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

Niu Huang, Ashish Nagarsekar, Guanjun Xia, Jun Hayashi,* and Alexander D. MacKerell, Jr.*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, 20 Penn Street, Baltimore, Maryland 21201

Received September 18, 2003

The protein p56 lymphoid T cell tyrosine kinase (Lck) is predominantly expressed in T lymphocytes where it plays a critical role in T-cell-mediated immune response. Lck participates in phosphotyrosine-dependent protein−protein interactions through its modular binding unit, the Src homology-2 (SH2) domain. Accordingly, virtual screening methods combined with experimental assays were used to identify small molecular weight nonpeptidic compounds that block Lck SH2 domain-dependent interactions. Virtual screening included scoring normalization procedures and postdocking structural clustering that is shown to facilitate the selection of active compounds. By targeting the well-defined hydrophobic binding pocket known to impart specificity on Lck−protein interactions (i.e., pY + 3 site), inhibitors of the Lck SH2 domain were discovered that omit the phosphotyrosine (pY) or related moieties. The 34 out of 196 computationally selected compounds were shown to inhibit Lck SH2 domain association with phosphorylated immunoreceptor tyrosine based activation motifs peptide. Twenty-four of the active compounds were further tested for their ability to modulate biological function. Thirteen of these compounds showed inhibitory activity in mixed lymphocyte culture assay. Flourescence titration experiments on four of these active compounds further verified their binding to the SH2 domain. Because of their simple chemical structures, these small organic compounds have the potential to act as lead compounds for the development of novel immunosuppressant drugs.

Introduction

The protein p56 Lck (lymphoid T cell tyrosine kinase) is a member of the Src family of tyrosine kinases and is predominantly expressed in T lymphocytes and natural killer cells where it plays a critical role in T-cell-mediated immune responses.[1,2] p56 Lck is responsible for the phosphorylation of conserved tyrosine residues of CD3 chains, called immunoreceptor tyrosine-based activation motifs (ITAMs), the first step required for T cell activation signaling cascades.[3,4] Failure of the p56 Lck SH2 domain to bind to ITAMs of CD3 will hamper the T cell receptor (TCR) proximal activation process and suppress the downstream T cell activation signaling cascades.[3,5] Lck participates in phosphotyrosine (pY) dependent protein−protein interactions through its modular binding units, called Src homology-2 (SH2) domains.[6] Accordingly, ligands that are able to block Lck SH2 domain-dependent protein−protein interactions may ultimately find therapeutic utility as immunosuppressants and in the treatment of T cell leukemias, lymphomas, and autoimmune diseases such as rheumatoid arthritis.[2,7]

A phosphopeptide library screen has identified a preferred pY containing peptide binding sequence Ac-pY-E-E-I (Figure 1a) for the Lck SH2 domain.[8] This tetrapeptide is an attractive lead structure for the rational design of agents to compete with the SH2 domain's natural ligands. Unfortunately, the tetrapeptide Ac-pY-E-E-I has several undesirable features that hinder its ability to elicit a response in cell-based assays of T cell activation. First, the phosphate group, an essential element for peptide binding to the SH2 domain, is metabolically unstable to phosphatases present in cells, and second, the five negative charges at physiological pH and the high peptidic character may limit its ability to reach efficacious concentrations inside the cell. Because of the conservation of the pY binding site, a pY or similar functional group is strictly required to maintain the peptide binding.[1] Attempts to design SH2 inhibitors with high receptor binding affinity, chemical stability, and minimally charged phosphate group replacements (Figure 1b−d) have met with limited success.[9−11] Accordingly, novel approaches for the identification of p56 Lck SH2 domain inhibitors that avoid the problems associated with the strategies applied to date are required.

High-resolution X-ray structures of the Lck SH2 domain complexed with the pY-E-E-I peptide have provided a 3D molecular map revealing that the pY and Ile residues of the peptide are bound to two well-defined cavities, referred to as the pY and pY + 3 binding sites, where the interaction resembles a two-pronged plug engaging a two-hole socket.[1] This binding mode is consistent with experimental observation that SH2 affinity is strongly dependent on the pY and Ile side chains.[12] Moreover, site mutations of amino acid residues in the pY + 3 binding site switched the binding specificity,[13−15] which led to the proposal that the pY +

* To whom correspondence should be addressed. For J.H.: phone, 410-706-8295; fax, 410-706-0346; e-mail, jhayashi@rx.umaryland.edu. For A.D.M.: phone, 410-706-7442; fax, 410-706-5017; e-mail: amackere@rx.umaryland.edu.

10.1021/jm030470e CCC: $27.50
Published on Web 06/04/2004

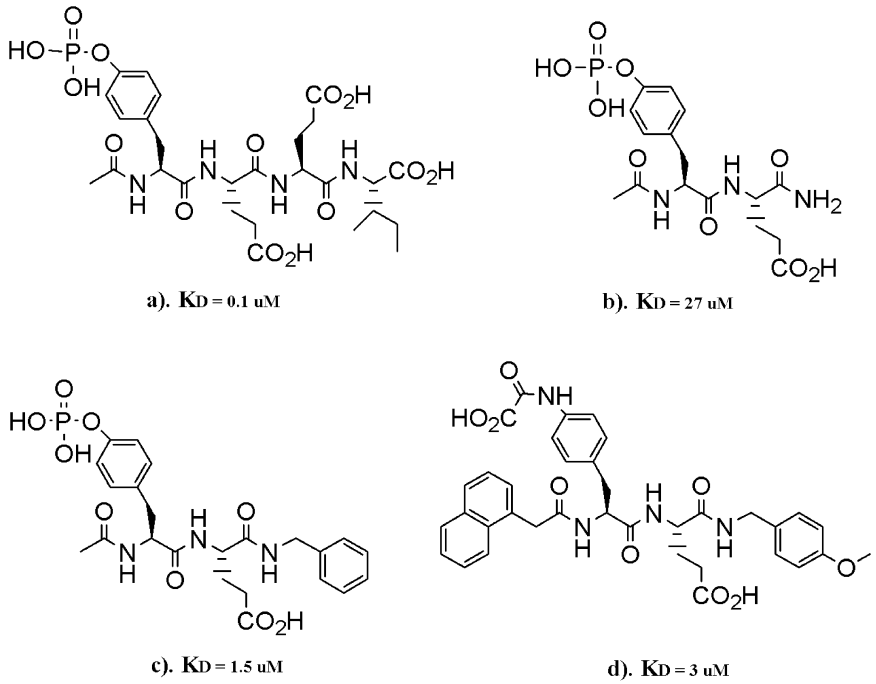

Figure 1. Peptide-based inhibitors of the Lck SH2 domain.

3 binding pocket is important for specific binding.[12] Thus, the pY + 3 site represents a novel target site for the application of rational drug design approaches for identifying nonpeptidic, specific inhibitors of the p56 Lck SH2 domain.

Traditionally, it takes 12−15 years and costs as much as 500 million dollars to develop a new drug and bring it to market.[16] Computer screening of compounds promises to speed up the discovery of new drugs and dramatically reduce costs by identifying compounds with a high probability of binding to a target protein on the basis of a variety of criteria, including receptor-based properties such as binding affinity, selectivity, and ligand-based physicochemical properties such as molecular size and solubility.[17,18] Structure-based virtual database screening is designed to identify novel compounds complementary to a putative binding site on an enzyme or receptor.[19] From the computer screening, a small number of "hits" from a large chemical library are selected for experimental analysis. This approach has successfully been applied to identify thrombin inhibitors,[20] CD4 blockers,[21] HIV integrase inhibitor,[22] and growth hormone antagonists,[23] among others.

In the present study, we apply virtual screening techniques followed by experimental assays to identify small molecular weight (MW) nonpeptidic compounds targeting the pY + 3 binding site that are potent inhibitors of the Lck SH2 domain and omit the phosphotyrosine or related moieties.

Experimental Section

3D Database System Generation. A 3D database of 2 000 000 commercially available compounds was built via a procedure described previously.[24] Briefly, this procedure involves database file format conversion, initial 3D geometry generation, hydrogen addition, charge assignment, and force field optimization using SYBYL6.4.[25] This strategy is similar to the 2D−3D building program CORINA.[26] An in-house 3D database system, CHAOS (CHemoinformatics Assisted cOmputer Screening), and an integrated computational platform have been set up to maximize the performance of our computer screening studies. The major data source contains 2D SDF format files and 3D MOL2 format files, and the database management module includes utility programs developed on the UNIX system using PERL and C/C++ programming languages, which are used to integrate, analyze, and retrieve data.

Primary Database Screening. The 3D structure of the Lck SH2 domain in complex with the phosphotyrosine peptide Ac-pY-E-E-I [1] was retrieved from the Protein DataBank (PDB identifier: 1lkk).[27] The coordinates, with all water molecules removed, served as the starting point for docking and subsequent calculations. Charges and hydrogens were added to the protein by SYBYL. All docking calculations were carried out with DOCK[28] using flexible ligands based on the anchored search method.[29] The solvent-accessible surface[30] was calculated with the program DMS[31] using a probe radius of 1.4 Å. Sphere sets, required for initial placement of the ligand during database screening, were calculated with the DOCK associated program SPHGEN. Docking targeted the pY + 3 binding site of the Lck SH2 domain because of that site making a significant contribution to SH2 domain binding affinity and specificity. Thus, only those spheres within 6 Å of the pY + 3 binding site and within 3 Å of the crystallographically determined location of the pY + 3 Ile residue were selected for the search. Ligand−receptor interaction energies were approximated by the sum of electrostatic and van der Waals components as calculated by the GRID method[28,32] implemented in DOCK using default values. The GRID box dimensions were 36 × 35 × 29 Å^3 based on the edges being 6 Å beyond the position of the crystallographic Ac-pY-E-E-I peptide. To avoid identifying compounds that bind to the pY binding site, phenolphosphate was maintained in this site, and the rest of the peptide ligand was deleted. To avoid improper electrostatic interactions between docked ligands with this added moiety, a total charge of zero was assigned to it.

Database screening initially selected compounds that contain 10 or less rotatable bonds and between 10 and 40 nonhydrogen atoms. Ligand flexibility was considered by dividing each compound into a collection of nonoverlapping rigid segments. Individual rigid segments with five or more heavy atoms (e.g., aromatic rings) were selected as "anchors". Each anchor was docked separately into the binding site in 200 different orientations, based on different overlap of the anchor

atoms with the sphere set, and was energy-minimized. The remainder of each molecule was built onto the anchor in a stepwise fashion until the entire molecule was built, with each step corresponding to a rotatable bond. At each step the dihedral about the rotatable bond, which was connecting the new segment to the previously constructed portion of the molecule, was sampled in 10° increments and the lowest energy conformation was selected. During the build-up procedure, selected conformers were removed on the basis of energetic considerations and maximization of diversity of the conformations being sampled, as previously described.[29,33] The orientation with the most favorable interaction energy was finally selected.

Recent studies in our laboratory show that the DOCK energy score based approach is biased toward the selection of high molecular weight compounds because of the contribution of the compound size to the energy score. Such biasing behavior was observed to depend on the shape and chemical properties of the binding pocket. Hence, a computationally efficient procedure[24] was developed via normalization of the energy score by the number of heavy atoms N or by a selected power of N in each respective compound. This normalization approach shifts the MW distribution of selected compounds into better agreement with that of the entire database. In the present study, the $N^{1/2}$ normalized vdW attraction interaction energy value,

$$\mathrm{IE}_{\mathrm{norm.vdW}} = \frac{\mathrm{IE}_{\mathrm{vdW}}}{N^{1/2}}$$

was applied to select small MW compounds with the best complementarities to the pY + 3 binding site. By decrease of the number of high MW compounds selected, it is anticipated that absorption and disposition properties of those compounds would be improved.[34,35] Moreover, lower MW compounds are better suited for later lead optimization efforts.[36] From this procedure, a total of 25 000 compounds were selected.

Secondary Database Docking. The selected 25 000 compounds based on $N^{1/2}$ normalized vdW attraction interaction energy were subject to secondary screening that was performed by applying a more rigorous docking method that included simultaneous energy minimization of the anchor fragment during the iterative build-up procedure. Two sets of 1000 compounds were selected on the basis of (1) the total interaction energy and (2) the $N^{1/2}$ normalized total interaction energy scores.

Diversity Analysis and Final Compound Selection. To facilitate the selection of chemically diverse compounds for biologically assay, structural clustering was applied. This was performed by dividing each set of 1000 compounds from the secondary dock run into chemically dissimilar clusters by applying the Tanimoto similarity indexes[37] using the program MOE.[38] The clustering procedure started with the calculation of the molecular fingerprints, followed by the calculation of the pairwise Tanimoto similarity matrix $\mathbf{A}(i,j)$ containing the similarity metric between the molecular fingerprints of compounds i and j. From $\mathbf{A}(i,j)$, a binary matrix $\mathbf{B}$ was created such that $\mathbf{B}(i,j)$ has the value 1 if $\mathbf{A}(i,j)$ is equal to or greater than S, or 0 otherwise, where S is the predefined similarity threshold used to determine if two fingerprints are similar. Then, the rows of the $\mathbf{B}$ matrix were treated as fingerprints, where two molecules belong to the same cluster if the Tanimoto coefficient of their corresponding rows in $\mathbf{B}$ is greater than or equal to T, the predefined overlap threshold. This results in two molecules being clustered together if they are similar to the same set of molecules. Cluster similarity and overlap thresholds of 60 and 45, respectively, were used in the present study.

Compounds for biological assay were selected from the dissimilar sets. This was performed by individually analyzing the clusters and selecting compounds from each cluster based on several criteria such as adequate solubility (ClogP $\leq$ 5), molecular weight ($\leq$500 Da), the number of the hydrogen bond donors and acceptors ($\leq$10), and chemical stability.

In Vitro Assay for the Inhibition of Binding of the CD3ζ ITAM2 Peptide to the SH2 Domain of p56 Lck. Phosphopeptide H-CAEApYSEIG(Nle)-OH corresponding to the ITAM2 of CD3ζ was synthesized (Anaspec, San Jose, CA) and conjugated to maleimide activated agarose beads through the N terminus cysteine residue using the Sulfolink coupling kit (Pierce, Rockford, IL). p56 Lck has been shown to bind to this ITAM residue with high affinity.[8]

a. Preparation of Jurkat Cell Lysate. Jurkat cells (human T cell lymphoma) were grown in RPMI-1640 medium containing glutamine (BioSource, Camarillo, CA) supplemented with 10% iron-supplemented calf serum (sCS) (HyClone, Logan, UT). Confluent cells were harvested, washed three times in chilled, sterile phosphate buffered saline (PBS, BioSource), and lysed (5×10^5 mL^{-1}) in buffer containing 1% Triton X-100, 1 mM EDTA, 0.2 mM sodium orthovanadate, 500 μM AEBSF, 150 nM aprotinin, 1 μM E-64, and 1 μM leupeptin (Protease Cocktail Set I, Calbiochem, San Diego, CA). Cell lysate was sonicated on ice (VibraCell, Sonics and Materials, Danbury, CT) with 6 $\times$ 5 s bursts with a 1 min cooling period between bursts. The lysate was centrifuged at 20000g at 4 °C, and the supernatant was used in the assay as a source of p56 Lck.

b. Addition of Putative Inhibitors. Putative inhibitors to be tested were freshly dissolved in dimethyl sulfoxide (DMSO, cell culture grade, Sigma) at 10 mM before the assay, and 10 μL of this was added to 1 mL of cell lysate in 1.5 mL microcentrifuge tubes. The tubes were placed on an end-over-end rotator at 4 °C for 30 min. An amount of 20 μL of a 50% slurry of the ITAM-conjugated agarose beads in PBS was then added to each tube, and the tubes were rotated overnight.

c. Western Blot Analysis. The beads were washed three times in 800 μL of lysis buffer. Excess lysis buffer was aspirated off, and 10 μL of 2$\times$ sample loading buffer containing 3% β-mercaptoethanol was added to the beads. After being immersed in a boiling water bath for 2 min, sample tubes were centrifuged and the supernatant was loaded onto 8% Tris-HCl SDS−PAGE precast minigels (GeneMate, ISC Bioexpress, Kaysville, UT). Electrophoresis was run at 100 V for 1.5 h in a Protean III minigel unit (Bio-Rad, Hercules, CA). The gels were blotted onto immobilon-P membrane (Millipore, Bedford, MA) in a mini-trans-blot electrophoretic transfer cell (Bio-Rad). The membrane was blocked with PBS solution containing 5% powdered milk for 1 h, exposed to primary antibody (rabbit anti-Lck, Upstate Biotechnologies, Lake Placid, NY) at 1:5000 dilution in 1% w/v powdered milk in PBST [PBS containing 0.5% v/v Tween 20], and washed three times for 5 min each in PBST. The membrane was then exposed to secondary antibody (horseradish peroxidase labeled goat antirabbit IgG (KPL, Gaithersburg, MD) at 1:5000 dilution in 1% milk-PBST and washed three times for 15 min each in PBST. Immunoblots were visualized with SuperSignal West Femto Maximum Sensitivity substrate (Pierce) in the EpiChemi^3 darkroom equipped with a cooled CCD camera (UVP, Upland, CA).

Physical Characterization of Active Compounds. All the compounds showing activity in the in vitro assay were subjected to physical characterization. Molecular weights of the compounds were confirmed by mass spectrometry. Briefly, the compounds were dissolved in methanol with 0.1% acetic acid and injected on a LCQ mass spectrometer (Finnigan Mat, San Jose, CA), and spectra were acquired using the Thermo Finnigan software. The purity of the compounds was confirmed by thin-layer chromatography. Briefly, compounds were dissolved in 5% methanol/chloroform and run on silica gel GHLF plates and visualized under ultraviolet light in an EpiChemi^3 darkroom and also under iodine vapor.

Mixed Lymphocyte Culture. C57BL/6J and Balb/c mice (Jackson Laboratories, Bar Harbor, ME) were killed by cervical dislocation, and mesenteric lymph nodes and spleens were collected under sterile conditions. A single cell suspension of these tissues from each strain of mice was prepared in RPMI 1640 medium, and after being washed twice, cells were resuspended (2×10^6 cells/mL) in RPMI 1640 supplemented with 10% CS. Equal volumes of cell suspensions from two

strains of mice were mixed, and 100 μL of the mix was dispensed into sterile 96-well plates. The 100 mM stock solutions of compounds were prepared in DMSO. Compound solution to be added to the culture was prepared in RPMI 1640 containing 1% BSA. One hundred microliters of the compound solution was added to each well to achieve final compound concentrations of 100, 10, 1, and 0.1 μM with maximum final DMSO concentrations of 0.1, 0.01, 0.001, and 0.0001%, respectively. Blank DMSO controls and mitomycin (25 μg/well) controls were run for each compound/DMSO concentration. The 96-well plates were incubated at 37 °C in a CO_2 incubator. Half (100 μL) of the culture medium was replaced, without disturbing the cells, with fresh medium containing the compounds (or vehicle or mitomycin controls) at 24 h. An amount of 1 μCu [^{3}H]thymidine (Perkin-Elmer, Boston, Massachusetts) in 30 μL of RPMI was added to each well at 42 h, and 6 h later cells were harvested on a cell harvester (Inotech, Rockville, MD) onto a glass fiber filter, type G-7 (Inotech). Filters were treated with methanol, washed three times with PBS, and finally washed again with methanol. Filter papers were air-dried, and the radioactivity was counted on a LS6500 scintillation counter (Beckman). Percent inhibition was calculated as

$$\frac{1 - ((\text{sample cpm} - \text{mitomycin cpm})}{(\text{DMSO cpm} - \text{mitomycin cpm}))} \times 100$$

Cell Viability. For each condition in the mixed lymphocyte assay, an additional well with identical conditions was prepared. The viability of cells in these wells was determined by trypan blue dye exclusion assay in order to assess the cytotoxicity of the compounds.

Expression and Purification of the p56 Lck SH2 Domain. The expression and purification of the Lck SH2 domain were performed according to Tong et al.(1996).[1] The gene corresponding to the p56 Lck SH2 domain (residues 126−226) was cloned into the pRSET-C vector (Invitrogen) with the restriction enzymes NdeI and EcoRI (BioLabs) and was overexpressed in BL 21 *E. coli* with IPTG induction. Cells were resuspended in lysis buffer (PBS, 1% (v/v) Triton X-100, and protease inhibitors cocktail (Calbiochem), pH 7.4). After sonication and 1 h of centrifugation at 40000*g*, the clarified supernatant containing the recombinant protein was applied onto an *o*-phosphate-L-tyrosine agarose affinity column (Sigma). The Lck SH2 protein was eluted with 20 mM *o*-phosphate-L-tyrosine in the elution buffer (200 mM NaCl, 50 mM Tris, pH 7.4). The purified SH2 protein was concentrated and desalted with the Amicon concentrator (Millipore). The protein was estimated to be more than 90% pure judged by the Coomassie Blue staining of SDS−PAGE gel. Protein concentration was determined by BCA (Pierce).

Fluorescence Titrations. Fluorescence spectra were recorded with a luminescence spectrometer LS50 (Perkin-Elmer). All reported fluorescence intensities are relative values and are not corrected for wavelength variations in detector response. The excitation wavelength was 270 nm, and fluorescence was monitored at 588 nm. In all the studies, the Lck SH2 protein and the compounds (6.67 mM in DMSO) were dissolved in Milli-Q water. Titrations were performed by adding increasing volumes of compounds (0.25, 0.5, 1, ..., 64 μL) while maintaining the recombinant protein concentration at 600 nM. Dissociation constants, K_D, were determined using reciprocal plots, $1/\nu$ vs $1/[I]$, where ν represents the percent occupied sites calculated assuming fluorescence quenching to be directly proportional to the percentage of occupied binding sites, [I] represents the concentration of the inhibitor compound, and the slope of the curve equals the K_D.[39]

Results and Discussion

Primary Database Screening. From the first database screen of two million compounds, a total of 25 000 compounds were selected on the basis of the normalized van der Waals attractive energy. Screening targeted the pY + 3 binding site of the Lck SH2 domain because that site makes a significant contribution to SH2 domain binding affinity and specificity.[12] In addition, the pY binding pocket was filled with a phenolphosphate moiety to avoid identifying compounds that bind to that site. Shown in Figure 2 is the Lck SH2 binding surface with spheres used in the docking filling the pY + 3 binding site. Use of van der Waals attractive interaction energy with the Lck SH2 identifies compounds that are structurally complementary with the pY + 3 binding region and avoids selecting compounds whose interactions are dominated by electrostatics. This is important in the present case because of the presence of several charged residues in the vicinity of the pY + 3 binding pocket.

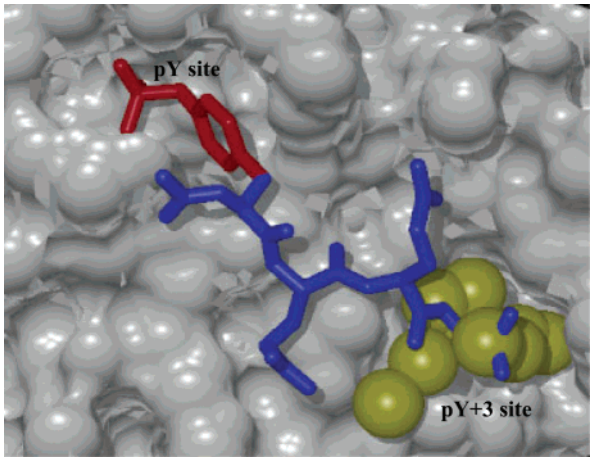

Figure 2. Lck SH2 domain binding surface: Lck SH2 domain (gray, CPK model), phosphotyrosine (red), the rest of phosphorylated peptide (blue), and spheres used in the docking filling the pY + 3 binding site (yellow balls). Structure data are from Tong et al.[1]

Presented in Figure 3A are histograms of the MW probability distributions for the top 25 000 compounds selected on the basis of vdW attractive energy and normalized vdW attractive energy scores. Score normalization based on $N^{1/2}$ leads to a significant downshift in the MW of the selected compounds. The mean MW for the normalized compounds is 345 Da compared to a value of 475 for the unnormalized compounds. This former value is in a good range with respect to druglike properties.[34−36,40] Importantly, a large number of compounds with MW less than 300 are selected by the normalization procedure, whereas most low MW compounds are excluded on the basis of vdW attractive energy scoring alone. The lower MW compounds are advantageous because they generally possess better leadlike properties because of their intrinsic simpler chemical structures, which are considered to be more suitable for further drug optimization.[35,36] Presented in Figure 3B are the histograms of the vdW attractive energy distributions for selected compounds based on the two scoring strategies. It is shown that approximately half of the compounds in the original list still appear in the normalized list, where high MW compounds anticipated to have poor leadlike properties were eliminated. Most of the compounds with very favorable interaction energies (less than −48 kcal/mol) were kept. In addition, the normalization procedure is observed to significantly speed up the secondary docking run because of the exclusion of larger compounds.

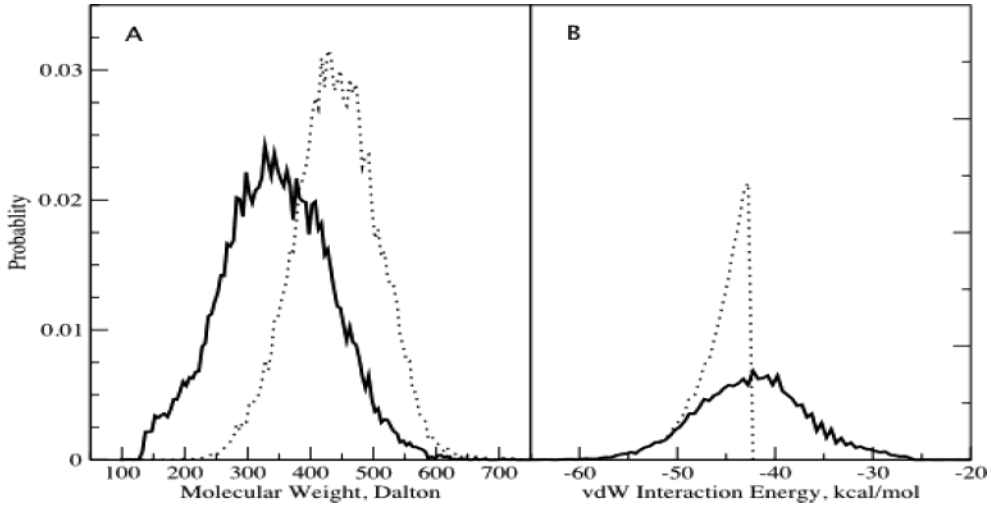

Figure 3. (A) Molecular weight distribution for the top 25 000 compounds selected on the basis of vdW attractive energy (dashed line) and normalized vdW attractive energy (black line). (B) vdW attractive interaction energy probability distribution for the top 25 000 compounds selected on the basis of vdW attractive energy score (dashed line) and normalized score (black line).

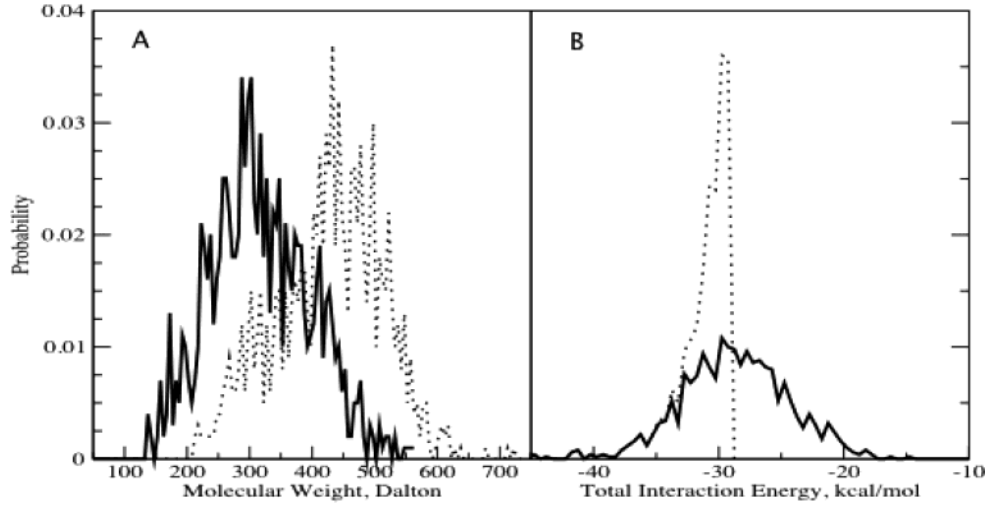

Figure 4. (A) Molecular weight distribution for the top 1000 compounds selected on the basis of total interaction energy score (dashed line) and normalized energy score (black line). (B) Total interaction energy probability distribution for the top 1000 compounds selected on the basis of total interaction energy score (dashed line) and normalized score (black line).

Secondary Database Docking. The selected 25 000 compounds based on the normalized vdW attractive interaction energy were subject to secondary docking by applying a more rigorous docking method that includes additional conformational searching and energy minimization over the initial screening. The total interaction energy score and corresponding normalized score were used to select compounds with both shape and physical complementary to the binding site for further diversity clustering and druglike property filtering analysis. Presented in parts A and B of Figure 4 are histograms of the MW and total interaction energy probability distributions, respectively, for the top 1000 compounds selected on the basis of the total interaction energy and normalized energy scores. Again, score normalization based on $N^{1/2}$ leads to a significant downshift in the MW of the selected compounds. The mean MW for the normalized compounds is 300 Da.

Final Compound Selection. Diversity is an important issue in virtual screening because the purpose of lead discovery is to identify hits covering different regions of chemical space in order to increase the chances of developing a drug candidate with an acceptable pharmacological profile. Hence, two sets of the top 1000 compounds were chosen on the basis of total interaction energy and normalized energy scoring and were individually subjected to clustering. The clustering similarity and overlap thresholds, 60 and 45, respectively, were chosen to produce a reasonable number of diverse clusters (i.e., ~100) for each clustering process, from which two compounds were selected from each cluster. The final compounds for biological assay were chosen from the clusters by considering several physicochemical property filters as described in Experimental Section. From the total energy and normalized total energy score clusters 210 and 218 compounds, respectively, were selected. Comparison of these two sets revealed a significant number of identical compounds; taking these into account led to a final total of 288 unique compounds being selected. Of these, a total of 196 were available from the commercial vendors and purchased for biological assay.

Inhibition of the Phosphorylated ITAM2 Peptide Association with Lck SH2 Domain. The 196 compounds were initially tested for their ability to inhibit p56 Lck SH2 domain association with phospho-

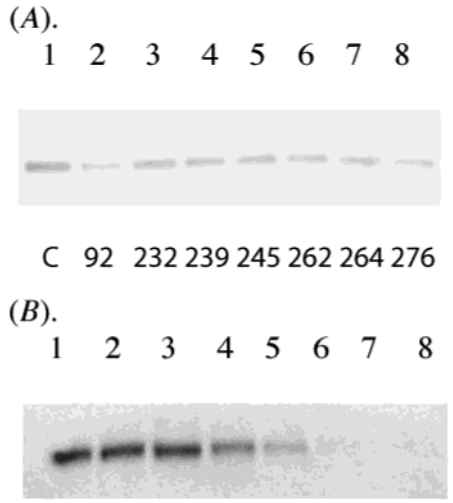

Figure 5. Immunoblots from compounds inhibiting p56 Lck SH2 domain association with phosphotyrosine-containing ITAM2 peptide. The inhibition of p56 Lck is reflected by the decreased intensity of the blots compared to the control. (A) (lane 1) control; (lanes 2−8) compounds **92**, **232**, **239**, **245**, **262**, **264**, and **276** at 100 μM, respectively (compound numbers are listed under the gel). (B) **(**lane 1) control; (lanes 2−8) compound **73** at 10, 20, 30, 40, 60, 80, and 100 μM, respectively.

tyrosine-containing C-terminal ITAM2 peptide in the affinity precipitation assay. Immunoblots from the inhibition assays for compounds **92**, **232**, **239**, **245**, **262**, **264**, **276**, and **73** are shown in Figure 5. The inhibition of p56 Lck binding to the peptide is reflected by the decreased intensity of the blots compared to the control. Figure 5A shows that compounds **92**, **232**, **239**, **245**, **262**, **264**, and **276** have significant inhibitory activities at 100 μM. Figure 5B shows a dose-dependent inhibition of p56 Lck coprecipitation by the inhibitor **73**; at 40 μM (lane 5) the compound significantly blocked p56 Lck association with the ITAM2 peptide. Among the 196 compounds, a total of 34 compounds (Figure 6) were shown to have significant inhibitory activity at 100 μM, yielding a 17% success rate. Of these, compounds **73** and **92** show strong inhibitory activity at 40 and 10 μM, respectively.

The 17% hit rate for the computationally selected compounds is encouraging. Clearly, the DOCK-based screening method is capable of identifying compounds possessing steric and electrostatic complementarity with the target binding site. However, as evidenced by the significant number of compounds not having activity in the in vitro assay, this approach cannot reliably identify the true hits from the false hits that appear to fit into the binding site merely based on the DOCK interaction energy score. This difficulty reflects the approximate docking method being applied, which is the only applicable approach for docking millions of compounds with currently available computational power. Presently, the risk of missing potential active compounds seems to be an unavoidable fact of database reduction. Under this circumstance, if we assume that errors in the scoring functions are not likely to be uniformly distributed,[19] the chance of success may be increased by picking the most diverse hits via "clustering" of the selected compounds. A family-based strategy, which includes preclustering of the database into chemically dissimilar families before docking and choosing only the highest scoring member of each well-docked family, was reported to improve the diversity of hits.[41] However, most clustering schemes are not capable of clustering a database containing millions of compounds, limiting the usefulness of a predocking scheme. Instead, a postdocking clustering strategy was employed in our screening. To show the value of this approach, we individually consider the two sets of 1000 compounds selected on the basis of (1) the total energy and (2) the normalized energy score. Following clustering, for the total energy set, 210 compounds were selected from which 125 were available for assay. Of these, 19 had activity. Of these 19, only 12 would have been identified if the top 210 compunds were selected on the basis of only the total energy score (i.e., without clustering). Similarly, with the normalized score set, 218 compounds were selected from which 150 were obtained for assay, yielding 26 active compounds. Of the active compounds, only 8 would have been identified if the 218 compounds were selected on the basis of only the normalized energy scores. Thus, it is clear that the use of diversity clustering leads to the selection of active compounds that would have been overlooked if energy scoring alone was the criteria for compound selection. Importantly, the selected compounds (Figure 6) possess diverse structural scaffolds with druglike or leadlike properties, thereby serving as lead structures for further drug optimization. Parallel development of lead compounds with different types of structural characteristics has been proposed to be critical in drug optimization in cases where one or more lead series encounter poor absorption, distribution, metabolism, and excretion/toxicity (ADME/T) problems.[42]

The applied energy normalization approach also appears to facilitate the identification of active compounds over the use of total energies alone. In this case, the active compounds from the normalized scoring with clustering can be compared to the top 288 compounds that would have been selected on the basis of the total energy score alone, where 288 represents the number of unique compounds selected from both scoring methods. In this case, only 8 of the 26 active compounds selected from the normalized score and clustering approach occur in the top 288 compounds from the total energy score. Thus, 18 active compounds would have been missed if the total energy scoring alone was the only criterion for selecting compounds for assay. This result indicates that the normalization procedure also facilitates the selection of a more diverse range of active compounds.

Previous studies have shown that neglecting the electrostatic component of ligand solvation biases DOCK ranked compounds toward highly charged species while neglecting the nonpolar component of ligand solvation biases toward larger MW compounds.[43] Our normalization procedure[24] biases the selection toward small MW compounds, which may compensate for the neglect of the nonpolar component of ligand solvation. In addition, problems due to neglect of the electrostatic component of ligand solvation in the present scoring approach may be diminished by ranking compounds using the normalized attractive vdW interaction energy in the primary database screening. These observations indicate that the stepwise docking and normalization procedures applied

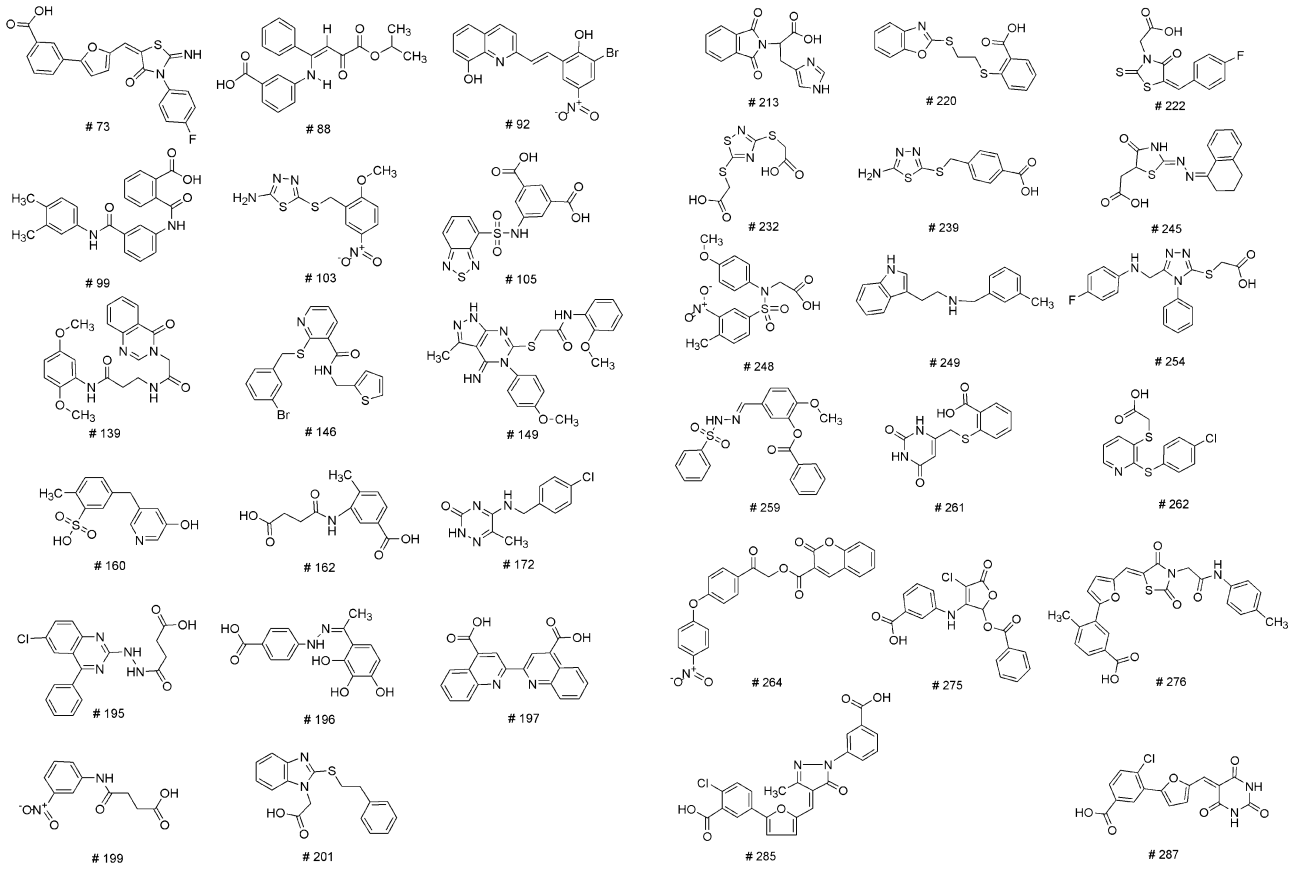

Figure 6. Chemical structures of 34 active compounds in the inhibition of p56 Lck SH2 domain association with phosphotyrosine-containing C-terminal ITAM2 peptide.

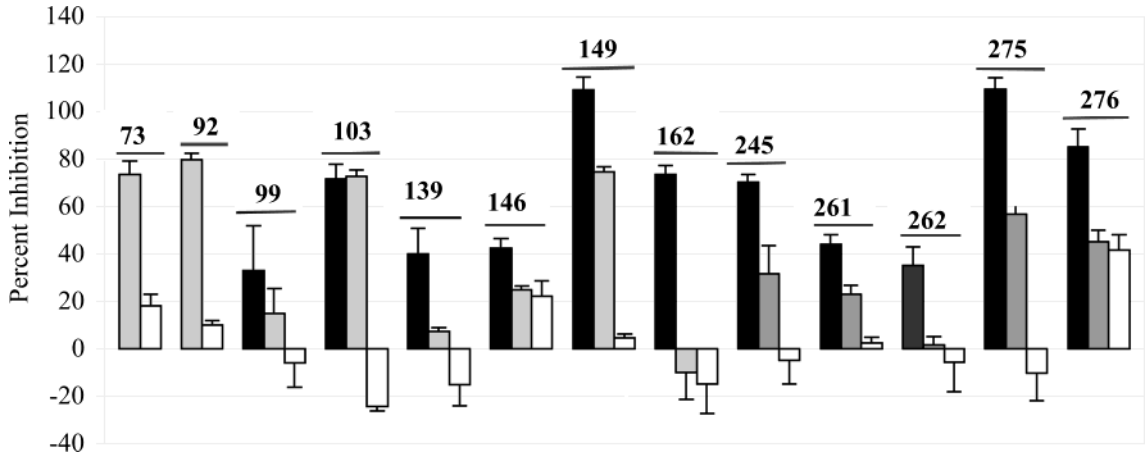

Figure 7. Percent inhibition of [^{3}H]thymidine uptake in mixed lymphocyte culture by 13 compounds selected by in vitro screening: (black bars) 100 μM; (gray bars) 10 μM; (white bars) 1 μM. Compounds **73** and **92** were not tested at 100 μM because of solubility issues. Each bar represents the mean ± standard deviation of three replicates.

in the present study facilitate the selection of compounds that would be missed if selection was based only on total interaction energy.

Inhibitory Activity in Mixed Lymphocyte Culture Assay. Although encouraging results were achieved in the in vitro biological testing, it is desirable to test if the identified compounds possess activity in in vivo cellular functional assays. Mixed lymphocyte culture assays were performed where lymphocytes from two different strains of mice with different histocompatibility antigens were mixed. Because of the difference in the histocompatibility antigens, resting T cells from both strains of mice will undergo blast transformation and propagate. The association of Lck to the ITAM residues of CD3 is essential for this process. Therefore, the modulation of the association of Lck to the CD3 ITAM can be quantified as downstream modulation of the levels of ^{3}H-TdR incorporation into DNA. Because of availability issues, 24 out of the 34 identified compounds were tested in vivo, with 13 compounds showing inhibitory activity at a 100 μM concentration (Figure 7). Thus, over 50% of the compounds identified by the in vitro assay actually showed activity in the cellular functional assay. These compounds represent the most promising leads for further development. For seven compounds, biphasic activity was observed in the mixed lymphocyte

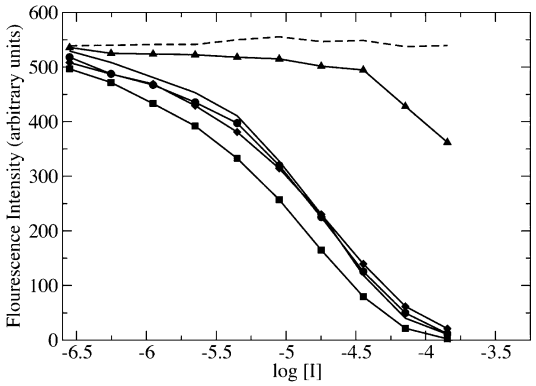

Figure 8. Fluorescence titrations of the Lck SH2 domain with selected compounds. The log [I] denotes the log of the concentration, *M*, of the compounds. Data are included for compounds **73** (line), **92** (circles), **103** (squares), **276** (diamonds), **102** (triangles), and DMSO alone (dashed). See Figure 1 of Supporting Information for the chemical structure of **102**.

culture assay, where an inhibitory activity was observed at the higher concentration (100 μM) and an activation occurred at lower concentrations (1 μM). Such effects may be due to the autoregulatory mechanism of p56 Lck.

The functional domains of p56 Lck include SH3, SH2, and the kinase domain oriented from the N to C termini. In the inactive state, the C terminal tyrosine (Y505) residue is phosphorylated. This pY505 binds to the SH2 domain, thereby folding the molecule and creating an inactive "closed" conformation. The intramolecular interaction between the C tail and the SH2 domain responsible for inactivation of Lck kinase is quite weak compared with the binding of the SH2 domain to optimal high-affinity ITAMs.[44−46] High-affinity inhibitors (e.g., **73** and **92**) may block the intermolecular interaction at all tested concentrations, leading to inhibition in the mixed lymphocyte culture assay. Biphasic compounds (e.g., **99** and **139**) at lower concentrations may only block the intramolecular interaction, thereby facilitating the unfolding of p56 Lck and leading to the augmentation of activation in the mixed lymphocyte culture assay, while at higher concentrations, they may block the intermolecular interaction, leading to inhibition in the mixed lymphocyte culture assay. While such biphasic activities are of interest to understand the regulation of p56 Lck and other kinases, the emphasis of the proposed study is the identification of inhibitors.

Verification of Lck SH2 Domain Binding via Fluorescence Titrations. To further validate that the active compounds identified via competition for the phosphorylated ITAM2 peptide were binding directly to the SH2 domain, fluorescence titration experiments were undertaken. These experiments take advantage of the presence of a tryptophan in the protein. By use of an excitation wavelength of 270 nm, an emission maximum was obtained at 585 nm; this wavelength was monitored in the titrations. Addition of the compounds to the Lck SH2 domain was shown to quench fluorescence upon binding (Figure 8). Four compounds, **73**, **92**, **103**, and **276**, shown to inhibit ITAM2 binding effectively quench the protein fluorescence. On the basis of double reciprocal plots, the respective dissociation constants K_D were 12, 6, 4, and 6 μM. An additional compound, **102**, was included as a negative control (see Figure 1 in Supporting Information) because it did not inhibit ITAM binding at 100 μM. Consistent with that result, quenching of fluorescence was not seen until a concentration of 70 μM. Thus, fluorescence titration experiments confirm that the identified inhibitors do bind to the Lck SH2 domain.

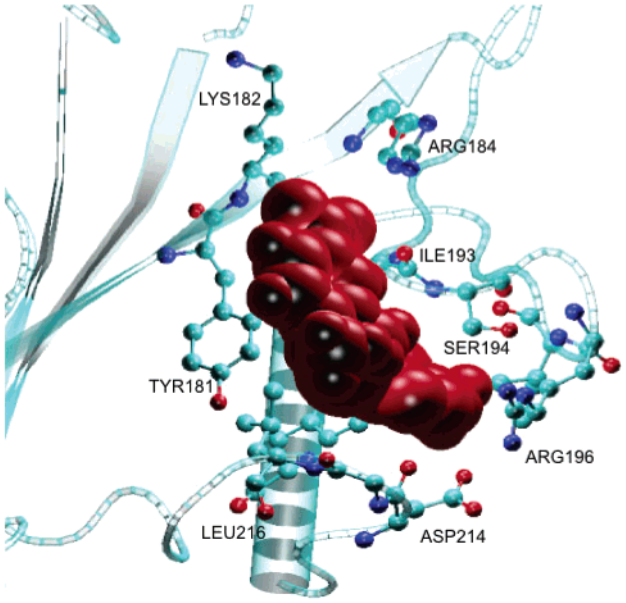

Figure 9. Predicted binding mode of compound **103**. Lck SH2 domain (gray cartoon model), amino acid residues in pY + 3 binding site (stick and ball model), and docked compound (red space-fill model). The BG loop contains residues ASP214−LEU216, and the EF loop contains residues ILE193−ARG196. The image was generated using the program VMD.[47]

Predicted Binding Mode between Active Compounds and Lck SH2 Domain. The biological testing combined with the applied screening approach strongly support the hypothesis that the inhibitors exert their biological activity via binding to the pY + 3 binding site of the Lck SH2 domain. The pY + 3 binding site for the Ile residue is primarily formed by two loops (EF and BG), and amino acid residues on EF and BG loops are important for differential specificity.[1,7] As shown in Figure 9, the predicted interaction between compound **103** and the Lck SH2 domain shows binding to occur in the pY + 3 binding site, with multiple binding contacts with both the EF (ILE193-ARG196) and BG (ASP214-LEU216) loops. It should be noted that the presence of three positively charged residues (LYS182, ARG189, and ARG196) leads to several of the identified compounds (Figure 6) having negatively charged moieties. Alternatively, the possibility that some of the inhibitors are interacting with the pY binding pocket cannot be excluded on the basis of the current data. Structural studies will be required to confirm the proposed binding orientation.

Conclusion

Inhibitors of Lck−protein interactions that block Lck SH2 domain-dependent interactions may ultimately find therapeutic utility as immunosuppressant agents and in the treatment of T-cell-based leukemia, lymphomas, and autoimmune diseases. Accordingly, a combined computer screening and experimental study was undertaken to identify low molecular weight nonpeptidic

compounds that inhibit interaction of p56 Lck with its substrate proteins. The screening strategy targeted the well-defined (pY + 3) binding site known to impart affinity and specificity on Lck−protein interactions. This unique strategy was developed to overcome current limitations in the development of novel Lck inhibitors based on phosphopeptide mimetics.

During this study, a procedure for 3D database conversion was developed and used to generate an in-house 3D database system containing over 2 million commercially available compounds. Building upon current screening approaches, modified scoring approaches have been incorporated to facilitate screening of a large database containing millions of compounds.[24] An important aspect of this is the ability to bias selection toward lower MW compounds, thereby identifying compounds more appropriate for lead optimization. Importantly, ligand-based postdocking structural clustering leads to the selection of diverse compounds, many of which would have been missed via selection based solely on the total interaction energy.

Out of 196 computationally selected compounds, 34 were shown to inhibit the p56 Lck SH2 domain association with ITAM2 phosphotyrosine peptide. Twenty-four of the 34 compounds were further tested for their ability to modulate biological function, with 13 of those compounds having inhibitory activity in the mixed lymphocyte culture assay. Four of those compounds were then subjected to fluorescence titration experiments, validating the Lck SH2 domain as their target. Overall, these data strongly support the hypothesis that computer-based database searching is an effective tool to identify small molecular weight compounds that bind to the SH2 domain of p56 Lck, thereby leading to inhibition of T cell activation. The identified compounds are anticipated to act as lead compounds for the development of novel immunosuppressant drugs.

Acknowledgment. The authors thank Drs. Yongping Pan and I-Jen Chen for helpful discussions and Dr. Gary Strahan for computer support and acknowledge financial support from the NIH (Grant GM CA95200), the Computer-Aided Drug Design Center, School of Pharmacy, University of Maryland, Baltimore, and a Dunning fellowship to N.H. from the Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore.

Appendix

Abbreviations. 3D, three-dimensional; MW, molecular weight; Lck, lymphoid T cell tyrosine kinase; SH2, src *Homology*-2; pY, phosphotyrosine; ITAMs, immunoreceptor tyrosine based activation motifs; TCR, T cell receptor.

Supporting Information Available: Figure of chemical structure of **102**. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tong, L.; Warren, T. C.; King, J.; Betageri, R.; Rose, J.; et al. Crystal structures of the human p56lck SH2 domain in complex with two short phosphotyrosyl peptides at 1.0 Å and 1.8 Å resolution. *J. Mol. Biol.* **1996**, *256*, 601−610.
(2) Broadbridge, R. J.; Sharma, R. P. The Src homology-2 domains (SH2 domains) of the protein tyrosine kinase p56lck: structure, mechanism and drug design. *Curr. Drug Targets* **2000**, *1*, 365−386.
(3) Straus, D. B.; Weiss, A. Genetic evidence for the involvement of the lck tyrosine kinase in signal transduction through the T cell antigen receptor. *Cell* **1992**, *70*, 585−593.
(4) Weiss, A.; Littman, D. R. Signal transduction by lymphocyte antigen receptors. *Cell* **1994**, *76*, 263−274.
(5) Straus, D. B.; Chan, A. C.; Patai, B.; Weiss, A. SH2 domain function is essential for the role of the Lck tyrosine kinase in T cell receptor signal transduction. *J. Biol. Chem.* **1996**, *271*, 9976−9981.
(6) Pawson, T.; Gish, G. D. SH2 and SH3 domains: from structure to function. *Cell* **1992**, *71*, 359−362.
(7) Cody, W. L.; Lin, Z.-W.; Panek, R. L.; Rose, D. W.; Rubin, J. R. Progress in the development of inhibitors of SH2 domains. *Curr. Pharm. Des.* **2000**, *6*, 59−98.
(8) Cousins-Wasti, R.; Ingraham, R. H.; Morelock, M. M.; Grygon, C. A. Determination of affinities for lck SH2 binding peptides using a sensitive fluorescence assay: comparison between the pYEEIP and pYQPQP consensus sequences reveals context-dependent binding specificity. *Biochemistry* **1996**, *35*, 16746−16752.
(9) Sawyer, T. K. Src homology-2 domains: structure, mechanisms, and drug discovery. *Biopolymers* **1998**, *47*, 243−261.
(10) Beaulieu, P. I.; Cameron, D. R.; Ferland, J. M.; Gauthier, J.; Ghiro, E.; et al. Ligands for the tyrosine kinase p56lck SH2 domain: discovery of potent dipeptide derivatives with monocharged, nonhydrolyzable phosphate replacements. *J. Med. Chem.* **1999**, *42*, 1757−1766.
(11) Lee, T. R.; Lawrence, D. S. SH2-directed ligands of the Lck tyrosine kinase. *J. Med. Chem.* **2000**, *43*, 1173−1179.
(12) Kuriyan, J.; Cowburn, D. Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 259−288.
(13) Songyang, Z.; Cantley, L. C. Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends Biochem. Sci.* **1995**, *20*, 470−475.
(14) Marengere, L.; Songyang, Z.; Gish, G. D.; Schaller, M. D.; Parsons, J. T.; et al. SH2 domain specificity and activity modified by a single residue. *Nature* **1994**, *369*, 502−505.
(15) Songyang, Z.; Gish, G.; Mbamalu, G.; Pawson, T.; Cantley, L. C. A single point mutation switches the specificity of group III Src homology (SH2) domains to that of group I SH2 domains. *J. Biol. Chem.* **1995**, *270*, 26029−26032.
(16) Brennan, M. B. Drug Discovery Filtering Out Failures Early in the Game. *Chem. Eng. News* **2000**, *78*, 63−73.
(17) Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L. D.; et al. A novel approach for the virtual screening and rational design of anticancer compounds. *J. Med. Chem.* **2000**, *43*, 1975−1985.
(18) Baxter, C. A.; Murray, C. W.; Waszkowycz, B.; Li, J.; Sykes, R. A.; et al. New approach to molecular docking and its application to virtual screening of chemical databases. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 254−262.
(19) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening—an overview. *Drug Discovery Today* **1998**, *3*, 160−178.
(20) Massova, I.; Martin, P.; Bulychev, A.; Kocz, R.; Doyle, M.; et al. Templates for design of inhibitors for serine proteases: application of the program DOCK to the discovery of novel inhibitors for thrombin. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2463−2466.
(21) Li, S.; Gao, J.; Satoh, T.; Friedman, T. M.; Edling, A. E.; et al. A computer screening approach to immunoglobulin superfamily structures and interactions: discovery of small non-peptidic CD4 inhibitors as novel immunotherapeutics. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 73−78.
(22) Chen, I.-J.; Neamati, N.; Nicklaus, M. C.; Orr, A.; Anderson, L.; et al. Identification of HIV-1 integrase inhibitors via three-dimensional database searching using ASV and HIV-1 integrases as targets. *Bioorg. Med. Chem.* **2000**, *8*, 2385−2398.
(23) Deborah, A. L.; Willian, V. M.; Linda, K. J. Application of virtual screening tools to a protein−protein interaction: database mining studies on the growth hormone. *Med. Chem. Res.* **1999**, *9*, 579−591.
(24) 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.
(25) *SYBYL*, version 6.7; Tripos Associates: St. Louis, MO, 2002.
(26) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537.
(27) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; et al. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(28) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505−524.
(29) Leach, A. R.; Kuntz, I. D. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comput. Chem.* **1992**, *13*, 730−748.

(30) Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221*, 709−713.
(31) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. The MIDAS display system. *J. Mol. Graphics* **1988**, *6*, 13−27.
(32) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1984**, *28*, 849−857.
(33) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms used in automated molecular docking. *J. Comput. Chem.* **1997**, *18*, 1175−1189.
(34) Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 251−264.
(35) 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.
(36) Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. I. The design of leadlike combinatorial libraries. *Angew. Chem., Int. Ed.* **1999**, *38*, 3743−3748.
(37) Godden, J. W.; Xue, L.; Bajorath, J. Combinatorial preferences affect molecular similarity/diversity calculations using binary fingerprints and Tanimoto coefficients. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 163−166.
(38) *MOE*; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2002.
(39) Marshall, A. G. *Biophysical Chemistry: Principles, Techniques, and Applications*; John Wiley & Sons: New York, 1978.
(40) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery development settings. *Adv. Drug Delivery Res.* **1997**, *23*, 3-25.
(41) Su, A.-l.; Lorber, D. M.; Weston, G. S.; Baase, W. A.; Matthew, B. W.; et al. Docking molecules by families to increase the diversity of hits in database screening: computational strategy and experimental evaluation. *Proteins* **2001**, *42*, 279−293.
(42) Kubinyi, H. Structure-based design of enzyme inhibitors and receptor ligands. *Curr. Opin. Drug Discovery Dev.* **1998**, *1*, 4−15.
(43) Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. Ligand solvation in molecular docking. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 4−16.
(44) Payne, G.; Stolz, L. A.; Pei, D.; Band, H.; Shoelson, S. E.; et al. The phosphopeptide-binding specificity of Src family SH2 domains. *Chem. Biol.* **1994**, *1*, 99−105.
(45) Songyang, Z.; Shoelson, S. E.; Chaudhuri, M.; Gish, G.; Pawson, T.; et al. SH2 domains recognize specific phosphopeptide sequences. *Cell* **1993**, *72*, 767−778.
(46) Liu, X.; Brodeur, S. R.; Gish, G.; Songyang, Z.; Cantley, L. C.; et al. Regulation of c-Src tyrosine kinase activity by the Src SH2 domain. *Oncogene* **1993**, *8*, 1119−1126.
(47) Humphrey, W.; Dalke, A.; Schulten, K. VMD−visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

JM030470E