Identification of Small Molecular Weight Inhibitors of Src Homology 2 Domain-Containing Tyrosine Phosphatase 2 (SHP-2) via in Silico Database Screening Combined with Experimental Assay

Wen-Mei Yu,[§,†] Olgun Guvench,[§,‡] Alexander D. MacKerell, Jr.,*[,‡] and Cheng-Kui Qu*[,†]

Department of Medicine, Division of Hematology/Oncology, Case Comprehensive Cancer Center, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, 20 Penn Street, Baltimore, Maryland 21201

Received March 3, 2008

Virtual screening methods combined with experimental assays were used to identify low molecular weight inhibitors for Src homology 2 domain-containing phosphatase 2 (SHP-2) that is mutated and hyperactivated in Noonan syndrome and a significant portion of childhood leukemias. Virtual screening included multiple conformations of the protein, score normalization procedures, and chemical similarity considerations. As the catalytic core of SHP-2 shares extremely high homology to those of the related SHP-1 phosphatase and other tyrosine phosphatases, in order to identify selective inhibitors, we chose to target an adjacent protein surface pocket that is predicted to be important for binding to phosphopeptides and that has structural features unique to SHP-2. From a database of 1.3 million compounds, 9 out of 165 computationally selected compounds were shown to inhibit SHP-2 activity with IC_{50} values of $\approx 100\ \mu M$. Two of the active compounds were further verified for their ability to inhibit SHP-2-mediated cellular functions. Fluorescence titration experiments confirmed their direct binding to SHP-2. Because of their simple chemical structures, these small organic compounds have the potential to act as lead compounds for the development of novel anti-SHP-2 drugs.

Introduction

Src homology 2 (SH2) domain-containing phosphatase 2 (SHP-2),[a] a ubiquitously expressed SH2 domain-containing protein tyrosine phosphatase (PTP), plays a critical role in diverse cell signaling processes.[1–3] SHP-2 contains two tandem SH2 domains at the N-terminus and a PTP domain at the C-terminus, with flexible polypeptide linker regions connecting the three domains. The 2.0 Å X-ray crystal structure of the self-inhibited form of SHP-2 reveals the formation of an intramolecular protein−protein interface between the N-terminal SH2 (N-SH2) domain and the PTP domain.[4,5] This self-interaction is characterized by the binding of a loop on the backside of the N-SH2 domain to the catalytic pocket of the phosphatase domain, thereby blocking substrate access to the catalytic site. Numerous interdomain hydrogen bonds exist in this conformation; some of them are direct and some are bridged by water molecules. Polypeptide ligands with phosphotyrosine (pY) residues activate SHP-2 by binding the tandem SH2 domains, which disrupts the N-SH2:PTP interface, leading to exposure of the PTP catalytic site. Thus, the recognition of pY-peptides by the SH2 domains is normally coupled with the activation of SHP-2 phosphatase capability. In most circumstances, SHP-2 plays an overall positive role in transducing signals initiated from growth factors/cytokines and extracellular matrix proteins.[1–3] Despite extensive studies over the past decade, the signaling mechanisms of SHP-2 are still not well understood. For example, the molecular basis for the positive role of its catalytic activity in the Erk pathway remains elusive. Part of the reason for this is the lack of SHP-2 specific inhibitors that can be used as research tools to probe SHP-2 signaling.

Consistent with its overall positive role in cell signaling, genetic lesions in the SHP-2 gene (*PTPN11*) that cause hyperactivation of its catalytic activity have been identified in the developmental disorder Noonan syndrome[6] and various childhood leukemias, including juvenile myelomonocytic leukemia (JMML), B cell acute lymphoblastic leukemia (ALL), and acute myeloid leukemia.[7,8] Fifty percent of the patients with Noonan syndrome, 35% of JMML, and 6% of B cell-ALL cases harbor SHP-2 mutations. Moreover, activating mutations of SHP-2 have also been identified in sporadic solid tumors.[9] The SHP-2 mutations found in these diseases are associated with changes in amino acids located at the interface formed by the N-SH2 and PTP domains in the self-inhibited SHP-2 conformation. Therefore, it is thought that these mutations cause a decrease in the affinity of the binding between the N-SH2 and PTP domains, leading to the gain of function (GOF) by allowing access to the phosphatase catalytic site on the enzyme. Remarkably, the SHP-2 mutations appear to play a causal role in the development of these diseases because SHP-2 mutations and other JMML-associated Ras or Neurofibromatosis 1 mutations are mutually exclusive.[6–9] Furthermore, recent studies have shown that single SHP-2 GOF mutations are sufficient to induce cytokine hypersensitivity in hematopoietic progenitor cells and Noonan syndrome and JMML-like myeloproliferative disease in mice.[10–14]

* To whom correspondence should be addressed. Phone: 216-368-3361; Fax: 216-368-1166; E-mail: cxq6@case.edu (C.K.Q.). Phone: 410-706-7442; Fax: 410-706-5017; E-mail: alex@outerbanks.umaryland.edu (A.D.M).

† Department of Medicine, Division of Hematology/Oncology, Case Comprehensive Cancer Center, Case Western Reserve University.

‡ Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore.

§ These authors contributed equally to this work.

[a] Abbreviations: ADME, absorption/distribution/metabolism/excretion; ALL, acute lymphoblastic leukemia; SHP-2: Src homology 2 domain-containing tyrosine phosphatase 2; PTP: protein tyrosine phosphatase; pY: phosphotyrosine; GOF: gain-of-function; EGF: epidermal growth factor; JMML: juvenile myelomonocytic leukemia; FBS: fetal bovine serum; CADD: computer-aided drug design; IE: interaction energies; TC: tanimoto coefficient; vdW: van der Waals.

10.1021/jm800229d CCC: $40.75
Published on Web 11/14/2008

The direct connection between activating mutations of SHP-2 and Noonan syndrome and childhood leukemias points to SHP-2 as a target for mechanism-based therapeutics for these diseases. While a recently reported compound effectively inhibits SHP-2 catalytic activity, it also cross-inhibits the related SHP-1 phosphatase with the same efficiency[15] owing to the homology in the targeting sites between SHP-2 and SHP-1. Thus, there are currently no reports of selective SHP-2 inhibitors. The discovery of SHP-2 specific inhibitors would not only facilitate research on SHP-2 signaling in model systems but could also lead to the development of new drugs for SHP-2 associated diseases. The availability of three-dimensional (3D) crystal structural information on SHP-2[4,5,16] makes it possible to apply target-based computer-aided drug design (CADD) methods to find SHP-2 phosphatase inhibitors. In the present study, using CADD screening of a virtual database of compounds against the 3D structure of SHP-2 followed by experimental assays, we have identified low molecular weight compounds that selectively inhibit SHP-2 catalytic activity. Moreover, biological assays revealed that the lead compounds were effective in inhibiting SHP-2 mediated cellular functions.

Experimental Section

CADD in Silico Screening. The 3D structure of SHP-2 in the unphosphorylated state[5] (PDB ID 2SHP) was retrieved from the Protein Data Bank.[17] Following deletion of the SH2 domains, the Reduce algorithm[18] was used to place hydrogen atoms and optimize adjustable groups (OH, SH, NH_3^+, Met-CH_3, and Asn, Gln, and His side chain orientation). To prepare the structure for docking, partial charge and Lennard-Jones parameters from the CHARMM force field[19] including the CMAP backbone correction energy[20] were applied. All docking calculations were carried out with DOCK[21] using flexible ligands based on the anchored search method.[22] The solvent accessible surface[23] was calculated with the program DMS[24] using a surface density of 2.76 surface points per Å^2 and a probe radius of 1.4 Å^2. Sphere sets were calculated with the DOCK associated program SPHGEN. From the full sphere set, sphere clusters in the SHP-2 docking site important for binding pY peptides were identified. Residues involved in intermolecular interactions with residues at the pY + 5 position of pY peptides were used to select the docking site. Though no cocrystal structures of SHP-2 exist with PTP-bound pY-peptide, such structural information is available for the close homologue SHP-1.[25] We predicted the likely location of the SHP-2 PTP pY + 5 binding pocket based alignment with the SHP-1 PTP−pY-peptide cocrystal and selected spheres in this pocket. Specifically, spheres within 6 Å of residues 255, 258, 261, 498, and 503 were selected, resulting in a set containing 12 spheres and located as shown in Figure 1. The selected sphere set acted as the basis for initial ligand placement during database searching. The GRID method[26] within DOCK was used to approximate the ligand−receptor interaction energy during ligand placement by the sum of the electrostatic and van der Waals (vdW) components. The GRID box dimensions were 38.5 Å × 37.6 Å × 38.9 Å, centered around the sphere set to ensure that docked molecules were within the grid.

A database of 1.3 million compounds was used for the initial virtual screening. This database of low molecular weight, commercially available compounds had been created in our laboratory by converting files obtained from the vendors in the 2D SDF format to the 3D MOL2 format through a procedure that included geometry generation, addition of hydrogens and charges, and force field optimization using SYBYL6.4 along with in-house programs.[27,28] The compounds that were screened had between 10 and 40 heavy atoms and less than 10 rotatable bonds. During the docking procedure, each compound was divided into nonoverlapping rigid segments connected by rotatable bonds. Segments with more than five heavy atoms were used as anchors, each of which was docked into the binding site in 250 orientations and minimized. The

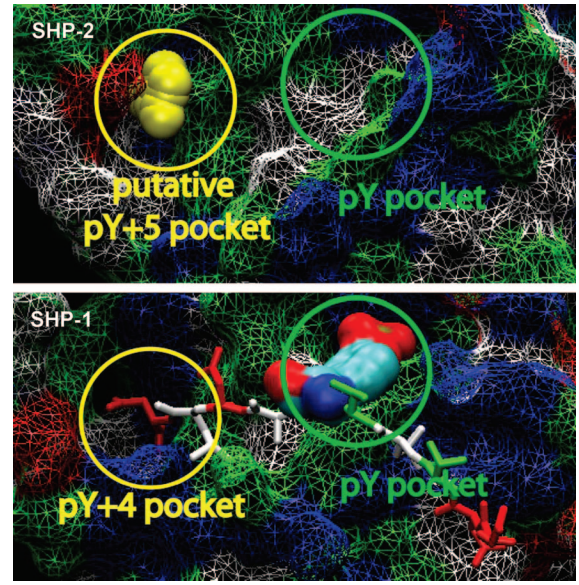

Figure 1. Location of the docking site on the SHP-2 catalytic domain. Top: Sphere set (yellow) in the putative pY + 5 pocket used for docking to the crystal structure of the SHP-2 PTP domain (wireframe). The locations of the residues used in selecting the sphere set are such that residues 255, 258, and 261 lie to the left of the spheres, 498 is behind the spheres at the base of the pocket, and 503 is to the right of the spheres. Bottom: Crystal structure of the homologous SHP-1 PTP domain and bound pY peptide (PDB ID 1FPR),[25] with the pY residue shown as a molecular surface. Amino acids are colored according to type: red = acidic, blue = basic, green = polar, white = hydrophobic, cyan = aromatic. Molecular graphics were prepared with VMD.[29]

remainder of the molecule was built around the anchor in a stepwise fashion by adding other segments connected through rotatable bonds. At each step, the dihedral of the rotatable bond was sampled in increments of 10° and the lowest energy conformation was selected. During primary docking, each rotatable bond was minimized as it was created without reminimizing the other bonds. Pruning of the conformational orientations ensured conformational diversity and more favorable energies.[30,31] Energy scoring was performed with a distant-dependent dielectric, with a dielectric constant of 4, and using an all atom model. Once the whole molecule was built, it was minimized. The conformation of each molecule with the most favorable interaction energy was selected and saved.

After the primary docking, compounds were chosen for the secondary docking based on their normalized vdW attractive interaction energy scores (see Results and Discussion). Compound selection based on the DOCK energy score favors compounds with higher molecular weight (MW) because their size contributes to the energy score. To minimize this size bias, an efficient procedure was applied in which the DOCK interaction energies (IE) are normalized by the number of heavy atoms N raised to a power x.[27]

$$IE_{norm} = IE/N^x \quad (1)$$

Normalization of the vdW attractive energies was done with x = 0, 0.33, 0.5, 0.66, and 1, and the MW distributions of the top 50000 compounds in each category were analyzed to choose the normalized set with a peak closest to MW = 300 (see Results and Discussion).

The top 50000 compounds obtained from the primary database search were screened in a more rigorous and computationally expensive docking procedure, referred to as secondary docking. The procedure described for primary docking was followed with the additional step of minimizing all rotatable bonds simultaneously during the stepwise building of the molecule. Additionally, docking was done to multiple conformations of the PTP domain. In addition to the crystallographic conformation, three

additional conformations were used to generate docking grids. These conformations were representative snapshots from the final 4 ns of a 5 ns all-atom explicit-water molecular dynamics trajectory of the PTP domain and were selected by conformational clustering[32] of the heavy atoms in residues 255, 258, 261, 398, and 503 (Figure 1) after root-mean-squared alignment of the protein C_α atoms to the crystal structure. The resultant snapshots were at the 1.4, 2.4, and 3.8 ns time points and were representative of the diverse binding site conformations sampled during the molecular dynamics. A new docking sphere set was created for each of these conformations as described above for the crystal structure. Each compound was docked to the crystallographic PTP conformation as well as each of the three molecular dynamics snapshots and was scored based on the most favorable of the four docking energies. The total interaction energies (electostatic + vdW) of the 50000 compounds were normalized, as described above, and the top 1000 compounds from the normalized distribution with a peak closest to MW = 300 (see Results and Discussion) were selected and subjected to chemical diversity analysis.

Compound Selection Based on Chemical Diversity. Chemical similarity clustering of the top 1000 compounds identified during secondary docking was performed to maximize the chemical diversity of the final compounds selected for biological assay. Clustering calculations were performed using the program MOE (Chemical Computing Group, Inc.). The Jarvis−Patrick algorithm,[33] as implemented in MOE, was used to cluster the compounds using the MACC_BITS fingerprinting scheme and Tanimoto coefficient (TC).[34] The algorithm first calculates the MACC_BITS fingerprints that encodes the 2D structural features of each compound as a sequence of 0's and 1's. Then, the pairwise similarity matrix between each compound was calculated based on the TC values.[35] TC is a metric that provides a similarity score for two compounds by dividing the fraction of features common to both molecules by the total number of features, where the features are defined by the MACC_BITS fingerprints. The similarity matrix is converted into a second matrix in which each TC value is replaced by a 0 or 1, representing similarity values below and above the threshold value (*S*) provided by the user, respectively. The rows of the new matrix were treated as fingerprints and the "TC" value between each is calculated. Molecules with values above the selected overlap threshold (*T*) were put in the same cluster. To generate clusters of reasonable sizes (i.e., 3−10 compounds), an iterative procedure was followed in which the *T* value was gradually decreased by 10, and for each *T* value, three *S* values (*T*-10, *T*-20, *T*-30) were calculated.

Compounds were chosen from the individual clusters for experimental assay with emphasis on compounds with druglike physical characteristics as defined by Lipinski et al.[36] Properties considered were the MW, number of hydrogen donors (NHD) and acceptors (NHA), and the logP values as calculated by MOE. However, exceptions were made when all compounds in a cluster had one or more physical characteristics beyond the range defined by Lipinski et al.[36] The resultant chemically diverse list of approximately 200 compounds selected via CADD were purchased from ChemDiv (San Diego, CA), ChemBridge (San Diego, CA), or Specs (Wakefield, RI) and dissolved in dimethyl sulfoxide (DMSO) at a stock concentration of 25, 50, or 100 mM. The purity of the active compounds was verified by mass spectrometry.

In Vitro and In Vivo Phosphatase Assay. Candidate compounds identified by CADD were screened using the in vitro PTP assay kit (Sigma, St. Lois, MO). SHP-2 PTP domain GST fusion protein purified in house was used as the enzyme and a phosphopeptide corresponding to the surrounding sequence of $pTyr^{1146}$ in the insulin receptor (Thr-Arg-Asp-Ile-Tyr[PO_3H_2]-Glu-Thr-Asp-Tyr-Tyr) was used as the substrate. The assay determined free phosphate generated by dephosphorylation of the PTP substrate using the Malachite Green reagent. Briefly, 0.5 μg of GST-SHP-2 PTP was incubated in 40 μL of assay buffer (25 mM Tris-HCl, pH 7.4, 50 mM NaCl, 5 mM DTT, and 2.5 mM EDTA) with test compounds at various concentrations or DMSO at room temperature for 30 min. The PTP substrate was then added to a final concentration of 0.2 mM. The system was incubated at 30 °C for 30 min. Finally, 50 μL of Malachite Green solution was added and OD_{620} was measured after 5 min. In the phosphatase assay for SHP-1, CD45, and TC-PTP, procedures were similar, with the exception that GST-SHP-1, GST-CD45 cytoplasmic domain, and GST-TC-PTP purchased from Biomol International, L.P. (Plymouth Meeting, PA) were used instead as the enzymes.

For the in vivo phosphatase assay, mouse embryonic fibroblasts were used to assess the activities of the test compounds in inhibiting epidermal growth factor (EGF)-induced activation of SHP-2 catalytic activity. WT and mutant fibroblasts derived from activating mutation of SHP-2 D61G knock-in mice ($SHP\text{-}2^{D61G/D61G}$)[10] were starved in serum-free medium overnight and the tested compounds (100 μM) were added in the last 1 h. Cells were then stimulated with EGF (50 ng/mL) for 10 min. The cells were harvested and lysed in RIPA buffer (50 mM Tris-HCl pH 7.4, 1% NP-40, 0.25% Na-deoxycholate, 150 mM NaCl, 1 mM EDTA, 1 mM NaF, 10 μg/mL leupeptin, 10 μg/mL aprotin, and 1 mM PMSF). Whole cell lysates (500 μg) were immunoprecipitated with 1 μg of anti-SHP-2 antibody. Immunoprecipitates were washed three times with HNTG buffer (20 mM Hepes pH 7.5, 150 mM NaCl, 1% Glycerol, and 0.1% Triton X-100) and assayed for the catalytic activity using the phosphatase assay as described above.

Ba/F3 Cell Proliferation Assay. To test the effects of the active compounds on IL-3-induced cellular response, Ba/F3, an IL-3 dependent murine pro-B lymphoma cell line, was used for the assay. Ba/F3 cells were seeded into 96-well plates (2×10^4 cells/well) in the RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS), recombinant mouse IL-3 (1 ng/mL), and test compounds (100 and 200 μM). Three days later, cell number was determined using the MTS cell proliferation kit (Promega Life Science, Madison, WI).

Fluorescence Titrations. For all experiments, purified SHP-2 PTP domain GST-fusion protein was diluted into 20 mM Tris-HCl, pH 7.5. Fluorescence spectra were recorded with a luminescence spectrometer LS50 (Perkin-Elmer, Boston, MA). Titrations were performed by increasing the test compound concentration while maintaining the SHP-2 protein concentration at 3 μM. Contributions from background fluorescence of the inhibitors were accounted for by subtracting the fluorescence of the inhibitors alone from the protein-inhibitor solution. The excitation wavelength was 295 nm and fluorescence was monitored from 360 to 500 nm. All reported fluorescence intensities are relative values and are not corrected for wavelength variations in detector response.

Results and Discussion

Selection of the Docking Site in the SHP-2 Catalytic Domain. SHP-2 has over 60% sequence identity with the homologous SHP-1 phosphatase, although the latter protein plays an opposing role to SHP-2 in hematopoietic cell signaling.[1–3] To identify SHP-2 PTP specific inhibitors, a structural alignment between SHP-2 and SHP-1 was performed in order to search for a potential drug-docking pocket in the SHP-2 catalytic domain that would be specific for that protein. Such a site must be structurally different than that present in SHP-1, allowing for specific interactions to occur between the compounds and SHP-2 that could not occur between the compounds and SHP-1 or other phosphatases. Not surprisingly, the residues that perform the pY hydrolysis in both SHP-2 and SHP-1, i.e., the catalytic site, are identical both in sequence and in spatial location.[5,25,37,38] Thus, any SHP-2 inhibitor that binds to the catalytic site would be expected to have cross-reactivity with SHP-1, as well as with other PTP-domain containing proteins that also utilize this spatial arrangement of amino acids for catalysis.

Using the program SPHGEN, potential binding sites on the SHP-2 protein were identified and the sites in the SHP-2 PTP domain were analyzed in detail. This led to the identification of a putative pocket in the surface of the protein that appeared

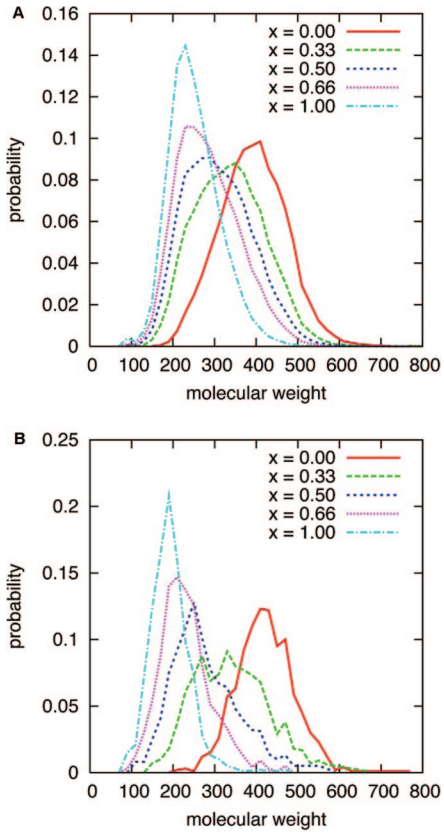

Figure 2. Molecular weight distribution histograms of the top scoring compounds after normalization of docking energies. (A) Distributions of the top 50000 compounds after primary docking. (B) Distributions of the top 1000 compounds after secondary docking. Scoring is based on the normalized interaction energy IE_{norm}, which is calculated as IE/N^x, where N is the number of non-hydrogen atoms in each ligand and x is an exponent of N and ranges from 0 (no normalization) to 1 (full normalization).

to be different from that in SHP-1 in both structure and amino acid composition. As demonstrated by a cocrystal structure, in the case of SHP-1, this pocket binds the acidic C-terminal aspartic acid at the pY + 4 position of a pY peptide.[25,38] In contrast, SHP-2 PTP has been shown in assays to preferentially bind to pY peptides with a basic residue at the pY + 5 position,[39] and our protein alignment suggests that the chosen binding pocket is a logical place for this specific interaction to be occurring. Figure 1 illustrates the pY-peptide binding clefts of the two proteins and the particular pocket in the SHP-2 cleft used as the target binding site for the CADD screening.

Primary Database Screening. Compound selection from the primary database screen was based on the vdW attractive energy, as previously performed.[28] Use of this term selected compounds that have favorable steric interactions with the protein and avoids the selection of compounds whose interaction with the protein are dominated by one or two electrostatic interactions. Use of the vdW attractive term for scoring does not exclude the selection of compounds for which favorable electrostatic interactions with the protein occur as the total energy (i.e., electrostatic and vdW terms) was used for the actual docking process (i.e., posing). Use of the vdW attractive energy without any normalization yielded a MW distribution for the top scoring 50000 compounds with a peak at 410 Da (Figure 2A). Based on this histogram, approximately 10% of those compounds were above a MW of 500 Da. As druglike compounds typically have MWs below 500 Da[36] and lead compounds, the target of the present study, have even lower MWs,[40] the goal was to choose a distribution with a peak at 300 Da via the previously developed MW normalization procedure.[27] Using N, $N^{0.66}$, $N^{0.5}$, and $N^{0.33}$ normalization, the peaks of the histograms of the top scoring 50000 compounds were at MW values of 230, 230, 290, and 350 Da, respectively; thus, $N^{0.5}$ normalization was chosen. The MW probability distribution of the entire database screened in the present study is centered at 360 Da, as is the average of the World Index Database.[27] The resultant low MW of the lead compounds identified by normalized scoring allows the addition of functional groups during future lead optimization efforts.

It should be noted that significant overlap of compounds occurred for the different normalization schemes. Of the 50000 compounds selected via $N^{0.5}$ normalization, 85% compounds were common in the $N^{0.33}$ set, 87% in the $N^{0.66}$ set, 67% in the N set, and 58% in the set of nonnormalized (i.e., N^0) compounds. Thus, it may be assumed that compounds with highly favorable interaction orientations with the protein binding site were not being excluded by the normalization procedure.

Secondary Database Screening. Secondary screening of the 50000 compounds selected from the primary screen involved a more exhaustive and computationally demanding conformational search of the docked molecules. As compounds with good steric fit were selected from the primary screen, electrostatic as well as vdW interaction energies were used for the selection of compounds from the secondary screen. Additionally, each compound was docked not only to the crystal PTP conformation but also to three other conformations from molecular dynamics simulation, and the most favorable interaction energy was taken as that compound's score. The inclusion of additional conformations helps to account for the conformational heterogeneity of protein surfaces in a solvated environment and aims to address a potential limitation of representing the protein as a rigid body during grid-based docking. The total interaction energies were normalized using different powers of N, and the MW distributions of the top 1000 compounds were determined for each normalization scheme (Figure 2B). For the top 1000 compounds selected via the N, $N^{0.66}$, $N^{0.5}$, and $N^{0.33}$ normalization, the distribution peaks were at 190, 210, 250, and 330 Da, respectively (Figure 2B). The peak for the top 1000 compounds without normalizing the energies was 410 Da (Figure 2B). The top 1000 scoring compounds in the set obtained after $N^{0.33}$ normalization was chosen to select compounds in the target MW range discussed above while avoiding molecules that were too small and lacking adequate structural diversity for lead or druglike candidates.

Final Compound Selection. The final list of compounds to be submitted for experimental testing should be chemically diverse to increase the probability of identifying unique leads.[25] To facilitate selection of diverse compounds, a clustering algorithm that distributes the compounds into groups (clusters) of structurally similar compounds with compounds across clusters being dissimilar was applied. One or more compounds are then selected from each cluster for experimental validation. The selection process for choosing a compound from a particular cluster was based on Lipinski's Rule of 5: MW of less than 500 Da and the number of hydrogen bond donors, the number of hydrogen bond acceptors and logP values each less than 5, although exceptions were made for clusters that did not contain members fulfilling these criteria. Compounds fulfilling these empirical rules have an increased likelihood of demonstrating good absorption/distribution/metabolism/excretion (ADME) profiles[36] which is an important consideration when the ultimate goal is to develop a therapeutic. A total of 236 compounds were thus selected from the list of 1000, 165 of which were

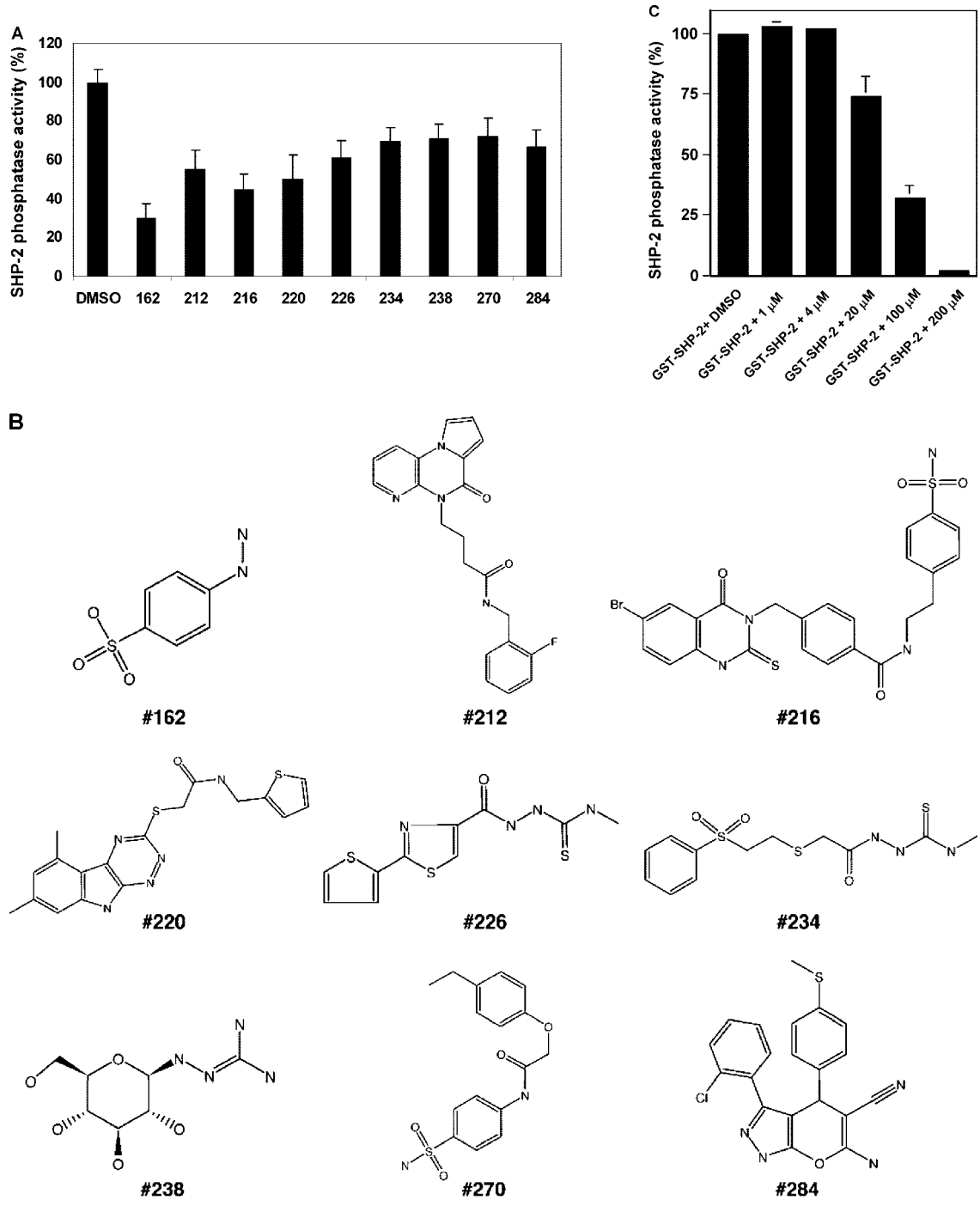

Figure 3. Inhibitory effects of the nine active compounds on the phosphatase activity of SHP-2. (A) Candidate compounds identified by CADD were screened using the in vitro tyrosine phosphatase assay as described in the Experimental Section. Compounds were dissolved in DMSO at 100 μM. Shown are results of the nine active compounds. (B) Chemical structures of the nine active compounds. (C) The inhibitory effect of compound no. 162 on SHP-2 activity was tested at the indicated concentrations using the in vitro phosphatase assay.

subsequently obtained from commercial sources and subjected to biological assays.

In Vitro Experimental Assay of the Candidate Compounds Identified by CADD. Candidate compounds identified by CADD were screened by the in vitro tyrosine phosphatase assay as described in the Experimental Section. The GST-SHP-2 PTP domain fusion protein was incubated with the compounds at room temperature for 30 min before the phosphopeptide substrate was added to the assay systems, allowing compounds to bind to target sites in SHP-2. Each test compound was dissolved in DMSO. The initial concentration of the test compounds was 100 μM. In the experimental assay, DMSO was included as the negative control and sodium orthovanadate (100 μM), a general nonspecific tyrosine phosphatase inhibitor, was used as the positive control. From the 165 commercially available compounds tested, we identified nine compounds that inhibit SHP-2 catalytic activity with various efficiencies (Figure 3A). The structures of the nine active compounds are shown in Figure 3B. Except no. 238, which does not seem to be bioavailable, the other eight compounds are druglike. At a concentration of 100 μM, the five most active compounds (nos. 162, 216, 220, 212, and 226) inhibited 70%, 55%, 50%, 45%, and 40% of SHP-2 enzymatic activity, respectively. The inhibitory effects of the compounds were dose-dependent. For

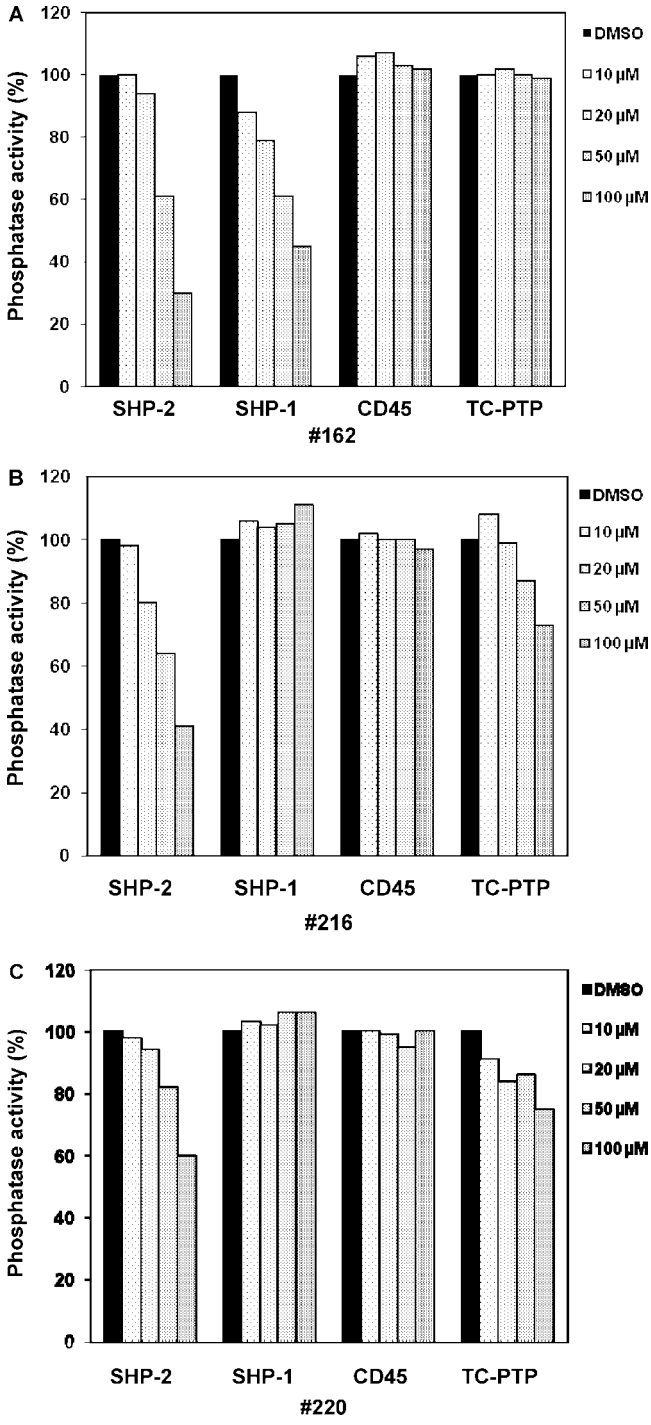

Figure 4. Functional specificity of the top three active compounds. Compound nos. 162 (A), 216 (B), and 220 (C) at the indicated concentrations were subjected to the phosphatase assays using SHP-2, SHP-1, CD45, and TC-PTP as enzymes. DMSO was used as negative controls.

example, at 200 μM, no. 162 essentially blocked the catalytic activity of SHP-2 (Figure 3C).

We next determined functional specificities of the three most active compounds, i.e., the activities on related tyrosine phosphatases, in particular those phosphatases that are highly expressed in hematopoietic cells, such as SHP-1, CD45, and TC-PTP. As shown in Figure 4A, compound no. 162 equally inhibited SHP-2 and SHP-1 phosphatases, but it did not inhibit the other two phosphatases tested. Compounds nos. 216 and 220 selectively inhibited SHP-2 vs SHP-1. Both of them also

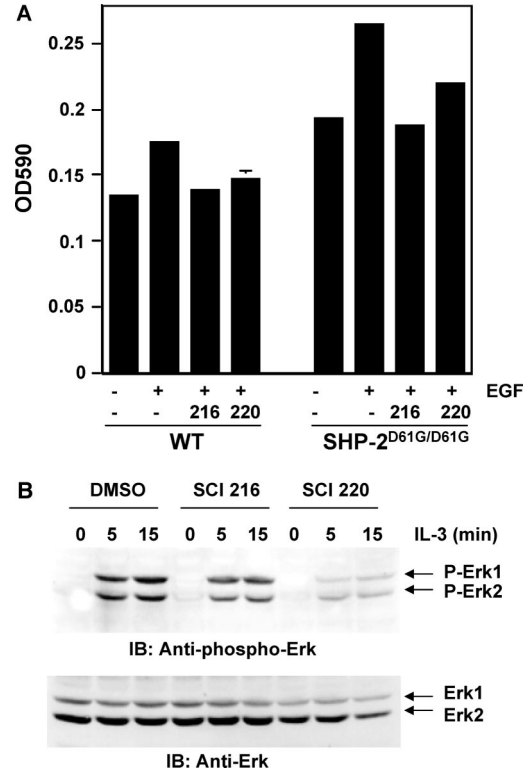

Figure 5. In vivo effects of active compounds on SHP-2 activity and SHP-2-mediated signaling. (A) WT and mutant embryonic fibroblasts derived from activating mutation SHP-2 D61G knock-in mice were starved in serum-free medium overnight and were treated with the indicated compounds in the last 1 h. Cells were then stimulated with EGF (20 ng/mL) for 10 min. Whole cell lysates were prepared and immunoprecipitated with anti-SHP-2 antibody. Immunoprecipitates were washed and then subjected to the in vitro phosphatase assay as described in the text. (B) Ba/F3 cells were deprived of IL-3 for 8 h. The cells were treated with the indicated compounds 1 h prior to the stimulation with IL-3 (2 ng/mL). Cells were harvested after the indicated periods of time. Whole cell lysates were prepared and examined for Erk activation with antiphospho-Erk immunoblotting. The blot was striped and reprobed with antipan Erk antibody to examine protein loading.

showed slight inhibition of TC-PTP (Figure 4B and 4C). It is worth mentioning that although compound no. 162 did not show selectivity between SHP-2 and SHP-1, this compound is still useful for future studies of SHP-2 function in nonhematopoietic cells because expression of SHP-1 phosphatase is restricted to hematopoietic and epithelial cells.

In Vivo Effects of the Active Compounds on SHP-2-Mediated Cellular Function. We next tested whether the active compounds are effective in cells. Compound nos. 216 and 220 that selectively inhibit SHP-2 but not SHP-1 were used to treat WT and mutant mouse embryonic fibroblasts with an activating mutation D61G in SHP-2. Activation of SHP-2 in response to EGF stimulation was determined using the immunocomplex phosphatase assay. Compared to WT SHP-2, catalytic activity of SHP-2 D61G mutant in the mutant cells was increased both at the basal level and following EGF stimulation (Figure 5A), consistent with previous studies.[10,41] In the presence of the inhibitors, EGF-induced activation of SHP-2 was blocked in both WT and SHP-2$^{D61G/D61G}$ cells (Figure 5A). We also examined IL-3-induced activation of Erk kinases in hematopoietic cells that requires SHP-2 catalytic activity.[42] Preincubation of Ba/F3 cells with the inhibitors suppressed Erk activation by IL-3 (Figure 5B), in agreement with previous results that SHP-2 catalytic activity is required for optimal

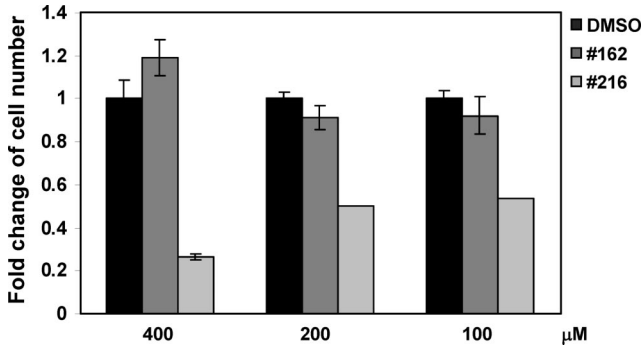

Figure 6. Effects of SHP-2 inhibitors on cell proliferation. Ba/F3 cells were cultured in IL-3 (1 ng/mL) containing medium supplemented with compound nos. 162 and 216 (100 or 200 μM). DMSO was included at negative controls. Cell numbers were determined 72 h later using the MTS assay.

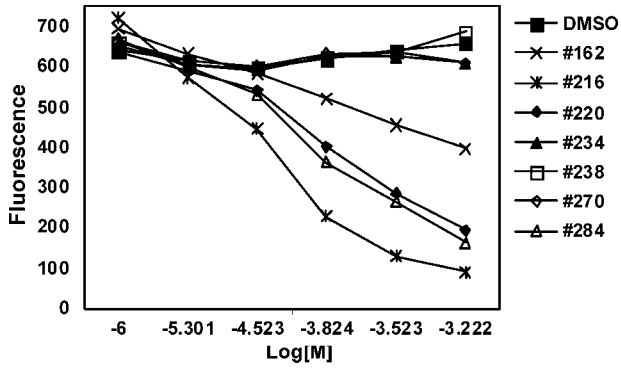

Figure 7. Effects of active compounds on SHP-2 fluorescence. Fluorescence titration of SHP-2 was done as described in the Experimental Section. The fluorescence is plotted against the log concentration in mol/L (Log[M]) for each compound.

activation of Erk kinases.[42] These data suggest that the two inhibitors tested are effective in inhibiting SHP-2 in cells.

To further determine the effects of the active compounds on cellular function mediated by SHP-2, we treated Ba/F3 cells with nos. 162 and 216 (no. 220 showed modest acute cytotoxicity, and was therefore excluded from the test) and assessed cell proliferation response to IL-3. As shown in Figure 6, treatment of Ba/F3 cells with no. 162 did not significantly disturb cell growth, possibly because no. 162 does not have selectivity in between SHP-2 and SHP-1 phosphatases. However, treatment with no. 216 greatly decreased IL-3-stimulated cell growth, consistent with the overall positive role of SHP-2 catalytic activity in cellular response to IL-3.[42]

Verification of the Binding between Active Compounds and SHP-2 via Fluorescence Titrations. To validate that the active compounds identified were binding directly to the SHP-2 protein, we determined whether the active compounds directly interacted with the SHP-2 PTP domain using fluorescence quenching and taking advantage of the four tryptophans in the PTP domain, with Trp248 located 15 Å from the putative pY + 5 pocket and Trp423 located 8 Å from the phosphatase active site. Of the nine active compounds identified by the in vitro phosphatase assay, nos. 216, 284, and 220 showed strong quenching of SHP-2 fluorescence, while quenching occurred to a small extent at the higher concentrations with no. 162. The other three compounds (nos. 234, 238, and 270) did not show significant quenching (Figure 7). Compound nos. 212 and 226 showed strong fluorescence at the excitation wavelength of 295 mm. Therefore, their binding to SHP-2 protein could not be determined using this measurement. Thus, the fluorescence quenching experiments indicate that at least four active compounds identified do directly bind to SHP-2, thereby leading to their biological activities.

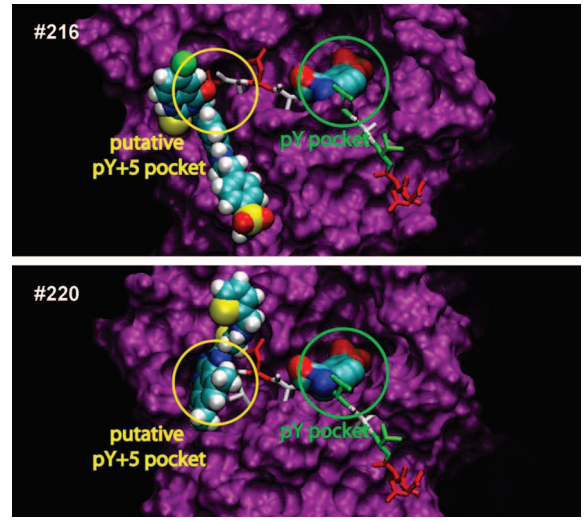

Figure 8. Predicted binding modes of compound nos. 216 and 220. Top: no. 216 bound to the 2.4 ns molecular dynamics conformation of SHP-2 PTP. Bottom: no. 220 bound to the crystallographic conformation [PDB ID 2SHP] of SHP-2 PTP. The crystallographic conformation of a SHP-1 bound pY-peptide [1FPR] after alignment of the SHP PTP domains illustrates that both inhibitors bind to the putative pY + 5 pocket of SHP-2 PTP and not to the pY-binding pocket (the pY-peptide pY residue is shown as a molecular surface and other residues as lines colored by residue type: red = acidic, green = polar, white = hydrophobic).

Predicted Binding Mode between Active Compounds and the SHP-2 PTP Domain. The placement of the sphere sets for docking in the putative pY + 5 pocket of the SHP-2 PTP pY-peptide binding group constrains the docking procedure to place compounds in this region of the protein. The SHP-2 vs SHP-1 specificity exhibited by compound nos. 216 and 220 suggest that this strategy helped in discovering compounds that targeted this pocket. Analysis of the docked poses of these two compounds from the secondary docking shows that both bind this pocket and neither bind to the pY pocket (i.e., catalytic site), which is essentially identical in SHP-1 and SHP-2 (Figure 8). Also, while no. 220 bound most favorably to the crystallographic conformation of SHP-2 PTP, no. 216 bound most favorably to the molecular dynamics snapshot taken at 2.4 ns. These two PTP conformations exhibit different surfaces adjacent to the putative pY + 5 pocket, and the two compounds bind in contrasting poses accordingly. In the case of no. 220, a second shallower pocket immediately adjacent to the pY + 5 pocket serves to accommodate the portion of the compound that is not in the pY + 5 pocket (Figure 8, bottom). In contrast, the region of no. 216 not in the pY + 5 pocket extends in the opposite direction on the surface of the protein. It binds to a groove that is created by normal thermal fluctuations in the PTP residues during the course of the molecular dynamics simulation (Figure 8, top). This groove is not well-defined in the crystallographic conformation, hence the preferential binding of no. 216 to the molecular dynamics conformation vs the crystallographic conformation. Although the inclusion of additional noncrystallographic conformations generated by explicit-water molecular dynamics and used for the secondary docking significantly increases the computational cost of the CADD procedure, in

the case of SHP-2 PTP it has led to the discovery of one of the more promising lead compounds.

Conclusions

We have reported the identification of SHP-2 selective inhibitors using CADD combined with experimental assay. Activating mutations of SHP-2 phosphatase are directly associated with the pathogenesis of Noonan syndrome and childhood leukemias.[6–8] Accordingly, the availability of SHP-2 selective inhibitors would greatly facilitate the development of new therapeutic drugs for these diseases. However, a big challenge in SHP-2 inhibitor development has been the specificity of the compounds. The active site of SHP-2 is almost identical to that of the highly related SHP-1 phosphatase. Therefore compounds that target the catalytic core of SHP-2 often inhibit SHP-1 phosphatase. As SHP-1 plays an opposing role in hematopoietic cell processes due to the specificity determined by its N-terminal SH2 domains,[1,3] it is vitally important to develop inhibitors with high selectivity for SHP-2 over SHP-1. To achieve this goal, in the present study we chose to target a pY-peptide binding site in SHP-2 sufficiently distant from the catalytic site and having unique features in both amino acid composition and structure. Using CADD in combination with experimental assay, we have identified active compounds that selectively inhibit SHP-2 but not SHP-1 by disrupting the interaction between pY-peptide substrates and the SHP-2 catalytic domain.

Nine of the 165 computationally selected compounds showed various extents of inhibitory effects on SHP-2 catalytic activity in the in vitro phosphatase assay (Figure 3A). Among the three most active compounds, two have selectivity for SHP-2 over SHP-1. Additional biological analyses verified that these two compounds (nos. 216 and 220) were effective in inhibiting SHP-2-mediated signaling and cellular functions. Four of the nine compounds were verified to directly bind to the SHP-2 protein using fluorescence titrations. However, further optimization of the compounds is necessary to improve their potency. Future studies will look at the mode of activity of these compounds as well as expand the CADD screen to identify additional molecules, with the aim of having a diverse set of lead compounds for the development of novel SHP-2 inhibitors for experimental and clinical purposes.

Acknowledgment. This work was supported by NIH HL082670 (C.K.Q. and A.D.M.) and F32 CA119771 (O.G.), and the University of Maryland Computer-Aided Drug Design Center. The authors also acknowledge a generous grant of computer time from the National Cancer Institute Advanced Biomedical Computing Center and thank Dr. Alba Macias and Dr. Shijun Zhong for helpful discussions.

References

(1) Neel, B. G.; Gu, H.; Pao, L. The "shp"ing news: SH2 domain-containing tyrosine phosphatases in cell signaling. *Trends Biochem. Sci.* **2003**, *28*, 284–293.
(2) Tonks, N. K. Protein tyrosine phosphatases: from genes, to function, to disease. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 833–846.
(3) Qu, C. K. Role of the SHP-2 tyrosine phosphatase in cytokine-induced signaling and cellular response. *Biochim. Biophys. Acta* **2002**, *1592*, 297–301.
(4) Eck, M. J.; Pluskey, S.; Trub, T.; Harrison, S. C.; Shoelson, S. E. Spatial constraints on the recognition of phosphoproteins by the tandem SH2 domains of the phosphatase SH-PTP2. *Nature* **1996**, *379*, 277–280.
(5) Hof, P.; Pluskey, S.; Dhe-Paganon, S.; Eck, M. J.; Shoelson, S. E. Crystal structure of the tyrosine phosphatase SHP-2. *Cell* **1998**, *92*, 441–450.
(6) Tartaglia, M.; Mehler, E. L.; Goldberg, R.; Zampino, G.; Brunner, H. G.; Kremer, H.; van der Burgt, I.; Crosby, A. H.; Ion, A.; Jeffery, S.; Kalidas, K.; Patton, M. A.; Kucherlapati, R. S.; Gelb, B. D. Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat. Genet.* **2001**, *29*, 465–468.
(7) Tartaglia, M.; Niemeyer, C. M.; Fragale, A.; Song, X.; Buechner, J. Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat. Genet.* **2003**, *34*, 148–150.
(8) Loh, M. L.; Vattikuti, S.; Schubbert, S.; Reynolds, M. G.; Carlson, E.; Jung, A.; Hahlen, K.; Hasle, H.; Licht, J. D.; Gelb, B. D. Mutations in PTPN11 implicate the SHP-2 phosphatase in leukemogenesis. *Blood* **2004**, *103*, 2325–2331.
(9) Bentires-Alj, M.; Paez, J. G.; David, F. S.; Keilhack, H.; Halmos, B.; Naoki, K.; Maris, J. M.; Richardson, A.; Bardelli, A.; Sugarbaker, D. J.; Richards, W. G.; Du, J.; Girard, L.; Minna, J. D.; Loh, M. L.; Fisher, D. E.; Velculescu, V. E.; Vogelstein, B.; Meyerson, M.; Seller, W. R.; Neel, B. G. Activating mutations of the Noonan syndrome-associated SHP2/PTPN11 gene in human solid tumors and adult acute myelogenous leukemia. *Cancer Res.* **2004**, *64*, 8816–8820.
(10) Araki, T.; Mohi, M. G.; Ismat, F. A.; Bronson, R. T.; Williams, I. R.; Kutok, J. L.; Yang, W.; Pao, L. I.; Gilliland, D. G.; Epstein, J. A.; Neel, B. G. Mouse model of Noonan syndrome reveals cell type- and gene dosage-dependent effects of Ptpn11 mutation. *Nat. Med.* **2004**, *10*, 849–857.
(11) Mohi, M. G.; Williams, I. R.; Dearolf, C. R.; Chan, G.; Kutok, J. L.; Cohen, S.; Morgan, K.; Boulton, C.; Shigematsu, H.; Keilhack, H.; Akashi, K.; Gilliland, D. G.; Neel, B. G. Prognostic, therapeutic, and mechanistic implications of a mouse model of leukemia evoked by Shp2 (PTPN11) mutations. *Cancer Cell* **2005**, *7*, 179–191.
(12) Chan, R. J.; Leedy, M. B.; Munugalavadla, V.; Voorhorst, C. S.; Li, Y.; Yu, M.; Kapur, R. Human somatic PTPN11 mutations induce hematopoietic-cell hypersensitivity to granulocyte-macrophage colony-stimulating factor. *Blood* **2005**, *105*, 3737–3742.
(13) Schubbert, S.; Lieuw, K.; Rowe, S. L.; Lee, C. M.; Li, X.; Loh, M. L.; Clapp, D. W.; Shannon, K. M. Functional analysis of leukemia-associated PTPN11 mutations in primary hematopoietic cells. *Blood* **2005**, *106*, 311–317.
(14) Yu, W. M.; Daino, H.; Chen, J.; Bunting, K. D.; Qu, C. K. Effects of a leukemia-associated gain-of-function mutation of SHP-2 phosphatase on interleukin-3 signaling. *J. Biol. Chem.* **2006**, *281*, 5426–5434.
(15) Chen, L.; Sung, S. S.; Yip, M. L.; Lawrence, H. R.; Ren, Y.; Guida, W. C.; Sebi, S. M.; Lawrence, N. J.; Wu, J. Discovery of a novel shp2 protein tyrosine phosphatase inhibitor. *Mol. Pharmacol.* **2006**, *70*, 562–570.
(16) Barford, D.; Neel, B. G. Revealing mechanisms for SH2 domain mediated regulation of the protein tyrosine phosphatase SHP-2. *Structure* **1998**, *6*, 249–254.
(17) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur. J. Biochem.* **1977**, *80*, 319–324.
(18) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285*, 1735–1747.
(19) MacKerell, A. D., Jr.; Bashford, D; Bellott, M; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S; Gao J; Guo, H; Ha, S; Joseph-McCarthy, D; Kuchnir, L; Kuczera, K; Lau, F. T. K.; Mattos, C; Michnick, S; Ngo, T; Nguyen, D. T.; Prodhom, B; Reiher, W. E.; Roux, B; Schlenkrich, M; Smith, J. C.; Stote, R; Straub, J; Watanabe, M; Wiorkiewicz-Kuczera, J; Yin, D; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(20) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
(21) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
(22) Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992**, *257*, 1078–1082.
(23) Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221*, 709–713.
(24) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. The MIDAS display system. *J. Mol. Graphics* **1988**, *6*, 13–27.
(25) Yang, J.; Cheng, Z.; Niu, T.; Liang, X.; Zhao, Z. J.; Zhou, G. W. Structural basis for substrate specificity of protein-tyrosine phosphatase SHP-1. *J. Biol. Chem.* **2000**, *275*, 4066–4071.
(26) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
(27) Pan, Y.; Huang, N.; Cho, S.; MacKerell, A. D., Jr. Consideration of molecular weight during compound selection in virtual target-

based database screening. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 267–272.
(28) Huang, N.; Nagarsekar, A.; Xia, G.; Hayashi, J.; MacKerell, A. D., Jr. Identification of non-phosphate-containing small molecular weight inhibitors of the tyrosine kinase p56 Lck SH2 domain via in silico screening against the pY + 3 binding site. *J. Med. Chem.* **2004**, *47*, 3502–3511.
(29) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38, 27−38.
(30) Leach, A. R.; Kuntz, I. D. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comput. Chem.* **1992**, *13*, 730–748.
(31) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms used in automated molecular docking. *J. Comput. Chem.* **1997**, *18*, 1175–1189.
(32) Kelley, L. A.; Gardner, S. P.; Sutcliffe, M. J. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* **1996**, *9*, 1063–1065.
(33) Jarvis, R. A.; Patrick, E. A. Clustering Using a Similarity Measure Based on Shared Near Neighbors. *IEEE Trans. Comput.* **1973**, *C22*, 1025–1034.
(34) Tanimoto, T. T. *IBM Internal Report*, Nov. 17, **1957**.
(35) Butina, D. Unsupervised database clustering on daylight's fingerprint and Tanimoto similarity: A fast and automated way to cluster small and large data sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 747–750.
(36) 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.* **2001**, *46*, 3–26.
(37) Yang, J.; Liang, X.; Niu, T.; Meng, W.; Zhao, Z.; Zhou, G. W. Crystal structure of the catalytic domain of protein-tyrosine phosphatase SHP-1. *J. Biol. Chem.* **1998**, *273*, 28199–28207.
(38) Yang, J.; Liu, L.; He, D.; Song, X.; Liang, X.; Zhao, Z. J.; Zhou, G. W. Crystal structure of human protein−tyrosine phosphatase SHP-1. *J. Biol. Chem.* **2003**, *278*, 6516–6520.
(39) Imhof, D.; Wavreille, A. S.; May, A.; Zacharias, M.; Tridandapani, S.; Pei, D. Sequence specificity of SHP-1 and SHP-2 Src homology 2 domains. Critical roles of residues beyond the pY + 3 position. *J. Biol. Chem.* **2006**, *281*, 20271–20282.
(40) Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308–1315.
(41) Keilhack, H.; David, F. S.; McGregor, M.; Cantley, L. C.; Neel, B. G. Diverse biochemical properties of Shp2 mutants. Implications for disease phenotypes. *J. Biol. Chem.* **2005**, *280*, 30984–30993.
(42) Yu, W. M.; Hawley, T. S.; Hawley, R. G.; Qu, C. K. Catalytic-dependent and -independent roles of SHP-2 tyrosine phosphatase in interleukin-3 signaling. *Oncogene* **2003**, *22*, 5995–6004.

JM800229D