ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2002) 469–477

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Coupling structure-based design with combinatorial chemistry: application of active site derived pharmacophores with informative library design

John E. Eksterowicz [a,*], Erik Evensen [a], Christian Lemmen [a], G. Patrick Brady [b], J. Kevin Lanctot [a], Erin K. Bradley [a], Eddine Saiah [c], Leslie A. Robinson [c], Peter D.J. Grootenhuis [c], Jeffrey M. Blaney [a]

[a] *DuPont Pharmaceuticals Research Laboratories, 150 California Street, Suite 1100, San Francisco, CA 94111, USA*
[b] *DuPont Pharmaceuticals Company, Experimental Station, P.O. Box 80500, Wilmington, DE 19880-0500, USA*
[c] *DuPont Pharmaceuticals Research Laboratories, 4570 Executive Drive, Suite 400, San Diego, CA 92121, USA*

Abstract

Protein structural information is combined with combinatorial library design in the following protocol. Active site maps are generated from protein structures. All possible 2-, 3- and 4-point pharmacophores are enumerated from the active site map and encoded as bit strings. The pharmacophores define a design space that can be used to select compounds using an informative library design tool. The method was evaluated against a collection of compounds assayed previously against a cyclin-dependent kinase target, CDK-2, starting with 23 X-ray co-crystal structures. Performance was assessed based on the number of active scaffolds selected after four rounds of iterative informative library design. The method selects compounds from 12 out of the 15 active scaffolds from the CDK-2 library and outperforms a two-dimensional similarity search and docking calculations.

Keywords: Combinatorial chemistry; Cyclin-dependent kinase; Informative library design; Pharmacophore; Site map; Structure-based design

1. Introduction

The high failure rate of drug candidates in the development phase can be largely attributed to problems with pharmacokinetics and safety [1], both of which are to some extent coupled to the chemical structure of the compound. Thus success in moving from discovery to the clinic requires a strategy to hedge risks by generating lead candidates from different chemical classes. The computational approach presented here is designed to generate multiple, promising, structurally diverse leads rapidly and thus is particularly useful in the early stages of a discovery project. The search for multiple structural series is accomplished by coupling protein structural information with combinatorial library design.

The protocol is as follows (Fig. 1): active site maps are generated from protein structures, all possible 2-, 3- and 4-point pharmacophores are enumerated from the site map and encoded as a bit string (signature) these pharmacophores define a space to be probed by compounds that are selected using the informative library design tool. The metric used to evaluate the success of the approach is the number of active scaffolds selected in the library design, with the number of active compounds as a secondary measure.

There is literature precedent for similar approaches. A number of software tools exist for mapping site points (e.g. GRID [2] and SITEPOINT [3]) and MCSS [4–8] techniques have been applied for generating site maps. The "design in receptor" method has been developed [9] and applied [10] to map active sites and design chemical libraries. In addition, methods have been developed to account for multiple protein conformations [11], including the creation of a dynamic pharmacophore model [12] from molecular dynamics simulations. And finally, an experimental and computational needle screening approach has also been developed for mapping active sites with molecular fragments [13].

The approach presented here builds on existing techniques and differs in three key areas: the nature of the site

* Corresponding author.
E-mail address: jeksterowicz@combichem.com (J.E. Eksterowicz).

1093-3263/02/$ – see front matter
PII: S1093-3263(01)00148-6

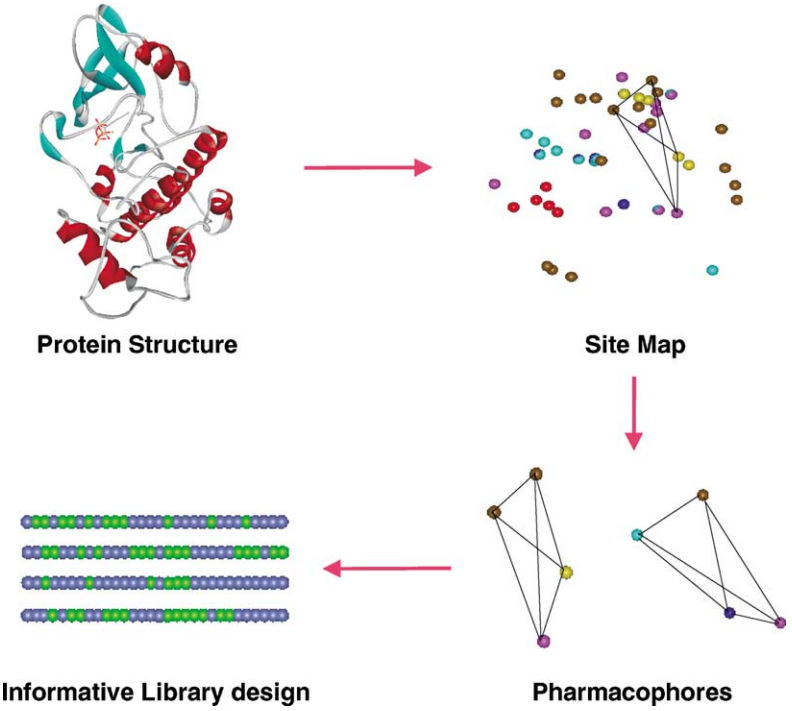

Fig. 1. The library design protocol. Starting with the protein structure, a site map is generated from the active site. Pharmacophores are enumerated and used to define the space for library design. Compounds are then selected with the informative design tool, such that the resulting subset will interrogate the target in different, but overlapping ways. The bit strings for four sample compounds are illustrated. A green dot indicates a bit is turned on (pharmacophore is present in the molecule).

map points, the number of protein structures considered and the specific approach to library design (i.e. compound selection). First, the algorithm for site map generation typically generates between 30 and 80 feature positions for each active site (ligand features are not considered) often yielding 10^5 pharmacophores. This is in contrast to other approaches which tend to generate <20 feature positions [10,11,14]. Second, multiple protein structures can be incorporated into the method in a straightforward manner. While this has been discussed before [11], incorporating the information from more than 20 X-ray structures is novel. Finally, this is the first application of an informative library design [15,16] approach to compound selection within the pharmacophore space defined by the active site. These differences are discussed in Section 2.

The method was assessed using a collection of compounds assayed previously against a cyclin-dependent kinase target, CDK-2. For this particular target 23 X-ray co-crystal structures [17–20] [1] were available as well as a set of approximately 3800 compounds [21,22] of known activity. Twenty-two distinct scaffolds were represented in the library, fifteen of which contained active compounds. Pharmacophores were generated for each structure and a union signature was created. The approach is evaluated in terms of the number of active scaffolds selected after four rounds of iterative library design and is compared to two-dimensional similarity searching and docking calculations.

[1] The remaining structures were generated internally and have not been released.

2. Methods

The protocol is illustrated in Fig. 1 and considered in more detail in the following sections.

2.1. Generating a site map

Feature points complementary to the active site are computed using an internally developed software tool. For example, a hydrogen bond donor feature is mapped in the proximity of a hydrogen bond acceptor in the protein active site. The collection of 3D coordinates and labels (acceptors, donors, negatives, positives, hydrophobes and aromatics) is called a site map. Technically, the site map is the union of three separately computed maps, ESMap which contains the electrostatic feature points (P, N, and H) HBMap with hydrogen-bonding feature points (D and A) and AroMap containing aromatic feature points (Ar).

The electrostatic feature map, ESMap, is computed by first using the sphere placement algorithm employed in the program PASS [23]. It generates an evenly-distributed set

of points (ProbeMap) in regions of buried volume along the protein surface. A subset of points in the ProbeMap comprises the P, N, and H feature points depending upon the local electrostatic character of the protein. The CVFF molecular mechanics force field is used to compute the electrostatic potential, ϕ_i, at each point i of ProbeMap, along with the mean potential $\langle\phi\rangle$ and mean magnitude $\langle|\phi|\rangle$ averaged over all points in ProbeMap [24]. The value of ϕ_i determines whether or not point i is included as a P, N, or H feature point, according to the following definitions

$$\phi_i > \langle\phi\rangle + 1.5^*\sigma(\phi), \quad i = \text{N feature point}$$

$$\phi_i > \langle\phi\rangle - 1.5^*\sigma(\phi), \quad i = \text{P feature point}$$

$$\langle|\phi|\rangle - 1.0^*\sigma(|\phi|) < |\phi_i| < \langle|\phi|\rangle + 1.0^*\sigma(|\phi|),$$
$$i = \text{H feature point}$$

Here $\sigma(X)$ denotes the standard deviation about the mean of quantity X. This normalizes the point assignments relative to the overall electrostatic environment of the active site. This presents non charge-neutral protein structures (which may result from counter ions not being resolved or present in the crystal structures) from skewing feature point assignments unreasonably.

The hydrogen-bonding feature map, HBMap, is determined by projecting complementary points outward from known hydrogen-bonding atoms of the protein. The resulting superset of points is filtered on the basis of steric clash, insufficient burial and minimal proximity of alike feature points. Statistical analysis of hydrogen-bonding geometries in the protein data bank (PDB) revealed two types of geometries, ideal and bifurcated [25]. For both types, hydrogen-bonding points are placed separately. Ideal hydrogen-bonding points are positioned on the basis of the mean angle and distance as observed in the PDB. Points that clash with the protein are removed. However, for robustness, small positional perturbations are applied to retain potentially important hydrogen-bonding positions. Bifurcated hydrogen-bonding points are computed heuristically by investigating full rings of points equally bifurcated between protein atoms that are considered moderate or strong hydrogen bond participants. Points on such rings are retained as bifurcated HB points if they do not violate steric clash, burial and mutual proximity conditions. To build the final HBMap, the surviving sets of ideal and bifurcated HB points are combined and subjected to filtration on the basis of mutual proximity.

The AroMap set of aromatic feature points is computed by repeatedly docking a benzene ring into the protein active site and retaining the centroids of the top-scoring configurations. The protein is represented using a polar-hydrogen CVFF force field. The docking is performed using internal code in local optimization mode [26]. One hundred separate local docking trials with different starting positions are performed. Any of the docked configurations whose score lies within an energy window of 5 kcal/mol of the minimum-energy configuration is included in AroMap. Again points are subjected to filtration on the basis of burial and mutual proximity.

2.2. Converting pharmacophores into a signature

Pharmacophores are generated on the basis of feature points in the active site by exhaustive enumeration of all 2-, 3-, and 4-point subsets of the feature points. The details of this procedure have been described elsewhere [11,27,28]. For all pairs of feature points their distance in 3D-space is precomputed. In order to arrive at a discrete representation of a pharmacophore, the distances are binned, applying a user-defined binning scheme (e.g. the 4-point pharmacophore bins are as follows 1.6–4.6, 4.6–7.6, 7.6–10.6, 10.6–16.6, 16.6–19.6, 19.6–22.6 Å). Chirality is denoted by encoding the handedness of 4-point pharmacophores. Each pharmacophore is mapped onto a unique address, such that any possible combination of up to four features and distances is represented. The address is taken for a binary representation of the pharmacophores, called a signature. The length of the signature is the highest possible address for an encoding of a 4-point pharmacophore. All bits in the signature are initially set to 0. In order to represent a pharmacophore the bit at the respective address in the signature is turned on (set to 1). For the representation of the active site all pharmacophores are exhaustively enumerated and the respective bits are turned on.

2.3. Union of signatures for multiple structures

The method can handle multiple protein active site conformations simultaneously. The binary union of multiple signatures yields a single bit string representing all pharmacophores present in any structure. Note that variations of this protocol are implemented easily as well. If instead of a binary union, the number of occurrences of a pharmacophore in any of the signatures is counted, then the binary union is achieved by turning all non-zero bits on. At the other extreme, turning on only those bits with a count equal to the number of signatures, yields the intersection of all signatures. Any consensus threshold c can be used to define the consensus representation of multiple active sites. That is, a pharmacophore is present in at least c of active site conformations. Note that this way of handling multiple active site snapshots is quite expedient. In principle thousands of active-site conformations, for example from a dynamics simulation, can be treated in this manner.

2.4. Molecular signatures

Ligand molecules are encoded as follows. First, conformers are generated for each compound using an internal tool, that generates a fairly complete conformational model

of the molecule [29–31][2]. Features are assigned using a substructure-based set of rules. Pharmacophores are enumerated from these three-dimensional feature positions following the same protocol as for the active site, thus ensuring compatibility of the binary encodings. However, multiple conformers need to be represented simultaneously here. This is done by wrapping the exhaustive enumeration of pharmacophores for a single conformer into an extra loop over all the conformers of a compound. That is, any pharmacophore on any conformer of a compound is represented by turning the respective bit in the signature on.

2.5. Molecular signature masking

With the binary representation of the active site and the binary representation of the molecules being defined analogously, the meaning of a bit at a certain address is the same (the same pharmacophore, within the tolerances of the distance binning). Therefore, representing a design space amounts to masking all molecule signatures by the active site signature. Masking a signature means taking the logical *and* of the bits of the site signature and the molecule signature. For a given molecule, bits representing pharmacophores not present in the active site are turned off, whereas the bits of the pharmacophores in the active site can be either on or off, depending on their presence or absence in the molecules. This way only the pharmacophore space defined by the active site is taken into account.

2.6. Informative library design

Informative library design is a molecule selection strategy that optimizes information return for a given virtual library [15,16,32]. The goal is to detect a set of features (pharmacophores) that determine activity against a particular target. Informative design aims at selecting a set of compounds such that the resulting subset will interrogate the target in different, but overlapping ways. Molecules are selected for synthesis and screening such that each pharmacophore in the design space has a unique pattern of occurrence in the molecules of the set. This unique 'code' enables the identification and retention of the important pharmacophores when the set of compounds is assayed, regardless of the actual experimental outcome. This is in contrast to diversity methods that seek to produce a unique pattern of pharmacophore occurrences in each molecule.

Given a design space, the algorithm seeks to optimize decoding as many pharmacophores as possible, with the smoothest distribution across the size of pharmacophore classes. A pharmacophore class refers to the subset of pharmacophores that all have the same code or pattern. Note that the optimum solution is a set of compounds that enables decoding each individual pharmacophore. However, this may not be possible due either to the source pool, bit correlation or to limited size of selection.

The cost function for an unconstrained optimization in terms of molecule selection is the entropy of the class distribution. The entropy is given by

$$H = -\sum_{i=1}^{C} \frac{|c_i|}{f} \ln \frac{|c_i|}{f} \quad (1)$$

where H is the entropy of the feature classes, C the number of distinct classes, f the number of features in the design space and $|c_i|$ is the size of class i. During the course of the optimization, molecules are selected, such as to maximize H.

3. Results

To assess the performance of the coupling of structure-based methods and informative library design described in the preceding section, retrospective design and analysis was performed using legacy data. In particular, up to four rounds of simulated compound selection and screening were carried out using a variety of compound selection or experimental design methods, in all cases a total of at least 1000 compounds were selected for "screening." Each protocol was evaluated in terms of the quality of the designs measured in two ways. The first comparison is the number of templates on which active compounds are discovered, protocols that find more templates with active compounds are considered better. This measures how each method performs in terms of finding alternate chemical classes. Finding these classes is important because they may be needed as backup series for development. Second, methods are compared in terms of whether they lead to a refined model, as measured by increased library enrichment. Methods that yield a predictive model for ligand activity (e.g. a set of pharmacophores correlated to activity) are desirable.

3.1. Dataset

A dataset derived from a lead discovery project for CDK2 [21,22], a cell cycle regulatory protein chosen as an anticancer target, was selected for evaluating the methods. This dataset provides a source pool of approximately 3800 compounds synthesized in a targeted virtual library with CDK2 inhibitor activity determined. There are 22 distinct molecular scaffolds represented. Active compounds are found on

[2] For this study, a maximum of 100 conformers per molecule were generated using an in-house tool called conformational analysis (CONAN). Briefly, CONAN is similar to a deterministic search algorithm in that the final list of low-energy confomations lies on predefined torsion grid. It is distinct because it is a fragment-based algorithm designed to take advantage of the nature of combinatorial libraries. Each molecule is decomposed into overlapping fragments and the conformations of these fragments are analyzed separately and stored in a database, permitting their reuse. The conformational models of the fragments are used to construct conformations for the entire molecule through the use intersection of the overlapping substructures. This intersection process quickly generates a complete description of the low-energy space of a molecule.

15 of these scaffolds. The most active compound has an IC_{50} of approximately 100 nM, molecules having an IC_{50} <500 nM were identified on two scaffolds. Compounds with an IC_{50} lower than 25 μM or percent inhibition >50% (if no IC_{50} was available) were considered active. These cut-offs yield a total of 161 active molecules. Twenty-three X-ray crystal structures were available from the protein data bank (PDB) and in-house crystallography [17–20]. Structures were used as available from the PDB or the crystallography group, except they were aligned to the coordinate frame of the ATP-bound structure (accession code 1hck) and ligand coordinates were removed. That is, no minimization or additional refinement was performed for any structure.

3.2. Site map and signature generation

A site map and signature were generated for each CDK2 co-crystal structure using the protocol defined in Section 2. The site map for the ATP structure is shown in Fig. 2 and was generated in the absence of the ligand. The number of site map points for each X-ray structure varies between 36 and 86, yielding signatures ranging between approximately 31,000 and 199,000 bits in length. The union signature, approximately 967,000 bits, is the sum of all 23 signatures.

3.3. Informative design experiment

Iterative informative design was applied to select molecules that would test systematically and efficiently pharmacophores in the structure-derived design space. In the first round of design, 500 compounds were selected for maximal informational entropy in the structure-derived designed space (Section 2) from the source pool and "screened" by looking up their activity. To assess the effect of accounting for induced fit, the first round selection was performed using design spaces derived from one (PDB accession code 1hck, an ATP-bound structure) or all 23 structures. Significant side chain reorganization occurs in the CDK2 active site depending on the size and nature of the ligand, Fig. 3 illustrates an example. The composite design space yielded slightly better results than the single structure design space in terms of number of active scaffolds identified (10 versus 9). Therefore, the remaining rounds of design were carried out in the composite space. Pharmacophores present in more than 10 inactive compounds and no active compounds were removed from the design space. For the next round, 250 compounds were selected from the remaining source pool, pharmacophores were removed if present in more than five inactive and no active compounds. In round 3, 125 compounds were selected and assayed, pharmacophores were removed according to a cut-off of three inactives, zero actives. In the final round 125 compounds were selected from those remaining and assayed. Fig. 4 compares the number of active scaffolds selected in round 1 with the total number selected after four rounds. The average two-dimensional daylight pairwise Tanimoto similarity between all selected active compounds is 0.27. In sum, 1000 compounds were selected and screened, of these 74 were active representing

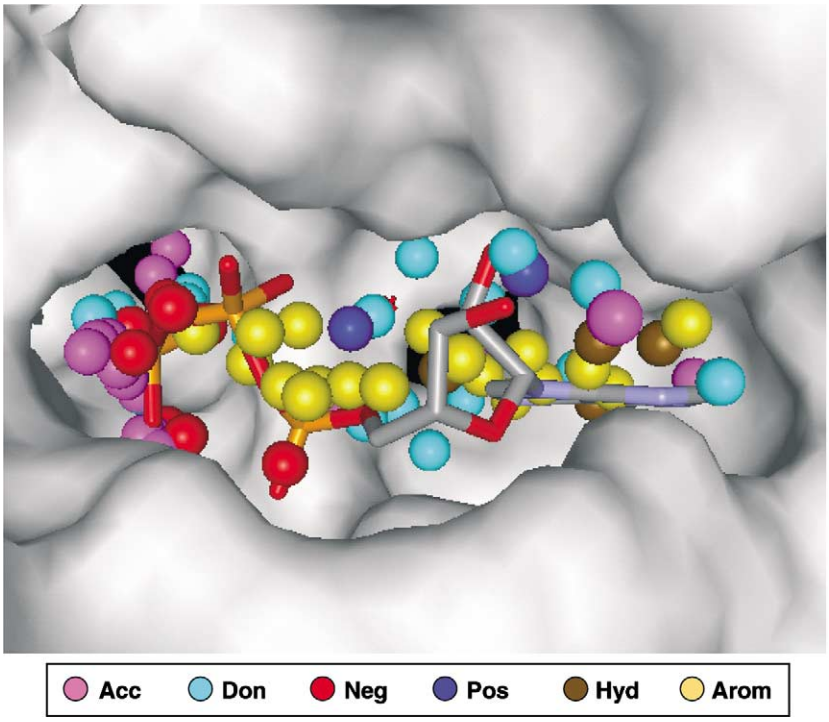

Fig. 2. Site map calculated with DPCSiteMap for the ATP-bound structure of CDK2. The ligand is removed in the site map generation but shown here to allow for comparison of the site map feature positions with the ligand orientation.

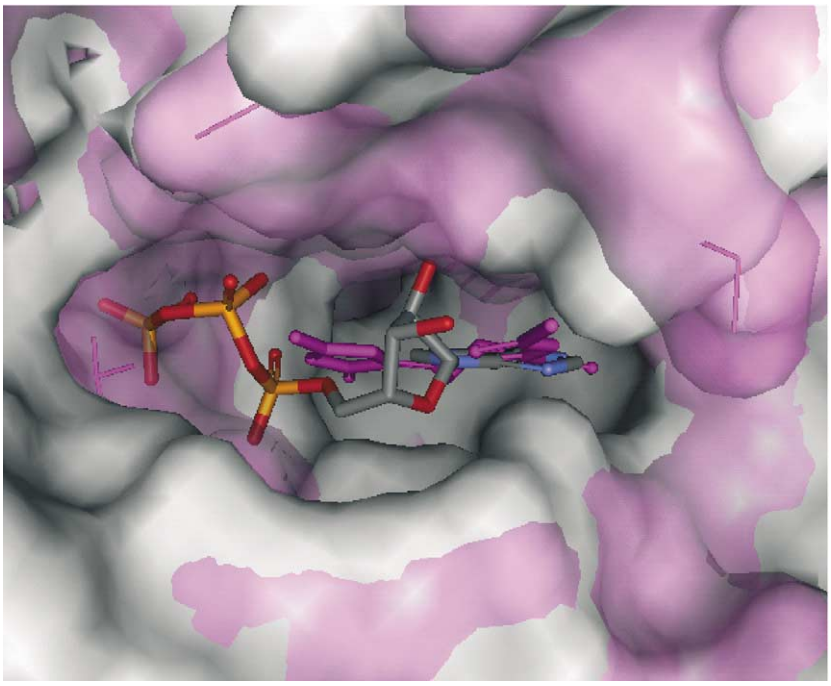

Fig. 3. Aligned co-crystal X-ray structures for CDK2 with ATP and hymenialdisine (HMD) bound. The ATP structure is shown in the white surface with ATP colored by atom type. The HMD structure is shown in purple for both the protein surface and the ligand.

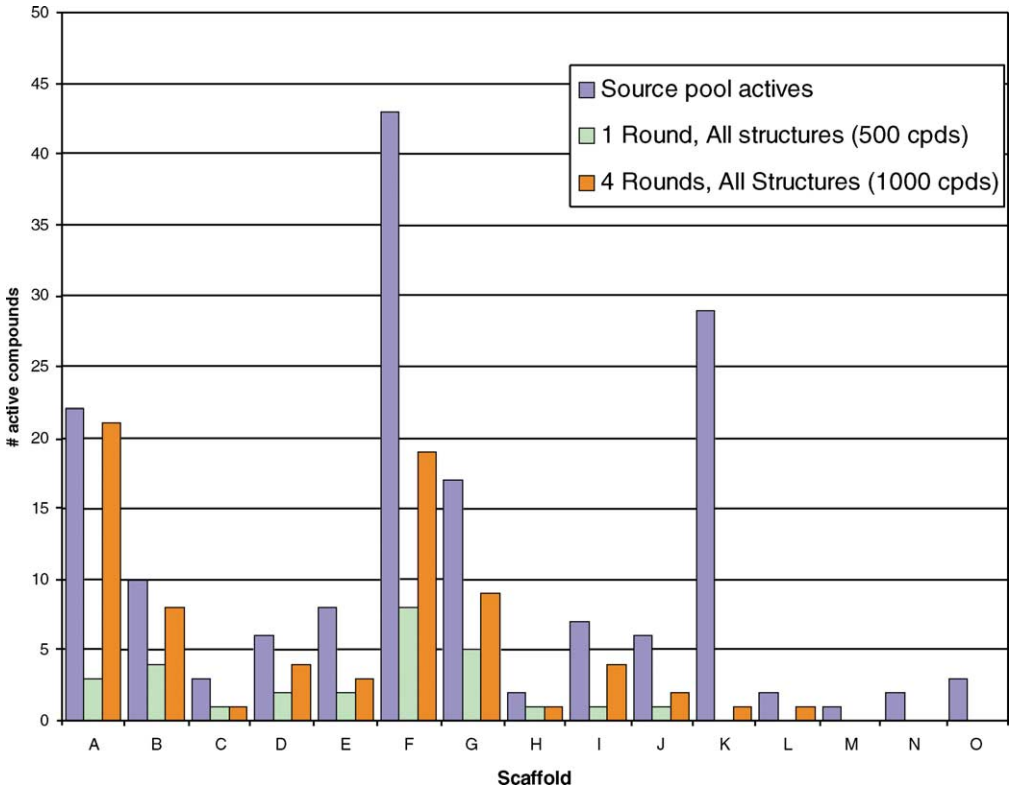

Fig. 4. Library design results. The 15 active scaffolds are labeled alphabetically. The blue bars illustrate the distribution of actives in the source pool. The green bars show the actives selected after one round of library design and the orange bars show the final results after four rounds.

Table 1
Library design experiment summary

		Number of compounds selected	Design space (thousands of bits) (K)	Number of scaffolds sampled	Number of scaffolds with active compounds	Number of actives selected	Enrichment
Source pool		N/A	N/A	22	15	161	N/A
Informative design							
Single structure	1HCK	500	171	20	9	27	
	1CKP	500	94	19	9	24	1.3
Informative design							
Composite space (23 structures)	Round 1	500	970	20	10	28	1.3
	Round 2	250	825	19	7	16	1.6
	Round 3	125	757	16	5	16	3.3
	Round 4	125	717	19	9	14	3.2
	Rounds 1 and 2	750	N/A	21	11	44	1.4
	Rounds 1–3	875	N/A	21	11	60	1.6
	Rounds 1–4	1000	N/A	21	12	74	1.7
Comparison studies	DockIt/PLP	1000	N/A	19	10	48	1.1
	2D	1003	N/A	N/A	7	46	1.1

12 scaffolds. In other words, in screening approximately a quarter of the source pool nearly half of the actives (enrichment for all selected compounds = 1.7) and 80% of the active scaffolds were found. Note, however, that the enrichments of libraries in rounds three and four are >3.0, indicating that the method converges to a model that correlates with activity. Table 1 summarizes the details of each round of experiments.

3.4. Comparison studies

The structure-derived pharmacophore/iterative design method presented here is compared to other accepted compound selection strategies: docking and two-dimensional similarity. Fig. 5 illustrates the results. In the docking comparison, the virtual library was docked to a single structure (1hck) of the target protein using the program DockIt [33] and compounds that have docked conformations with the best PLP score were selected [34]. It has been shown that docking against multiple protein structures and or using consensus scoring can improve hit rates, however these approaches would involve much greater preparation and computational effort and were beyond the scope of this study. Because the docking method and scoring function do not utilize screening data to refine their model, iterative design is not applicable. Therefore, 1000 compounds were selected in one pass. The compounds selected by DockIt/PLP represent 66% of the active scaffolds with an enrichment of

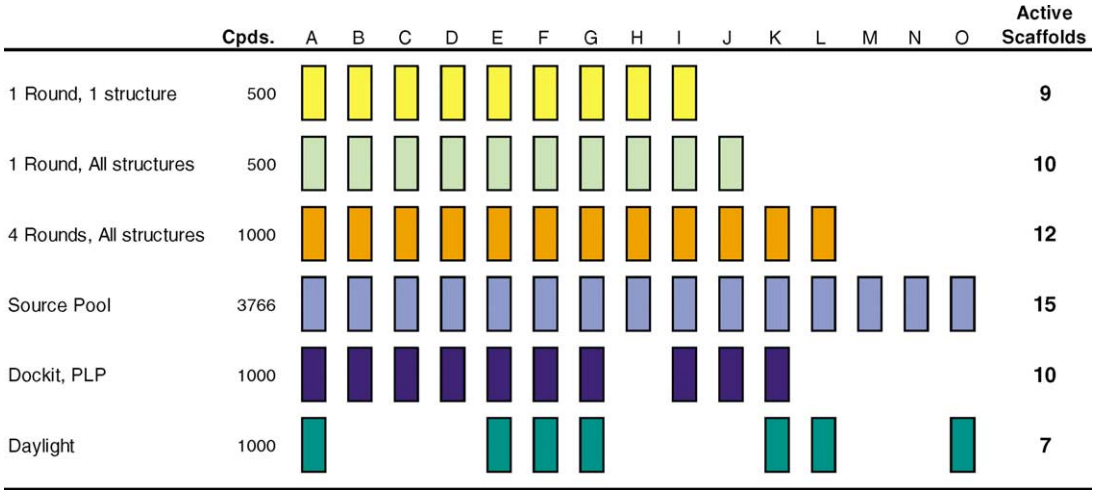

Fig. 5. Comparison of active scaffolds discovered with various informative design experiments, docking and two-dimensional similarity calculations. The 15 active scaffolds in the source pool are labeled alphabetically (A–O) and are illustrated with blue boxes. The number of compounds selected with each experiment is also shown.

1.1. For two-dimensional similarity, 170 compounds with the highest daylight Tanimoto similarity (i.e. nearest neighbors) to 19 of the crystallographic ligands were selected; after removing duplicates this yielded 1003 compounds picked in a single pass [35]. Compounds chosen based on two-dimensional similarity to the crystallographic ligands included 46 actives on seven scaffolds, with an enrichment of 1.1.

4. Discussion

The experiments presented above illustrate several points. First, structure-derived pharmacophores define a reasonable space for designing or selecting compounds for a particular target. Next, sampling this space using iterative informative design is an advantageous method for selecting compounds for synthesis and screening in comparison to docking or two-dimensional similarity searching (Fig. 5). In part this is because it provides a protocol of incorporating screening data, thereby refining the target specific design space based on experimental feedback. Third, considering induced fit, or the protein's conformational profile in response to a variety of ligands yields a more complete picture of the pharmacophore space that is compatible with a given target. While this can expand dramatically the number of pharmacophores to be considered, the larger design space appears not to be a liability in the experiments outlined above. This maybe attributed, in part, to the protocol for reducing the design space—which is to remove from consideration only pharmacophores that are sampled exclusively and relatively deeply by inactive compounds. The results show that by screening 875 compounds (the first three rounds) the protocol reduces the design space by more than 25%; that is, it rules out more than a quarter of the structure-derived pharmacophores. This is appropriate because the protein structure(s) define the pharmacophores that should be tested but does not overly limit the space.

Fig. 5 illustrates how the structure-derived pharmacophore space definition coupled with informative design approach performs in comparison to docking or similarity selection on one of the major comparison criteria, acquiring or discovering compounds on a large number of scaffolds. The figure shows that the method outlined above finds 50% more active scaffolds than does two-dimensional similarity selection and 20% more than docking [3]. Moreover, cumulative enrichment for the informative design approach is 1.7 (enrichment in the final (round 4) selection is 3.2) in comparison to a 1.1 for both docking and similarity. This compares favorably to random selection of compounds, which would yield enrichment of 1.0 by definition. In an additional comparison to random, repeated random selection experiments using varying random number seeds yield between 10 and 14 active scaffolds (average = 11) whereas, informative design assures discovery of the largest number of active scaffolds. While informative design does not select explicitly molecules for predicted activity, the rising enrichment in the successive rounds (Table 1) implies the method is ruling out parts of space proven to be uncorrelated with activity and focusing on sampling active pharmacophore space.

[3] While the docking calculations presented here do not identify any scaffolds not already picked by informative design the two-dimensional daylight method does identify one additional scaffold. This suggests that improved results be achieved by combined application of these methods.

5. Summary

The present work shows protein structural information successfully integrated with informative library design. The design space is defined with the relevant pharmacophores from protein structural data with the ligands removed (to provide an unbiased approach). Informative design chooses compounds which sample the design space pharmacophores with unique patterns. The combination of informative and structure-based design results in the selection of compounds from many scaffolds but also allows for model refinement. Creating a composite design space from all available X-ray data and using an iterative approach to library design selects the most active scaffolds among other common approaches (docking and two-dimensional similarity searching). During multiple iterations the method is able to focus in on the pharmacophores that are most relevant for activity, as demonstrated by the increase in the number of scaffolds selected and the higher enrichment in the later rounds of design. This technique may be most useful in the early stages of a drug discovery effort when the primary focus is on rapidly identifying multiple lead series.

Acknowledgements

The authors would like to thank the following for providing the cyclin-dependent kinase data, Angelo Castellino, Soan Cheng, Joe Cohen, Klaus Gubernator, Kelly Jenkins, Dan Kassel, Michele Ramirez-Weinhouse, Dan Rogers, Robyn Rourick, Jayashree Srinivasan, Melissa Wayland, Ron Xu. In-house protein crystallography was performed by Jodi Muckelbauer, Anzhi Wei and Chong-Hwan Chang. In addition, the authors would like to thank Dennis Underwood for helpful discussions.

References

[1] T. Kennedy, Managing the drug discovery process, Drug Discovery Today 2 (1997) 436–444.
[2] GRID, Molecular Discovery Limited, London, UK.
[3] J.E.J. Mills, P.M. Deam, Three-dimensional hydrogen-bond geometry and probability information from a crystal survey, J. Comput.-Aided Mol. Des. 10 (1996) 607–622.
[4] E. Evensen, D. Joseph-McCarthy, M. Karplus, MCSS version 2.1, 1997, Harvard University, Cambridge.

[5] A. Miranker, M. Karplus, Functionality maps of binding sites: a multiple copy simultaneous search method, Proteins 11 (1991) 29–34.
[6] D. Joseph-McCarthy, B.E. Thomas III, J.C. Alvarez, Pharmacophore-based molecular docking, in: Proceedings of the 221st ACS National Meeting on Book of Abstracts, San Diego, 2001, CINF-041.
[7] B.E. Thomas, D. Joseph-McCarthy, J.C. Alvarez, Pharmacophore-based molecular docking. Pharmacophore perception, development and use in drug design (Iul biotechnology series, 2) in: O.F. Guner (Ed.), International University Line, La Jolla, 2000, pp. 353–367.
[8] A.K. Ghose, V.N. Viswanadhan, J.J. Wendoloski, Adapting structure-based drug design in the paradigm of combinatorial chemistry and high-throughput screening: an overview and new examples with important caveats for newcomers to combinatorial library design using pharmacophore models or multiple copy simultaneous search fragments, ACS Symp. Ser. 719 (1999) 226–238.
[9] C.M. Murrary, S.J. Cato, Design of libraries to explore receptor sites, J. Chem. Inf. Comput. Sci. 39 (1999) 46–50.
[10] J.S. Mason, D.L. Cheney, Library design and virtual screening using multiple 4-point pharmacophore fingerprints, Pac. Symp. Biocomput. 2000, pp. 576–587.
[11] J.S. Mason, I. Morize, I.R. Menard, D.L. Cheney, C. Hulme, R.F. Labaudiniere, New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures, J. Med. Chem. 42 (1999) 3251–3264.
[12] H.A. Carlson, K.M. Masukawa, K. Rubins, F.D. Bushman, W.L. Jorgensen, R.D. Lins, J.M. Briggs, J.A. McCammon, Developing a dynamic pharmacophore model for HIV-1 integrase, J. Med. Chem. 43 (2000) 2100–2114.
[13] H.-J. Boehm, M. Boehringer, D. Bur, H. Gmuender, W. Huber, W. Klaus, D. Kostrewa, H. Kuehne, T. Luebbers, N. Meunier-Keller, F. Mueller, Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening, J. Med. Chem. 43 (2000) 2664–2674.
[14] J.S. Mason, D.L. Cheney, Ligand-receptor 3-D similarity studies using multiple 4-point pharmacophores, Pac. Symp. Biocomput. (1999) 456–467.
[15] S.L. Teig, Informative libraries are more useful than diverse ones, J. Biomol. Screening 3 (1998) 85–88.
[16] D.A. Barnum, J. Greene, S. Teig, The design of informative libraries, in: Proceedings of the 215th National Meeting of the American Chemical Society, Dallas, TX, 1998.
[17] U. Schulze-Gahmen, H.L. De Bondt, S.-H. Kim, High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: bound waters and natural ligand as guides for inhibitor design, J. Med. Chem. 39 (1996) 4540–4546.
[18] N.S. Gray, L. Wodicka, A.-M.W.H. Thunnissen, T.C. Norman, S. Kwon, F.H. Espinoza, D.O. Morgan, G. Barnes, S. LeClerc, L. Meijer, S.-H. Kim, D.J. Lockhart, P.G. Schultz, Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors, Science 281 (1998) 533–538.
[19] L. Meijer, A.M.W.H. Thunnissen, A.W. White, M. Garnier, M. Nikolic, L.H. Tsai, J. Walter, K.E. Cleverley, P.C. Salinas, Y.Z. Wu, J. Biernat, E.M. Mandelkow, S.H. Kim, G.R. Pettit, Inhibition of cyclin-dependent kinases, GSK-3.beta. and CK1 by hymenialdisine, a marine sponge constituent, Chem. Biol. 7 (2000) 51–63.
[20] A.A. Russo, P.D. Jeffrey, A.K. Patten, J. Massague, N.P. Pavletich, Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex, Nature 382 (1996) 325–331.
[21] L. Robinson, Recent applications of solution phase parallel synthesis to kinas inhibitor libraries, in: Proceedings of the San Diego Combinatorial Chemistry Symposium, San Diego, CA, 2000.
[22] E. Saiah, E. Bradley, A. Castellino, J. Cohen, K. Gubernator, K. Jenkins, M. Ramirez-Weinhouse, L. Robinson, D. Rogers, J. Srinivasan, M. Wayland, R. Xu, in preparation.
[23] G.P. Brady Jr., P.F.W. Stouten, Fast prediction and visualization of protein binding pockets with PASS, J. Comput.-Aided Mol. Des. 14 (2000) 383–401.
[24] P. Dauber-Osguthorpe, V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest, A.T. Hagler, Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydrofolate reductase-trimethoprim, a drug-receptor system, Proteins: Struct.-Funct.-Genet. 4 (1988) 31–47.
[25] J.A. Ippolito, R.S. Alexander, D.W. Christianson, Hydrogen bond stereochemistry in protein structure and function, J. Mol. Biol. 215 (1990) 457–471.
[26] G.P. Brady Jr., Prediction of ligand-binding modes via energy-based genetic algorithm docking, in: Proceedings of the 218th ACS National Meeting on Book of Abstracts, New Orleans 1999, COMP-067.
[27] M.J. McGregor, S.M. Muskal, Pharmacophore fingerprinting. 1. Application to QSAR and focused library design, J. Chem. Inf. Comput. Sci. 39 (1999) 569–574.
[28] J.H. Van Drie, R.A. Nugent, Addressing the challenges of combinatorial chemistry: 3D databases, pharmacophore recognition and beyond, SAR QSAR Env. Res. 9 (1998) 1–21.
[29] A. Smellie, R. Stanton, R. Henne, S. Teig, J. Comp. Chem., Submitted.
[30] A. Smellie, S. Teig, Conformational analysis by intersection: ring conformations, in: Proceedings of the 217th National Meeting of the American Chemical Society, Anaheim, CA 1999.
[31] S.L. Teig, A.S. Smellie, CombiChem Inc., Method and apparatus for conformationally analyzing molecular fragments, W09859306, 1998.
[32] J.L. Miller, in preparation.
[33] DockIt, 2000, Metaphorics, Santa Fe, NM.
[34] D.K. Gehlhaar, G.M. Verkhivker, P.A. Rejto, C.J. Sherman, D.B. Fogel, L.J. Fogel, S.T. Freer, Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming, Chem. Biol. 2 (1995) 317–324.
[35] Daylight, 1995, Daylight Chemical Information Systems, Santa Fe, NM.