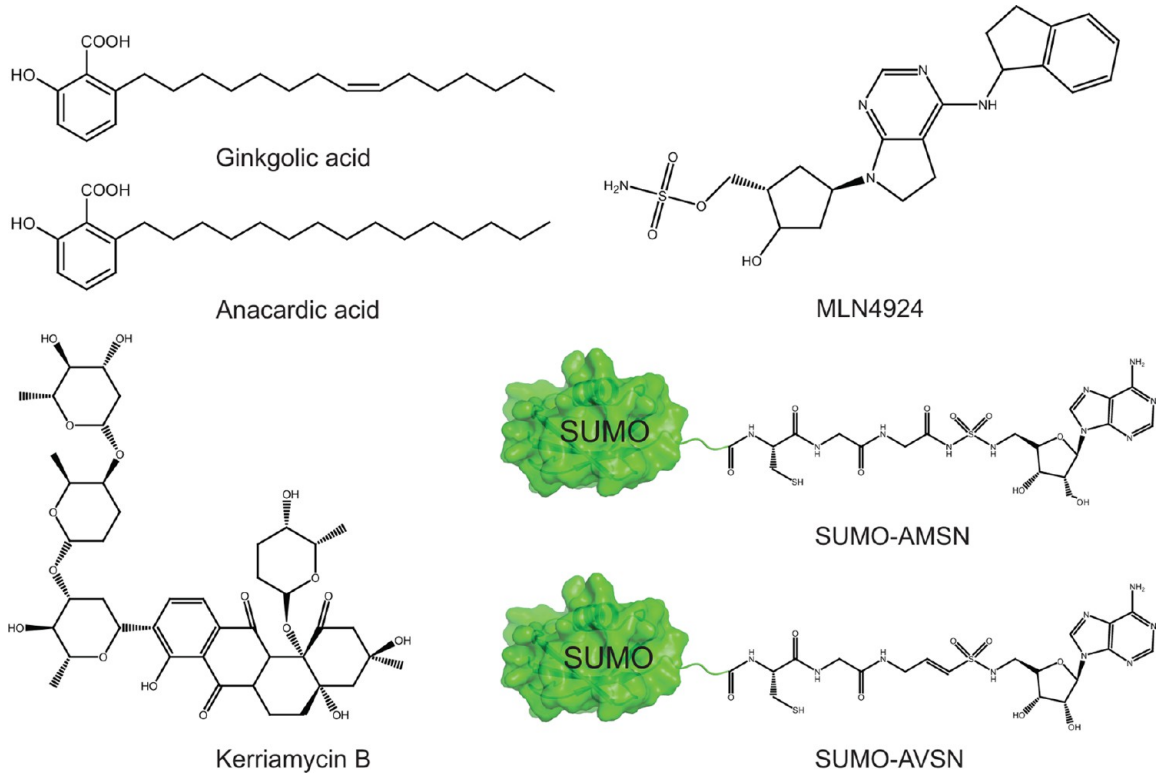


Figure 2. Previously reported inhibitors of sumoylation activating enzyme 1 (SUMO E1).

activating enzyme in the sumoylation pathway known to date.[11] SUMO E1 is a heterodimer consisting of two subunits SAE1 and SAE2 that are also known as Aos1 and Uba2.[11] SUMO E1 is known to tightly regulate the SUMO modifications. The SAE2 subunit of SUMO E1 is autosumoylated at Lys236, which inhibits its interaction with SUMO E2, preventing the transfer of SUMO from E1 to E2 and thereby affecting SUMO conjugation to target proteins.[25] There is evidence for direct involvement of SUMO E1 against different cancer. Lee and Thorgeirsson[18] reported direct correlation between higher levels of SUMO E1 enzyme and lower survival rates in patients with hepatocellular carcinoma. Furthermore, Kessler et al.[26] using a genome wide RNA interference screening approach unveiled the role of SUMO E1 in the maintenance of the Myc tumorigenic phenotype. They reported that SUMO E1 is essential for Myc-dependent tumor growth in mice, and there is correlation between low SUMO E1 levels and longer metastasis-free survival of patients with Myc-high human breast cancers. They have also identified that the inactivation of SUMO E1 initiates synthetic lethality with c-Myc and promotes mitotic catastrophe and cell death in cancer cells and not in normal cells.

The above studies suggest that SUMO E1 is an attractive target for drug development against a variety of human diseases

dx.doi.org/10.1021/ci300618e | J. Chem. Inf. Model. 2013, 53, 809−820

especially against cancer. However, until now only a few inhibitors of SUMO E1 have been reported in the literature (Figure 2).[13,27−29] These include three natural products, ginkgolic acid, anacardic acid, and kerriamycin B, identified by Fukuda et al.[28,29] These three natural products inhibit in vitro protein sumoylation with IC_{50} values of 3.0, 2.2, and 11.7 μM for ginkgolic acid, anacardic acid, and kerriamycin B, respectively. All of these inhibitors also inhibited the in vivo sumoylation by directly binding to SUMO E1 and inhibiting the formation of E1-SUMO intermediate. These inhibitors were also selective to SUMO E1 and do not affect ubiquitination. Soucy et al.[27] reported an AMP mimic MLN4924 as a selective and highly potent inhibitor of NEDD8-activating enzyme with an IC_{50} of 4.7 nM. MLN4924 also showed inhibition toward SUMO E1 and Ubiquitin E1 with IC_{50} values of 8.2 and 1.56 μM, respectively. Lu et al.[13] designed semi-synthetic mechanism based protein inhibitors (SUMO-AMSN and SUMO-AVSN) of SUMO E1 and Ubiquitin E1. They used a chemical approach that involves the synthesis of SUMO and Ubiquitin derivatives that are selective for their cognate E1s. These derivatives mimic either adenylate intermediate or the tetrahedral intermediate generated during thioester bond formation.

Some of the above-mentioned inhibitors were effectively used as probe molecules to the study biological mechanism of sumoylation.[12] However, it is desirable to identify drug-like chemical scaffolds that are suitable for chemical optimization and development into therapeutic agents against numerous human diseases like cancer. Motivated by the need of novel chemical scaffolds for the development of SUMO E1 inhibitors, we have carried out structure-based virtual screening targeting of the ATP binding site of SUMO E1. Our study identified phenyl urea containing compounds as a novel class of SUMO E1 inhibitors. The in vitro sumoylation and SUMO E1 inhibitory activity reported in this study suggests that the phenyl urea scaffold could be used as a starting point for the development of novel therapeutic agents against cancer, especially c-Myc driven human cancers such as breast cancer and medullablastoma, by targeting SUMO E1 and the sumoylation pathway.

MATERIALS AND METHODS

Preparation of Ligand and Receptor Structures for Molecular Docking. Maybridge screening collection was downloaded from the vendor Web site and converted into a three-dimensional format using OMEGA.[30] SUMO E1 crystal structure (PDB code: 1Y8Q) was downloaded from the RCSB Protein Data Bank[31] (http://www.rcsb.org) and used for virtual screening. To prepare receptor structure for molecular docking, hydrogens were added, bond orders were assigned, and all heteroatoms were removed. Protonation states of charged residues were determined using Protonate3D in MOE2011.10.[32] The pdbqt files for the receptor and ligands used in Autodock-Vina[33] docking were accomplished using the python scripts from Autodock Tools package.[34] For RosettaLigand docking, the receptor was prepared by repacking the side chains to avoid pre-existing clashes according to Rosetta energy function. The "ligand_rpkmin" protocol of the Rosetta suite[35] was used for repacking the side chains. RosettaLigand uses pre-generated ligand conformations; therefore, an ensemble of a maximum of 200 conformations for each ligand was generated using OMEGA.[30] Partial charges to the ligands for RosettaLigand docking were assigned according to AM1BCC the semi-empirical method.[36]

Molecular Docking of Maybridge Screening Library. Maybridge small molecule library, consisting of 77 931 compounds, was docked into the ATP binding site of SUMO E1 using a two-step docking procedure. First, the whole library was docked using Autodock-Vina,[33] which is a fast docking program to remove compounds with non-compatible geometries and energetics toward the ligand-binding site. Autodock-Vina uses an "iterated local search global optimizer" algorithm to evaluate the interactions between a ligand and its receptor. The search space for docking was restricted in a cubic box of 40 Å × 40 Å × 40 Å centered on the ATP position in the SUMO E1 binding site. Autodock-Vina calculations were carried out using a default value of 8 for exhaustiveness of the global search. Twenty poses were generated for each ligand that were ranked by Autodock-Vina's empirical scoring functions, and the pose with lowest docking score value was selected as the best pose. Fast docking programs generally do not consider full protein flexibility; therefore, top-ranking compounds from the first phase were redocked using RosettaLigand, a fully flexible ligand and protein docking program.[37−39] RosettaLigand employs a stochastic conformational search to identify low-energy protein ligand complexes. RosettaLigand docking starts with the placement of pregenerated ligand conformations into the receptor binding site. Amino acid side-chain rotamers are then placed around the ligand in receptor binding sites, which is followed by optimization of the randomly sampled flexible ligand poses using a Metropolis Monte Carlo simulated annealing algorithm. The docked poses were ranked using an energy function that includes terms for van der Waals attractive and repulsive forces, electrostatic interactions between pairs of amino acids, and solvation assessing the effects of side-chain−side-chain interactions and side-chain−ligand interactions, statistical energy derived from the probability of observing a side-chain conformation in the PDB, and an orientation-dependent hydrogen bonding potential. For each ligand, we generated approximately 5000 docked poses, and the top 5% of the best scoring structures were re-ranked by ligand−protein interface scores (InterfaceDelta term in RosettaLigand scores), which is the difference in the energy of protein in ligand bound and unbound states. The pose with the lowest ligand−protein interface score was then chosen as the best docked pose.

Molecular Dynamics Simulation and Binding Free Energy Calculations. Molecular mechanics Poisson−Boltzmann surface area (MM-PBSA)[40,41] and molecular mechanics generalized Born surface area (MM-GBSA)[42−44] methods as implemented in AMBER10[45] were used to calculate ligand-binding free energy, which was used for ranking of compounds selected after the docking run. To generate a conformational ensemble for the binding free energy calculation molecular dynamics simulation was carried out for compounds selected after docking. In every case, the proteins were immersed in the rectangular truncated octahedron filled with 8 Å TIP3P water molecules and neutralized by adding Na^+ or Cl^- ions. Initially, the protein system was minimized by 500 steps of steepest descent followed by 2000 steps of conjugate gradient. After minimization, the system was gradually heated in the canonical ensemble from 0 to 300 K over 50 ps and then equilibrated for 200 ps. Finally, a short 2 ns MD simulation was performed under a constant temperature of 300 K. The choice of 2 ns MD simulation was based on previous reports where short

molecular dynamics simulations were enough to achieve good correlation between experimental affinities and predicted ligand-binding free energies.[46−50] Moreover, 2 ns simulation time is a reasonable trade-off between computational time and accuracy for a large protein system in a virtual screening setup (824 amino acids for SUMO E1 protein). Additionally, a retrospective analysis revealed good correlation between MM-PBSA and MM-GBSA values calculated using 2 and 10 ns molecular dynamics trajectory (Figure S1, Supporting Information). The long-range electrostatic interactions were dealt by employing the partical mesh Ewald (PME) method.[51] All hydrogen atoms were constrained using the SHAKE algorithm,[52] and the time step was set to 2 fs. All simulations were performed by the SANDER module of Amber10[45] with the AMBER force field (ff03). Binding free energy is calculated using the following equation

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}) \quad (1)$$

The above equation can be conceptually summarized as

$$\Delta G_{\text{binding}} = \Delta G_{\text{MM}} + \Delta G_{\text{polar solvation}} + \Delta G_{\text{nonpolar solvation}} - T\Delta S \quad (2)$$

$$\Delta G_{\text{MM}} = \Delta E_{\text{vdw}} + \Delta E_{\text{elec}} + \Delta E_{\text{int}} \quad (3)$$

where ΔG_{MM} is the molecular mechanics binding free energy between small molecules and SUMO E1. The ΔE_{vdw}, ΔE_{elec}, and ΔE_{int} account for differences in van der Waals energy, electrostatic energy, and internal energy, respectively. $\Delta G_{\text{polar solvation}}$ is the polar contribution to solvation free energy and is computed by solving the Poisson−Boltzmann equation in MM-PBSA or by using Onufriev's generalized Born model[53] in MM-GBSA. $\Delta G_{\text{nonpolar solvation}}$ is the nonpolar contribution to solvation free energy and is estimated using the following equation

$$\Delta G_{\text{nonpolar solvation}} = \gamma \text{SASA} + b \quad (4)$$

where SASA is the solvent accessible surface area calculated using the LCPO method[54] for MM-PBSA and the ICOSA method[45] for MM-GBSA. The γ and b are empirical constants with default values of 0.0072 and 0, respectively. The $T\Delta S$ represents the solute entropy that was not considered in this study as we were mainly interested in the relative ranking of virtual screening hits rather than absolute binding free energies. Moreover, the computation of entropy is expensive especially in a high throughput environment, and it has been reported that the current computational methods (e.g., normal-mode analysis) are not precise enough to improve the accuracy of binding free energy calculations.[46,55] Additionally, there are several studies reported in the literature where good correlation between experimental binding affinities and predicted binding free energies have been found, even ignoring the entropic term.[47,56,57]

Hardware. The molecular docking calculations and molecular dynamics simulations were performed on a 97.4 TFLOPS Intel Xeon 5570-based Massively Parallel PC Cluster of the RIKEN Integrated Cluster of Clusters (RICC). The results were analyzed on a Dell Precision T5400 workstation with a 2.0 GHz Intel Xeon CPU.

Compounds. Tested compounds were purchased from Maybridge via a local distributor. The vendor had verified that each compound had more than 95% purity by liquid chromatography−mass spectrometry and NMR experiments. Compounds were maintained as DMSO stock solution.

In Vitro Sumoylation Assay and SUMO-1 Thioester Bond Formation. The in vitro sumoylation reaction was performed as previously described with minor modifications.[28] Briefly, the reaction mixtures containing 0.1 μg of His and T7-tagged RanGAP1-C2, 0.3 μg of GST-Aos1/Uba2, 0.01 μg of His-tagged Ubc9, and 0.1 μg of His-tagged or His-tagged and GFP-fused SUMO-1 were incubated for 2 h at 30 °C. Samples were separated by 10% SDS-PAGE followed by either immunoblotting using an anti-T7 antibody or an anti-SUMO-1 antibody for detecting sumoylated RanGAP1-C2 or exposing using ImageQuant Las 4000 (GE Healthcare) for detecting the fluorescence of GFP-fused sumoylated RanGAP1-C2. The in vitro SUMO-1 thioester bond formation assay was previously described.[28]

RESULTS AND DISCUSSION

In Silico Screening. The hierarchical in silico screening procedure employed in this study involves two stages of docking followed by prioritization of top-ranking compounds using molecular dynamics simulation and binding free energy calculations to select compounds for biological testing. The in silico screening procedure is summarized in Figure 3. Initially,

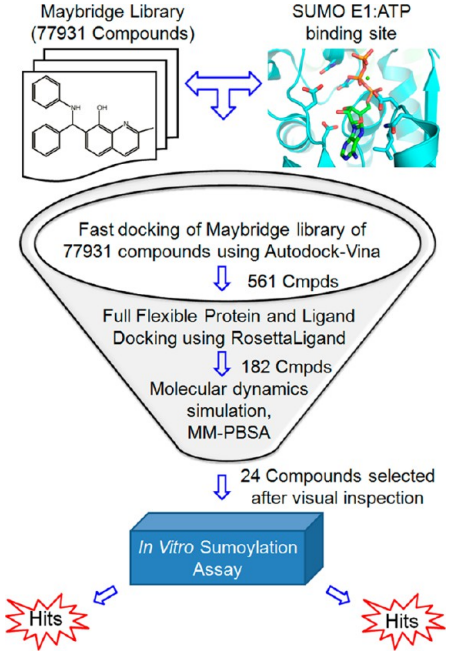


Figure 3. In silico screening scheme used to identify inhibitors of SUMO E1. Initially, Maybridge screening library was docked to SUMO E1 using a two-step docking strategy, which was followed by prioritization using a MM-PBSA binding free energy calculation and visual inspection. Compounds selected after in silico screening were acquired and tested using in vitro sumoylation assay.

77 931 compounds in the Maybridge small molecule database were docked to SUMO E1 using Autodock-Vina[33] to eliminate compounds with incompatible geometry and energetics. The ATP binding cleft, consisting of Arg21, Val23, Gly24, Ala25, Gly26, Gly27, Ile47, Asp48, Leu49, Asp50, Ser55, Asn56, Arg59, Gln60, Lys72, Asp94, Ser95, Ile96, Met97, Ala115, Leu116, Asp117, Asn118, Ala121, Lys346, and Asp347 amino acid residues, was selected as the target area for molecular

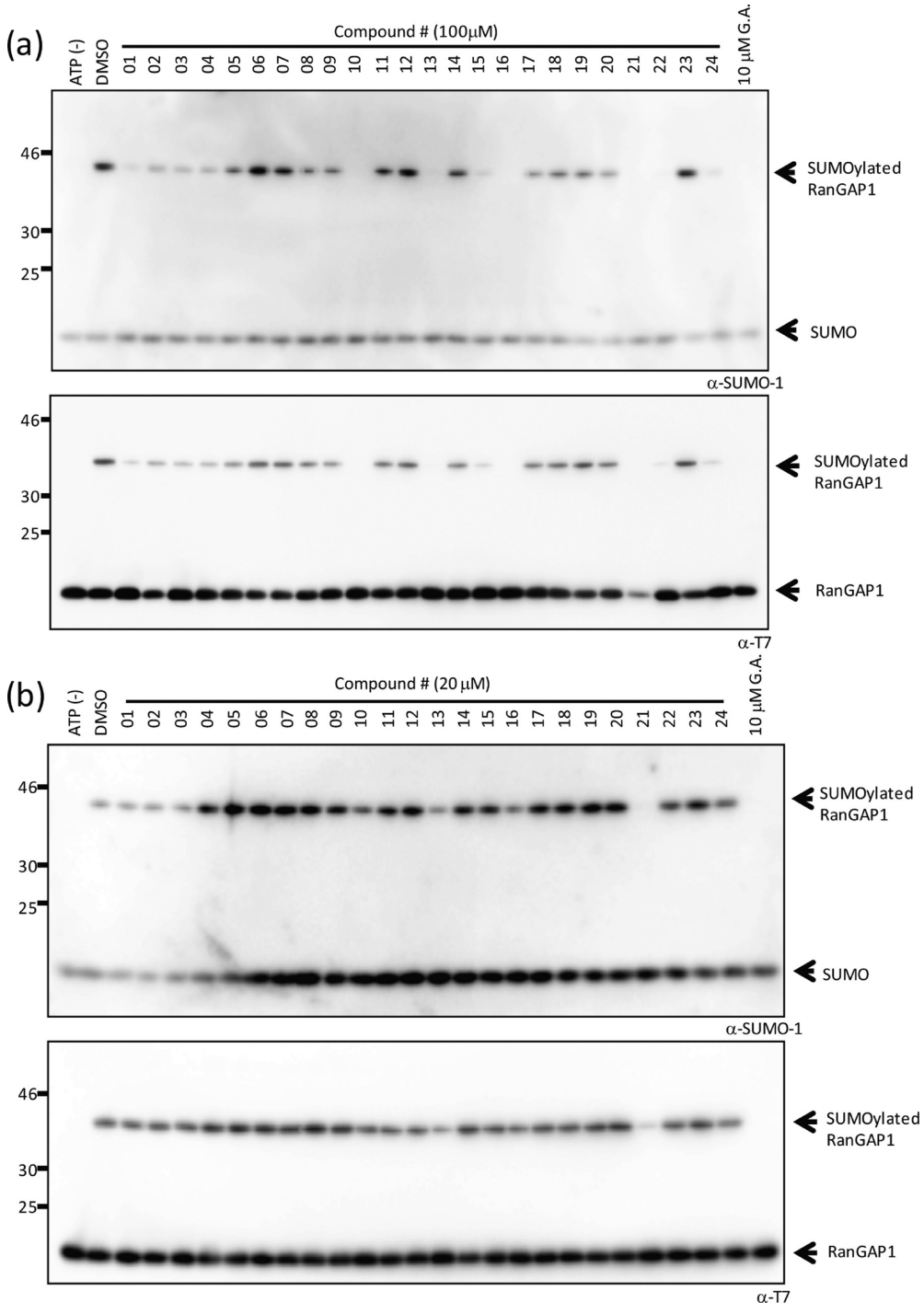


Figure 4. Inhibitory activities of in silico-selected compounds against in vitro protein sumoylation. Amounts of 100 μM (a) or 20 μM (b) of in silico-selected compounds were added to the sumoylation reaction mixture in the presence of 2 mM ATP, and sumoylated RanGAP1-C2 was detected by immunoblotting using an anti-SUMO-1 or anti-T7 antibody. Ten μM of ginkgolic acid (G.A.) was used as a positive control.

docking. The compounds were ranked using the Autodock-Vina scoring function,[33] and top-ranking compounds that met the score cutoff of −10.0 kcal/mol (561 compounds) were subjected to full flexible docking using RosettaLigand,[37−39] which takes into account the protein flexibility along with ligand flexibility. The ligand−protein interface scores (InterfaceDelta term in RosettaLigand scores), which is the difference in energy of protein in ligand bound and unbound states, were used to rank compounds, and the 182 top-ranking compounds were selected for further prioritization. In order to select compounds on the basis of their predicted binding free energy, a short 2 ns molecular dynamics simulation in an explicit water environment was then carried out to rerank top scoring compounds from RosettaLigand docking using MM-PBSA and MM-GBSA predicted binding free energy. It is important to note that the MM-PBSA and MM-GBSA predicted energies do not represent absolute binding free energy, and they are used here to rank virtual screening hits based on relative affinities. In general, MM-PBSA and MM-GBSA are reasonable approaches in predicting relative binding free energy differences, and good

dx.doi.org/10.1021/ci300618e | J. Chem. Inf. Model. 2013, 53, 809−820

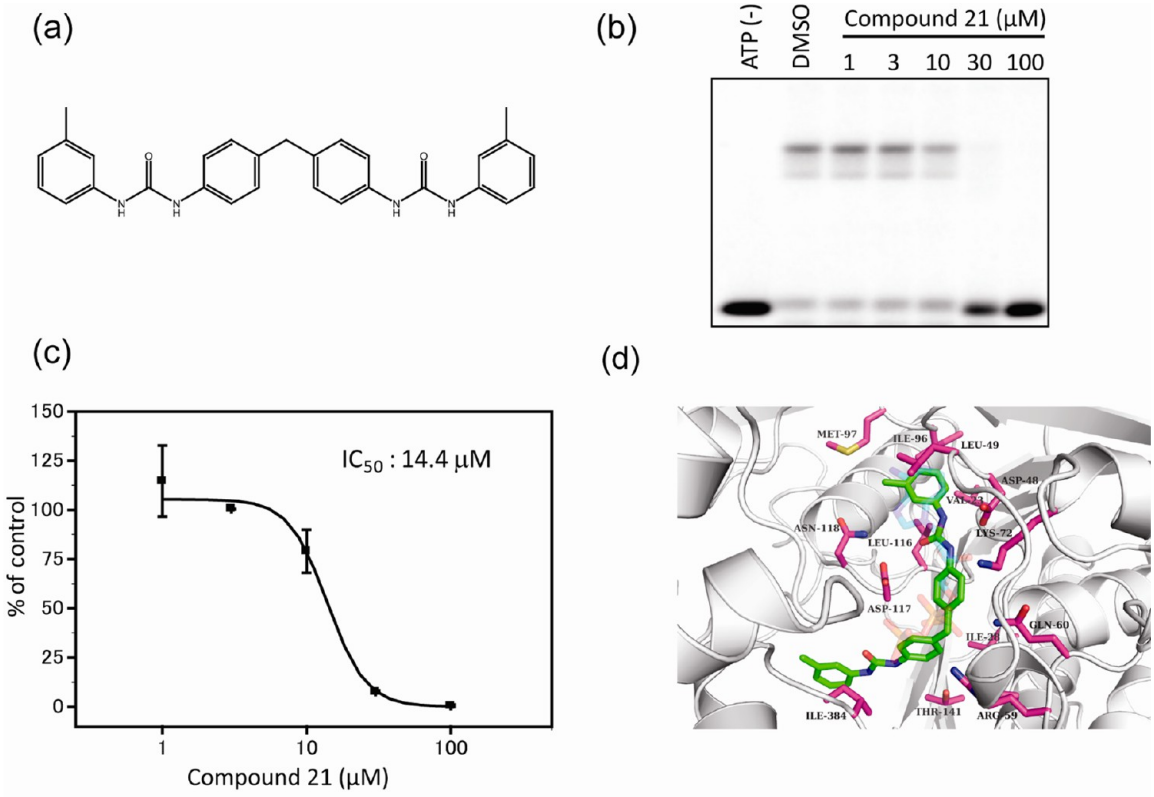


Figure 5. Compound **21** inhibited protein sumoylation in vitro. (a) Chemical structure of compound **21**. (b and c) Inhibition of in vitro sumoylation by compound **21** at indicated concentrations (1−100 μM) was added to the reaction mixture, and then the fluorescent levels of GFP-fused sumoylated RanGAP1-C2 was detected and measured using ImageQuant Las 4000 (GE Healthcare). Error bars show standard deviations from three independent assays, and the IC_{50} value was calculated using 50% inhibition compared with a control sample. (d) Docking-predicted binding mode of compound **21** (green atom color) with ATP (cyan atom color with high transparency) in SUMO E1 ATP binding site.

correlations between the experimental affinities and predicted binding free energies have been observed.[46,50,58−60] However, both MM-PBSA and MM-GBSA lack the accuracy for estimation of absolute binding free energy,[61] and even in some cases, this approach fails to reproduce relative binding free energies.[62−64] The accuracy of prediction using MM-PBSA and MM-GBSA varies and heavily depends upon the protein, ligand, computational protocol, etc. Although, a reasonable correlation was observed between MM-PBSA and MM-GBSA values (Figure S2, Supporting Information), instead of relying on single approach, we followed the consensus approach using both MM-PBSA and MM-GBSA to rerank the top-ranking compounds from docking. Two lists were prepared by ranking the docking solution based on either MM-PBSA or MM-GBSA. The MM-PBSA predicted binding free energy spans between 2.1 and −45 kcal/mol, while the MM-GBSA predicted binding free energy ranges from −3.3 to −61.9 kcal/mol for docking solutions. Lower ranked 25% compounds from each ranked list were removed, and a new list was prepared by choosing compounds that are common in both ranked lists. The 100 compounds from this ranked list were then visually inspected for geometric and electrostatic complementarities between protein and ligand. The compounds forming good interactions with the protein were prioritized. To select the most diverse compounds for biological assay, similar compounds were clustered together, and one compound for each class was selected. Finally, 24 compounds were selected to be purchased from commercial vendors for biological assay.

In Vitro Sumoylation Assay. The 24 compounds selected and acquired after in silico screening were tested for their ability to inhibit protein sumoylation using an in vitro sumoylation assay with RanGAP1-C2, a C-terminal fragment of RanGAP1, as a substrate. Ginkgolic acid, which was earlier reported as selective inhibitor of SUMO E1,[28] was used as a positive control. The results of the in vitro sumoylation assay revealed that 11 compounds (**1**−**4**, **10**, **13**, **15**, **16**, **21**, **22**, and **24**) inhibited at least 60% of RanGAP1 sumoylation at the concentration of 100 μM with 4 compounds (**10**, **13**, **16**, and **21**) inhibiting at least 80% of RanGAP1 sumoylation (Figure 4). The compounds were further tested to check their ability to inhibit RanGAP1 sumoylation at a concentration of 20 μM. The inhibitory potency of most of the above compounds diminished at this concentration except compounds **10**, **13**, and **21**, which still show more than 40% inhibition of RanGAP1 sumoylation. As shown in Figure 4, the strongest inhibition was observed for compound **21** as it almost completely inhibited the RanGAP1 sumoylation even at a lower concentration of 20 μM. To calculate the half maximal inhibitory concentration (IC_{50}), compound **21** was retested in a dose-dependent experiment varying the concentration in the range from 1 μM to 100 μM. The IC_{50} was determined from three independent experiments and reported as mean ± SD. The inhibition of RanGAP1 sumoylation changes in a dose-dependent manner, and the IC_{50} value for compound **21** was found to be 14.4 ± 1.3 μM. The IC_{50} for positive control ginkgolic acid under the same assay condition was determined and found consistent with previous reports.[28] Although, the

dx.doi.org/10.1021/ci300618e | J. Chem. Inf. Model. 2013, 53, 809−820

potency of compound **21** is lower than ginkgolic acid, compound **21** is more amenable to medicinal chemistry optimization for improving potency and pharmacokinetic properties. Figure 5 represents the chemical structure of compound **21** (Figure 5a) along with dose-dependent inhibition of RanGAP1 sumoylation (Figure 5b,c). Further, we sought to confirm that the target of compound **21** is SUMO E1. The complex of SUMO E1 with biotinylated SUMO-1 via the thioester bond can be detected in the presence of ATP under non-reducing conditions, but this band disappeared after addition of the reducing agent DTT (Figure 6). The formation

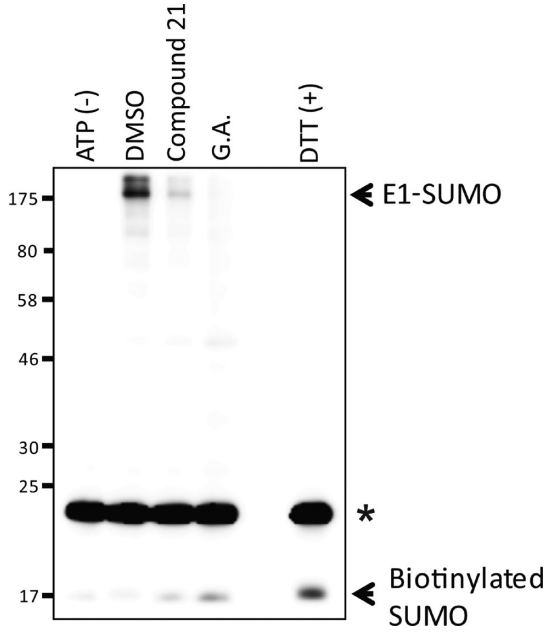


Figure 6. Impairment of the thioester bond formation between E1 and biotinylated SUMO-1 by compound **21**. Compound **21** at 100 μM was added to a reaction mixture containing 0.1 μg of biotinylated SUMO-1 and 1 μg of GST-Aos1/Uba2 in the presence or absence of 2 mM ATP. After the mixtures had been incubated at 37 °C for 20 min, they were separated by SDS-PAGE, followed by analysis using avidin-conjugated horse radish peroxidase. Addition of either 10 μM of ginkgolic acid (G.A.), a SUMO E1 inhibitor, or 1 mM of DTT to the reaction completely abolished the complex formation of biotinylated SUMO-1 and GST-Aos1/Uba2. The asterisk represents a non-specific band.

of the E1-biotinylated SUMO-1 intermediate was blocked by compound **21** as well as ginkgolic acid, a selective SUMO E1 inhibitor (Figure 6). These results suggested that compound **21** inhibits protein sumoylation by targeting SUMO E1. To gain insight into the mechanism of inhibition of SUMO E1 by compound **21** and to further guide experiments to improve potency and other desirable physicochemical properties, the docking-predicted binding mode was analyzed. As shown in Figure 5d, compound **21** occupies the ATP binding site, and the binding is mainly driven by a network of hydrophobic contacts between methyl-substituted phenyl and side chains of Val23, Leu49, Ile96, Leu116, and Ile384. Compound **21** also forms hydrogen bonds with the active site residues. Among the two urea moieties in compound **21**, one forms a hydrogen bond with the Asp48 side chain carboxyl and the other with the main chain carbonyl of Thr141. Docking of compound **21** gave a hint that it interferes with ATP binding and thereby prevents the formation of the E1-SUMO thioester intermediate.

Similarity Search for Expansion of Active Hits. The structure-based virtual screening to identify small molecule inhibitors of SUMO E1 yielded compound **21** as a micromolar inhibitor. Consequently, in order to improve the inhibitory activity, compounds with similar structures were retrieved by carrying out a similarity search against a ZINC database,[65] which is a publically available collection of ~18 million commercially available molecules that can be purchased. Compound **21** was used as a query to identify similar compounds. MACCS structural keys[66] and a Tanimoto coefficient[67] cutoff of 0.85 were used to retrieve similar compounds. About 110 similar compounds were obtained that were docked using RosettaLigand to the SUMO E1 ATP binding site. The best docked poses were again subjected to molecular dynamics simulation to facilitate the calculation of MM-PBSA and MM-GBSA binding free energy for the prioritization of similar compounds for biological assay. The docking score, MM-PBSA and MM-GBSA binding free energy, and predicted binding pose of compound **21** in its best docked conformation was used to select compounds displaying good scores and similar interaction patterns. On the basis of this analysis, another set of 37 molecules were selected for biological assay. The structures of these compounds along with docking scores and various energy components of MM-PBSA and MM-GBSA predicted binding free energy are given in Tables S1 and S2 of the Supporting Information. Most of these compounds can be subclassified into two classes based on the similarity they share with the core structure of the query molecule. The first class consists of 10 compounds belonging to the N,N″-[methylenebis(4,1-phenylene)]bis(N′-phenylurea) scaffold that has a maximum similarity with compound **21.** The N,N″-1,4-phenylenebis(N′-phenylurea) scaffold constitutes the second class of molecules and is the major class among the analogues of compound **21** with 20 compounds. Other compounds were also chemically related with slight differences from these two classes. These 37 compounds were acquired, and their inhibitory activities against in vitro sumoylation were tested. As shown in Figure 7, at least four compounds (**25**, **26**, **30**, and **37**) had more than 50% inhibitory activities at 20 μM. Chemical structures of these four compounds were shown in Figure 8. The dose-dependent experiment that yielded IC_{50} values for these four commercially available analogues of compound **21**, confirmed the inhibition with IC_{50} ranging approximately from 10 to 40 μM with compound **25** displaying the highest inhibition with an IC_{50} value of 11.1 ± 3.1 μM (Table 1).

Computational Analysis of Inhibitor Binding Mode and Druggability. As discussed earlier, compound **21** is predicted to bind in the ATP binding site of SUMO E1 and interacts via hydrogen bonding and hydrophobic contacts with Asp48, Thr141, Val23, Leu49, Ile96, Leu116, and Ile384. The molecular docking of compounds **25**, **26**, **30**, and **37** in the ATP binding site of SUMO E1 demonstrated that all of these compounds displayed a similar binding mode as compound **21** and share similar interactions. Figure 9 presents the best-ranked pose of compounds **25**, **26**, **30**, and **37** identified by RosettaLigand. The closest binding mode to compound **21** is shown by compound **37**, which makes the same hydrogen bond and hydrophobic contacts with side chains of Asp48, Thr141, Val23, Leu49, Ile96, Leu116, and Ile384 (Figure 9d). The docking of compounds **21**, **25**, **26**, **30**, and **37** (Figures 5d and 9) showed that one of phenylurea protrudes into the hydrophobic patch near Leu49 and Ile96, while other

dx.doi.org/10.1021/ci300618e | J. Chem. Inf. Model. 2013, 53, 809−820

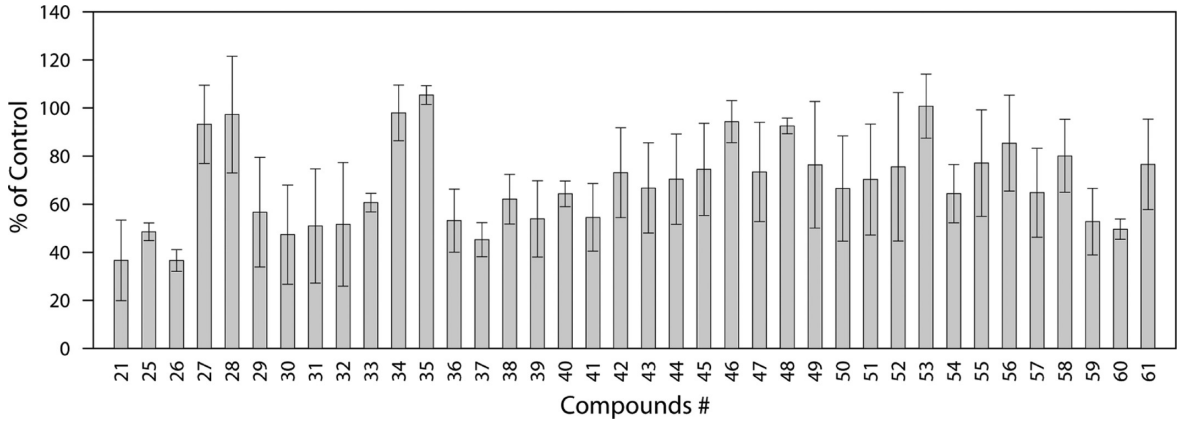


Figure 7. Inhibitory activities of commercially available analogues of compound **21** against in vitro protein sumoylation. Compound **21** or its analogues were added to the sumoylation reaction mixture, and the fluorescent levels of GFP-fused sumoylated RanGAP1-C2 were detected as described in Figure 5. Error bars show the standard deviations from three independent assays, and the average of percent of control is shown as gray bars.

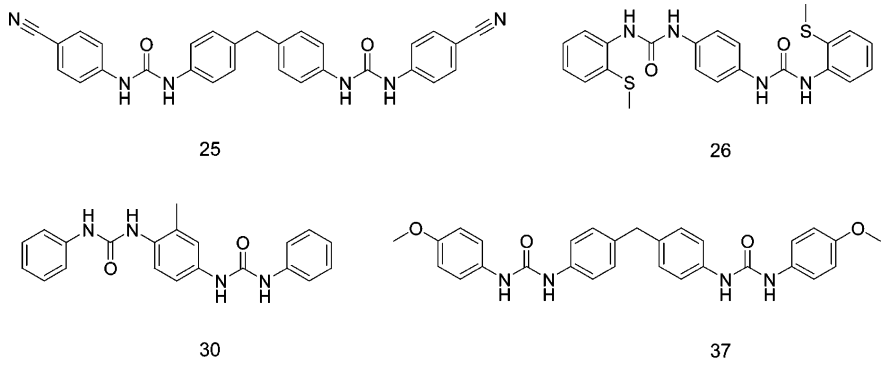


Figure 8. Chemical structures of potent analogues of compound **21**.

Table 1. IC_{50} Values of Compound 21 Analogues against in Vitro Sumoylation[a]

compound	IC_{50} (μM)
21	14.4 ± 1.3
25	11.1 ± 3.1
26	13.0 ± 0.4
30	40.3 ± 7.4
37	11.7 ± 5.3

[a]The IC_{50} value was calculated by measuring the fluorescent intensity of bands corresponding to the GFP-fused sumoylated RanGAP1-C2. Means ± SD from three independent experiments are shown.

phenylurea is placed near Ile384 and Thr141, close to the entrance of the ATP binding pocket. Both of these phenylurea make significant contacts and are important for the activity as compounds lacking symmetric architecture (compounds **27** and **35**) were devoid of any activity (Figure 7 and Table S1, Supporting Information). As nitrogens in the urea group of reported compounds act as hydrogen bond donors and are involved in hydrogen bonds with Ile384 and Thr141, any substitution here would affect the potency adversely as can be seen in compounds **28**, **49**, **50**, and **58** (Figure 7 and Table S1, Supporting Information). The presence of two urea moieties in our compounds mimics the protein backbone of the SUMO protein and maximizes the protein ligand contacts. Along with intermolecular interactions, molecular symmetry may also have some role in ligand binding for the reported chemical series. It is expected that there will be smaller entropic loss due to the symmetric nature of compounds **21**, **25**, **26**, **30**, and **37**, which contributes positively toward ligand binding. However, detailed investigation is required to elucidate the role of molecular symmetry in binding of these reported compounds. To gain further insight into the binding mechanism of active hits, the docking-predicted binding mode of compounds **21**, **25**, **26**, **30**, and **37** was analyzed by 10 ns molecular dynamics simulation. The overall stability of predicted binding pose of compound, **21**, **25**, **26**, **30**, and **37** was analyzed by investigating the trajectory from 10 ns molecular dynamics simulation. Figure 10 presents the root-mean-square deviation (RMSD) of heavy atoms of binding site residues bound to compound, **21**, **25**, **26**, **30**, and **37** over a period of 10 ns. As shown in Figure 10a, there is no significant deviation in SUMO E1 binding site residues, and the binding site was stable throughout the simulation. The average RMSD values of binding site residues of SUMO E1 bound with compounds **21**, **25**, **26**, **30**, and **37** were found to be 0.80, 0.83, 0.89, 1.02, and 0.76, respectively, which suggest stable binding sites. The ligand-binding poses were also stable with average RMSD values of 2.05, 1.76, 1.15, 1.84, and 2.17 for compounds **21**, **25**, **26**, **30**, and **37**, respectively, over a period of 10 ns (Figure 10b). The results from the molecular dynamics simulation indicate that docked conformation of compounds **21**, **25**, **26**, **30**, and **37** was stable, and important hydrogen bond and hydrophobic contacts were preserved.

As our initial goal was to identify compounds with better physicochemical properties, the drug-likeness of compounds

dx.doi.org/10.1021/ci300618e | J. Chem. Inf. Model. 2013, 53, 809−820

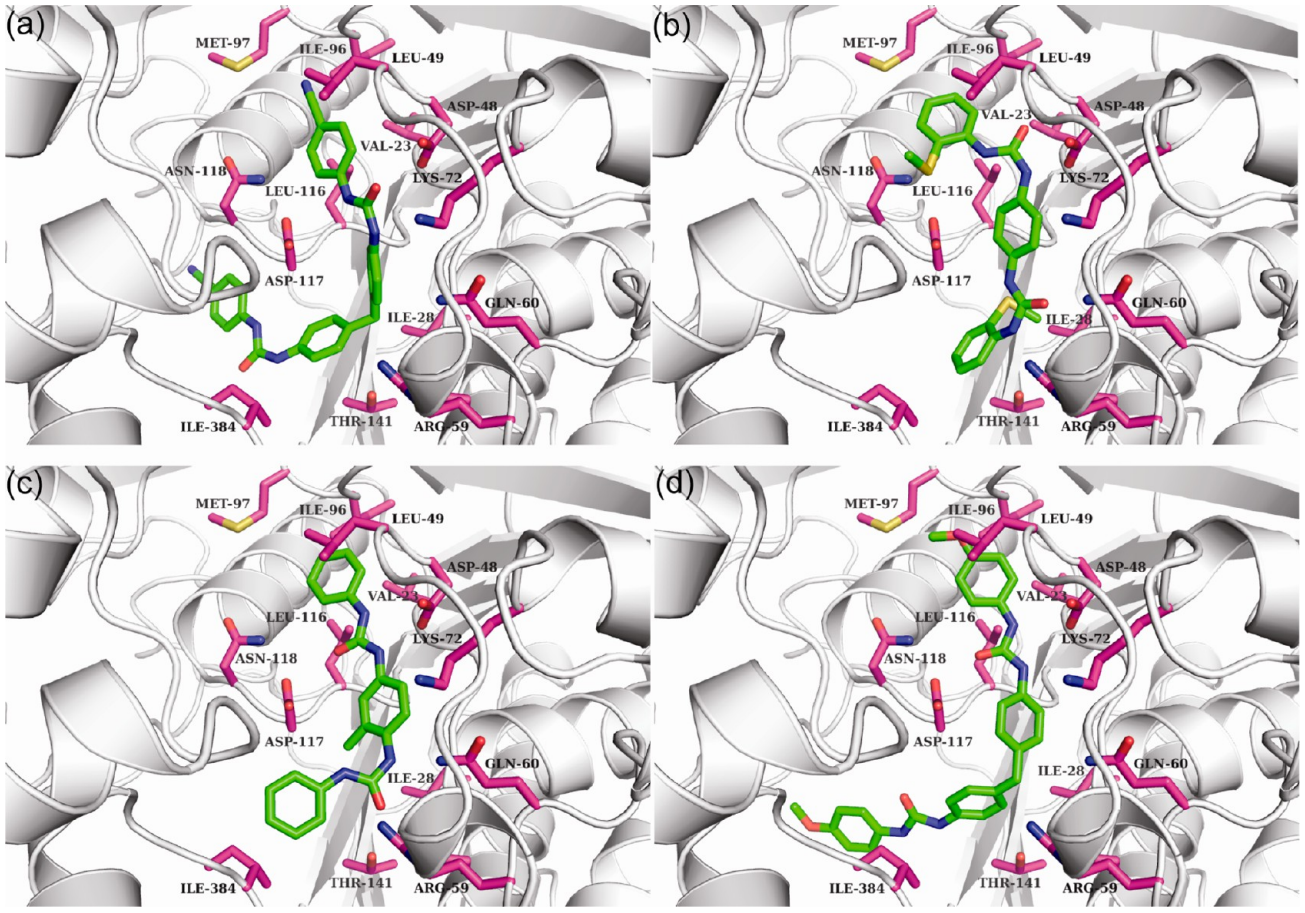


Figure 9. Docking-predicted binding mode of (a) compound **25**, (b) compound **26**, (c) compound **30**, and (d) compound **37** shown in green atom color sticks in the SUMO E1 ATP binding site.

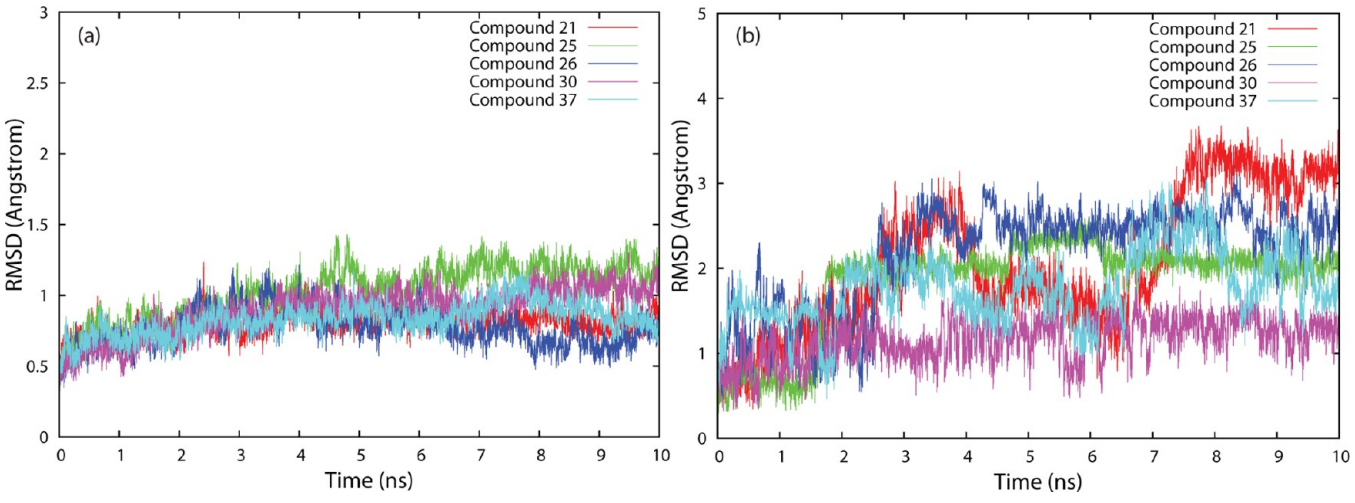


Figure 10. Time evolution of root-mean-square deviation (RMSD) of (a) binding site residues and (b) ligand.

21, **25**, **26**, **30**, and **37** was accessed according to Lipinski's drug-likeness criteria.[68] The physicochemical properties were calculated using QikProp[69] and compared with previously reported SUMO E1 inhibitors (Table 2).[27−29] None of the in silico-screened compounds violated Lipinski's "rule of five", whereas there were violations in the cases of ginkgolic acid, anacardic acid, and kerriamycin B. MLN4924, an AMP mimic, satisfies all of the Lipinski's drug-likeness criteria. However, selectivity is difficult to achieve with this type of compound because of its similarity with the natural substrate. Although, our compounds fulfills Lipinski's drug-likeness rule[68] and could be potential starting points for development of potent SUMO E1 inhibitors, there are some concerns that need to be addressed while optimizing for potency and druggability. The physicochemical properties values of compounds **21**, **25**, **26**, **30**, and **37** are within the limits but are on the higher side, especially for molecular weight and solubility. Moreover, compounds **21**, **25**, **26**, **30**, and **37** possess symmetric

dx.doi.org/10.1021/ci300618e | J. Chem. Inf. Model. 2013, 53, 809−820

Table 2. Comparison of Physicochemical Properties of in Silico-Screened Compounds with Previously Known Inhibitors of SUMO E1

compound	activity	molecular weight	HBD	HBA	Log P (o/w)	Lipinski violations
ginkgolic acid	3.0	346.5	1	2	6.2	1
anacardic acid	2.2	348.5	1	2	6.9	1
kerriamycin B	11.7	844.9	6	25	0.92	3
MLN4924	8.2	445.5	4	9	2.2	0
compound 21	14.4	464.5	4	4	4.9	0
compound 25	11.1	486.5	4	7	2.9	0
compound 26	13.0	438.5	4	5	3.6	0
compound 30	40.3	360.4	4	4	2.8	0
compound 37	11.7	486.5	4	5	2.9	0

architecture, and molecular symmetry has significant effect on the solubility of these compounds.[70] A symmetric compound tends to form stable crystals and therefore is sparingly soluble when compared with unsymmetrical structures. Therefore, optimization efforts should be focused toward reducing molecular symmetry, molecular weight, and increasing solubility. One approach to simultaneously reduce molecular symmetry and increase solubility is out-of-plane substitution with ionizable functional groups such as carboxylic acids or amines, which will disrupt crystal packing and leads to better solubility.

CONCLUSION

Using a hierarchical docking-based virtual screening approach, we have identified several symmetric phenylurea-containing compounds as a new class of SUMO E1 inhibitors. These compounds were shown to inhibit sumoylation of protein substrate RanGAP1 in an in vitro assay. Among these inhibitors, compound **21** was shown to target SUMO E1 by blocking the formation of the E1-biotinylated SUMO-1 thioester intermediate. Molecular docking-predicted compounds **21**, **25**, **26**, **30**, and **37** bind in a similar manner and share similar interactions. Although the activity of these reported inhibitors is only in the moderate micromolar range and not more than previously reported ginkgolic acid and anacardic acid, these compounds are more amenable for medicinal chemistry optimization and represent novel starting points for the development of potent and drug-like inhibitors of SUMO E1.

ASSOCIATED CONTENT

Supporting Information

Two additional tables on (1) structures, docking scores and inhibitory activities of compound **21** analogues, and (2) MM-PBSA and MM-GBSA predicted binding free energy components and standard deviations for compound **21** and analogues. Two additional figures on (1) correlation plot between MM-PBSA and MM-GBSA predicted binding free energy values calculated using 2 and 10 ns molecular dynamics trajectory and (2) correlation between MM-PBSA and MM-GBSA predicted binding free energy for virtual screening hits. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* E-mail: kamzhang@riken.jp.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank RIKEN Integrated Cluster of Clusters (RICC) at RIKEN for the supercomputing resources used for the study. We thank members of Zhang initiative research unit, RIKEN, for help and discussions. This work was supported in part by the Initiative Research Unit program and the Chemical Genomics Research Project, RIKEN ASI, CREST Research Project, Japan Science and Technology Corporation, and a Grant-in-Aid for Scientific Research (C) and on Priority Area "Cancer" from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

REFERENCES

(1) Baba, D.; Maita, N.; Jee, J. G.; Uchimura, Y.; Saitoh, H.; Sugasawa, K.; Hanaoka, F.; Tochio, H.; Hiroaki, H.; Shirakawa, M. Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature* **2005**, *435* (7044), 979−982.

(2) Geiss-Friedlander, R.; Melchior, F. Concepts in sumoylation: A decade on. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (12), 947−956.

(3) Johnson, E. S. Protein modification by SUMO. *Annu. Rev. Biochem.* **2004**, *73*, 355−382.

(4) Palancade, B.; Doye, V. Sumoylating and desumoylating enzymes at nuclear pores: underpinning their unexpected duties? *Trends Cell Biol.* **2008**, *18* (4), 174−183.

(5) Yeh, E. T. H. SUMOylation and De-SUMOylation: Wrestling with life's processes. *J. Biol. Chem.* **2009**, *284* (13), 8223−8227.

(6) Zhao, J. Sumoylation regulates diverse biological processes. *Cell. Mol. Life Sci.* **2007**, *64* (23), 3017−3033.

(7) Mukhopadhyay, D.; Dasso, M. Modification in reverse: the SUMO proteases. *Trends Biochem. Sci.* **2007**, *32* (6), 286−295.

(8) Hay, R. T. SUMO-specific proteases: A twist in the tail. *Trends Cell Biol.* **2007**, *17* (8), 370−376.

(9) Drag, M.; Salvesen, G. S. DeSUMOylating enzymes–SENPs. *IUBMB Life* **2008**, *60* (11), 734−742.

(10) Zuo, Y.; Cheng, J.-K. Small ubiquitin-like modifier protein-specific protease 1 and prostate cancer. *Asian J. Androl.* **2009**, *11* (1), 36−38.

(11) Lois, L. M.; Lima, C. D. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *EMBO J.* **2005**, *24* (3), 439−451.

(12) Olsen, S. K.; Capili, A. D.; Lu, X.; Tan, D. S.; Lima, C. D. Active site remodelling accompanies thioester bond formation in the SUMO E1. *Nature* **2010**, *463* (7283), 906−912.

(13) Lu, X.; Olsen, S. K.; Capili, A. D.; Cisar, J. S.; Lima, C. D.; Tan, D. S. Designed semisynthetic protein inhibitors of Ub/Ubl E1 activating enzymes. *J. Am. Chem. Soc.* **2010**, *132* (6), 1748−1749.

(14) Capili, A. D.; Lima, C. D. Taking it step by step: mechanistic insights from structural studies of ubiquitin/ubiquitin-like protein modification pathways. *Curr. Opin. Struct. Biol.* **2007**, *17* (6), 726−735.

(15) Sarge, K. D.; Park-Sarge, O.-K. Sumoylation and human disease pathogenesis. *Trends Biochem. Sci.* **2009**, *34* (4), 200−205.

(16) Sarge, K. D.; Park-Sarge, O.-K.; Kwang, W. J. SUMO and Its Role in Human Diseases. In *International Review of Cell and Molecular Biology*; Academic Press: 2011; Vol. *288*, Chapter 4, pp 167−183.

(17) Pandolfi, P.; Vogt, P.; Seeler, J. S.; Bischof, O.; Nacerddine, K.; Dejean, A. SUMO, the Three Rs and Cancer. In *Acute Promyelocytic Leukemia*; Springer: Berlin, Heidelberg: 2007; Vol. *313*, pp 49−71.

(18) Lee, J.-S.; Thorgeirsson, S. S. Genome-scale profiling of gene expression in hepatocellular carcinoma: Classification, survival prediction, and identification of therapeutic targets. *Gastroenterology* **2004**, *127* (5, Supplement 1), S51−S55.

(19) McDoniels-Silvers, A. L.; Nimri, C. F.; Stoner, G. D.; Lubet, R. A.; You, M. Differential gene expression in human lung adenocarci-

nomas and squamous cell carcinomas. *Clin. Cancer Res.* **2002**, *8* (4), 1127−1138.
(20) Mo, Y.-Y.; Yu, Y.; Theodosiou, E.; Rachel, Ee, P. L.; Beck, W. T. A role for Ubc9 in tumorigenesis. *Oncogene* **2005**, *24* (16), 2677−2683.
(21) Wang, L.; Banerjee, S. Differential PIAS3 expression in human malignancy. *Oncol. Rep.* **2004**, *11* (6), 1319−1324.
(22) Cheng, J.; Bawa, T.; Lee, P.; Gong, L.; Yeh, E. T. Role of desumoylation in the development of prostate cancer. *Neoplasia* **2006**, *8* (8), 667−676.
(23) Bawa-Khalfe, T.; Cheng, J.; Lin, S. H.; Ittmann, M. M.; Yeh, E. T. SENP1 induces prostatic intraepithelial neoplasia through multiple mechanisms. *J. Biol. Chem.* **2010**, *285* (33), 25859−25866.
(24) Xu, Y.; Li, J.; Zuo, Y.; Deng, J.; Wang, L. S.; Chen, G. Q. SUMO-specific protease 1 regulates the in vitro and in vivo growth of colon cancer cells with the upregulated expression of CDK inhibitors. *Cancer Lett.* **2011**, *309* (1), 78−84.
(25) Truong, K.; Lee, T.; Chen, Y. SUMO modification of the E1 Cys domain inhibits its enzymatic activity. *J. Biol. Chem.* **2012**, *287* (19), 15154−15163.
(26) Kessler, J. D.; Kahle, K. T.; Sun, T.; Meerbrey, K. L.; Schlabach, M. R.; Schmitt, E. M.; Skinner, S. O.; Xu, Q.; Li, M. Z.; Hartman, Z. C.; Rao, M.; Yu, P.; Dominguez-Vidana, R.; Liang, A. C.; Solimini, N. L.; Bernardi, R. J.; Yu, B.; Hsu, T.; Golding, I.; Luo, J.; Osborne, C. K.; Creighton, C. J.; Hilsenbeck, S. G.; Schiff, R.; Shaw, C. A.; Elledge, S. J.; Westbrook, T. F. A SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. *Science* **2012**, *335* (6066), 348−353.
(27) Soucy, T. A.; Smith, P. G.; Milhollen, M. A.; Berger, A. J.; Gavin, J. M.; Adhikari, S.; Brownell, J. E.; Burke, K. E.; Cardin, D. P.; Critchley, S.; Cullis, C. A.; Doucette, A.; Garnsey, J. J.; Gaulin, J. L.; Gershman, R. E.; Lublinsky, A. R.; McDonald, A.; Mizutani, H.; Narayanan, U.; Olhava, E. J.; Peluso, S.; Rezaei, M.; Sintchak, M. D.; Talreja, T.; Thomas, M. P.; Traore, T.; Vyskocil, S.; Weatherhead, G. S.; Yu, J.; Zhang, J.; Dick, L. R.; Claiborne, C. F.; Rolfe, M.; Bolen, J. B.; Langston, S. P. An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* **2009**, *458* (7239), 732−736.
(28) Fukuda, I.; Ito, A.; Hirai, G.; Nishimura, S.; Kawasaki, H.; Saitoh, H.; Kimura, K.; Sodeoka, M.; Yoshida, M. Ginkgolic acid inhibits protein SUMOylation by blocking formation of the E1-SUMO intermediate. *Chem. Biol.* **2009**, *16* (2), 133−140.
(29) Fukuda, I.; Ito, A.; Uramoto, M.; Saitoh, H.; Kawasaki, H.; Osada, H.; Yoshida, M. Kerriamycin B inhibits protein SUMOylation. *J. Antibiot. (Tokyo)* **2009**, *62* (4), 221−224.
(30) *OMEGA*, version 2.4.3; OpenEye Scientific Software, Inc.: Santa Fe, NM, 2012. www.eyesopen.com (accessed April 5, 2013).
(31) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235−242.
(32) *Molecular Operating Environment (MOE)*, version 2011.10; Chemical Computing Group, Inc.: Montreal, Quebec, Canada, 2011.
(33) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455−461.
(34) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785−2791.
(35) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y.-E. A.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; PopoviÄ‡, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P.; Michael, L. J.; Ludwig, B. Rosetta3: An object-oriented software suite for the simulation and design of macromolecules. In *Methods in Enzymology*, Academic Press: San Diego, 2011; Vol.*487*, Chapter 19, pp 545−574.
(36) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21* (2), 132−146.
(37) Davis, I. W.; Baker, D. RosettaLigand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **2009**, *385* (2), 381−392.
(38) Davis, I. W.; Raha, K.; Head, M. S.; Baker, D. Blind docking of pharmaceutically relevant compounds using RosettaLigand. *Protein Sci.* **2009**, *18* (9), 1998−2002.
(39) Meiler, J.; Baker, D. ROSETTALIGAND: Protein−small molecule docking with full side-chain flexibility. *Proteins* **2006**, *65* (3), 538−548.
(40) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **1998**, *120* (37), 9401−9409.
(41) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33* (12), 889−897.
(42) Bashford, D.; Case, D. A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.
(43) Tsui, V.; Case, D. A. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56* (4), 275−291.
(44) Simonson, T. Macromolecular electrostatics: Continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **2001**, *11* (2), 243−252.
(45) Case, D. A. D., T., A.; Cheatham, I., T.E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvry, I.; Wong, K. F.; Paesani; J. Vanicek, F.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER*, version 10; University of California: San Francisco, CA: 2008.
(46) Hou, T.; Wang, J.; Li, Y.; Wang, W. 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), 69−82.
(47) Hou, T.; Wang, J.; Li, Y.; Wang, W. 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), 866−877.
(48) Yang, T.; Wu, J. C.; Yan, C.; Wang, Y.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Virtual screening using molecular simulations. *Proteins* **2011**, *79* (6), 1940−1951.
(49) Gao, C.; Herold, J. M.; Kireev, D. Assessment of free energy predictors for ligand binding to a methyllysine histone code reader. *J. Comput. Chem.* **2012**, *33* (6), 659−665.
(50) Rastelli, G.; Rio, A. D.; Degliesposti, G.; Sgobba, M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2009**, *31* (4), 797−810.
(51) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(52) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327−341.
(53) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55* (2), 383−394.

(54) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20* (2), 217−230.
(55) Oehme, D. P.; Brownlee, R. T.; Wilson, D. J. Effect of atomic charge, solvation, entropy, and ligand protonation state on MM-PB(GB)SA binding energies of HIV protease. *J. Comput. Chem.* **2012**, *33* (32), 2566−2580.
(56) Wichapong, K.; Lawson, M.; Pianwanit, S.; Kokpol, S.; Sippl, W. Postprocessing of protein−ligand docking poses using linear response MM-PB/SA: application to Wee1 kinase inhibitors. *J. Chem. Inf. Model.* **2010**, *50* (9), 1574−1588.
(57) Brown, S. P.; Muchmore, S. W. Rapid estimation of relative protein−ligand binding affinities using a high-throughput version of MM-PBSA. *J. Chem. Inf. Model.* **2007**, *47* (4), 1493−1503.
(58) Anisimov, V. M.; Ziemys, A.; Kizhake, S.; Yuan, Z.; Natarajan, A.; Cavasotto, C. N. Computational and experimental studies of the interaction between phospho-peptides and the C-terminal domain of BRCA1. *J. Comput. Aided Mol. Des.* **2011**, *25* (11), 1071−1084.
(59) Srivastava, H. K.; Sastry, G. N. Molecular dynamics investigation on a series of HIV protease inhibitors: Assessing the performance of MM-PBSA and MM-GBSA approaches. *J. Chem. Inf. Model.* **2012**, *52* (11), 3088−3098.
(60) Okimoto, N.; Futatsugi, N.; Fuji, H.; Suenaga, A.; Morimoto, G.; Yanai, R.; Ohno, Y.; Narumi, T.; Taiji, M. High-performance drug discovery: Computational screening by combining docking and molecular dynamics simulations. *PLoS Comput. Biol.* **2009**, *5* (10), e1000528.
(61) Singh, N.; Warshel, A. Absolute binding free energy calculations: On the accuracy of computational scoring of protein−ligand interactions. *Proteins: Struct., Funct., Bioinf.* **2010**, *78* (7), 1705−1723.
(62) Anisimov, V. M.; Cavasotto, C. N. Quantum mechanical binding free energy calculation for phosphopeptide inhibitors of the Lck SH2 domain. *J. Comput. Chem.* **2011**, *32* (10), 2254−2263.
(63) Mikulskis, P.; Genheden, S.; Rydberg, P.; Sandberg, L.; Olsen, L.; Ryde, U. Binding affinities in the SAMPL3 trypsin and host-guest blind tests estimated with the MM/PBSA and LIE methods. *J. Comput.-Aided Mol. Des.* **2012**, *26* (5), 527−541.
(64) Sulea, T.; Cui, Q.; Purisima, E. O. Solvated Interaction Energy (SIE) for scoring protein−ligand binding affinities. 2. Benchmark in the CSAR-2010 scoring exercise. *J. Chem. Inf. Model.* **2011**, *51* (9), 2066−2081.
(65) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **2012**, *52* (7), 1757−1768.
(66) *MACCS Keys*; MDL Information Systems, Inc.: San Leandro, CA
(67) Rogers, D. J.; Tanimoto, T. T. A computer program for classifying plants. *Science* **1960**, *132* (3434), 1115−1118.
(68) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23* (1−3), 3−25.
(69) *QikProp, version 3.4*, Schrödinger, LLC, New York, NY, 2011.
(70) Pinal, R. Effect of molecular symmetry on melting temperature and solubility. *Org. Biomol. Chem.* **2004**, *2* (18), 2692−2699.