ELSEVIER

Journal of Chromatography B, 715 (1998) 93–102

JOURNAL OF CHROMATOGRAPHY B

Protein affinity map of chemical space[1]

Lawrence M. Kauvar*, Hugo O. Villar, J. Richard Sportsman, Deborah L. Higgins, Donald E. Schmidt, Jr

Terrapin Technologies, Inc., 750 Gateway Boulevard, So. San Francisco, CA 94080, USA

Abstract

Affinity fingerprinting is a quantitative method for mapping chemical space based on binding preferences of compounds for a reference panel of proteins. An effective reference panel of <20 proteins can be empirically selected which shows differential interaction with nearly all compounds. By using this map to iteratively sample the chemical space, identification of active ligands from a library of 30 000 candidate compounds has been accomplished for a wide spectrum of specific protein targets. In each case, <200 compounds were directly assayed against the target. Further, analysis of the fingerprint database suggests a strategy for effective selection of affinity chromatography ligands and scaffolds for combinatorial chemistry. With such a system, the large numbers of potential therapeutic targets emerging from genome research can be categorized according to ligand binding properties, complementing sequence based classification.

Keywords: Affinity fingerprinting; Chemical space; Proteins

1. Introduction

In the conventional use of robotic systems for high throughput screening (HTS), a single pharmaceutical target is challenged sequentially with members of a large library of chemicals in a search for hits, defined as agents binding the target better than a designated threshold affinity [1]. A drawback of this approach is that preparing sufficient amounts of the target for HTS can be time consuming, and the assays which are compatible with robotic assay may not be fully relevant to the physiological state of interest. Furthermore, supplies of rare compounds are consumed in a process that is designed so that the vast majority of the compounds are nearly always scored as uninteresting. In an alternative approach described here, HTS is used to establish affinity fingerprints of the compounds, defined as their distinctive patterns of binding to a small reference panel of proteins. This panel is chosen empirically and represents in a generic fashion the non-covalent binding modes by which small molecules interact with proteins [2]. Such affinity fingerprints, in turn, facilitate novel approaches to increasing the yield of useful pharmaceutical information derived from HTS [3].

In particular, affinity fingerprints provide a set of molecular descriptors that are qualitatively distinct from descriptors based on properties calculated from the structural formulae of small organic molecules, such as hydrophobicity and various indices of steric and electronic features. Computed physicochemical properties of small organic molecules have provided insights into the mechanistic basis of protein binding, particularly in conjunction with high resolution data

*Corresponding author.

[1]Presented at the 2nd Conference on Affinity Technology, Arlington, VA, USA, Sept. 29–30, 1997.

0378-4347/98/$19.00
PII: S0378-4347(98)00045-0

from X-ray crystallography [4]. The use of such descriptors has become so pervasive that they are now increasingly being used not only to generate measures of molecular similarity for a given target but also to provide a framework for defining the diversity of collections of compounds [5]. Since the properties that can be computed in reasonable timeframes are necessarily limited, however, it is important to have an independent way of estimating the quality of the information generated.

Affinity fingerprinting can provide such an independent measure of biochemical diversity by replacing the indirect aspect of computed molecular descriptors with an empirically derived profile of protein binding preferences. The utility of data derived from implementing this idea arises from the statistically directed choice of reference proteins as being representative of proteins in general with regard to the property of interest, namely ligand binding. Once an appropriate panel is available, and a large database of affinity fingerprints has been assembled, then a variety of techniques can be applied for classifying the compounds with regard to utility for particular applications. The examples presented here are derived from work using Terrapin's proprietary TRAP™ reference panel applied to the company's library of 30 000 structurally varied compounds. Full details regarding a representative non-proprietary data set from an early version of the technology have been previously published [2].

2. Experimental

2.1. Selecting a reference panel

The criteria defining a protein as useful for inclusion in the reference panel are based on measured binding properties, rather than on amino acid sequence or catalytic function (e.g., Enzyme Commission classification). The goal is to account for the major binding modes involved in protein recognition of small molecules, and in this sense the reference panel is best thought of as a set of chemical reagents rather than as a sample of proteins chosen by the more familiar definitions of protein families.

Thus, a proposed panel member is considered useful if its binding properties show statistical independence from the proteins already in the panel. In practice, the selection of a reference panel is an iterative process [6]. A small group of proteins is first tested against a small number of compounds, and the redundant proteins eliminated from consideration. Empirically, most randomly selected proteins are uncorrelated in binding properties, when considered pairwise, consistent with expectations based on the unique structure and function of every protein. The statistical independence needed for inclusion in the reference panel, however, is a more stringent test than simply pairwise orthogonality. The new protein must not only be different from each of the existing panel members considered individually, but it must also lack any multicollinearities with them. That is, it must be different from any algebraic combination of the other proteins. If the binding properties of a new protein are simply the average of two other panel members, for example, then it is different from each considered individually but the information it provides is not sufficiently unique to merit inclusion in the panel. As the panel grows, it becomes increasingly difficult to find a new protein that is substantially uncorrelated with all possible combinations of the previously chosen panel members. Expanding the number of compounds tested from a few hundred to several thousand tends to create new opportunities for proteins to enter the panel, but after several such iterations it again becomes increasingly difficult to find new proteins worth including.

In addition to the statistical criteria, other pragmatic considerations come into play when implementing the system using robotic assay systems typical of HTS. Specifically, the proteins need to be reproducibly available in adequate quantities, and must be readily assayable by robotic methods. For optimal effectiveness, the actual half maximal binding data for each compound to each protein is needed over a wide dynamic range, requiring titration in replicates. The HTS demands are thus quite intense when compared to screening a single target for high affinity hits in a binary positive/negative format.

2.2. Fluorescence polarization assays

One assay method that has been particularly useful for creating an affinity fingerprint database is fluores-

cence polarization [7]. A fluorescent tracer is first identified which binds to a single site on the reference protein with suitable affinity. The degree to which a chemical library compound can displace this tracer is then a measure of how well the library compound binds to the reference protein. The displacement is indicated by a change in polarization. When free in solution, the tracer tumbles more rapidly than its fluorescence lifetime and thus molecules excited by polarized light will have assumed randomized orientations before emitting light, eliminating the polarization information. By contrast, when bound to the protein, the tracer tumbles much more slowly, and polarized excitation yields polarized emission.

Fluorescence polarization has been widely used in commercial immunoassay systems, and has proven to be a robust assay format [8]. It depends on a rapidly attained equilibrium, and allows for a homogeneous assay which does not require any washing steps. Both properties are very convenient for robotic systems used in HTS. Further, with suitable choice of tracers, the incidence of serious interferences by colored library compounds is only a few percent; and unlike a direct fluorescence assay, the polarization assay can be normalized for quenching and background effects. Implementation for HTS has been enabled by commercialization of a 96-well microplate reader able to make fluorescence polarization measurements [9]. With this system, Terrapin has been able to collect 50 000 data points in a day, allowing ~10 000 compounds to be fully fingerprinted with high accuracy in a month.

2.3. Fingerprint database

Using fluorescence polarization assays, assembling a suitable reference panel can be accomplished by cyclically applying HTS to steadily larger numbers of proteins and compound sets. The panel used for the experiments described here comprises a proprietary set of 18 reference proteins and associated tracers. It is similar in character to those in a preliminary panel which has been previously described in full detail [2]. In particular, the panel members do not have any recognizable functional or sequence homology to the target proteins for which predictions are made, reinforcing the view that the reference proteins are chemical reagents for characterizing properties of small ligands rather than individual mimics of the target.

A compound collection of 30 000 chemicals, selected for structural diversity by conventional measures, has been fingerprinted using the 18-member panel. Again, the structures of representative compounds and their fingerprints against the prototype reference panel have been previously published [2]. Taking advantage of the convenient aspects of fluorescence polarization assays, measurements were made over a wide dynamic range. For most of the compounds, considering data across the whole panel, the range of observable binding is from 50 n*M* to 500 μ*M*, with 75% of the values being above 300 μ*M*. The vast majority of these data are thus outside the range normally considered useful in a direct screen of a particular target against a compound library, for which the threshold defining a hit is typically around 10 μ*M*. For the purposes of defining and analyzing fingerprints, however, the gradations among weak affinity data have proven to be quite informative. A second feature facilitating a reference panel of manageable size is the use of proteins, such as nonspecific xenobiotic detoxification enzymes, that are more broadly cross-reactive than typical pharmaceutical targets. In short, the reference panel is a set of chemical reagents for probing the surface features of chemical compounds rather than being a collection of proteins which individually provide target mimics.

3. Results

3.1. Target surrogates

Once a reference panel has been chosen, finding hits for a new protein can be accomplished by comparing limited empirical data on the target itself to the large amount of data represented by the affinity fingerprint database. A variety of mathematical techniques can be used in making such comparisons [2]. A simplified illustration of how the process operates in a prospective screening application is shown in Fig. 1, with data drawn from an actual example conducted as part of a confidential collaboration with a client company.

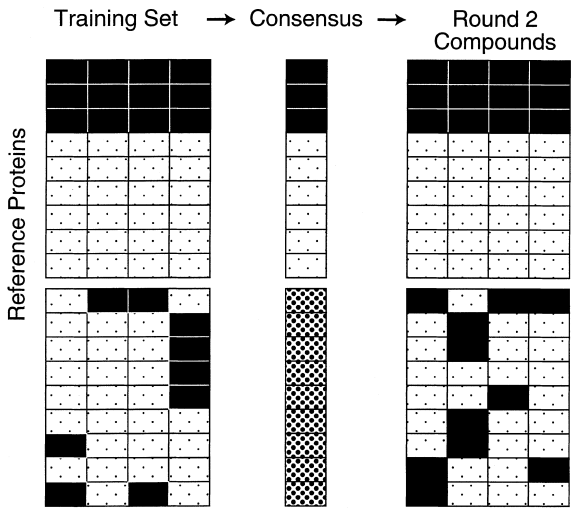

Fig. 1. A training set of 75 compounds, chosen for diversity in their affinity fingerprints against a standardized 18-member reference panel of proteins, was tested in a biological assay. The fingerprints of the four most active compounds (IC_{50} in the 10–50 μM range) defined a consensus pattern which was then used to select 25 additional compounds from the fingerprint database of 30 000 compounds. Four additional active compounds were thereby identified, including one with an IC_{50} of 3 μM. The fingerprints are shown using an exaggerated gray scale: binding at 300 μM or better (black) vs. no binding (white); the order of the reference proteins has been arranged to clarify the consensus features.

In the first step, a ''training set'' of 75 compounds was chosen from Terrapin's library of 30 000 compounds based on quantitatively assessed diversity in the fingerprints [10]. These compounds were shipped to the client for assay, and the IC_{50} binding values for active compounds reported back. Based on this information, the computer was ''trained'' to recognize the fingerprint features shared by the active compounds to the exclusion of inactive compounds, which then guided selection of compounds for the next iteration. In the diagram, an exaggerated gray scale is used to indicate the degree of binding of the active compounds to the reference panel: black indicates readily measureable binding, while white indicates no reliable binding; in reality, of course, all the data are handled numerically with measurement reliability equivalent to having >12 gray scale levels defining binding potency rather than just the two levels shown here to clarify the significance of the fingerprints. Furthermore, the order in which the reference proteins are listed has been adjusted to highlight the consensus features of the four most active compounds from the training set. With these simplifications, the consensus fingerprint features are obvious upon inspection. In most instances, more sophisticated pattern recognition techniques [11] are applied in order to identify consensus features in the numerical data.

Although none of the strongest compounds from the initial set would have been considered a hit in a conventional HTS program, their consensus fingerprint served as a basis for selecting a second set of compounds expected to be enriched for compounds that share characteristics with the best of the training set, and are thus most likely to exhibit strong binding to the target. Among the compounds selected in this manner, several were in fact found to be active. One of the 25 second iteration compounds had an IC_{50} of 3 μM, and was a structurally reasonable starting point for medicinal chemistry optimization. In short, a legitimate hit was found from a library of 30 000 compounds after physically testing ~100 compounds. The subset of training compounds showing the best binding varies with the particular target, as does the subset of reference proteins whose binding to the successful training compounds correlates with that of the target. The results illustrated in Fig. 1, however, are typical of the overall phenomenology seen with a variety of targets.

The successes of the fingerprint matching concept in prospective screening applications suggest that these novel parameters for characterizing chemical structures are recognizing key features relevant to biochemical activity.

Fingerprint matching as a means of characterizing compounds can also be studied retrospectively using compounds with similar bioactivity. In Fig. 2, based on an early version of the reference panel, a collection of structurally diverse dopaminergic D2 antagonists can be seen to have systematic trends in affinity for three proteins unrelated to each other or to the D2 receptor: D-amino acid oxidase, arginase and butyryl cholinesterase. For a broader sampling of compounds, potency against any one of these proteins independently does not correlate with affinity for the D2 receptor. However, within the fingerprinted collection of 30 000 compounds, few examples other than the known D2 antagonists show affinity for all

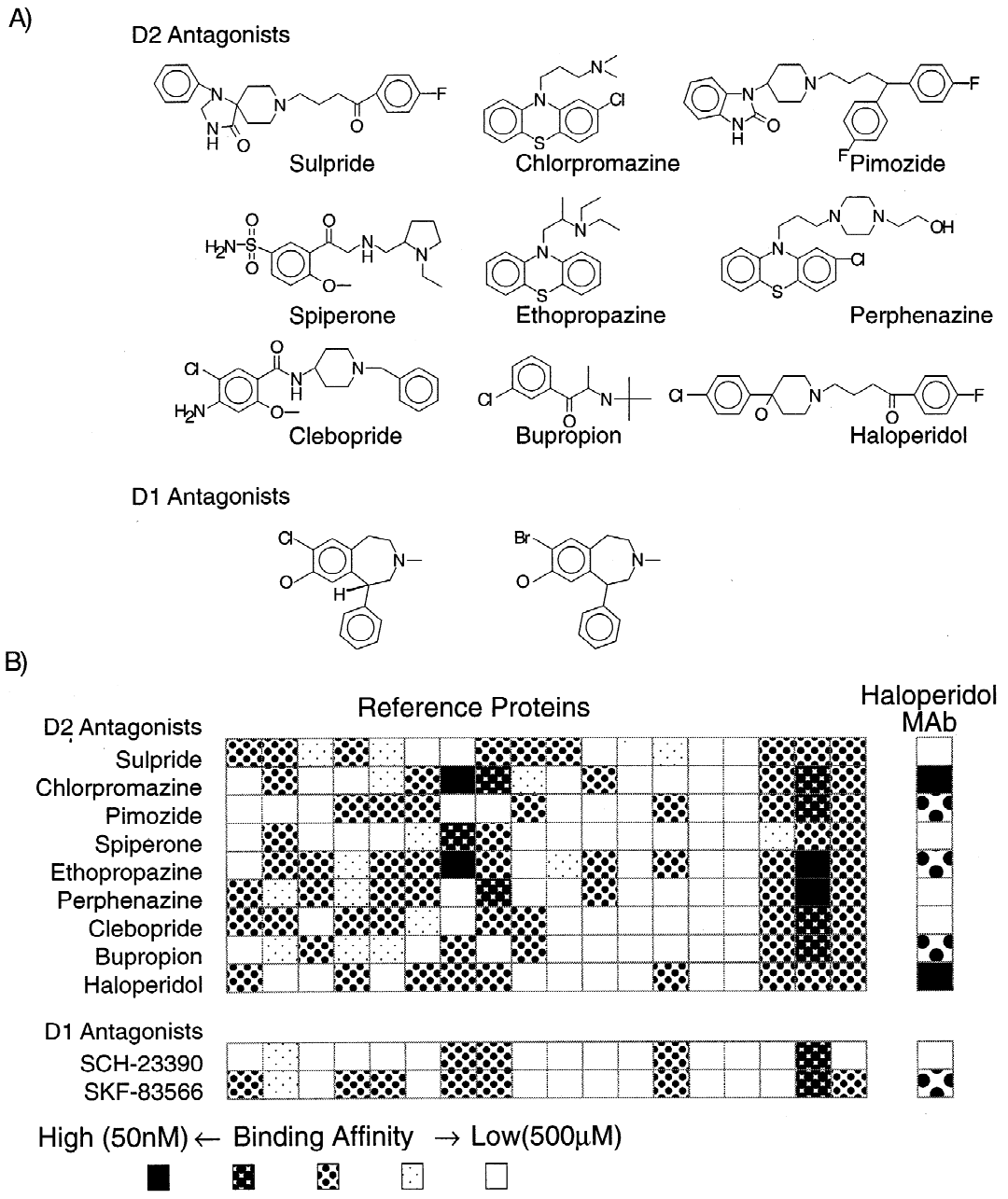

Fig. 2. (A) Structurally diverse dopamine D1 and D2 receptor antagonists. (B) Binding patterns of the D2 antagonists to the three proteins placed at the end of the reference panel show consistencies that are more correlated with activity than is binding to an antibody prepared against one compound (haloperidol). D1 antagonists do not share this fingerprint feature. The gray scale is evenly spaced from 50 n*M* to 500 μ*M*.

three. Even antagonists at the related D1 receptor show differences in the key fingerprint features. Although the affinities for the diagnostic reference proteins are not very high on an absolute scale, they are well within the working range of the fingerprinting assays.

Previous attempts to create surrogates of a target protein have been reported, typically using antibodies [12]. An antibody to haloperidol [13], a prototypical D2 antagonist, was thus also studied. Affinity for this surrogate was far less predictive of bioactivity against the D2 receptor than was the consensus fingerprint feature. Furthermore, a variety of other proteins examined also showed no particular correlations in their affinities for these compounds, beyond what would be expected from the product of their individual hit rates against random chemicals [3]. Comparable trends have been observed for targets as diverse as cyclooxygenase [6], monoamine oxidase and the putative ion channel believed to be the target of loop diuretic agents. In none of these cases was there any homology at the level of sequence or catalytic function between the target and the diagnostic indicators in the reference panel.

3.2. Classifying targets

The overall phenomenology observed using affinity fingerprints as molecular descriptors is consistent with published data on drug cross-reactivity [14]. For a dozen well known drugs which were surveyed, there are documented cross-reactions to unrelated proteins across a wide range of potency. This result is consistent with the familiar finding that drugs have side effects, and suggests a route to reducing clinical failure of drugs arising from cross-reactions. That is, given two compounds of equal efficacy in hitting the intended target, the more useful compound to test as a drug is the one with weaker cross-reactions to proteins implicated in toxicity. To implement this strategy, it would be helpful to survey a large number of proteins for their ligand binding preferences. Affinity fingerprinting can aid in this effort by providing a first approximation of the binding preferences of proteins, following direct assay of only a small number of compounds.

Classification of proteins based on their ligand binding properties has been reported in a tentative manner in the past. For example, the family of glutathione transferases is now known to include at least four major families at the DNA level; three of these families were initially identified by multivariate clustering of substrate and inhibitor specificity data along with immunological similarities [15]. Similarly, drugs binding rhinovirus capsid proteins have been used to classify 100 serotypes into two broad categories; a more refined analysis found that the median activity of a candidate drug against a properly chosen subset of 17 serotypes was an accurate predictor of median activity against all 100 [16].

To expand the scope of protein classification by ligand binding properties, it is desirable to identify a typing set of compounds. While pursuing this goal, a class of compounds termed ‘‘master keys’’ was identified whose existence had not been suspected from analysis of the compound library using structural descriptors. These compounds may prove useful as typing compounds as well as for other purposes, such as initiation points for combinatorial chemistry and as diversified chromatographic ligands for protein fractionation.

Fig. 3 illustrates the concept using the gray scale representation from Fig. 2. Fingerprints for a random assortment of compounds are presented along with the fingerprints for a set of master key compounds. Visual inspection confirms the quantitative analysis that identified the master keys as compounds which show highly differential binding to the panel, just as do the random compounds, but which on average have binding affinities half a log to one log unit tighter than the random compounds. Such compounds are rare within the extensive database collected so far, well under 1%. Building on the familiar lock and key metaphor for drugs binding to proteins, these compounds are termed master keys because they bind well to multiple proteins. The physicochemical properties which account for this remarkable property are not clear. The master keys are not extremely hydrophobic, for instance, as this property leads to tight binding to all proteins in the reference panel.

3.3. Application to difficult targets

To date, >40 pharmaceutically relevant targets have been explored using the affinity fingerprinting

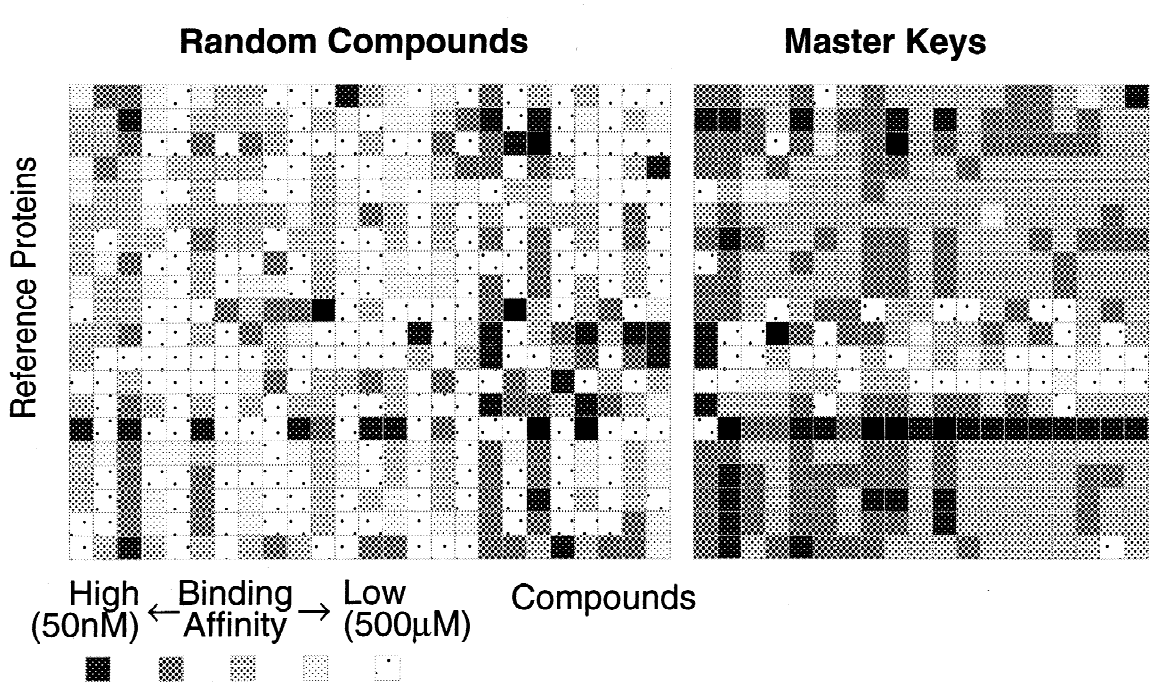

Fig. 3. Fingerprints for a random sampling of compounds on the left are compared to fingerprints for Master Keys on the right. Both sets show differential binding to the reference proteins, but the Master Keys bind on average 0.5 to 1 log unit more tightly. The gray scale is evenly spaced from 50 n*M* to 500 μ*M*.

system as a way to find hits, including a wide spectrum of protein types [10]. For 70% of the targets, a hit was identified using 20 μ*M* as the potency threshold defining a hit, with a 50% success rate using 10 μ*M* as the threshold; of course, each target generated unique hits. These rates are in the range obtained across a broad spectrum of targets using direct HTS, and extrapolation from that HTS experience implies that expanding the database of fingerprinted compounds should improve the hit rate at more stringent thresholds. In all cases, the fingerprinting system identified its hits after fewer than 200 compounds had been physically assayed against the target. Taking into consideration the synthetic tractability of the hit, the difficulty of finding hits for the target in previous screening efforts, and the actual potency attained, about half of the targets surveyed have yielded interesting starting points for further medicinal chemistry work.

For example, in an internal Terrapin project, a totally non-peptide primary hit which activates the human insulin receptor has been identified by this approach and is now being studied with regard to improving a variety of chemical characteristics. In assays measuring glucose uptake into fat cells in response to insulin, addition of this compound can sensitize the response to insulin by ten-fold. In experiments using diabetic mice, injection of a compound derived from the initial hit showed utility in lowering blood glucose.

4. Discussion

In practice, it is at the level of screening that large databases of chemical interactions with proteins tend to accumulate. It is also at this point, early in the drug discovery process, that the right choice of a target and associated lead can save the most time and money compared to a choice that ends up failing after extensive investment in research and clinical development. Affinity fingerprinting provides a common framework for collecting and analyzing such data.

4.1. Expanding opportunities for screening

Three advantageous aspects of the affinity fingerprinting approach, applicable to any modern drug discovery effort, can be illustrated in concrete form using the discovery of insulin receptor activators. First, because the number of compounds that were directly assayed to find the initial hit was small, use of the cloned insulin receptor was not necessary; instead, small amounts of the natural receptor, partially purified from normal human cells, were sufficient. Second, the primary assay measured tyrosine kinase activity leading to autophosphorylation of the receptor [17]. This assay has several steps, requiring precise timings, and includes numerous washing steps. Such assays are typically difficult to implement reliably in HTS systems. The advantage of using this cumbersome assay is that it reflects an essential feature of the biological system. By contrast, numerous groups have previously looked unsuccessfully for insulin mimics via competitive binding against radioactively labelled insulin, an assay quite amenable to HTS but which does not directly indicate biologically relevant activity.

These two features of affinity fingerprinting are important for extending the use of screening to earlier stages of research. The small numbers of compounds needed for physical assay relaxes the requirement for preparing large amounts of target; thus, effective screening is enabled using partially purified protein from cell extracts or the small amounts of protein that can be readily obtained from transient transfection of cloned genes. Likewise, the ability to use complex, but physiologically meaningful, assays is particularly useful for emerging targets in which the complex assay defines an otherwise poorly characterized activity. For example, an understanding of a biological pathway may indicate promising targets, known initially only as activities in partially purified cell extracts [18]. Adapting such physiologically relevant assays to a format compatible with HTS can be a lengthy process.

The third point of general interest is that the discovery of a pharmacologically active agent can be of great help in prioritizing research on the large numbers of potential drug targets emerging from genomics and other basic research. In the case of the insulin receptor, the physiological effects observed were a pleasant surprise. Surveys of human diabetics had not implicated inherited receptor defects as a significant cause of clinically manifested disease, raising doubts as to the utility of the receptor as a therapeutic intervention point. Further, currently used insulin secretagogues, which are effective in lowering blood glucose, work by raising blood insulin levels, possibly contributing to long term complications. Accordingly, the utility of an orally active mimic of insulin, in the sense of a direct agonist, had been questioned as compared to agents hitting subsequent steps in the signal transduction cascade or even more remote steps in the regulation of glycogen metabolism. The observation that it is possible to find agents which sensitize the receptor itself, leading to desirable consequences, thus creates renewed interest in this target.

The enormous effort put into genomics in the past few years is beginning to yield a steady stream of new targets for evaluation. Even in a ''simple'' organism like yeast, however, a function for most of the newly cloned genes remains mysterious [19]. Given the rate at which new genes are being implicated in various diseases, the molecular biology paradigm may be a mixed blessing with regard to improving the efficiency of drug discovery. The flood of information on genes needs to be converted into information about probable pharmaceutical utility. Sequence comparison to known proteins can provide some general indication of function, but it is not sufficient to predict physiology. Although a chemical hit is not the only tool which can be effective for validating a target, it is clearly one that is informative when available. Other approaches, such as gene knockouts and antisense DNA, or antibody and peptide based probes, are also useful but can be at least as expensive and time consuming to generate as is a chemical hit.

In principle, the tens of thousands of potential targets uncovered by genomics could each be studied by current HTS systems, replicated many-fold. Aside from the practical financial constraints, a fundamental technical obstacle to such indefinite replication is the availability of compounds. Since HTS screening is designed to reject $>99.99\%$ of all compounds tested, limited supplies of individual compounds are inevitably expended primarily on futile screens. A second obstacle to simply replicating the current HTS paradigm is the limited supplies of protein available for most new targets. Even in cases in

which target discovery is driven by molecular biology initially, preparing adequate quantities of purified recombinant protein can require substantial effort.

4.2. Ligand optimization

Once found, a hit must be optimized to create patentable drugs which can be manufactured at reasonable cost. This is not a straightforward process in many cases. Peptides, for example, have utility in defining the biological role of a potential target [20], but have been of limited use as guideposts to inexpensively synthesized compounds with appropriate stability and bioavailability [21]. By providing a reliable indicator of chemical similarity, as viewed from the vantage point of protein binding sites, affinity fingerprinting should be useful in the optimization phase of drug discovery.

Building on insights from X-ray crystallography, optimization is increasingly benefiting from use of computable parameters to characterize molecular properties [4]. For evaluating the theories of binding underlying such parameters, the database of affinity fingerprints provides a useful set of uniformly collected measurements which span the several orders of magnitude in potency needed to provide realistic comparison to theory. More specifically, it is surprising from a mechanistic perspective that unique affinity fingerprints for tens of thousands of compounds can be created with a reference panel of under 20 proteins. Although theoretical models for olfactory function had suggested this might be the case [22], it is not readily predicted from the protein structure and genetic studies which have estimated that there are at least several hundred primitive exons, believed to have provided the building blocks for all proteins [23]. Insights into the structural basis of this empirical finding should benefit from computer docking of the same set of compounds to a variety of proteins. Initial work using the DOCK program [24] confirms the underlying premise of affinity fingerprinting, namely that information can be transferred from one protein binding site to another which is unrelated by conventional measures. At present, however, computational docking programs are more effective for describing steric and enthalpic factors involved in binding than for modeling the equally important entropic factors arising from restrictions in rotational freedom and from exchanges of solvent molecules between the separated and complexed ligand and protein. The fact that conventional physicochemical parameters do not correlate with a property such as that embodied in the master keys indicates the gap between the theoretical and empirical parameters for characterizing chemical structures. One promising approach to elucidating the mechanistic basis of the novel fingerprint parameters is analysis of statistical regularities in binding sites as compared to bulk protein [25].

A second means by which affinity fingerprinting can aid in optimization is by use of the master keys as combinatorial scaffolds to aid in accelerating the exploration of structures around a hit. For reasons of synthetic practicality, structure/activity relationship (SAR) studies are normally limited to making variants of the actual hit. If this structure is difficult to make, then the SAR work goes slowly. Even if analoging is feasible, SAR work is likely to be most effective when used to explore the entire region of fingerprint space surrounding the hit, rather than just those directions accessible by chemical modifications of the hit.

For the purpose of saturating a region of fingerprint space, combinatorial chemistry can be applied to any convenient scaffolds in the neighborhood of the hit. The guiding principle in this approach is that the intermediate stages of drug optimization do not need to resemble either the initial hit or the final drug. They only need to provide information defining trends which can be overlaid on the actual hit(s) in order to define the pharmacophore. The advantage of master keys as scaffolds is that small variations in these intrinsically ''generic'' binding agents are likely to have large effects on potency and specificity. The use of affinity fingerprints to select diverse subsets for intensive investigation should also be helpful in combinatorial chemistry generally, which typically generates large numbers of similar compounds that must then be winnowed out in secondary screens [26].

4.3. New chromatographic ligands

Purification and analysis of large numbers of proteins should benefit from improvements in separation technology. In particular, sequential purification of proteins on sorbents that are diverse in character has been the foundation of most purifica-

tion procedures, and indeed of analytical chemistry generally [27]. Identification of additional broad spectrum chromatographic ligands that are orthogonal in their binding properties would thus be useful. In the past, most affinity chromatography work has been viewed as different from other purification methods, with an emphasis on very high affinity ligands as special purpose reagents for purifying one protein. Sequential use of moderate affinity ligands, however, has been shown to be useful for purifying a variety of proteins, under very mild elution conditions likely to preserve biological activity. This principle has been demonstrated using peptide ligands which differ in their protein binding profiles [28].

Variants of Cibacron Blue have also been explored as ligands addressing the need for moderate affinity, orthogonal separations [29]. The advantages of rigid organic compounds over peptides are substantial with regard to stability in the chromatographic context, where the protein mixture being analyzed often contains proteases; moreover, harsh washing conditions are often used to regenerate the column. An affinity fingerprint database provides information useable for selecting additional robust small molecule ligands to enrich the repertoire of reagents for general protein separations work.

Purifying proteins with a diverse set of ligands provides an immediate route to selection of other compounds which bind the protein, since the diverse sorbents can act as a training set. More generally, as proteins become purified by any method, even in minute amounts, information about their interaction with a training set of compounds can be collected. Assay of ~100 compounds against thousands of proteins can in principle be accomplished using robotic HTS systems that are currently being used to test single proteins against hundreds of thousands of compounds. The resulting qualitative increase in data on ligand binding preferences for large numbers of proteins should be usable to estimate side effects, to choose optimal combinatorial chemistry scaffolds, and to select an improved reference panel for more accurate mapping of chemical space.

References

[1] J. Devlin (Editor), High Throughput Screening: The Discovery of Bioactive Substances, Marcel Dekker, New York, 1997.
[2] L.M. Kauvar, D.L. Higgins, H.O. Villar, J.R. Sportsman, A.E. Engqvist-Goldstein, R. Bukar, K.E. Bauer, H. Dilley, D.M. Rocke, Chem. Biol. 2 (1995) 107.
[3] L.M. Kauvar, Pharm. News 3 (1996) 12.
[4] H.J. Bohm, G. Klebe, Angew. Chem. Int. Ed. 35 (1996) 2588.
[5] E.J. Martin, D.C. Spellmeyer, R.E. Critchlow, J.M. Blaney, Rev. Computat. Chem. 10 (1997) 75.
[6] L.M. Kauvar, Pharm. Mfr. Intl. 8 (1995) 25.
[7] J.R. Sportsman, S.K. Lee, H. Dilley and R. Bukar, in J. Devlin (Editor), High Throughput Screening: The Discovery of Bioactive Substances, Marcel Dekker, New York, 1997, Ch. 22, p. 389.
[8] M.E. Jolley, J. Anal. Toxicol. 5 (1981) 236.
[9] Fluorolite FPM-2 Fluorescence Polarization Microtiter Reader, Jolley Consulting and Research, Grayslake, IL.
[10] L.M. Kauvar, Nature Biotechnol. 13 (1995) 347.
[11] H. van de Waterbeemd (Editor), Chemometrics in Molecular Design, Vol. 2, VCH, Weinheim, 1994.
[12] D.S. Linthicum, M.B. Bolger, P.H. Kussie, G.M. Albright, T.A. Linton, S. Combs, D. Marchetti, Clin. Chem. 34 (1988) 1676.
[13] M.A. Sherman, D.S. Linthicum, M.B. Bolger, Mol. Pharmacol. 29 (1986) 589.
[14] F.S. LaBella, Biochem. Pharm. 42 (1991) 51.
[15] B. Mannervik, P. Ålin, C. Guthenberg, H. Jensson, M.K. Tahir, M. Warholm, H. Jörnvall, Proc. Natl. Acad. Sci. USA 82 (1985) 7202.
[16] K. Andries, B. Dewindt, J. Snoeks, R. Willebrords, R. Stokbroekx, P.J. Lewi, Antiviral Res. 16 (1991) 213.
[17] A. Grupe, J. Alleman, I.D. Goldfine, M. Sadick, T.A. Stewart, J. Biol. Chem. 270 (1995) 22085.
[18] V.N. Reddy, M.L. Mavrovouniotis and M.N. Liebman, in T.F. Kumonsinski and M.N. Liebman (Editors), Molecular Modeling (ACS Symposium Series, Vol. 576), ACS Books, Washington, DC, Ch. 14, p. 221.
[19] V. Smith, D. Botstein, P.O. Brown, Proc. Natl. Acad. Sci. USA 92 (1995) 6479.
[20] D.J. Kenan, D.E. Tsai, J.D. Keene, Trends Biochem. Sci. 19 (1994) 57.
[21] L.M. Kauvar, Nature Biotechnol. 14 (1996) 709.
[22] D. Lancet, E. Sadovsky, E. Seidemann, Proc. Natl. Acad. Sci. USA 90 (1993) 3715.
[23] R.L. Dorit, L. Schoenbach, W. Gilbert, Science 250 (1990) 1377.
[24] H. Briem, I.D. Kuntz, J. Med. Chem. 39 (1996) 3401.
[25] H.O. Villar, L.M. Kauvar, FEBS Lett. 349 (1994) 125.
[26] M.H. Lyttle, Drug Dev. Res. 35 (1995) 230.
[27] L.M. Kauvar, in D.A. Kurtz, J.H. Skerritt and L. Stanker (Editors), New Frontiers in Agrochemical Immunoassay, AOAC, Washington, DC, 1995, Ch. 21, p. 305.
[28] K. Benedek, A. Varkonyi, B. Hughes, K. Zabel, L.M. Kauvar, J. Chromatogr. 627 (1992) 51.
[29] S.J. Burton, S.B. McLoughlin, C.V. Stead, C.R. Lowe, J. Chromatogr. 435 (1988) 127.