ELSEVIER

Molecular scaffold-based design and comparison of combinatorial libraries focused on the ATP-binding site of protein kinases

Florence L. Stahura,* Ling Xue,* Jeffrey W. Godden,* and Jürgen Bajorath†

**MDS Panlabs, Computational Chemistry & Informatics, Bothell, Washington, USA*
†MDS Panlabs and Department of Biological Structure, University of Washington, Seattle, Washington, USA

Compound libraries were designed to target specifically the ATP cofactor-binding site in protein kinases by combining knowledge- and diversity-based design elements. A key aspect of the approach is the identification of molecular building blocks or scaffolds that are compatible with the binding site and therefore capture some aspects of target specificity. Scaffolds were selected on the basis of docking calculations and analysis of known inhibitors. We have generated 75 molecular scaffolds and applied different strategies to compute diverse compounds from scaffolds or, alternatively, to screen compound databases for molecules containing these scaffolds. The resulting libraries had a similar degree of molecular diversity, with at most 12% of the compounds being identical. However, their scaffold distributions differed significantly and a small number of scaffolds dominated the majority of compounds in each library. © *1999 by Elsevier Science Inc.*

Keywords: target-specific library design, protein kinase inhibitors, molecular scaffolds, combinatorial chemistry, computational docking, cluster analysis

INTRODUCTION

Compound libraries designed to produce specific inhibitors of therapeutic target proteins have generated significant interest in drug discovery research. The probability of finding molecules with desired specificity may be increased in different ways. Combinatorial chemistry provides the opportunity to generate large libraries of compounds for biological testing. Compounds in diverse screening libraries are designed to cover uniformly a chemical space as defined by selected molecular descriptors.[1] Such libraries are used for testing against multiple targets and consist of many compounds, e.g., ~100 000 or more. On the other hand, the major goal in focused library design is to increase the probability that compounds have a specific biological activity.[2] Thus, compared with diverse screening libraries, focused libraries are typically smaller in size, e.g., ~1 000–10 000 compounds, since their scope is target specificity rather than chemical diversity. The computational design of diverse chemical libraries is an intensely studied topic.[3,4] Design of libraries can be reaction or product oriented. In reaction-based design, chemical transformations are encoded and applied to collections of reactants,[5] while design in product space combines, systematically or randomly, molecular scaffolds and R groups.[5,6] Reaction-based design is well suited to generate chemically diverse compound collections. The design of focused or target-specific libraries is less well described than diversity design.

In this contribution, we describe an approach to product-based design of compound libraries that specifically target the ATP-binding site in protein serine/threonine and tyrosine kinases. The approach investigated here depends on the ability to identify molecular scaffolds that are compatible with the targeted binding site. Thus, we describe how relevant scaffolds can be identified. In addition, much remains to be learned about selection criteria that are suitable to bias compound collections toward specific biological activities or targets. Therefore, we have generated and compared alternative libraries generated from the same set of molecular scaffolds. This has helped to enhance our understanding of how scaffold distributions and

Color Plates for this article are on pages 51–52.

Address reprint requests to: Jürgen Bajorath, New Chemical Entities, Inc., 18804 North Creek Pkwy. S., Bothwell, WA 98011 USA. Tel.: 425-424-7292; Fax: 425-424-7299. E-mail: jbajorath@nce-mail.com

© 1999 by Elsevier Science Inc.
655 Avenue of the Americas, New York, NY 10010

1093-3263/99/$–see front matter
PII S1093-3263(99)00015-7

compound compositions of focused libraries differ when the design approach is modified.

METHODS

Scaffold design

Scaffolds were generated from known compounds by deleting substituents from core structures, as shown in Results and Discussion. We generated such scaffolds from known kinase inhibitors or compounds with shape compatibility to the ATP-binding site in kinase structures, as suggested by docking calculations. Docking was performed using DOCK 4.0[7,8] and the X-ray structure of $p56^{lck}$ (PDB identification code "3lck").[9] The parameters used for the docking calculations are listed in Table 1. Twenty-eight residues were selected to represent the ATP-binding site. Seventeen of these 28 residues are identical in the ATP-binding sites of cAMP-dependent kinase,[10] a serine/threonine kinase, and $p56^{lck}$, a tyrosine kinase. To identify recurrent molecular scaffolds, compounds from pharmaceutically relevant databases were docked both rigidly and by anchored search.[11] In anchored search docking, a fragment of the ligand is docked initially, and the remaining parts are progressively added.[11] Compounds were ranked on the basis of contact score and AMBER force field energy score.[12] As compound sources, a subset of Optiverse™ (OV)[13] and Maybridge™ (MB)[14] were used, each containing ~60 000 compounds. Three-dimensional conformations of database compounds were generated by a 2D/3D conversion function of MOE[15] that calculated a single low-energy conformation for each compound. In each docking calculation, the top scoring 500 compounds were selected and the fit of each compound was analyzed by computer graphics. Recurrent molecular fragments were identified and used as scaffolds.

Library generation

The selected molecular scaffolds were used to compute three alternative compound libraries. Library1 (2 500 molecules) was generated by compound sampling using QuaSAR-CombiDesign function of MOE.[15] These calculations proceeded as follows. For addition of R groups to scaffolds, a maximum number of four substitution points per scaffold was defined. The R group database contained 1 500 R groups, which were isolated from OV and MB compounds by a previously described algorithm.[16] Initially, combinations of scaffolds and R groups were randomly generated and the following molecular descriptors calculated. Fifty-seven SSKey-like descriptors[17] and the number of aromatic bonds, the fraction of rotatable single bonds, and the number of hydrogen bonding acceptors were used. For 100 randomly selected compounds, a principal components analysis (PCA) of molecular descriptors was performed and the top three principal components were selected. This was followed by Monte Carlo (MC) simulated annealing calculations to sample 2 500 diverse compounds in principal component space. The calculations started at a normalized temperature (T) value of 1, performed 7 500 MC steps at each T value, and scaled T with factor 0.95 from one iteration to the next until T was smaller than 10^{-6}. These calculations randomly exchange compounds between the source database and the sample while optimizing an entropy-based metric, used

Table 1. DOCK parameters[a]

Parameter	Setting	Parameter	Setting
flexible_ligand	Yes (no)	bump_filter	Yes
orient_ligand	Yes	bump_maximum	0
score_ligand	Yes	contact_score	Yes
minimize_ligand	No	(contact_cutoff_distance	4.5)
multiple_ligands	Yes	(contact_clash_overlap	0.75)
parallel_jobs	No	contact_clash_penalty	50
(anchor_search	Yes)	chemical_score	No
(multiple_anchors	Yes)	energy_score	Yes
(anchor_score	10)	(energy_cutoff_distance	10)
(torsion_drive	Yes)	(distance_dielectric	Yes)
(clash_overlap	0.5)	(dielectric_factor	4)
(configurations_per_cycle	25)	(attractive_exponent	6)
flexible_bond_maximum	10	(repulsive_exponent	12)
match_receptor_sites	Yes	atom_model	u
random_search	No	vdw_scale	1
automated_matching	Yes	electrostatic_scale	1
maximum_orientations	500	ligands_maximum	100 000
write_configurations	No	initial_skip	0
(intramolecular_score	Yes)	interval_skip	0
intermolecular_score	Yes	heavy_atoms_minimum	0
gridded_score	Yes	heavy_atoms_maximum	100
grid_version	4	rank_ligands	Yes

[a] All parameters used for the DOCK calculations are listed. Parameters were used for anchored search docking (i.e., flexible ligand set to "yes"). In the case of rigid docking (i.e., flexible ligand set to "no"), parameters in parentheses were not present or set to internal default values.

as a diversity measure, to cover the PCA-based molecular descriptor space. The sampling procedure implemented in MOE is distinct from cell partitioning methods,[18] except that both approaches aim to reduce the dimensionality of descriptor space.

As an alternative approach to library generation, we used an ISIS substructure search[19] to screen OV and MB for molecules that contained designed scaffolds. Those identified were filtered to include only compounds with a molecular mass of 700 kDa or less, a logP range between −2 and 6, and no more than 10 rotatable bonds. Approximately 24 000 compounds were identified. DOCK was used to screen this database on the p56^{lck} tyrosine kinase structure, only by anchored search docking. The top 5 000 molecules ranked by contact scoring and the top 5 000 ranked by force field energy scoring were selected and compared. Of these 10 000 compounds, approximately 2 500 molecules were identical and selected to represent Library2.

The source database (with ~24 000 molecules) was also analyzed by PCA-based compound classification,[20] using the QuaSAR-Cluster function of MOE. These calculations partition molecules into subsets of similar compounds based on PCA. The set of sixty 2D molecular descriptors was used as described above. Molecules were positioned in Euclidean space defined by the top three principal components of their descriptors. Each axis was divided into intervals and a molecule was assigned a letter code depending on the axis intervals into which the coordinates fell. A representative subset of compounds was selected by combining all unique molecules with one randomly chosen representative of each cluster. The library so generated contained 2 379 compounds and was called Library3.

To calculate a representative subset of all library compounds, Library1–3 were combined and duplicates were eliminated, yielding a total of 8 810 molecules. This set was partitioned by use of the QuaSAR-Cluster function of MOE as described above, resulting in a diverse subset of 1 086 compounds.

Library analysis

The compound libraries were compared in different ways. First, identical molecules were identified. Second, the overlap between different libraries was calculated by similarity searching, using a conventional 2D fingerprint[21] with a 0.85 cutoff value[22] for the Tanimoto coefficient (TC). A fingerprint is a bit map that expresses molecular properties as a characteristic binary pattern.[23] The TC is a measure of similarity and is defined as $N_{A\&B}/(N_A + N_B - N_{A\&B})$ where N_A and N_B are the number of bit sets on (i.e., 1) in bit strings A and B, respectively, and $N_{A\&B}$ is the number of bits set on that are in common.[23] TC values range from 0 to 1 and reflect increasing molecular similarity. Average TC values were calculated by dividing the sum of all TC values for pairwise comparison of compounds in a library by their number. To compare diversity profiles of libraries, distributions of TC values were analyzed by a histogram method. In these calculations, the TC range from 0 to 1 was divided into 100 intervals of 0.01, and TC values within each interval were plotted.

RESULTS AND DISCUSSION

Protein kinases as targets

Protein kinases catalyze the phosphorylation of serine/threonine or tyrosine residues by transferring a phosphate group from the MgATP cofactor and are divided into families, i.e., serine/threonine kinases and tyrosine kinases.[24] Members of both families are critical components of the cellular signal transduction apparatus.[25] Defects in kinase pathways are implicated in a variety of disease states, and protein kinases are therefore important targets for therapeutic intervention.[26] Despite differences in specificity, protein kinases generally share the MgATP cofactor requirement for catalysis and have largely conserved cofactor-binding sites.[27] Cofactor-binding sites of kinases have been targeted, and specific kinase inhibitors have been obtained by generating ATP analogs.[28] This suggests that subtle differences in the binding sites are sufficient to discriminate between analogs and confer ligand specificity. We use the cofactor-binding sites of protein kinases as a model system for molecular design and concentrate our efforts on cAMP-dependent kinase, a serine/threonine kinase,[10] and p56^{lck}, a tyrosine kinase[9] that is predominantly expressed by T lymphocytes, where it mediates cellular immune responses.[29] Three-dimensional structures of both enzymes have been determined. The structure of cAMP-dependent kinase was obtained both in the presence and absence of ATP and the structure of p56^{lck} was determined without ATP. Color Plate 1 shows a comparison of the ATP-binding sites in p56^{lck} and cAMP-dependent kinase. Here, we have primarily focused on p56^{lck}. The enzyme belongs to the Src family,[24] which includes the most intensely studied kinases.

Scaffold and library design

The approach to focused compound libraries described here is primarily based on the identification of molecular scaffolds that are compatible with a particular binding site. Different strategies employed to design kinase-specific libraries are summarized in Figure 1. Since the ATP-binding sites in protein kinases are largely conserved, yet sufficiently different to confer specificity of inhibitors,[28,30,31] diverse compounds based on a number of carefully defined scaffolds should be an attractive resource for inhibitor identification. Thus, although we use

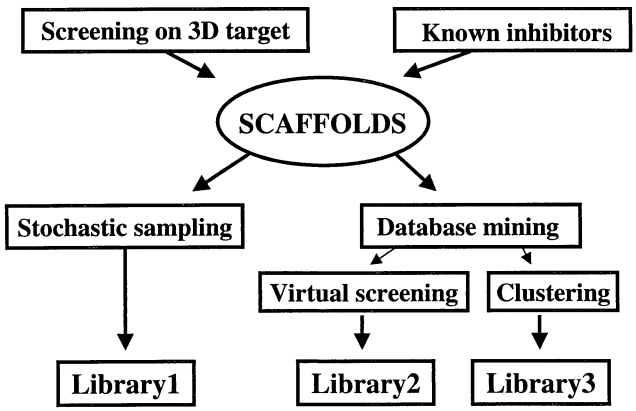


Figure 1. Schematic representation of focused library design strategies.

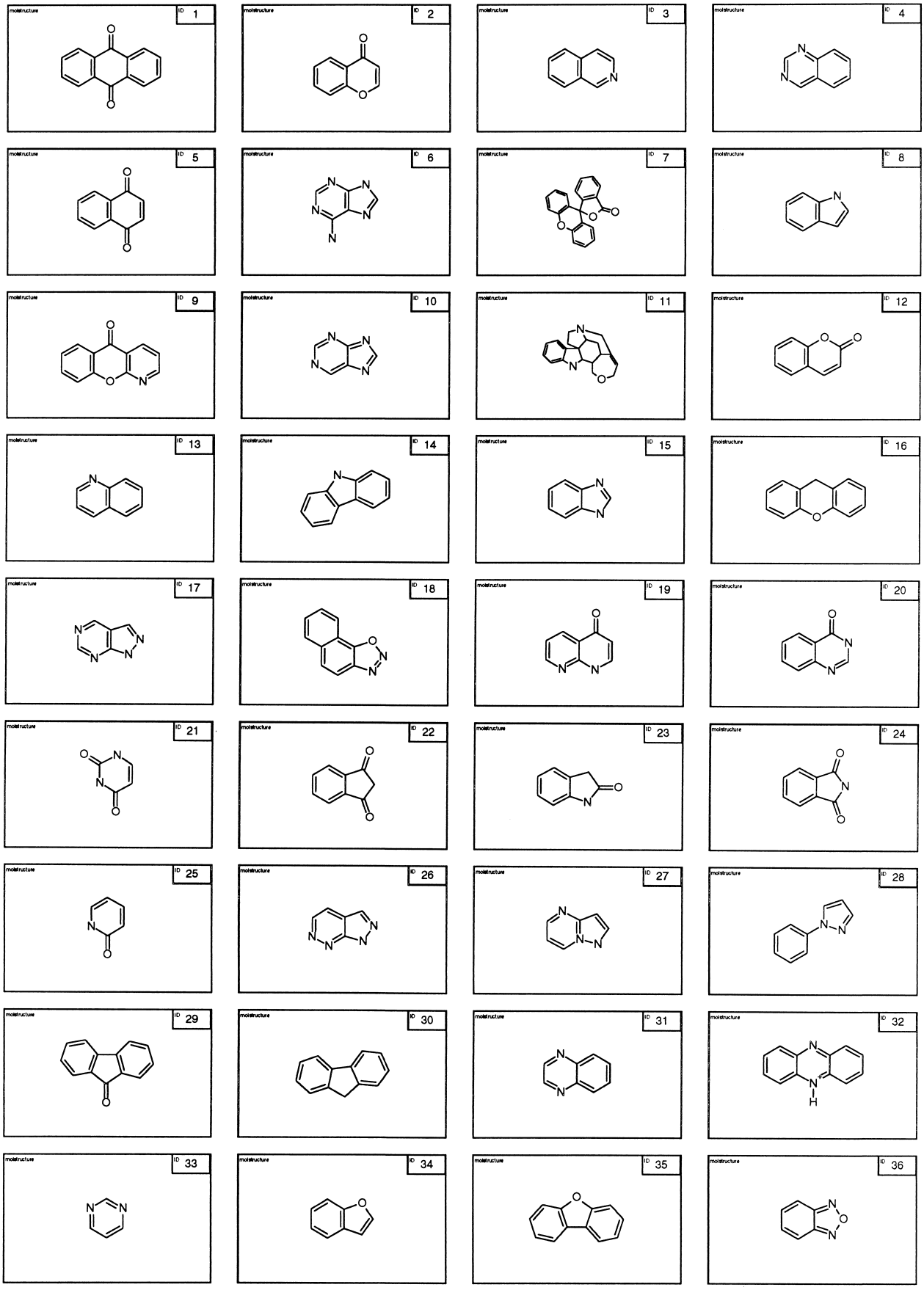


Figure 2. Molecular scaffolds. Shown here is the complete set of scaffolds that were designed and used in this study.

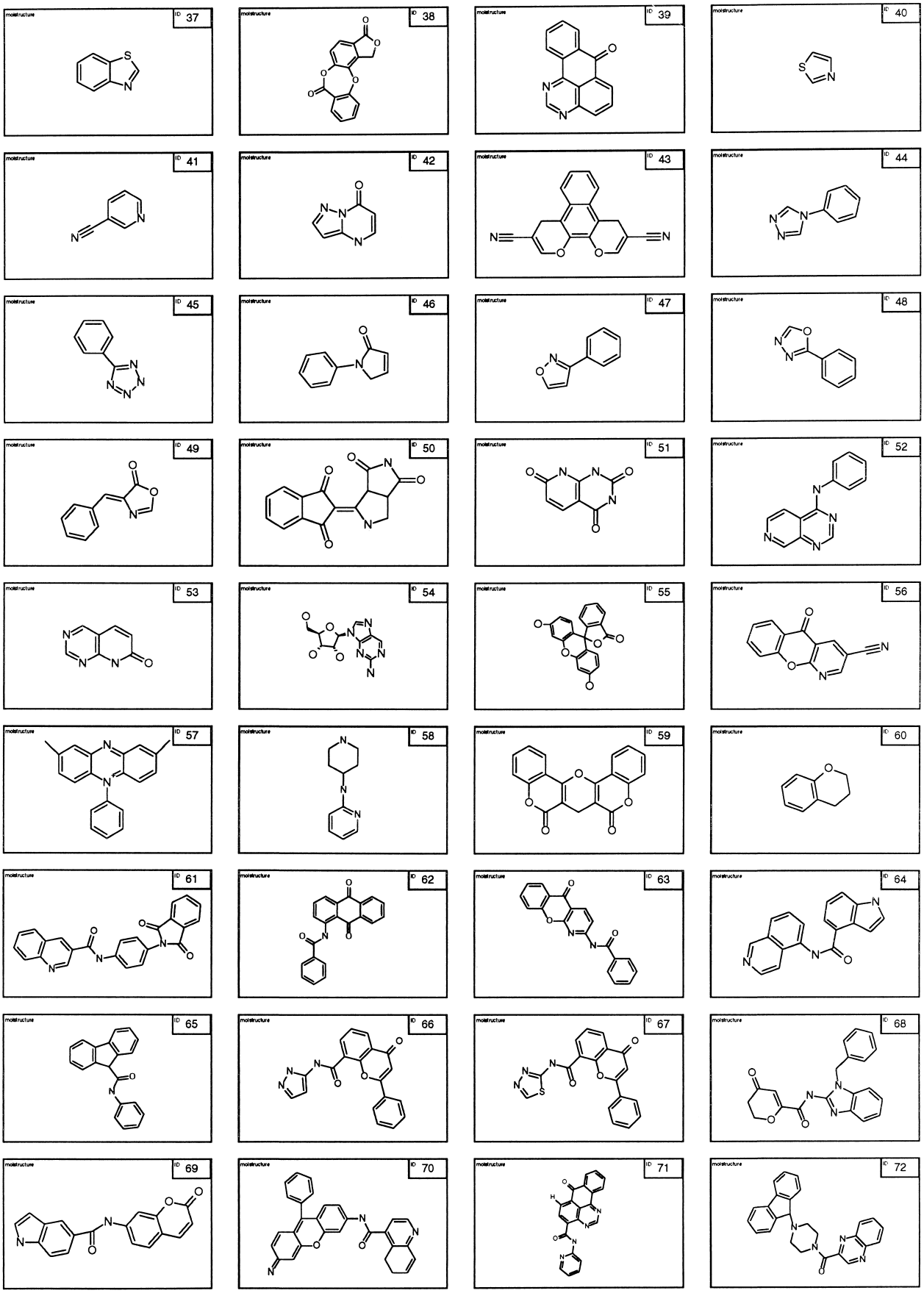


Figure 2. (continued)

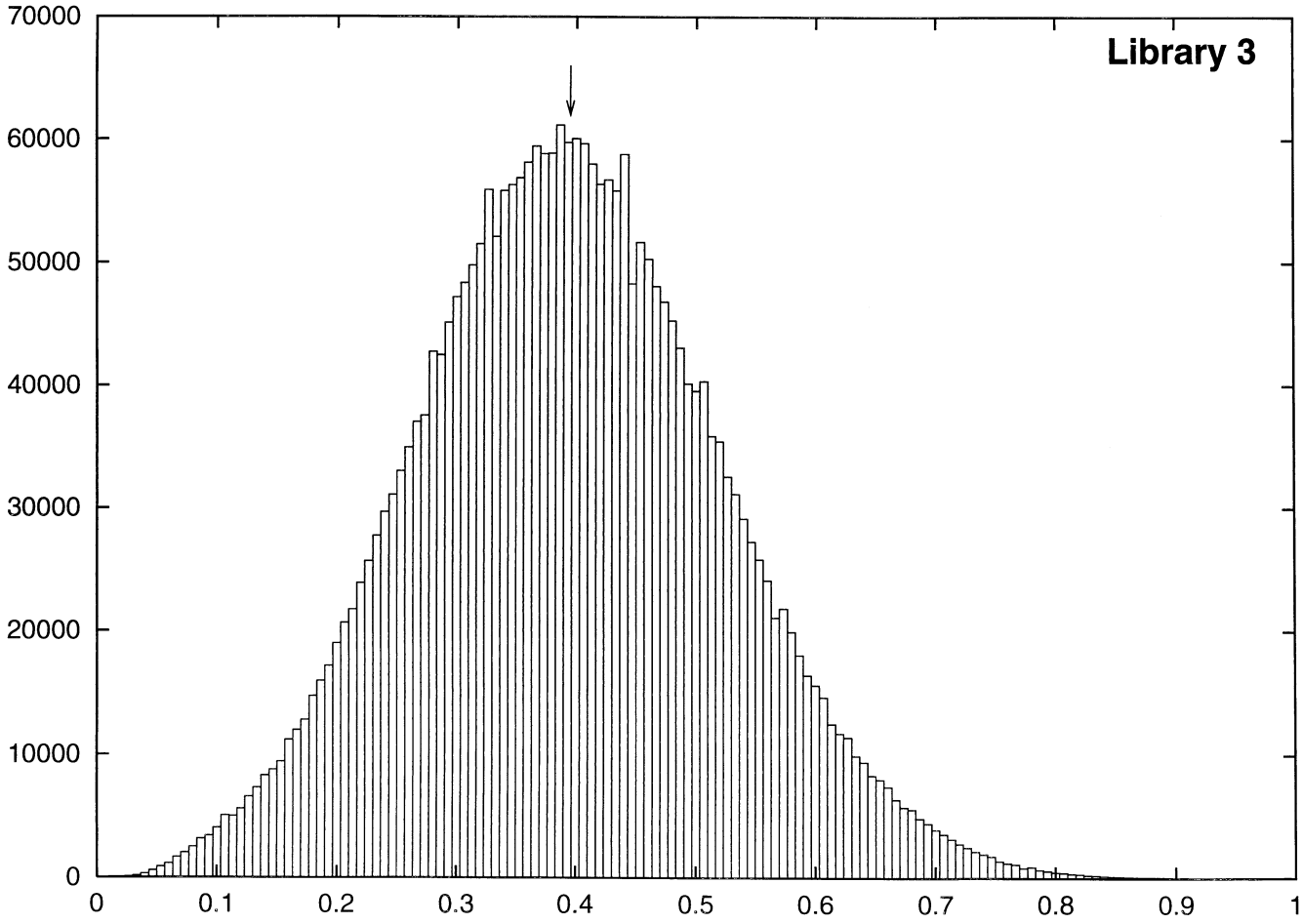


Figure 3. Diversity profile. The histogram represents Library3. The number of TC values for pairwise compound comparison within each 0.01 interval is shown. An arrow indicates the average TC value of the distribution. The diversity profiles were very similar for the three compared libraries.

$p56^{lck}$ as a model system, our calculations provide a basis for a "design on one—test on many" effort to identify kinase inhibitors.[31] Our scaffold design strategy exploits 3D information of the ATP-binding site by molecular docking and makes use of small molecular information obtained from known inhibitors. The scaffolds thus designed were used to compute diverse compound collections to increase further the probability of finding compounds with kinase-inhibitory activity.

A number of $p56^{lck}$ inhibitors have been reported in the literature.[32–34] In addition, X-ray structures of inhibitors bound to the ATP-binding site of a few kinase domains have been reported,[35–37] making this enzyme family an excellent model system for molecular design based on the use of both small and macromolecular structure information. A database of 75 scaffolds was designed from compounds identified by docking calculations (66 scaffolds) and based on comparison of known inhibitors (9 scaffolds). Figure 2 shows the complete set of scaffolds used to generate the different libraries. Representative scaffolds and compounds/inhibitors from which they were generated are shown in Color Plate 2. Damnacanthal (compound **4** in Color Plate 2) is the most potent and selective inhibitor of $p56^{lck}$, with an IC_{50} of 17 nM.[32] In contrast, genistein (compound **1**) is a weak inhibitor with broad tyrosine kinase specificity.[33] Two isoquinoline derivatives[34] yielded the same scaffold but show quite different IC_{50} values for $p56^{lck}$. Compound **2** has an IC_{50} value >100 μM for $p56^{lck}$, and compound **3** has an IC_{50} value of 0.5 μM. This illustrates that compounds containing the same scaffold but different substituents can have significantly different inhibitory activity and supports the design approach described here. Interestingly, some scaffolds identified by docking also occurred in known inhibitors. For example, several top-scoring compounds were found to contain a fragment of damnacanthal (Color Plate 2).

Diversity and compound composition

Average TC values were similar for the three libraries compared in this study (0.43, 0.50, and 0.40 for Library1, -2, and -3, respectively). Moreover, the diversity profiles of all three libraries were similar, showing unimodal TC distributions with the majority of TC values close to the average. A representative profile is shown in Figure 3. Thus, on the basis of these calculations, the molecular diversity of the libraries was comparable. Similarity searches were carried out with a TC cutoff value of 0.85 (at this level, compounds are similar in terms of their structure and functional properties[22]) and 1.0 (compounds

Table 2. Compound similarity[a]

Libraries compared	1	>0.85
Library1 and 2	0	~62
Library1 and 3	0	~57
Library2 and 3	~12	~78

[a] The percentage of identical (TC = 1) or very similar (TC >0.85) compounds was calculated for pairwise comparison of libraries.

are identical, or nearly identical). The results are shown in Table 2. The libraries did not contain many identical compounds. Few, if any, identical compounds were found in Library1 when compared with Library2 or Library3. Library2 and Library3 generated from a common subset of approximately 24 000 molecules share only 12% of their compounds. When a TC cutoff value of 0.85 was applied, similarity overlaps between 57 and 78% were observed.

Scaffold distributions

The occurrence of molecular scaffolds in the libraries was also analyzed. Color Plate 3 shows the distribution of 72 unique scaffolds (three of 75 scaffolds differed only in the position of substitution points) in each library. We found significant differences in the distribution of scaffolds in the compound libraries. Some scaffolds were preferentially selected in one of the libraries, while other scaffolds were used very little, if at all, in our libraries. In each library, a number of scaffolds (13–16) were not found. Two scaffolds (number 52 and 67 in Color Plate 3) did not occur in any of the libraries. The origin of scaffold 52 was a known inhibitor, while scaffold 67 was isolated from a docked compound. Color Plate 3 also demonstrates that some, but not all, of the observed scaffold preferences correlate with a high frequency of occurrence of these scaffolds in OV and MB (which were used as source databases for the assembly of Library2 and Library3). OV and MB compounds show substantial differences in the distribution of their molecular building blocks and R groups.[16] However, significant scaffold preferences were also detected in Library1 (Color Plate 3); these are not related to scaffold distributions in source databases but result from diversity sampling of scaffold/R group combinations. Therefore, Library1 also provides a control for observed scaffold distributions. Figure 4 shows the 10 most frequently occurring scaffolds. Together these scaffolds accounted for ~72% of all generated library compounds. Three of the "top 10" scaffolds were also found in the top 10 list of each library (Figure 4). These three scaffolds alone accounted for ~29% of all the compounds. Thus, regardless of the design protocol, a few scaffolds occur with high frequency, and in each library a small number of scaffolds dominates the majority of compounds.

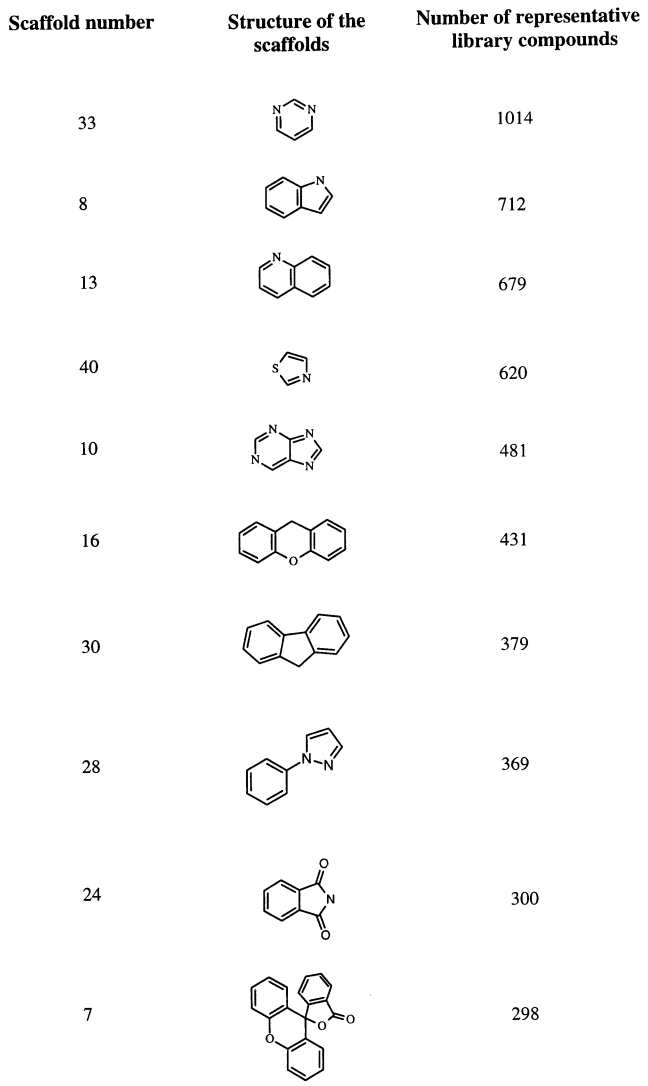


Figure 4. Structures of the 10 most frequently occurring scaffolds. The number of library compounds containing each scaffold is reported. Three scaffolds (33, 13, and 16) are in the top 10 list of each library and account for 29% of all compounds.

Target specificity

An important question concerns how target specificity is achieved in focused library design. The "design on one—test on many" idea discussed above implies that library compounds should share features that make them compatible with the targeted site, here the cofactor-binding site of protein kinases, and that favor binding. This has been addressed at the level of scaffold design where molecular fragments were selected either on the basis of their shape complementarity to the binding site or, alternatively, isolated from known active compounds. Selected scaffolds already capture some aspects of target specificity since they are compatible with the ATP-binding site of kinases rather than binding sites of other protein families. However, these scaffolds presumably do not bind kinases because the majority of scaffolds lack functional groups. Thus, binding specificity is obtained only by decorating molecular scaffolds with R groups, i.e., transforming preferred molecular building blocks into actual compounds that can form specific interactions with binding sites. These interactions are likely to differentiate between members of the kinase family. Thus, focusing of the library involves two aspects: compatibility and specificity. Ultimately, only experimental testing can confirm desired differences in specificity. However, we have designed

Table 3. DOCK screening of scaffolds and a representative subset of focused library compounds[a]

Docking calculation	Total number of scaffolds in the top 500 molecules	Highest rank of any scaffold in the top 500 molecules
Rigid body docking and contact scoring	8	25
Rigid body docking and energy scoring	4	126
Anchored search docking and contact scoring	11	27
Anchored search docking and energy scoring	7	285

[a] A total of 1 158 compounds (72 scaffolds and a subset of 1 086 focused library compounds) were docked to the ATP-binding site of $p56^{lck}$ and the 500 best scoring compounds in each of the four docking calculations were selected.

a computer experiment that at least illustrates the underlying idea. A representative subset of 1 086 library compounds was combined with our set of 72 scaffolds and screened in four different docking calculations on the ATP-binding site of $p56^{lck}$. The results are summarized in Table 3. Regardless of the docking and scoring protocol, many library compounds achieve better shape complementarity or intermolecular interactions with the binding site than the best scoring scaffold and a maximum of 11 scaffolds was found among the 500 top scoring molecules. This is consistent with the idea that specificity is achieved at the level of library compounds rather than scaffolds.

CONCLUSIONS

We have described an approach to design focused combinatorial libraries. In this approach, a critical step is the identification of molecular scaffolds that are compatible with a targeted binding site and therefore capture some aspects of target specificity. Scaffold selection was based on virtual screening and analysis of known active compounds. Different computational strategies were applied to generate alternative focused libraries on the basis of these scaffolds. In one case, diverse compounds were computed and in two others, compounds containing the selected scaffolds were sampled and filtered by structure-based screening and clustering techniques. A priori, each of these strategies is considered valid to generate a target-specific compound library. However, when alternative libraries were compared, we detected significant differences in compound composition and scaffold distribution. A small subset of scaffolds dominated the majority of compounds in each library. For compound libraries assembled by database-screening techniques, this was largely a consequence of scaffold distributions in the source databases. Interestingly, Library1 does not depend on a source database but also contains some overrepresented scaffolds. Since a number of our originally designed scaffolds were underrepresented or not present, the generated libraries did not contain the full information content of the originally designed ensemble of scaffolds. This can be achieved only if scaffolds are more evenly distributed in library compounds. However, we cannot conclude that libraries with biased scaffold distributions are less focused than libraries with even distributions. One may argue that enrichment of scaffold types during the design process is desired to increase the target focus of a library. Furthermore, most of the target specificity must be encoded in library compounds rather than scaffolds. Specificity is largely determined by the way scaffolds are structurally diversified and transformed into compounds.

ACKNOWLEDGMENTS

We thank Chris Williams (Chemical Computing Group, Inc.) for help with MOE. Furthermore, we thank our referees for constructive comments that helped to improve the article.

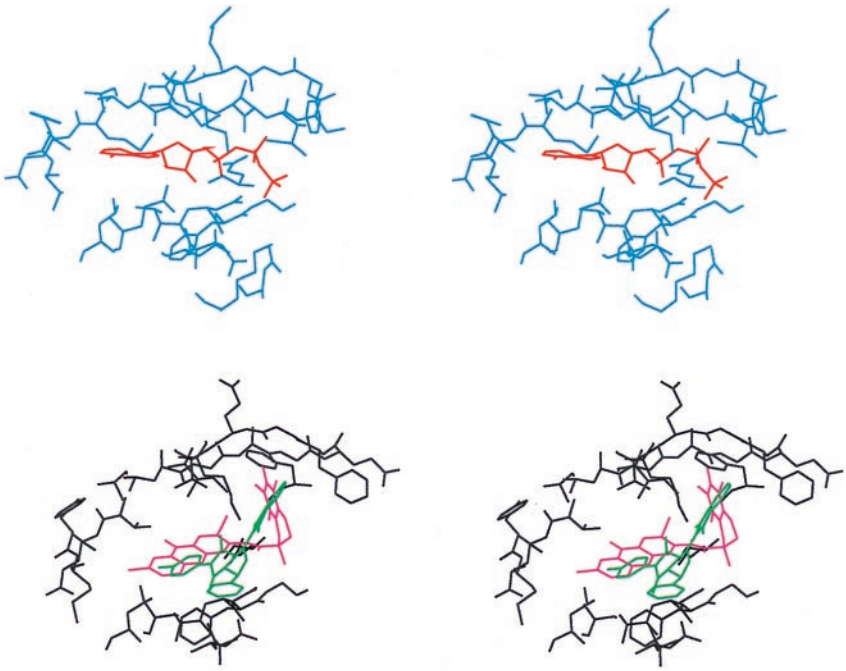
REFERENCES

1 Bures, M.G., and Martin, Y.C. Computational methods in molecular diversity and combinatorial chemistry. *Curr. Opin. Chem. Biol.* 1998, **2**, 376–380
2 Li, J., Murray, C.W., Waszkowycz, B., and Young, S.C. Targeted molecular diversity in drug discovery: Integration of structure-based design and combinatorial chemistry. *Drug Discov. Today* 1998, **3**, 105–112
3 Kubinyi, H. Combinatorial and computational approaches in structure-based drug design. *Curr. Opin. Drug Discov. Dev.* 1998, **1**, 16–27
4 Kauvar, L.M., and Laborde, E. The diversity challenge in combinatorial chemistry. *Curr. Opin. Drug Discov. Dev.* 1998, **1**, 66–70
5 Cramer, R.D., Patterson, D.E., Clark, R.D., Soltanshahi, F., and Lawless, M.S. Virtual compound libraries: A new approach to decision making in molecular discovery research. *J. Chem. Inf. Comput. Sci.* 1998, **38**, 1010–1023
6 Gillet, V.J., Willett, P., and Bradshaw, J. The effectiveness of reactant pools for generating structurally-diverse combinatorial libraries. *J. Chem. Inf. Comput. Sci.* 1997, **37**, 731–740
7 Meng, E.C., Gschwend, D.A., Blaney, J.M., and Kuntz, I.D. Orientational sampling and rigid-body minimization in molecular docking. *Proteins* 1993, **17**, 266–278
8 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langidge, R., and Ferrin, T.E. A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* 1982, **161**, 269–288
9 Yamaguchi, H., and Hendrickson, W.A. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Science* 1996, **384**, 484–489
10 Zheng, J., Knighton, D.R., Ten Eyck, L.F., Karlsson, R., Xuong, N., Taylor, S., and Sowadski, J.M. Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry* 1993, **32**, 2154–2161
11 Leach, A.R., and Kuntz, I.D. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comput. Chem.* 1992, **13**, 730–748
12 Weiner, S.J., Kollman, P.A., Nguyen, D.T., and Case, D.A. An all atom force field for simulation of proteins and nucleic acids. *J. Comput. Chem.* 1986, **7**, 230–252
13 Garr, C.D., Peterson, J.R., Schultz, L., Oliver, A.R., Underiner, T.L., Cramer, R.D., Ferguson, A.M., Lawless, A.S., and Patterson, D.E.J. Solution phase synthesis

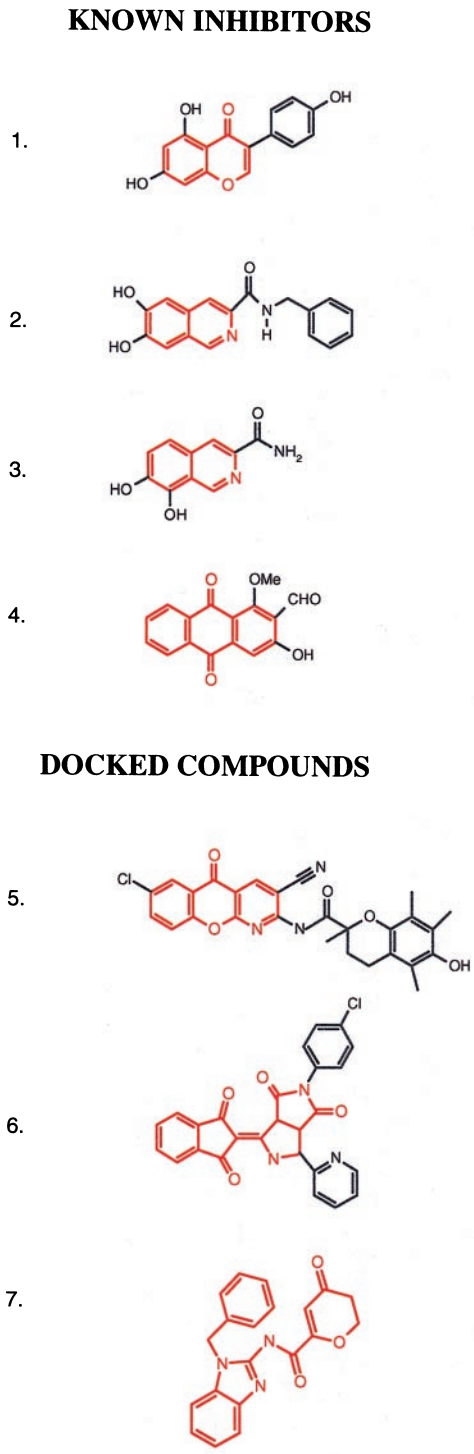
of chemical libraries for lead discovery. *J. Biomol. Screen.* 1996, **1**, 179–186
14 *Maybridge 97*. Daylight Chemical Information Systems, Irvine, California, 1997
15 *MOE, Molecular Operating Environment.* Chemical Computing Group, Inc., Montreal. http://www.chemcomp.com/fdept/prodinfo.htm
16 Xue, L., and Bajorath, J. Distribution of molecular scaffolds and R-groups isolated from large compound databases. *J. Mol. Model.* 1999, **5**, 97–102
17 Brown, R.D., and Martin, Y.C. Use of structure–activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sci.* 1996, **36**, 572–584
18 Pearlman, R.S. *Novel Software Tools for Addressing Chemical Diversity*. http://www.netsci.org/Science/Combichem/feature08.html, 1996
19 MDL Information Systems. *ISIS.* MDL Information Systems, Inc., San Leandro, California
20 Labute, P. *QuaSAR-Cluster: A Different View of Molecular Clustering*. http://www.chemcomp.com/article/cluster.html, 1998
21 Bush, B.L., and Sheridan, R.P. PATTY: A programmable atom typer and language for automatic classification of atoms in molecular databases. *J. Chem. Inf. Comput. Sci.* 1993, **33**, 756–762
22 Patterson, D.E., Cramer, R.D., Ferguson, A.M., Clark, R.D., and Weinberger, L.E. Neighborhood behavior: A useful concept for validation of molecular descriptors. *J. Med. Chem.* 1996, **39**, 3049–3059
23 Flower, D.R. On the properties of bit string-based measures of chemical similarity. *J. Chem. Inf. Comput. Sci.* 1998, **38**, 379–386
24 Hanks, S.K., Quinn, A.M., and Hunet, T. The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science* 1988, **241**, 42–52
25 Sandy, J.R. Signal transduction. *Br. J. Orthod.* 1998, **25**, 269–274
26 Showalter, H.D., and Kraker, A.J. Small molecular inhibitors of the platelet-derived growth factor receptor, the fibroblast growth factor receptor, and Src family tyrosine kinases. *Pharmacol. Ther.* 1997, **76**, 55–71
27 Taylor, S.S., and Radzio-Andzelm, E. Three protein kinase structures define a common motif. *Structure* 1994, **2**, 345–355
28 Denny, W.A. ATP site-directed kinase inhibitors. *IDrugs* 1998, **1**, 10–12
29 Weil, R., and Veillette, A. Signal transduction by the lymphocyte-specific tyrosine protein kinase $p56^{lck}$. *Curr. Top. Microbiol. Immunol.* 1996, **205**, 63–87
30 Cohen, P., and Goedert, M. Engineering protein kinases with distinct nucleotide specificities and inhibitor sensitivities by mutation of a single amino acid. *Chem. Biol.* 1998, **5**, R161–R164
31 Gray, N.S., Wodicka, L., Thunnissen, A.-M.W.H., Norman, T.C., Kwon, S., Espinoza, F.H., Morgan, D.O., Barnes, G., LeClerc, S., Meijer, L., Kim, S.-H., Lockhart, D.J., and Schultz, P.G. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* 1998, **281**, 553–558
32 Faltinek, C.R., Schroeder, J., Mauvais, P., Miller, D., Wang., S., Murphy, D., Lehr, R., Kelley, M., Maycock., A., Michne, W., Miski, M., and Thunberg., A.L. Damnacanthal is a highly potent, selective inhibitor of $p56^{lck}$ tyrosine kinase activity. *Biochemistry* 1995, **34**, 12404–12410
33 Akiyama, T., and Ogawara, H. Use and specificity of genistein as inhibitor of protein-tyrosine kinase. *Methods Enzymol.* 1991, **201**, 362–370
34 Burke, T.R., Jr., Lim, B., Marquez, V.E., Li, Z.H., Bolen, J.B., Stefanova, I., and Horak, I.D. Bicyclic compounds as ring-constrained inhibitors of protein-tyrosine kinase $p56^{lck}$. *J. Med. Chem.* 1993, **36**, 425–432
35 Mohammadi, M., McMahon, G., Sun, L., Tang, C., Hirth, P., Yeh, B. K., Hubbard, S.R., and Schlessinger, J. Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. *Science* 1997, **276**, 955–960
36 Schulze-Gahmen, U., Brandsen, J., Jones, H.D., Morgan, D.O., Meijer, L., Vesely, J., and Kim, S. Multiples modes of ligand recognition: Crystal structures of cyclin-dependent protein kinase 2 in complex with ATP and two inhibitors, olomoucine and isopentenyladenine. *Proteins* 1995, **22**, 378–391
37 Engh, R., Girod, A., Kinzel, V., Huber, R., and Bossemeyer, D. Crystal structures of catalytic subunit of cAMP-dependent protein kinase in complex with isoquinolinesulfonyl protein kinase inhibitors H7, H8, and H89. *J. Biol. Chem.* 1996, **271**(42), 26157–26164

Florence L. Stahura, Ling Xue, Jeffrey W. Godden, and Jürgen Bajorath

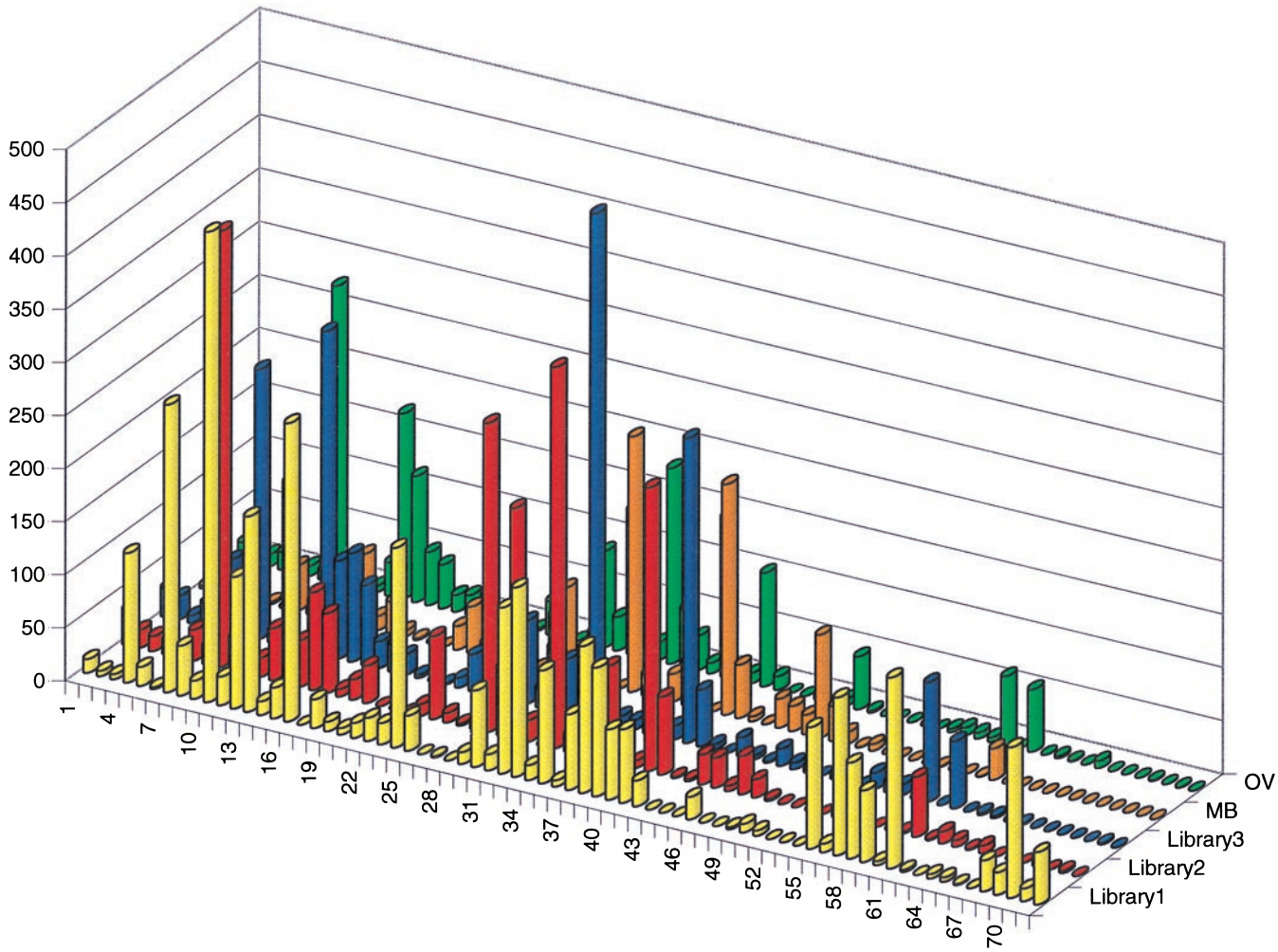
Molecular scaffold-based design and comparison of combinatorial libraries focused on the ATP-binding site of protein kinase



Color Plate 1. Stereo comparison of the ATP-binding sites in p56^{lck} and cAMP-dependent kinase. *Top:* cAMP-dependent kinase site is shown in blue with bound ATP (red) as determined by X-ray crystallography. *Bottom:* p56^{lck} site is shown in black with two docked compounds (green, magenta).



Color Plate 2. Illustrated here, using representative examples, is how scaffolds were isolated from known inhibitors and docked compounds. In each case, the scaffold is colored in red. Two of the compounds are also shown in Color Plate 1 (compound **5** in magenta and compound **6** in green). The scaffolds shown in red are the exact query molecules used in substrucuture searching. The scaffolds contain heteroatoms (N, O) that, in many cases, define points for R group attachment in diversity-based compound generation.



Color Plate 3. Scaffold distributions. The number of compounds containing each original scaffold is shown for Library1 (yellow), Library2 (red), and Library3 (blue) and, as reference, for MB (gold) and OV (green).