Chemoproteomics as a basis for post-genomic drug discovery

Paul Beroza, Hugo O. Villar, Michael M. Wick and Graeme R. Martin

The large number of small organic compounds now available for drug-lead screening has led to numerous methods for classifying molecular similarity and diversity, the aim being to restore a balance between the quantity and drug-like quality of compounds in small-molecule libraries. Whereas structural and physicochemical attributes continue to be emphasized in compound selection for drug-lead screening, chemoproteomics – the use of biological information to guide chemistry – offers a highly efficient alternative to small-molecule characterization that can accelerate drug discovery in the post-genomic era.

Paul Beroza*
Hugo O. Villar
Michael M. Wick
and Graeme R. Martin
Telik, 750 Gateway Blvd
South San Francisco
CA 94080, USA
*e-mail: pberoza@telik.com

▼ Historically, human medicines have been derived from three main sources. First, there are plant and animal products (e.g. aspirin, morphine and hirudin), which have innate biological significance as they are part of the chemical repertoire of other species. Second, there are derivatives of human endogenous ligands that, because of their inherent selectivity and tolerability, serve as good starting points for medicinal chemistry. The histamine H_2 receptor antagonists (e.g. cimetidine and ranitidine) and the more recently developed serotonergic 5-HT_{1B} receptor agonists (e.g. sumatriptan and zolmitriptan) testify to the success of this approach. Third, there are the synthetic and semi-synthetic chemicals, exemplified by the phosphodiesterase inhibitor sildenafil, which are used to probe biological systems and possess properties that confer therapeutic benefit. With the advent of molecular biology, and the accompanying development of combinatorial chemistry and HTS technologies, this last source of drugs has become dominant. Consequently, strategies aimed at industrialization of the drug discovery process have been implemented in an attempt to increase the number of new medicines without increasing the associated costs of discovery and development.

The path to small-molecule drugs

The current trend for high volume chemical synthesis and biological screening must overcome the considerable logistic difficulties imposed by the large number of compounds that could be synthesized. An upper bound of 10^{100} has been posited [1], and an estimate restricted to small molecules (<30 non-hydrogen atoms) yields a lower bound of 10^{60} [2]. Although filters based on physicochemical properties (such as those imposed by Lipinski's 'rule-of-five' [3]) can reduce the number of potential compounds by several orders of magnitude, the remaining number of molecules is still far too large for even the most modern synthesis and screening technologies [4].

An entirely random process for drug-lead identification cannot be efficient. Computational methods for molecular classification and design address this problem by selecting subsets of molecules for biological assay on the basis of their similarity to, and diversity from, other molecules. Often, one molecule is chosen to represent a family of molecules that are deemed to be similar, and its biological activity is assumed to represent that of the group [5,6]. Prioritization of molecules for synthesis is also often determined by computational assessment of chemical diversity [7,8]. Consequently, molecular classification methods based on similarity and diversity have a key role in drug-lead discovery.

Traditionally, classification methods have focused on physicochemical and structural features of molecules. Physicochemical descriptors evolved from work on quantitative structure–activity relationships (QSAR) and have proven especially useful during lead optimization efforts concentrated on a single chemical scaffold [9]. Molecular descriptors more closely tied to chemical structure have been used to assess diversity and similarity

1359-6446/02/$ – see front matter ©2002 Elsevier Science Ltd. All rights reserved. PII: S1359-6446(02)02371-1

between molecules for the design of chemical libraries [10–13], and for building computational models for biological activity [14–17]. Although there is a great variety of such descriptors, the relative merits of which continue to be debated, they have in common the use of physicochemical properties and/or chemical structure to evaluate molecules.

Similar classification methods have also been applied to known drugs to identify regions in chemical space that might be relevant to medicinal chemistry [18,19]. There is, however, a risk of circularity in these approaches as the features regarded as important for drug-like molecules are derived from compounds that act on the known therapeutic target classes. In the post-genomic era of drug discovery, this chemical heritage could hinder the discovery of drugs for new types of targets because novel pharmacophores will probably emerge as the repertoire of target classes expands.

Despite the widespread adoption of structure as a means to segregate and discriminate different molecular types, there are limitations to this approach. These are often evident in lead optimization efforts, when small changes in compound structure are accompanied by significant changes in biological activity. For example, an isosteric replacement of a hydrogen atom in a naphthyl-fused diazepine compound causes a 20-fold change in its binding affinity to the benzodiazepine receptor [20]. In an extreme case, substitution of an atom with a radioactive isotope can alter the bioactivity of a molecule [21]. In addition, many structural descriptors do not recognize chirality, although changes in chirality can result in different biological properties, as in the case of the enantiomeric pair dextrorphan and levorphanol [22–24], and as probably most widely documented in the case of thalidomide [25]. Although there are many examples in which small structural changes cause drastic changes in biological properties, this is generally the exception rather than the rule. A structural descriptor that is sensitive to such changes in chemical structure is more likely to be sensitive to the minor (and much more common) changes that do not affect bioactivity. Indeed, the idea that minor changes in structure should not significantly alter bioactivity is an important premise in structure-based similarity methods [26].

Minor structural changes do not usually alter biological activity; however, in some cases, they can alter it substantially. This relates to the 'energy landscape' of protein–ligand interactions; that is, the relationship between structural changes and changes in interaction energy. To account for this observation, the energy landscape must comprise surfaces that are smooth (representing small structural changes that have little effect) and surfaces that are rough (for small changes with drastic effects). This empirical observation has been noted in computational ligand–protein docking calculations [27] and is reminiscent of the energy landscape of protein folding, which has also been shown to have a complex energy profile [28]. As a result, analyses of biological activity based on structural information alone are unlikely to be sufficient to capture differences in the biological behavior of structurally similar ligands.

Biologically relevant chemical classification schemes

Physiologists and pharmacologists have used ligand–target binding information to classify biological targets and the ligands that interact with them for almost a century. However, extending this concept to a more global one of measuring ligand–protein interactions as the basis for chemical classification has not been previously attempted.

To explore this approach more comprehensively, we generated a large data collection, comprising affinity estimates of small molecules for different proteins [29]. To retain as much biologically relevant information as possible, all measurable interaction strengths were recorded; therefore, even modest affinities, which would be disregarded in most pharmacological assays, were used to describe molecules. As such, these affinity values resemble elements in a molecular descriptor more closely than they resemble pharmacological assay results.

As the database of protein–ligand binding affinities grew, patterns emerged that proved useful for drug discovery. An early analysis of the database suggested the existence of statistical relationships among affinity values, even when the proteins were unrelated by either structure or function [29,30]. In certain cases, these relationships adopted a linear form [Eqn 1], in which the affinity of a compound for protein T could be expressed as the weighted sum of its affinities for N other unrelated proteins:

$$pIC_{50}[T] = C_0 + \sum_{k=1}^{N} C_k pIC_{50}[k] \qquad \text{[Eqn 1]}$$

where affinity is expressed as pIC_{50} (the negative logarithm of IC_{50}), and the weighting factors (C_k) and independent term (C_0) are determined using linear least squares fitting.

One example of how information contained in two proteins can enable the prediction of information contained in a third, unrelated protein, emerged from the study of aldehyde dehydrogenase (ADH), snake venom phosphodiesterase I (PDE) and glutathione S-transferase rat 8–8 (R88). In this case, affinity values of 120 small molecules for ADH were related to the combined affinity of these chemicals for PDE and R88 [29], as shown in [Eqn 2]:

$$pIC_{50}[ADH] = 0.4 + 0.6\ pIC_{50}[PDE] + 0.2\ pIC_{50}[R88]$$
[Eqn 2]

with a correlation coefficient of 0.86 and a standard deviation of 0.50.

Similar relationships have been found in other datasets. For example, the interactions of human serum albumin (HSA) with 1500 randomly selected small molecules can be represented by a linear combination [Eqn 3] of the binding data for three other unrelated proteins: galactoside dehydrogenase (GAL), uricase (URI) and glutathione S-transferase A1–1 (A11):

$$pIC_{50}[HSA] = 0.61\ pIC_{50}[GAL] + 0.14\ pIC_{50}[A11] + 0.24\ pIC_{50}[URI]$$
[Eqn 3]

with a correlation coefficient of 0.88 and a standard deviation of 0.39. A scatter plot of fitted versus observed binding affinity for each compound is shown in Fig. 1. Models were cross-validated and found to be useful in predicting the affinities of much larger collections [29,30]. We have subsequently found numerous relationships of this type for a wide range of animal proteins, leading us to propose that information about small-molecule–protein interactions for one protein can be predicted by a linear combination of the data obtained for two or more unrelated proteins [29,30].

This important observation implies the presence of redundancies in the accumulated data. In assembling a database of information intended to have the widest use in small-molecule characterization and classification, it is advantageous to minimize these redundancies as much as possible. To achieve this, binding affinities have been determined for a diverse collection of small molecules with as many as 500 proteins, and a subset of proteins was identified that provided the most non-overlapping data with regard to their affinities for the set of small molecules. To identify the subset that retains most of the information contained in the overall dataset, procedures based on the Gram–Schmidt orthogonalization technique were used [30,31]. This protein panel, termed the reference panel, can adequately represent all the data collected. Thus, molecular descriptors constructed from this reference panel of proteins effectively represent the binding affinities for the entire set of proteins studied to date. Moreover, given the wide diversity of proteins that have been studied in this way, it is expected that these descriptors will also represent binding affinities for numerous untested proteins. In this respect, the reference panel represents a surrogate of the human proteome.

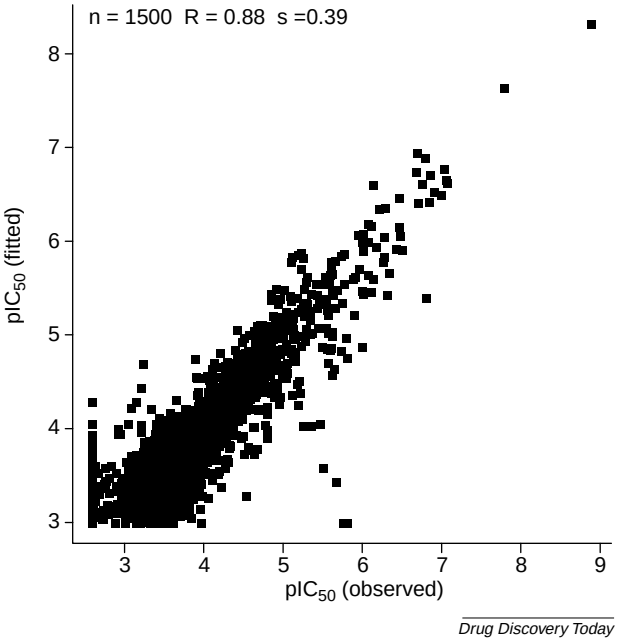

Figure 1. Scatter plot of fitted- versus observed binding affinities to human serum albumin (HSA) determined for 1500 randomly selected compounds. Affinity estimates were measured as IC_{50} values using fluorescence polarization to assess competitive displacement of a fluorophore from HSA, and compared to values calculated using the model for HSA binding presented in Eqn 3, resulting in a correlation coefficient of 0.88 and a standard deviation of 0.39.

The reference panel of proteins assembled using the methods described previously is not unique and could be substituted by another set of proteins within the collection. An early version of the reference panel that was constructed using similar procedures has been described elsewhere [29]. The main criterion for the panel is that it provides enough information to represent the entire matrix of small-molecule information. In other words, the process of reducing the matrix of data is more important than the final panel of reference proteins.

Building and analyzing an affinity fingerprint database

Having identified a suitable reference panel of proteins, it is possible to create a database of descriptors for collections of small molecules. Each molecule in the collection is assayed against each of the proteins in the reference panel, and the set of each compound's binding affinities is known as its affinity fingerprint. These fingerprints are unique for each chemical, and collectively constitute an affinity fingerprint database that is used for drug-lead data mining.

Affinity fingerprints can also be used to guide the expansion of a small-molecule compound library. As new compounds

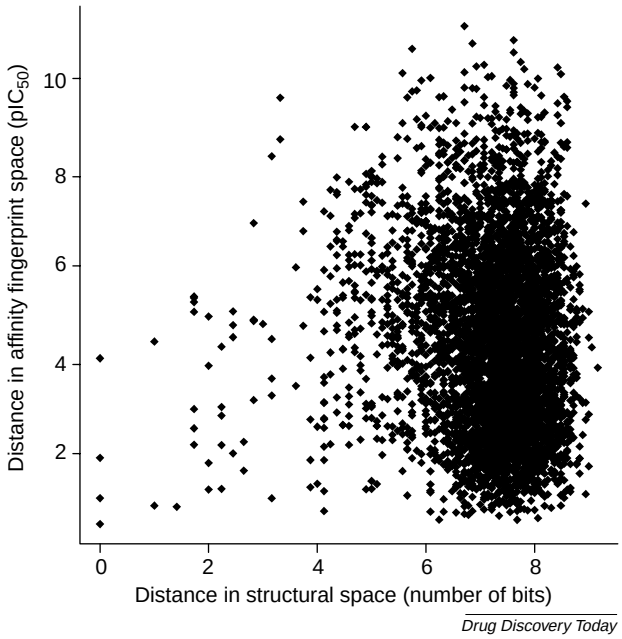

Figure 2. Scatter plot of the pairwise Euclidean distances between randomly selected pairs of compounds using structural descriptors (*x*-axis) and affinity fingerprints (*y*-axis). The lack of correlation suggests that the affinity fingerprints are complementary to structural descriptors.

are synthesized or acquired, they are fingerprinted and only added to the library if new descriptor information is being contributed. The value of a potential new library candidate is determined by calculating the number of other compounds in its affinity fingerprint neighborhood. When the number of near-neighbors is low, it signifies that the compound occupies a sparsely populated region in affinity fingerprint space and is, therefore, relatively rare. By synthesizing or acquiring compounds that display fingerprints similar to such unique compounds, underpopulated regions in the database can be filled. As already mentioned, small changes in chemical structure do not assure small changes in affinity fingerprint. However, small structural changes usually result in small changes in biological activity, so this is a reasonable approach to enriching the compound library.

Affinity fingerprints in molecular similarity and diversity

The affinity fingerprint for each compound can be thought of as a vector that describes the location of the molecule in a high-dimensional descriptor space. With the database in place, it becomes possible to explore the nature of this space and the molecules that populate it.

An important question is whether affinity fingerprints provide information that is significantly different from descriptors derived using molecular structure. Although structural descriptors have enabled the identification of useful SARs, cases in which small structural changes result in drastic changes in bioactivity remain commonplace. A descriptor that correlates with such changes would add significant value to structural descriptors. An essential condition, then, is that a novel descriptor be uncorrelated with structural descriptors. Figure 2 suggests that affinity fingerprints can meet this condition. The plot compares the representation of 100 randomly chosen molecules using two descriptors: MACCS keys (MDL Information Systems; http://www.mdli.com), a structural descriptor that records the presence or absence of predetermined structural moieties in a molecule, and affinity fingerprints, a biological descriptor that records a molecule's binding affinities for Telik's (http://www.telik.com) panel of reference proteins. Euclidean distances between each pair of compounds are calculated using both representations, and are shown in a scatter plot in which the *x*-axis represents inter-compound distance described by structural descriptors and the *y*-axis represents the distance described by affinity fingerprints. Although the graph shows that structurally similar compounds tend to have similar affinity fingerprints, there is no discernible relationship between the two sets of distances for highly dissimilar pairs of compounds. Similar results were found when affinity fingerprints were compared with physicochemical properties calculated from molecular structure. This demonstrates that diversity measured by affinity fingerprints is different from diversity measured by structural descriptors. Nevertheless, the same mathematical methods that are used to assess chemical diversity based on structure [11,32,33] can be applied in the space of affinity fingerprints [30].

Other important questions in the analysis of a descriptor space concern where the known drugs reside in the space, and the nature of the spatial relationship among molecules of a single pharmaceutical class. Ideally, drugs that act on the same pharmacological target would have a close spatial relationship. This could be overall fingerprint similarity or, more likely, close proximity in a hypervolume or subspace. For example, in the HSA example in Eqn 3, compounds that bind well to serum albumin would cluster in the hypervolume associated with the three proteins GAL, A11 and URI. The coordinates associated with the other proteins in the panel would not be as relevant; therefore, the overall fingerprint similarity of compounds that cluster in the hypervolume would not necessarily be high.

In some cases, commonalities in the affinity fingerprints of compounds from the same pharmacological class can be simple to notice. For example, in the case of certain

G-protein-coupled receptor ligands, several different dopamine D_2 receptor antagonists bind with measurable affinity to proteins as diverse as D-amino oxidase, arginase and butyryl cholinesterase [34]. In most cases, however, trends in the affinity fingerprints of compounds that share the same pharmacological profile are more subtle, and the identification of similarities requires the use of statistical methods [33–35].

A second example of chemical classification using affinity fingerprints is provided by non-steroidal anti-inflammatory drugs (NSAIDs) that inhibit cyclooxygenase-1 (COX-1). A random selection of NSAIDs (Fig. 3) can be shown to share commonalities in their protein-binding preferences [36]. For instance, the compounds have affinity for HF-2 but do not interact with P1–1, both enzymes of the GST family [29]. The same NSAIDs display affinity for HSA in competitive binding assays using fluorescence polarization as a detection technique [29,30]. The affinities of the NSAIDs for the panel proteins are much weaker than their affinities for their therapeutic target (COX-1). However, the repeated pattern of modest binding affinities for a given subset of panel proteins constitutes a statistically robust model that can be used to reliably identify molecules with COX-1 inhibitor activity, and to differentiate them from molecules lacking this activity. The molecular similarity recognized by a pharmacological target is captured by affinity fingerprints.

Importantly, because the affinity fingerprints are not based on structural information, compounds identified as similar by affinity fingerprints could show significant structural dissimilarity. For example, the two NSAIDS ketoprofen and piroxicam (Fig. 3) appear structurally distinct. The distance between structural descriptors for the two confirms chemical intuition: use of the MACCS structural descriptors shows that the molecules have a Tanimoto distance (i.e. the number of bits in common in the binary signature, divided by the total number of bits that are set in either compound's signature) of 0.53, which is larger than the average inter-compound distance in our compound collection. In affinity fingerprint space, these compounds have a Euclidean distance that is half the average inter-compound distance in our collection. Thus, they are very similar in this space, whereas in structural space they are distinct. This observation demonstrates that affinity fingerprints are able to find functional similarity between compounds exhibiting significant structural diversity.

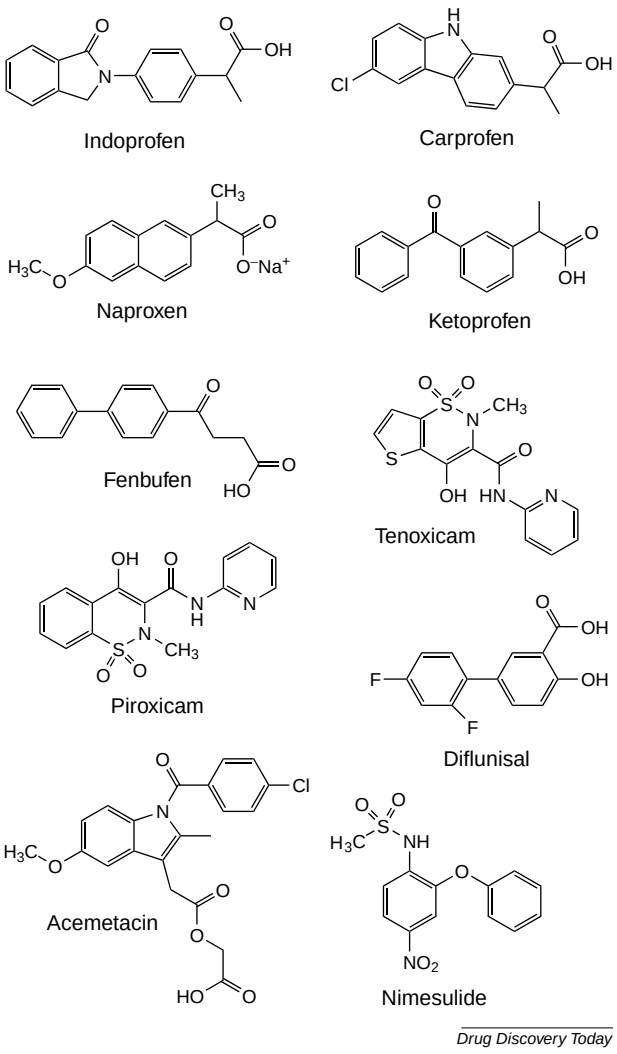

Figure 3. Chemical structures of non-steroidal anti-inflammatory drugs that have similar affinity fingerprints. The structural diversity of the molecules contrasts with their biological similarity, which is captured by the similarity in their affinity fingerprints.

Lead identification technology

Because affinity fingerprints are pharmacologically relevant descriptors and are complementary to structure-based descriptors, we use them as the basis for a chemoproteomics lead-identification platform: target-related affinity profiling (TRAP™). Using affinity fingerprints to guide compound selection for screening, bioactive small molecules with low micromolar affinity for the target can be found routinely after screening as few as 200 compounds. This platform has enabled extremely efficient lead discovery for a great diversity of therapeutic targets.

The process is iterative and begins with the assay of a 'training set' of ~70 compounds against a chosen target. These molecules are selected for their diversity in affinity fingerprint space [30] and represent the diversity of fingerprint patterns in the entire database. Thus, the initial screening set is constructed to maximize biological rather

Box 1. Chemical-omics explained

The term 'chemical genetics', first introduced by Schreiber and colleagues [a], describes the use of small molecules to selectively perturb gene function to provide insights into biological processes, in much the same way that the study of mutant alleles in genetics has provided similar information.

When conducted in a systematic way on families of related genes or gene products, this approach has been termed 'chemical genomics' or 'chemogenomics'. Small molecules with established activity at one member of the family are used to probe the biological function of other family members serving, at the same time, to guide the design of new, selective modulators of these putative targets. The term is now also used to describe this process on a genome-wide scale.

'Chemoproteomics' can be regarded as the application of chemogenomics specifically to protein targets. However, TRAP™ (target-related affinity profiling) technology offers a more explicit definition of the term, namely 'the use of biology to inform chemistry'. By determining the binding preferences of small molecules to a surrogate proteome, TRAP generates an empirical descriptor that relates chemistry to biology directly. A library of compounds annotated with biological information in this way enables rapid identification of diverse lead molecules unbiased by compound or target structural information.

Reference

a Schreiber, S.L. (1998) Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorg. Med. Chem.* 6, 1127–1152

than structural diversity. The training set is screened against the target at the highest concentration that the assay will tolerate, and any measurable activity or affinity is recorded. The initial activity demonstrated by the training set is usually weak, but it provides data about which elements of the fingerprints are associated with activity.

Fingerprint patterns associated with activity are identified using a variety of computational tools [30,33], and the database is mined for fingerprints that match these patterns. The compounds that possess these matching fingerprints are submitted for a second round of assays. The activity from this second set is usually better, but its value is largely in refining the 'draft' model for activity generated from the training set. The refined model from the combined assay results determines the third round of compound selection. Molecules selected in the third round typically have activity in the low micromolar range (<10 μM) and serve as the starting point for lead optimization.

TRAP classification of molecules combines functional diversity, pattern recognition and similarity analysis, to create an effective lead-discovery engine for diverse targets in a wide assortment of assays. Examples of the methods have been provided in detail for typical enzyme targets [30].

Discussion

As pharmaceutical drug discovery moves into the post-genomic era, the demand for lead compounds exceeds the ability of current technologies to identify them. One reason for this is that the design of new chemical entities relies heavily on the structural information that is represented in existing pharmaceutical agents, and these, in turn, have been developed against a limited repertoire of drug targets. The implication is that chemical and structural criteria currently used to classify compounds as 'drug-like' might not be entirely applicable to novel target classes that are emerging from a genome-wide approach to drug discovery. A more empirical view of small molecule classification, such as that imposed by chemogenomics and chemoproteomics (see Box 1), offers a possible solution to this dilemma.

Chemogenomics is most generally viewed as the process by which small molecules are used to gain an insight into the function of novel biological targets. Coined by Vertex Pharmaceuticals (http://www.vpharm.com) to describe an approach to identify selective modulators of different members of a single target family (kinases), several technology platforms have recently emerged to exploit this principle on a genome-wide scale [37,38]. Neogenesis (http://www.neogenesis.com) technology, ALIS (automated ligand identification system), uses putative protein targets as the bait for high-throughput multiplexed screening of small molecules. Other platforms, such as the one developed by Graffinity Pharmaceuticals AG (http://www.graffinity.com), use microarrays of 'tethered' chemicals, which serve as the bait for proteins of interest. TRAP differs from these technologies in several respects, most notably in its economy of scale and high efficiency of drug-lead discovery. The technology is most accurately described as a chemoproteomics technology because it uses protein-binding information to characterize molecules and subsequently uses these annotated molecules to probe biological systems. As a result, small-molecule leads are identified at biological targets in the absence of prior knowledge concerning chemical preferences of the target. Furthermore, this can be achieved in as few as 200 assays.

The ligand–protein interactions measured by TRAP to generate an affinity fingerprint collectively mimic those typically encountered by a drug acting on a macromolecular target. Hence, affinity fingerprints effectively serve as a set of coordinates that determine the location of chemicals in affinity fingerprint space. This highlights the major distinction between affinity fingerprints and more conventional

chemical descriptors. The latter sort molecules into related chemotypes, regardless of their 'biotype' (i.e. biological properties), whereas the affinity fingerprints of TRAP sort molecules according to their 'biotype' without particular reference to the chemotype. As with any empirical descriptor, a complete understanding of the physicochemical events that produce it presents a challenge. The linear nature of the surrogate receptor models suggests that each member of the protein panel captures unique types of surface interactions between small molecules and proteins. Because they are chosen to be orthogonal, these interactions can be weighted and summed independently, similar to how free energies of binding are separated into additive linear functions [39].

To be useful for lead identification, structural descriptors must be relevant to the physical and chemical preferences of the target. Ligand-based design methods, which use the characteristics of small molecules in the absence of information about how they interact with proteins, are at a distinct disadvantage in this respect. Recent work has addressed this problem by trying to identify which descriptors in a set are most relevant for receptor binding [32,40–42]. The advantage of affinity fingerprints is that they already contain protein-binding information and, therefore, are inherently relevant to the preferences of the target. As a result, the mining of the database for bioactive compounds can be initiated with much less information than that required by methods based on structural descriptors. Even a single weak hit provides an initial fingerprint. This then establishes coordinates for the bioactivity neighborhood from which to start the iterative search for ligands with an improved fit to the target. Another consequence of a descriptor consisting of protein-binding information is that it can be applied to either established or entirely novel targets with an equal probability of identifying bioactive lead molecules – a significant advantage with respect to existing approaches.

Other advantages are also inherent in drug-lead identification based on affinity fingerprints. As a relatively small number of assays (~200) is required to identify a bioactive molecule, there are clear benefits in time and costs associated with the development of high-throughput assay formats and the consumption of library compounds. Moreover, assay systems can be used that are more physiologically relevant and not amenable to HTS. TRAP also enables lead discovery to be guided by more robust criteria. Rather than registering interactions as 'hit' or 'miss', it is possible to determine a concentration–response relation, providing much more information about the nature of the target–ligand interaction. In addition, criteria that are typically not considered until later stages of discovery, such as selectivity, can be incorporated earlier in the lead-discovery process. In principle, any biological interaction between a small molecule and a protein could be modeled using affinity fingerprints, and we are currently investigating the use of affinity fingerprints for lead optimization, and for the prediction of toxicological effects, bioavailability, metabolism and interactions with other drugs.

Conclusion

As the number of new proteins that need to be evaluated as potential therapeutic targets grows, we run the risk of overwhelming traditional screening methodologies [43]. Our emphasis on methods that minimize the number of compounds that need to be screened offers a practical solution to problems arising in this target-intensive climate. Small-molecule leads identified using affinity fingerprints can serve as probes to validate the potential therapeutic value of a target, without the labor and expense associated with developing a reliable HTS assay and screening millions of compounds. From this perspective, technologies that bridge genomics, proteomics and chemistry are well-suited to post-genomic drug discovery.

References

1 Walters, W.P. *et al.* (1998) Virtual screening – an overview. *Drug Discov. Today* 3, 160–178
2 Bohacek, R.S. *et al.* (1996) The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* 16, 3–50
3 Lipinski, C.A. *et al.* (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26
4 Drews, J. (2000) Drug discovery today – and tomorrow. *Drug Discov. Today* 5, 2–4
5 Brown, R.D. and Martin, Y.C. (1996) Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sci.* 36, 572–584
6 Ferguson, A.M. *et al.* (1996) Designing chemical libraries for lead discovery. *J. Biomol. Screening* 1, 65–73
7 Van Drie, J.H. and Lajiness, M.S. (1998) Approaches to virtual library design. *Drug Discov. Today* 3, 274–284
8 Spellmeyer, D.C. and Grootenhuis, P.D.J. (1999) Recent developments in molecular diversity: computational approaches to combinatorial chemistry. *Annu. Rep. Med. Chem.* 34, 287–296
9 Kubinyi, H. (1995) The quantitative analysis of structure-activity relationships. In *Burger's Medicinal Chemistry and Drug Discovery* (Vol. 1) (Wolff, M.E., ed.), pp. 497–571, John Wiley and Sons
10 Blaney, J.M. and Martin, E.J. (1997) Computational approaches for combinatorial library design and molecular diversity analysis. *Curr. Opin. Chem. Biol.* 1, 54–59
11 Martin, Y.C. *et al.* (1998) Quantifying diversity. In *Chemistry and Molecular Diversity in Drug Discovery* (Gordon, E.M. and Kerwin, J.F.J., eds), pp. 369–385, Wiley-Liss
12 Willett, P. (2000) Chemoinformatics – similarity and diversity in chemical libraries. *Curr. Opin. Biotechnol.* 11, 85–88
13 Beroza, P. and Suto, M.J. (2000) Designing chiral libraries for drug discovery. *Drug Discov. Today* 5, 364–372
14 Cramer, R.D. *et al.* (1999) Prospective identification of biologically active structures by topomer shape similarity searching. *J. Med. Chem.* 42, 3919–3933

15 Bradley, E.K. *et al.* (2000) A rapid computational method for lead evolution: description and application to α(1)-adrenergic antagonists. *J. Med. Chem.* 43, 2770–2774
16 Bajorath, J. (2001) Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J. Chem. Inf. Comput. Sci.* 41, 233–245
17 Srinivasan, J. *et al.* (2002) Evaluation of a novel shape-based computational filter for lead evolution: application to thrombin inhibitors. *J. Med. Chem.* 45, 2494–2500
18 Ajay, A. *et al.* (1998) Can we learn to distinguish between 'drug-like' and 'nondrug-like' molecules? *J. Med. Chem.* 41, 3314–3324
19 Sadowski, J. and Kubinyi, H. (1998) A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* 41, 3325–3329
20 Zhang, W. *et al.* (1994) Synthesis of benzo-fused benzodiazepines employed as probes of the agonist pharmacophore of benzodiazepine receptors. *J. Med. Chem.* 37, 745–757
21 Grohmann, M. *et al.* (1986) Errors introduced by a tritium label in position 8 of catecholamines. *Naunyn Schmiedebergs Arch. Pharmacol.* 332, 34–42
22 Jaffe, J.H. and Martin, W.R. (1985) Opioid analgesics and antagonists. In *Goodman and Gilman's The Pharmacological Basis of Therapeutics* (Gilman, A.G. *et al.*, eds), pp. 491–531, Macmillan
23 Goldstein, A. *et al.* (1971) Stereospecific and nonspecific interactions of the morphine congener levorphanol in subcellular fractions of mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* 68, 1742–1747
24 Chou, Y.C. *et al.* (1999) Binding of dimemorfan to sigma-1 receptor and its anticonvulsant and locomotor effects in mice, compared with dextromethorphan and dextrorphan. *Brain Res.* 821, 516–519
25 Finnell, R.H. *et al.* (2002) Molecular basis of enviromentally induced birth defects. *Annu. Rev. Pharmacol. Toxicol.* 42, 181–208
26 Johnson, M. and Maggiora, G.M. (1990) *Concepts and Applications of Molecular Similarity*, Wiley
27 Rejto, P.A. and Verkhivker, G.M. (1996) Unraveling principles of lead discovery: from unfrustrated energy landscapes to novel molecular anchors. *Proc. Natl. Acad. Sci. U. S. A.* 93, 8945–8950
28 Onuchic, J.N. *et al.* (1997) Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48, 545–600
29 Kauvar, L.M. *et al.* (1995) Predicting ligand binding to proteins by affinity fingerprinting. *Chem. Biol.* 2, 107–118
30 Dixon, S.L. and Villar, H.O. (1998) Bioactive diversity and screening library selection via affinity fingerprinting. *J. Chem. Inf. Comput. Sci.* 38, 1192–1203
31 Topliss, J.G. and Edwards, R.P. (1979) Chance factors in studies of quantitative structure-activity relationships. *J. Med. Chem.* 22, 1238–1244
32 Brown, R.D. and Martin, Y.C. (1997) The information content of 2D and 3D structural descriptors relevant to ligand–receptor interactions. *J. Chem. Inf. Comput. Sci.* 37, 1–9
33 Dixon, S.L. and Villar, H.O. (1999) Investigation of classification methods for the prediction of activity in diverse chemical libraries. *J. Comput.-Aided Mol. Des.* 13, 533–545
34 Kauvar, L.M. *et al.* (1998) Protein affinity map of chemical space. *J. Chromatogr. B Biomed. Sci. Appl.* 715, 93–102
35 Willett, P. and Winterman, V.A. (1986) Comparison of some measures for the determination of intermolecular structural similarity. *Quant. Struct.-Act. Relat.* 5, 18–25
36 Kauvar, L.M. (1995) Affinity fingerprinting. *Biotechnology* 13, 965–966
37 Agrafiotis, D.K. *et al.* (2002) Combinatorial informatics in the post-genomics era. *Nat. Rev. Drug Discov.* 1, 337–346
38 Lenz, G.R. *et al.* (2000) Chemical ligands, genomics and drug discovery. *Drug Discov. Today* 5, 145–156
39 Bohm, H.J. (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* 7, 83–102
40 Matter, H. (1997) Selecting optimally diverse compounds from structure databases: a validation study of two-dimensional and three-dimensional molecular descriptors. *J. Med. Chem.* 40, 1219–1229
41 Pearlman, R.S. and Smith, K.M. (1999) Metric validation and the receptor-relevant subspace concept. *J. Chem. Inf. Comput. Sci.* 39, 28–35
42 Andrews, K.M. and Cramer, R.D. (2000) Toward general methods of targeted library design: topomer shape similarity searching with diverse structures as queries. *J. Med. Chem.* 43, 1723–1740
43 Woychik, R.P. *et al.* (1998) Functional genomics in the post-genome era. *Mutat. Res.* 400, 3–14

Don't miss *A Trends Guide to Imaging Technologies*

Free online July 5–August 5 2002:
just visit http://journals.bmn.com/supp

The convergence of imaging techniques with molecular biology, biochemistry and computing is a rapidly growing area. *A TRENDS Guide to Imaging Technologies* outlines the latest developments and applications within the life sciences. This collection features nine reviews that cover specific areas, such as molecular imaging and atomic scale imaging, and the latest developments in the application of imaging to neurosciences and in small animal models. These integrated approaches provide great promise for insights into biological processes and in the diagnosis of disease.