Lead Validation and SAR Development via Chemical Similarity Searching; Application to Compounds Targeting the pY+3 Site of the SH2 Domain of p56^{lck}

Alba T. Macias, Md. Younus Mia, Guanjun Xia, Jun Hayashi, and Alexander D. MacKerell, Jr.*

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

Received June 4, 2005

Compound selection based on chemical similarity has been used to validate active "parent" compounds identified via database searching as viable lead compounds and to obtain initial structure−activity relationships for those leads. Twelve parent compounds that have inhibitory activity against the SH2 domain of the p56 T-cell tyrosine kinase (Lck) are the focus of this study. Lck is involved in the T-cell mediated immune response, and inhibitors of Lck protein−protein interactions could potentially be used to develop novel immunosuppressants. Similarity searches for each parent compound were performed using 2D structural fingerprints on a database containing 1 300 000 commercially available compounds. The inhibitory activity of the selected compounds was assessed using enzyme immunoassay (EIA). In general, the most active parent compounds yield the most high activity similar compounds; however, in two cases low activity parent compounds (i.e. inhibitory activity < 25% at 100 μM) yielded multiple similar compounds with activities > 60%. Such compounds may, therefore, be considered as viable lead compounds for optimization. Structure−activity relationships were explored by examining both ligand structures and their computed bound conformations to the protein. Functional groups common to the active compounds as well as key amino acid residues that form hydrogen bonds with the active compounds were identified. This information will act as the basis for the rational optimization of the lead compounds.

INTRODUCTION

Chemical similarity searching is a methodology used in the virtual drug discovery process based on the idea that structurally similar compounds often have similar biological activity.[1−5] A common application of similarity searches is the identification of compounds similar to a known active compound when no other active compounds are known or when there is no available structural data of the receptor. A related application involves performing nearest-neighbor searches on high-throughput screening hits.[6] Another common goal of similarity searching is to increase the hit rate by identifying a diverse list of compounds to be experimentally tested.[7−9] Diversity based on dissimilarity increases the probability of identifying unique leads[9] and reduces the economic costs by eliminating the redundant testing of chemically similar compounds.

2D structural molecular fingerprints that contain information about molecular structure or properties are one method used to describe chemical similarity. Similarity can also be based on the 2D molecular graphs, 3D conformation, pharmacophore points, or a combination of physical and biological descriptors.[7,10−12] The performance of simple 2D fingerprints based on fragment bit-string data is comparable to fingerprints based on 3D pharmacophore points.[5,13−15] In the 2D structural molecular fingerprints the structural information is encoded into linear bit strings of data that enables the determination of molecular similarity based on metrics such as the Tanimoto coefficient (Tc). Tc provides a similarity 'score' by dividing the fraction of features common to both molecules by the total number of features.[16]

Tc based quantification of chemical similarity has previously been tested as a method to identify active compounds. Studies have shown that compounds with a Tc ≥ 0.85, based on Unity fingerprints,[17] had an 80% chance of being active in similar assays.[18] However, a recent study that evaluated the relationship between compound similarity and IC_{50} values found that there was only a 30% chance that two compounds with Tc ≥ 0.85 would have the same activity. This was attributed to methodological shortcomings as well as the fact that very small structural differences can affect the activity. Thus, further validation of the use of chemical similarity as a means to identify biological active chemical analogues of lead compounds is required.

We have previously identified inhibitors of the Lck SH2 domain by targeting the pY+3 site using virtual screening.[9] Lck is a tyrosine kinase that plays a critical role in T-cell-mediated immune response. A 3D database of 2 million compounds was screened using DOCK[19,20] in a two-step procedure following a protocol developed in our lab.[9,21,22] This protocol is initially used to identify compounds that are structurally complementary to the binding site on the target protein. Molecular fingerprints are then used to cluster the selected compounds into chemically dissimilar subsets, from which compounds are selected for experimental testing by choosing one to two compounds from each cluster, using Lipinski's Rule of 5 as a guide.[23] In the Lck study,[9] a total of 196 compounds were obtained for biological assay. In the initial assay using immunoblots, 34 compounds inhibited SH2 domain association with phosphorylated immunorecep-

* Corresponding author e-mail: amackere@rx.umaryland.edu.

10.1021/ci050225z CCC: $30.25
Published on Web 09/30/2005

tor tyrosine based activation motif (ITAM) peptide. Thirteen compounds out of these 34 showed inhibitory activity in mixed lymphocyte reaction (MLR) assay. Thus, database searching combined with 2D molecular fingerprints successfully identified a series of chemically dissimilar compounds with the desired biological activity.

In the present study we extend our previous work via the application of chemical similarity clustering to identify compounds similar to the active "parent" compounds. This effort was motivated by the desire to validate the relevance of the parent compounds as true lead compounds and to obtain initial structure−activity relationships (i.e.: pharmacophore information) for those leads. Twelve out of the 13 compounds, which showed inhibitory activity in both the competition (immunoblot) and mixed lymphocyte reaction assays were chosen for the present work; due to availability issues one of the original 13 compounds was not included in the present study. Compounds identified via chemical similarity were experimentally tested using a high-throughput EIA assay developed to rapidly quantify inhibition.

METHODS

The program MOE[24] was used to calculate the molecular fingerprints and to calculate Tc values with respect to the lead compounds. Structural keys based on MACC_BITS fingerprints were used for all clustering calculation, and the extent of similarity between compounds was quantified using the Tanimoto coefficient. Similarity searching was performed on a 2D database containing 1 300 000 commercially available compounds developed in our laboratory. Selected compounds were purchased from the respective vendors (Chembridge, Chemdiv, Specs) and used without further testing.

Biological activities were measured using a high-throughput EIA assay carried out using 96 well medium binding EIA plates (Costar). Wells were coated with 100 μL of human CD3 ζ chain ITAM 2 phosphopeptide conjugated to BSA (~10 pmole peptide equivalent) in PBS overnight at 4 °C and blocked with 300 μL of PBS containing 5% (wt/vol) powdered skim milk for 1 h at 37 °C. After washing with PBS containing 1% Tween 20 (PBST), 3 times, 100 μL of precalibrated bacterial lysate containing recombinant GST Lck SH2 domain fusion protein was added in the presence or absence of test compounds and incubated for 1 h at room temperature. After 3 extensive washings with PBST, 100 μL of PBS containing HRP-conjugated rabbit anti-GST antibody was added and incubated for 1 h. After extensive washing with PBST, 100 μL of TMB substrate was added, and absorbance at 620 nm was measured using a multiwell EIA plate reader (Anthos HTIII). The percent inhibition (% inhibition) was calculated based on the optical density (OD) using the formula: % inhibition = 100 − (ΔOD of test well /ΔOD of positive wells) × 100, where ΔOD was calculated by subtracting background OD (average OD of negative wells) from the test as well as positive control wells. Error analysis was performed on two or more measurements.

DOCK 4.0.1[19,20] was used to obtain bound conformations for the similar compounds for which experimental data was obtained. Each compound was docked to the same pY+3 protein binding site as used in our previous study.[9] As previously, 4-methylphenyl phosphate was inserted into the pY binding pocket to avoid sampling of this region during docking; however, docking tests in the absence of the 4-methylphenol phosphate showed the pY site to not be occupied by any inhibitors (not shown). During the docking procedure, each compound was divided into nonoverlapping rigid segments connected by rotatable bonds. The initial fragment from which the ligand was built had at least 5 heavy atoms, and each was docked into the binding site in 500 orientations and minimized. The 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 15°, and the lowest energy conformation was selected. All dihedrals were minimized at each step. Pruning of the conformational orientations ensured conformational diversity and more favorable energies. Energy scoring was performed with a distant-dependent dielectric, with a dielectric constant of 4 and using an all atom model. The conformation of each molecule with the most favorable interaction energy was selected and saved. Interatomic distance analysis was done using CHARMM[25] and hydrogen bonds were identified using FindHBond within Chimera[26] with a tolerance of 0.4 Å and 20° from the H-bond values in Mills and Dean.[27]

Table 1. Tanimoto Coefficient Values for the Set of Original Compounds[a]

	73	**92**	**276**	**245**	**146**	**162**	**275**	**149**	**139**	**262**	**99**	**103**
73	100											
92	**42**	100										
276	**67**	38	100									
245	57	33	**56**	100								
146	43	29	40	42	100							
162	37	27	46	**50**	26	100						
275	**54**	45	49	36	30	39	100					
149	**51**	40	**62**	57	40	34	30	100				
139	45	41	**58**	**51**	36	43	34	**67**	100			
262	34	33	35	**37**	48	34	35	32	24	100		
99	44	37	**50**	43	31	**69**	48	34	41	31	100	
103	41	49	44	45	38	24	28	**57**	42	31	28	100

[a] Compounds are in the order of their decreasing inhibitory activity.

RESULTS AND DISCUSSION

In our previous study the compounds selected for biological assay had been chosen from different structural clusters to obtain chemical diversity,[9] but the individual degree of dissimilarity among them had not been quantified. Accordingly, the similarity among the 12 selected compounds was calculated and is reported in Table 1. None of the Tc values for each compound pair is greater than 0.70 confirming that they have a degree of dissimilarity according to the MAC_BIT fingerprints. Examination of compound pairs with Tc values > 0.60 reveals four molecular pairs with high similarity (**73** and **276**, **276** and **149**, **162** and **99**, **149** and **139)**.

Both compounds **73** and **276** (Figure 1) contain the same central motif, a furan and an amide cyclic ring, which are connected to a substituted aromatic ring, agreeing well with chemical intuition of similarity. Compounds **99** and **162** both contain an aromatic ring attached to an amide, a methyl, and a carboxylic acid, although the connectivity of these groups

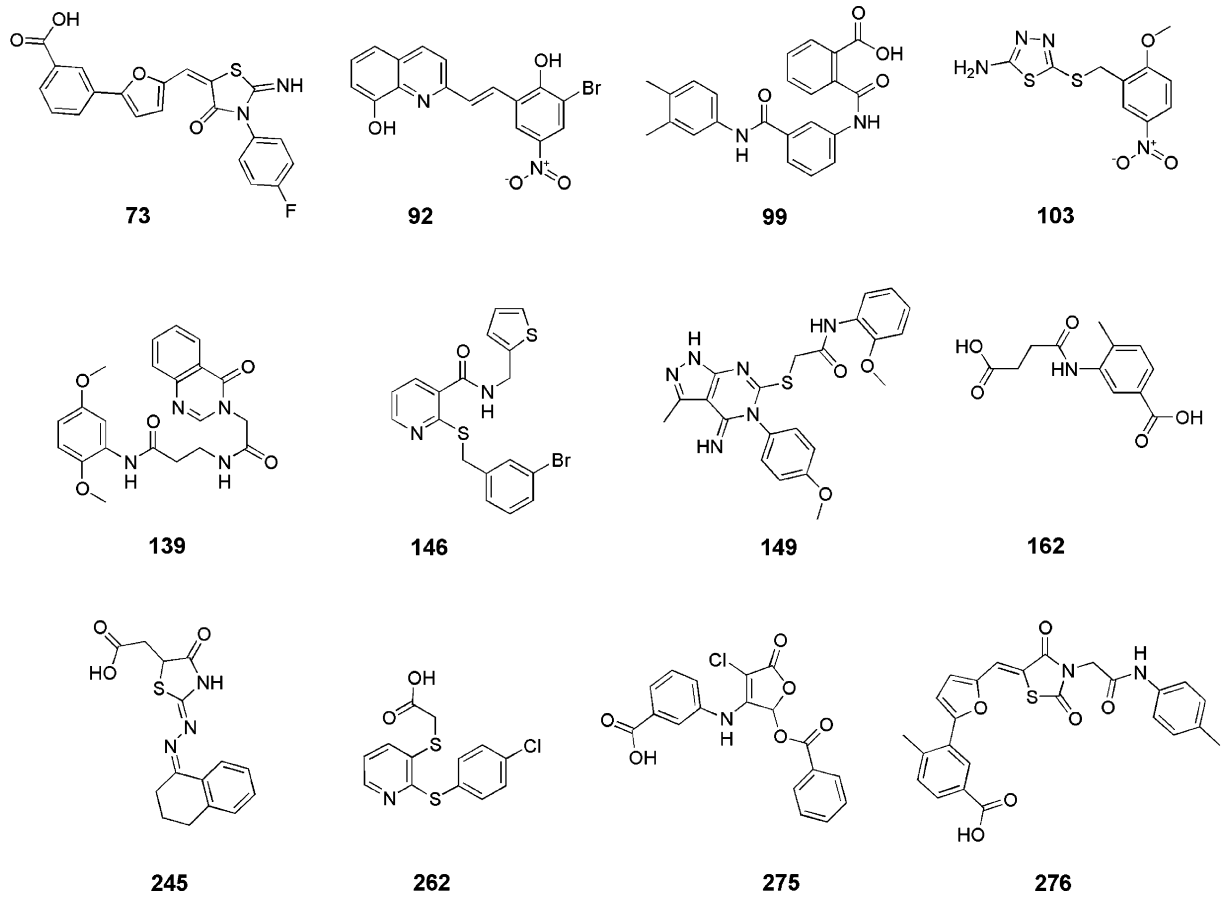

Figure 1. Original 12 compounds previously identified as possible inhibitors of the LCK SH2 domain.

is different revealing a limitation of similarity searches based on fingerprints. Compounds **149** and **139** also contain similar functional groups but different connectivity among them. Clustering the compounds into subsets using a lower similarity and overlap coefficients of 0.50 places compounds **73**, **276**, **245**, **149**, and **139** into the same subgroup, compounds **99** and **162** cluster into a second subgroup, and the rest of the compounds into subgroups containing only one compound.

Validation of Active Parent Compounds as Lead Compounds. Beyond the appropriate experimental verification of individual compounds, it is important to determine if an active parent compound is a member of a structural class that should be targeted for lead optimization. A simple way to do this without synthesizing new compounds is to search a chemical database of commercially available compounds for structurally similar entities. These compounds may then be purchased and subjected to biological assay. If a significant number of the new compounds identified as being similar to a given parent compound are active, it may be assumed that the parent compound will act as a good lead compound. Comparisons of the active compounds identified in the assay may also provide structural features that are important for their biological effect.

Similarity searching was performed individually on the 12 parent compounds shown in Figure 1. For each compound the Tc value was adjusted until approximately 30 hits were obtained. Since the availability of selected compounds through commercial sources is approximately 60−70% based on our experience, this would allow us to test approximately 20 similar compounds per original parent subgroup. Table 2 contains the Tc values used in the clustering for each compound, the number of similar compounds identified, and the number of compounds experimentally tested. In most cases it was possible to obtain approximately 20 compounds for each subgroup from commercial sources. In the case of **262**, only four of the requested compounds were available from the vendors. All the compounds obtained were subjected to biological assay.

Table 2. Tanimoto Coefficients Used during the Similarity Search and Number of Compounds Identified and Experimentally Tested in Enzyme Immunoassays

compd	Tc	no. identified	no. tested
73	87	33	24
92	84	20	14
99	93	37	23
103	85	20	18
139	90	26	23
146	80	22	13
149	94	30	29
162	91	30	17
245	84	26	21
262	80	22	4
275	89	23	14
276	90	29	20

Results from the solid-phase EIA inhibition assays for the original and similar compounds are presented in Figure 2 for compounds **73**, **92**, **99**, and **276**, and the results for all the compounds are summarized in Table 3. Activity plots for the remaining compounds are included in the Supporting Information. As shown, a significant number of the similar compounds had good activities for some of the parent compounds, while for others (i.e. **139**, **149**, **262**) none of

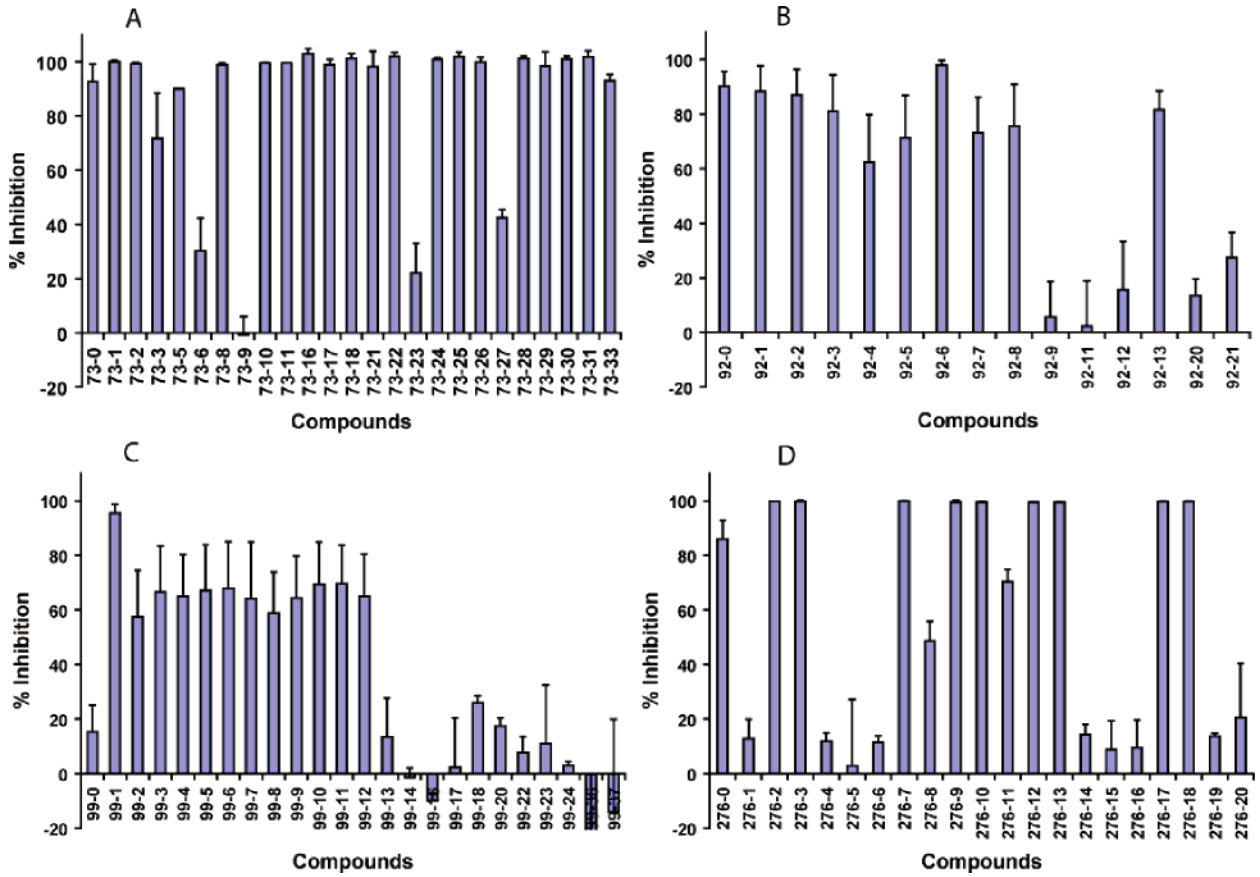

Figure 2. Experimental inhibition values for four sets of compounds similar to **73** (A), **92** (B), **99** (C), and **276** (D). The results are expressed as mean ± standard deviation inhibition of at least two experiments.

Table 3. Experimental Inhibition Values, Inh^E, for the Original Compounds and Summary of the Inhibitory Activities of the Similar Compounds

compd	original Inh^E of parent compd[a]	% of similar compds with $Inh^E > 20$	% of similar compds with $Inh^E > 40\%$	% of similar compds with $Inh^E > 60\%$	% of similar compds with $Inh^E > 80\%$	av Inh^E of similar compds ± SD
73	**95**	96	88	83	79	86 ± 30
92	**90**	71	64	64	36	56 ± 35
276	**86**	60	55	50	45	56 ± 43
245	**77**	33	19	14	5	23 ± 26
146	**67**	62	15	8	0	25 ± 22
162	**58**	24	11	0	0	19 ± 13
275	**60**	14	7	0	0	9 ± 17
149	**58**	17	0	0	0	12 ± 10
139	**55**	0	0	0	0	9 ± 6
262	**54**	0	0	0	0	11 ± 7
99	**15**	57	52	43	4	37 ± 35
103	**−2**	44	28	28	0	26 ± 29

[a] Compounds are in the order of their decreasing inhibitory activity.

the similar compounds were active at >40% inhibition. Notably, most of the better inhibitors were compounds similar to parent compounds which had inhibition levels > 80% (**73**, **92**, and **276**). Typically, parent compounds with activities of 60% or less had a relatively small number of active similar compounds.

Two interesting exceptions are **99** and **103**. Compound **99** only has 15% inhibitory activity, while a large number of similar compounds derived from **99** showed inhibitory activity over 60% (with one over 80%, Figure 2 and Table 3). Compound **103** was not active in the present study based on the solid-phase EIA assay; however, it did show inhibitory activity in the MLR assay as previously reported.[9] Several of the similar compounds derived from **103** had significant activity (>60% inhibition). Overall, the present results indicate that the most active parent compounds typically yielded the largest percentage of active similar compounds. However, the results also indicated the importance of following up on the parent compounds with lower activity since some of the similar compounds derived from them displayed significant inhibitory activity.

The observation that the similar compounds derived from less active parent compounds have higher inhibitory activity is not unexpected. Since compounds for biological testing from the initial screen are selected to maximize the chemical dissimilarity, it is likely that the compound actually selected

may not be optimal with respect to biological activity. Accordingly, when compounds similar to such parent compounds are obtained and tested, it is possible that more active compounds will be identified. Indeed, compounds with higher activity were obtained from the similarity screens of **99**, **103**, **146**, and **245**.

As discussed in the Introduction, the 85−80 rule states that compounds with Tc > 0.85 have an 80% chance of having similar activity to the lead compound.[18] Compounds similar to **73**, **99**, **139**, **149**, **162**, **275**, and **276** were chosen with Tc values greater than 0.85 (Table 2). Accordingly, the respective similar compounds should have activity similar to that of the lead compound. To assess this, averages of the experimental inhibition (Inh^E) for the similar compounds were calculated. As seen in Table 3, the 85−80 rule is not strictly followed, although the trend holds to a larger extent for the more active parent compounds. This observation is in agreement with Martin's recent study[5] and holds even though the fingerprints used in the two studies were different.

Structure−Activity Relationships. Beyond validation of the parent compounds as leads, similarity searching offers the potential to develop preliminary structure−activity relationships (SAR)/pharmacophores for a lead that may be used to facilitate the lead optimization process.[28,29] Such a SAR may be ligand based where only the structures of the ligands themselves are considered. Alternatively, a target-based pharmacophore can be developed via, for example, docking studies of all the similar compounds from which functional groups of importance on both the compounds and the target molecule can be identified. In the remainder of this section we apply both approaches; **276** is used for both the ligand- and target-based approaches as the range of activities of the similar compounds varied widely. Situations where the variability of the activities is not large, such as with **73**, **139**, or **149**, were not considered as the limited range is not conducive to SAR development. In cases such as **73**, where the majority of compounds are highly active, extending the similarity search to contain less similar compounds via use of a lower Tc value is recommended.

The parent compound (i.e. **276-0**) along with the similar compounds for **276** are shown in Figure 3. These compounds have been separated into those with >60% activity (Figure 3A) and those with less that 60% inhibitory activity (Figure 3B). Common to all the compounds is an amide linkage attached to a central five-membered heterocycle and to an aromatic ring. The amide's carbonyl group is attached to the five-membered ring's nitrogen, and the amide's nitrogen is attached to the aromatic ring. The highly active compounds all contain a furan ring linked via a double bond to the heterocycle, which is then linked to an aromatic ring that contains an acid group in the meta or para position. The only exception is **276-11**, which has a phenol moiety with the hydroxyl in the ortho position rather than the furan ring; however, this compound has the lowest activity among the highly active compounds (Figure 2). In the lower activity compounds (Figure 3B) the furan ring is omitted, with the exception of **276-5** and **276-8**. Many of the low activity compounds contain benzoic acid moieties, though in most cases the furan is omitted or exchanged with the pyrrole ring, which lacks the hydrogen bond acceptor of the furan. In addition, **276-5** and **276-8** contain ester moieties versus the acid on the terminal phenyl ring, which may also contribute to the decreased activity. The methyl group could be causing steric hindrance, or the lack of the negative charge could be affecting the binding. Interesting are **276**-(**6**, **14**, and **15**), which all contain a phenol group and lack the furan moiety, as in **276-11**. In these lower activity compounds the hydroxyl is meta or para versus ortho in the more active compound. This motif suggests that the hydroxyl in **276-11** may act as an acceptor, replacing that in the furan ring in the other more active compounds. Overall, these results indicate that beyond the heterocycle-amide-phenyl ring central core of the **276** series the presence of a furan ring 1,3 linked to benzoic acid moiety facilitates activity, though alternate functional groups with acceptor moieties may be considered to enhance the inhibitory activity. Thus, it is clear that the use of similarity searching in combination with HTS screening can allow for the development of a pharmacophore/SAR, laying the groundwork for rational lead optimization efforts.

Alternatively, the availability of the similar compounds can be used to identify interesting interactions between the inhibitors and the target protein. To identify relevant drug−protein interactions all the active compounds can be docked into the putative binding site of the protein with the resulting structures examined collectively to identify consensus interactions. Such consensus interactions may be assumed to be more representative of the experimental regimen versus the interactions observed for a single docked molecule. The set of compounds similar to **276** was examined comparing the interactions between the set of stronger inhibitors (>60% inhibition) and the set of weaker inhibitors (<60% inhibition) to provide insight into the development of a target-based pharmacophore. Pairwise interactions of 3.0 Å or less between the protein and all ligand atoms were considered in the determination of relevant protein residues. Residues which had at least five of these close interactions with ligand atoms were as follows: Arg134, Lys179, His180, Tyr181, Lys182, Arg184, Ile193, Ser194, Gly215, Leu216, and Cys217 as shown in Figure 4. These residues are in the BG and EF loops and βD strand of the Lck SH2 domain.

Table 4 summarizes the hydrogen bonds formed between the ligands associated with **276** and the protein in the docked conformations. One obvious difference between the strong and weak inhibitors is that strong inhibitors make more hydrogen bonds with the protein, usually through the carboxylic acid. Another difference between the two sets of compounds is that they interact with different protein residues. Residues Lys179, Lys182, and Arg184 make more hydrogen bonds with the strong inhibitors, while residues Arg134 and Arg184 hydrogen bond to the weak inhibitors, revealing different binding modes. The acid can interact with the same residue, Arg134 for some weak inhibitors, or with two different residues, Arg184 and Lys182 for some strong inhibitors. Figure 5A,B shows example binding modes of a strong and weak inhibitor, respectively. The predicted binding conformation of **276-13** illustrates an orientation common among several of the stronger inhibitors in which the compounds interact closely with Arg184 and Lys182. Alternatively, compound **276-8** illustrates the binding orientation common among several of the weak inhibitors that allows close interaction with Arg134.

The fact that docked compounds are not all perfectly superimposed orientations may be due to a variety of

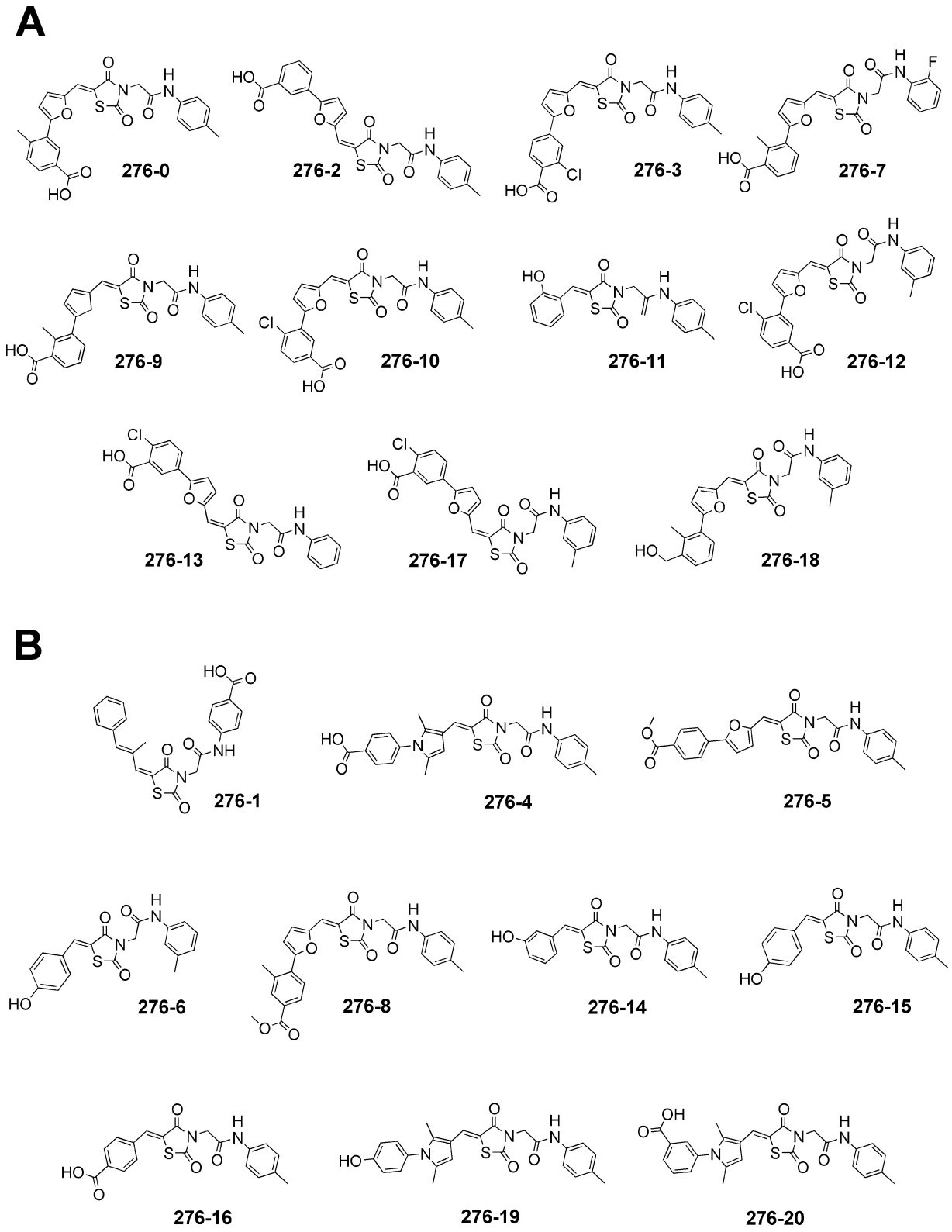

Figure 3. Structures of inactive compounds similar to **276**, with %Inh values greater than 60 (A) and with %Inh values less than 60 (B).

reasons: their chemical structure and size, the lack of a very well defined cavity or groove adjacent to the pY+3 hydrophobic cavity, and the inherent limitations of the docking method. Despite this, some predictions can be made based on frequent occurrences of common binding motifs. The ability of a compound to adopt a favorable binding conformation in which it can hydrogen bond to Lys179, Lys182, and Arg184 seems strongly related to its activity.

CONCLUSIONS

Structural similarity searches were performed for 12 active parent compounds that had previously shown inhibitory activity for Lck in both competitive and mixed lymphocyte culture essays. Based on the similarity search, a total of 220 new compounds were obtained and subjected to a competitive assay for Lck SH2 domain-ITAM2 phosphopeptide binding. Of these compounds 63 had inhibitory levels > 60%

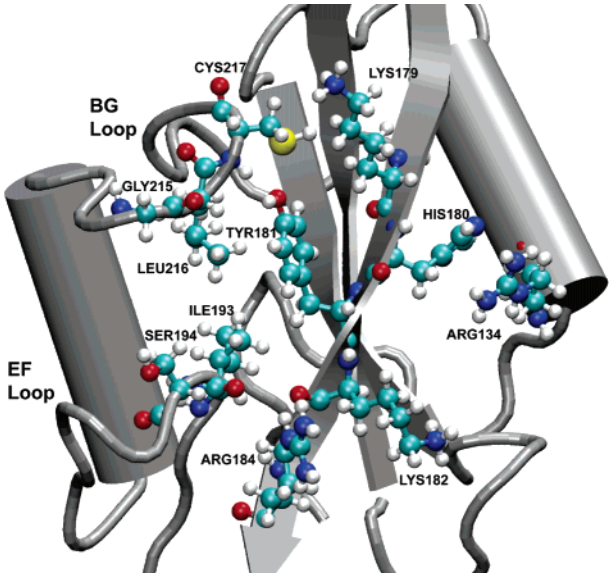

Figure 4. Detailed view of Lck residues which have frequent close contacts (<3 Å) with the predicted docked conformations of **276** and its similar compounds.

Table 4. Hydrogen Bonds between Docked Compounds and Protein Residues

residue	atom	compd	atom type	distance (Å)
		Strong Inhibitors: >60%		
ARG134	NH1	**276-10**	O acid	3.3
ARG134	NH2	**276-10**	O acid	3.4
LYS179	NZ	**276-2**	O amide	3.5
LYS179	NZ	**276-3**	O cyclic ketone	3.2
LYS179	NZ	**276-7**	O amide	3.2
LYS179	NZ	**276-10**	O cyclic ketone	3.2
HIS180	N	**276-10**	O acid	3.3
LYS182	N	**276-2**	O acid	3.4
LYS182	N	**276-7**	O acid	3.1
LYS182	N	**276-9**	O cyclic ketone	3.2
LYS182	N	**276-13**	O acid	3.0
LYS182	N	**276-17**	O acid	3.4
LYS182	N	**276-18**	O acid	3.1
LYS182	N	**276-0**	O acid	3.3
LYS182	NZ	**276-9**	O acid	3.2
LYS182	NZ	**276-12**	O acid	3.1
ARG184	NH2	**276-2**	O acid	3.1
ARG184	NH2	**276-11**	O cyclic ketone	3.5
ARG184	NH2	**276-13**	O acid	3.0
ARG184	NH2	**276-18**	O acid	3.2
ARG184	NH2	**276-0**	O acid	3.1
		Weak Inhibitors: <60%		
ARG134	NH1	**176-1**	O acid	3.4
ARG134	NH1	**176-1**	O acid	3.1
ARG134	NH1	**276-8**	O acid	3.1
ARG134	NH1	**276-16**	O ester ketone	3.5
ARG134	NH2	**176-1**	O acid	3.0
ARG134	NH2	**276-5**	O amide	3.5
ARG134	NH2	**276-16**	O acid	3.6
LYS179	NZ	**276-4**	O cyclic ketone	3.2
LYS179	NZ	**276-20**	O cyclic ketone	3.1
LYS182	N	**276-20**	O acid	3.2
ARG184	NH2	**276-4**	O acid	3.1
ARG184	NH2	**276-8**	O amide	3.0
ARG184	NH2	**276-16**	O amide	3.1
ARG184	NH2	**276-20**	O acid	3.3

at 100 μM, while 38 compounds inhibited at levels > 80%. In most cases, the highly active compounds were similar to the parents that themselves were highly active. However, in certain cases a number of active compounds (i.e. > 60% inhibition) were identified for low activity parent compounds

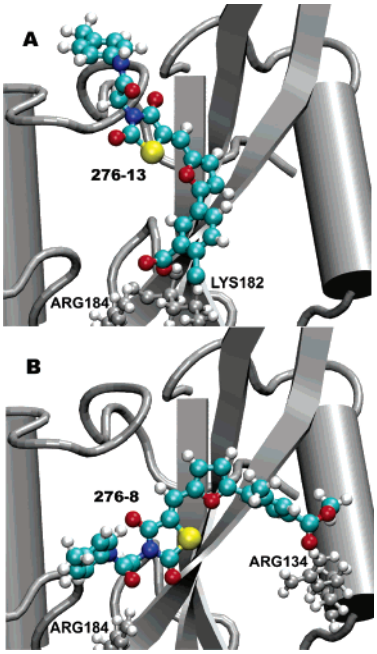

Figure 5. Docked conformations of a strong (A) and a weak (B) inhibitor. Compounds are shown in colored ball-and-stick representation. Lck protein is shown as a cartoon except those residues that form hydrogen bonds to the compounds which are shown in gray ball-and-stick representation.

(e.g. **99** and **103**). Thus, similarity searches can be used to validate parent compounds as lead compounds for optimization studies, including the "rescue" of parents compounds whose activity is relatively low.

The availability of the parent and similar compounds may then be used for the development of SAR/pharmacophore models. With **276** separation of the similar compounds into high and low activity classes followed by visual inspection allowed for the identification of a ligand-based pharmacophore. It should be noted that the availability of the similar compounds with their corresponding biological activity allows for more quantitative ligand-based analysis, such as 3D-QSAR/CoMFA based approaches.[30,31] Within the context of target-based drug design, docking methods may be used to pose all the parent and similar compounds in the putative binding site. While the availability of high-resolution X-ray or NMR data of the ligand−protein complex is the most desirable data, docking studies of multiple similar compounds may identify consensus interactions with the receptor that may be representative of the experimental regimen. Applying this approach with **276** led to the identification of residues Lys179, Lys182, and Arg184 of the Lck SH2 domain as being important for inhibitor−receptor interaction. Clearly, chemical similarity searching as used in the present work offers a cost-effective means to validate and rescue parent compounds identified in a database search, classifying them as viable lead compounds as well as produce data for the development of both ligand-based and target-based SAR models.

ACKNOWLEDGMENT

We would like to acknowledge funding from the NIH (CA095200) and the University of Maryland School of Pharmacy Computer-Aided Drug Design Center.

Supporting Information Available: %Inh^E plots of the similar compound for **103**, **139**, **146**, **149**, **162**, **245**, **262**, and **275**. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Johnson, M. A.; Maggiora, G. M. *Concepts and Applications of Molecular Similarity*; Wiley: New York, 1990.

(2) Willett, P. *Similarity and Clustering Techniques in Chemical Information Systems*; Research Studies Press: Letchworth, 1987.

(3) Dean, P. M. *Molecular Similarity in Drug Design*; Chapman & Hall: Glasgow, 1994.

(4) Van Drie, J. H.; Lajiness, M. S. Approaches to virtual library design. *Drug Discovery Today* **1998**, *3*, (6), 274−283.

(5) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **2002**, *45*(19), 4350−4358.

(6) Shanmugasundaram, V.; Maggiora, G. M.; Lajiness, M. S. Hit-directed nearest-neighbor searching. *J. Med. Chem.* **2005**, *48*(1), 240−248.

(7) Matter, H. Selecting optimally diverse compounds from structure databases: A validation study of two-dimensional and three-dimensional molecular descriptors. *J. Med. Chem.* **1997**, *40*(8), 1219−1229.

(8) Willett, P.; Winterman, V.; Bawden, C. Cluster Analysis Methods in Chemical Information Systems: Selection of Compounds for Biological Testing and Clustering of Substructure Search Output. *J. Chem. Inf. Comput. Sci*. **1986**, *26*, 109−118.

(9) Huang, N.; Nagarsekar, A.; Xia, G. J.; Hayashi, J.; MacKerell, A. D. Identification of nonphosphate-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*(14), 3502−3511.

(10) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors, In Methods and Principles in Medicinal Chemistry*; Wiley-VCH: New York, 2000; Vol. 11, p 667.

(11) Bajorath, J. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J. Chem. Inf. Comput. Sci*. **2001**, *41*(2), 233−245.

(12) Bender, A.; Glen, R. C. Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, *2*(22), 3204−3218.

(13) Brown, R. D.; Martin, Y. C. The information content of 2D and 3D structural descriptors relevant to ligand−receptor binding. *J. Chem. Inf. Comput. Sci.* **1997**, *37*(1), 1−9.

(14) Martin, Y. C.; Willett, P.; Lajiness, M.; Johnson, M.; Maggiora, G.; Martin, E.; Bures, M. G.; Gasteiger, J.; Cramer, R. D.; Pearlman, R. S.; Mason, J. S. Diverse viewpoints on computational aspects of molecular diversity. *J. Comb. Chem.* **2001**, *3*(3), 231−250.

(15) Horvath, D.; Jeandenans, C. Neighborhood behavior of in silico structural spaces with respect to in vitro activity spaces − A novel understanding of the molecular similarity principle in the context of multiple receptor binding profiles. *J. Chem. Inf. Comput. Sci.* **2003**, *43*(2), 680−690.

(16) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*(6), 983−996.

(17) Unity *Unity Chemical Information Software*; Tripos Associates: St. Louis, MO, 1996.

(18) Brown, R. D.; Martin, Y. C. An evaluation of structural descriptors and clustering methods for use in diversity selection. *SAR QSAR Environ. Res.* **1998**, *8*(1−2), 23−39.

(19) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; R., L.; Ferrin, T. E. A geometric approach in macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.

(20) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18*(9), 1175−1189.

(21) Pan, Y. P.; Huang, N.; Cho, S.; MacKerell, A. D. Consideration of molecular weight during compound selection in virtual target-based database screening. *J. Chem. Inf. Comput. Sci.* **2003**, *43*(1), 267−272.

(22) Chen, I. J.; Neamati, N.; Nicklaus, M. C.; Orr, A.; Anderson, L.; Barchi, J. J.; Kelley, J. A.; Pommier, Y.; MacKerell, A. D. Identification of HIV-1 integrase inhibitors via three-dimensional database searching using ASV and HIV-1 integrases as targets. *Bioorg. Med. Chem.* **2000**, *8*(10), 2385−2398.

(23) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*(1−3), 3−25.

(24) MOE *Molecular Operating Environment*, Montreal, Canada, 2002.

(25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(26) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF chimera − A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*(13), 1605−1612.

(27) Mills, J. E. J.; Dean, P. M. Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J. Comput.-Aided Mol. Des.* **1996**, *10*(6), 607−622.

(28) Klein, C.; Kaiser, D.; Kopp, S.; Chiba, P.; Ecker, G. F. Similarity based SAR (SIBAR) as tool for early ADME profiling. *J. Comput.-Aided Mol. Des.* **2002**, *16*(11), 785−793.

(29) Kibbey, C.; Calvet, A. Molecular property eXplorer: A novel approach to visualizing SAR using tree-maps and heatmaps. *J. Chem. Inf. Model.* **2005**, *45*(2), 523−532.

(30) Bohm, M.; Sturzebecher, J.; Klebe, G. Three-dimensional quantitative structure−activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J. Med. Chem.* **1999**, *42*(3), 458−477.

(31) Sheridan, R. P.; Feuston, B. P.; Maiorov, V. N.; Kearsley, S. K., Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR. *J. Chem. Inf. Comput. Sci*. **2004**, *44*(6), 1912−1928.

CI050225Z