Probes for chemical genomics by design

Edward D. Zanders, David S. Bailey and Philip M. Dean

Chemical genomics represents a convergence of biology and chemistry in the era of global approaches to target identification and intervention. The success of genomics has led to a bottleneck in target validation that could be overcome by using small diverse organic compounds to interfere with biological processes. Because of the limitations of existing compound collections, this diversity can only fully be exploited using *in silico* design techniques to guide the selection of molecules with optimal binding properties. Structure-based design is used to create structures *de novo* that can be synthesized for use as chemical probes and drug leads.

Edward D. Zanders*, David S. Bailey and Philip M. Dean
De Novo Pharmaceuticals
Compass House
Vision Park, Histon
Cambridge, UK CB4 9ZR
*tel: +44 1223 238 000
fax: +44 1223 238 088
e-mail: ed.zanders@denovopharma.com
Web sites:
http://www.denovopharma.com
http://www.denovodrugdesign.com

▼ Small molecule discovery for the development of new medicines is still the lifeblood of the pharmaceutical industry despite the introduction of effective protein therapeutics. By contrast, drug target validation at the discovery stage has often been undertaken with highly selective agents, such as antibodies, recombinant receptors and antisense polynucleotides, and genetic approaches, such as mutagenesis. Despite successes in this area, there are several limitations to these approaches, in part a result of the large size of the molecules used to perturb cellular systems *in vitro* or *in vivo*.

Chemical genomics

Target validation would benefit from a chemical biology approach through the use of small molecules (e.g. with the same degree of selectivity as antibodies) but having good cell permeability properties, for example. The scale of the problem is apparent when considering the number of drug targets within human and microbial genomes that are emerging from the sequencing projects, thus 'chemical biology' has become 'chemical genomics' to reflect the need for chemistry on a genomic scale. The ultimate aim of chemical genomics is to provide a small molecule for every protein encoded by the genome for use as probes of cellular function and possible drug leads [1,2]. A distinction should be made between probes that are just being used to explore target function, and those that are destined to become drug candidates. In the first case, there is a much less stringent requirement for good pharmacokinetic properties, and therefore more scope is available for creating molecules with a wide diversity of chemical groups.

This review will cover the main strategies for small-molecule discovery with an emphasis on their use as chemical probes for target validation, rather than as drug candidates. From a historical overview of the 20th Century chemical industry, we shall describe the screening methods used to isolate novel compounds and follow with a description of the emergent *in silico* design technologies that will dominate the field in the future.

Historical

Chemical genomics is essentially chemical biology brought up to date in the era of genome analysis. The field is as old as the period in which small molecules have been purified and used as medicines in a systematic way. Studies at the beginning of the 20th Century led to the notion that drugs bind to specific body constituents, leading to the concept of the 'receptor' as described by Paul Ehrlich [3]. Coincident with this was the growth of the chemical industry and the availability of many organic compounds derived from coal tar [4]. Some of these were found to be active in biological systems through observation of their selective interaction with cells and tissues (e.g. methylene blue and nerve cells) and helped to develop the receptor theory of small-molecule action. The foundations of the modern pharmaceutical industry are based on coal tar-derived compounds, as well as natural products from plants and microorganisms, the latter are equally useful as

1359-6446/02/$ – see front matter PII: S1359-6446(02)02325-5

probes for chemical genomic approaches and as medicines but are often difficult or expensive to produce in quantity.

From the early 20th Century to present day, drugs have been discovered in a serendipitous fashion through the systematic screening of compound collections. The collections available to drug companies are often biased towards those molecules that reflect the particular chemical and biological problems tackled by the individual organization and, therefore, are restricted in diversity [4]. Even with the advent of combinatorial libraries with greater diversity, the ability to occupy the full chemical space of protein binding sites is minuscule. Although there are effective medicines in clinical use for a wide range of conditions, the supply of high quality leads is slowing down and the number of targets is rising [5]. This situation is problematical for drug discovery, for obvious reasons; however, this is not the case with small molecule probes because these can be used to explore biological processes in cells and whole organisms without the need for an impeccable pharmacokinetic or safety profile. Indeed, the number of compounds with interesting biological activities that reside within the pages of suppliers' catalogues is impressive, even though many of these would never become drugs. These molecules must therefore be seen as probes of biological function, providing information on protein targets (assuming some degree of selectivity), or validating a chemical class as a possible drug of the future.

Screening for biologically active small molecules

The traditional pharmacology based approach to small-molecule discovery was mainly superseded in the 1990s by HTS campaigns, in which chance ligand–protein interactions would be picked up by changes in the behaviour of the pure protein *in vitro*, or in a whole cell or tissue environment [6]. The details of such screens and the relative merits of using a physiological readout, compared with a simple binding response, have been discussed and debated over the years, often with strongly held views in different 'camps'. The biologist, for example, might prefer the cellular screen that measures the response of the protein in its 'natural' environment. By contrast, the chemist might prefer a simple *in vitro* assay using purified protein in which it is possible to generate clean SAR data from closely related compounds in a way that might not be possible using the first approach. In reality, both methods have yielded useful results, as highlighted in the subsequent sections.

Phenotypic screens – forward chemical genomics

The analogy between forward genetics and forward chemical genetics has been reviewed by Stockwell [1]. In the former, a mutation in a particular target gene might lead to a measurable phenotype; in the latter, a compound screened in whole-cell assays might do the same thing. Chemical genomics is essentially a variant of chemical genetics, in which genomics technologies are used to identify the altered protein target. The advantages of such an approach are clear, because compounds that are overtly toxic or impermeant to cells are easily rejected.

One of the best examples of such a screen is the discovery of the immunosuppressant molecule FK506 in a search for inhibitors of interleukin-2 production by whole T-lymphocytes, published in 1987 [7]. This molecule has similar properties to cyclosporin A, a drug that is used clinically for the prevention of organ transplant rejection. Despite this, the mechanism-of-action of FK506 (inhibition of the protein phosphatase calcineurin) was not known in detail until 1991 and required a considerable amount of sophisticated biochemistry [8]. Other examples of functional screens yielding useful probes and/or drug candidates abound in the literature (reviewed in [1,2]), despite the problems outlined previously. As technologies improve and the basic 'parts list' of cells become more comprehensive, the time taken to identify protein targets in forward chemogenomic screens will decrease. Two areas of technology come to mind (Fig. 1) and are listed as follows:

- Global gene expression profiling: Modulation of a target by compounds could effect the transcription of a subset of genes that could give clues to the nature of the target. Several publications testify to the feasibility of this approach, which has many interesting possibilities in the field of lead optimization and toxicology, as well as basic biology [9,10].
- Genetic screens: The genetic approach relies upon the sensitivity of mutant strains to the compound in question, in model organisms such as yeast. This enables the construction of a detailed map of functional responses. The availability of deletion mutants that will eventually cover the entire genome of *Saccharomyces cerevisiae* provides an unprecedented opportunity to understand global compound effects as illustrated by the study of Chan *et al.*, who examined the targets of rapamycin [11]. This can also be applied to multi-cellular organisms, such as the zebrafish *Danio rerio* [12].

In conclusion, it is not hard to see that technological improvements will make target identification in forward chemical genomic (phenotypic) screens more efficient. Once the target is identified and purified, further small-molecule development can be undertaken.

Target-based screens – reverse chemical genomics

In contrast to phenotypic screens, purified natural or recombinant proteins are used in isolation to screen for

small-molecule inhibitors of receptor binding or intrinsic enzymatic activity. The biological activity of active molecules in whole cells is then assessed, where it is hoped that they are able to penetrate the cell membrane, remain stable and not cause sufficient toxicity to make them unusable as drugs. The list of compounds that fulfil these criteria is extensive, and many of those have become useful chemical probes. Interestingly, a large number have been screened against components of cellular signalling pathways but have not been progressed into drugs. This could be for several reasons, including the lack of total selectivity against the target kinase, a situation that might be ameliorated using *in silico* design approaches (see later). Representative examples of inhibitors that were identified for three different protein kinases of pharmaceutical interest (p38, MEK1, PI3 kinase) are shown in Fig. 2. Each has been used for extensive biological experiments in several systems, both *in vitro* and *in vivo*, and has proven valuable in assigning each target to key cell-signalling pathways [13–15].

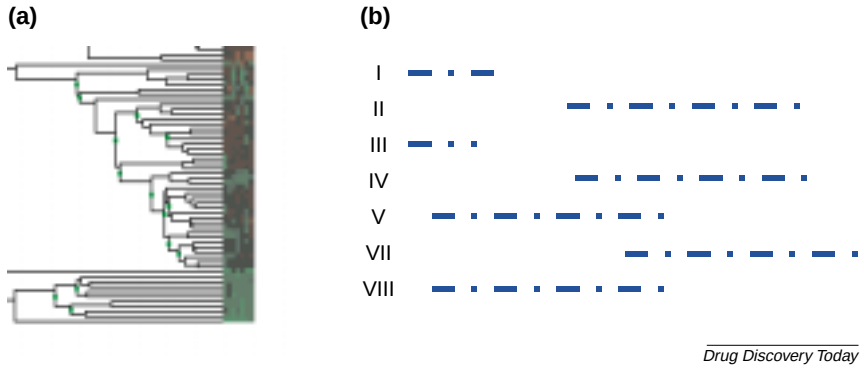

Figure 1. Forward chemical genomic screens**.** Technologies that assist in target identification for the whole-cell screening of compounds. **(a)** Changes in gene expression in response to drug treatment are measured using microarrays and displayed as red (up-regulation) or green (down-regulation). Further statistical analysis is used to display correlations between different groups of expressed genes as dendrograms (vertical axis). By using careful experimental designs, it is possible to gain useful information about the nature of the drug target from expression patterns alone [9,10]. **(b)** Schematic diagram of yeast chromosomes with deletions; the effects of compounds on individual mutants with known deletions could give indications about the nature and number of targets affected [11].

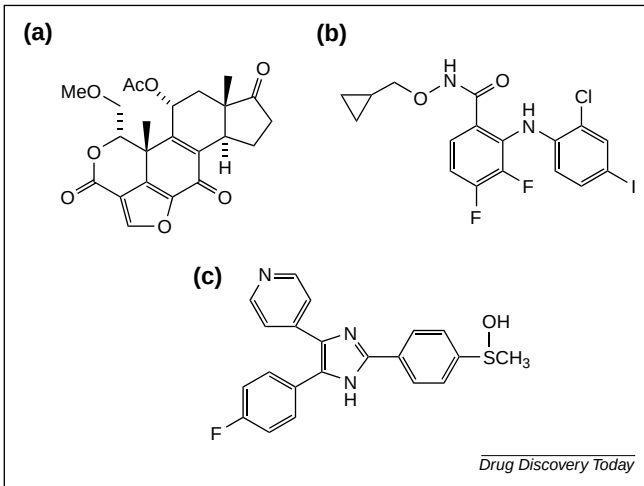

Figure 2. Examples of small-molecule kinase inhibitors used to probe cellular signalling pathways. Members of the protein kinase family are crucial elements in cellular signalling pathways, and are therefore of interest as drug targets for cancer, inflammation, etc. Many kinase inhibitors have been discovered from natural products or through medicinal chemistry programmes, and have been tested for selectivity against other kinases. The three examples shown inhibit key signalling pathways, and have been used extensively as probes for biological function. **(a)** Wortmannin (PI3 kinase); **(b)** PD184352 (MEK1); **(c)** SB203580 (p38 kinase).

In silico design of small-molecule probes and drug leads

Both chemogenomic screening approaches described in the previous sections are dependent on highly diverse small-molecule structures. Despite the introduction of chemical technologies that are designed to maximize such diversity [16], this is still insignificant compared with what is possible for molecules with a molecular weight <700 (of the order of 10^{180} [17]). A solution to this problem is to use *in silico* design before the synthesis of highly focussed libraries containing structures with a high probability of interacting with ligand-binding sites on the target protein. In addition, it is possible to eliminate potentially toxic compound designs before any synthesis and testing is undertaken.

The *in silico* design technology is sufficiently advanced to deal with both proteins with solved three-dimensional (3D) structures (structure-based design) and those only with information from existing ligands (ligand-based design). To create molecules that compare in selectivity with antibodies, for example, considerable attention has to be paid to design within gene families. In this way, it will be possible to obtain probes with varying degrees of selectivity, ranging from those affecting a wide range of related proteins to those that are highly specific. This combination of genomics, protein informatics and chemoinformatics is a primary manifestation of the chemogenomics approach to drug discovery and probe design.

Structure-based ligand design technologies

Site exploration

The growing field of structural genomics using X-ray crystallography, NMR and homology modelling has been discussed extensively [18,19] and is providing the raw material for future structure-based design initiatives. The coordinates from a 3D protein structure are used to identify the key amino acid residues that comprise the ligand-binding site. Bioinformatic methods, such as PROSITE [20], have been used to identify such sites by correlating primary sequence with 3D template motifs, but these lack the precision necessary for structure-based design. This precision is also lacking with sequence analysis, which can not reveal 3D structural information within complex site structures where significant folding creates the site architecture [21].

The nature of the site will determine the interaction energy between the protein and its ligand that has to be sufficiently strong for the complex to form, and to exist for a significant time. Maximizing the contact area between the ligand and the site on the protein surface increases the Van der Waals interactions, which is why the site should be a cavity or cleft, as opposed to flat or featureless. (This is the main reason why small-molecule inhibitors of protein–protein interactions are comparatively rare, despite their enormous pharmaceutical potential.)

Algorithms that search for surface cavities can be used to identify putative binding sites, using the molecular rules governing protein–ligand interactions derived from experimentally derived datasets of representative liganded structures (e.g. ReLiBase [22]).

Site analysis

Once the protein site has been explored, and suitable cavities found, the next stage is to create local site maps to identify important molecular determinants of ligand binding, such as hydrogen bonding and hydrophobic site points [23].

Hydrogen-bonding site points can be created on the surface of the site using a variety of algorithms [24] in an automated manner. Hydrophobic site points are less precise because these interactions are determined by summation of several component small interactions; hydrophobic site points are therefore regions rather than sharply defined points of interaction.

A typical site can contain 30 site points, although small drug-like molecules usually contain only a small number of corresponding ligand points. This fact has important consequences for drug design because it creates a combinatorial choice of points to which new ligands could bind, thereby increasing the number of possible strategies many times. Using five site points as an example, there are 140,000 possible subsets of sites available for design [25]. Furthermore, the small molecule might have too many opportunities for favourable binding; a single ligand can exhibit promiscuous binding modes because of similarities in the disposition of some subsets within the site. The identification of the subsets of site points where the choice is minimized is therefore crucial for drug design.

Design strategy

Selecting the site points that provide the greatest opportunity for creating active molecules is a key determinant of the design strategy. It is also important for design within similar sites of a related family of proteins where the objective is to obtain selectivity. Here, the selection of site points should be such that the set chosen for the target protein should be as different as possible from the other family members.

Strategies are determined in several ways, for example, in using existing ligand information from structural studies of protein–ligand co-crystals. Other approaches, such as interactions with key catalytic residues, would also be expected to work well. If there are overlaps in hydrogen-bonding probability, these can be identified and used as preferred site points from which to build molecules *de novo*.

De novo design

Once a design strategy has been established using a subset of hydrogen-bonding site points, or other features, the challenge is to build virtual molecules inside the binding cavity that are diverse, yet chemically tractable. Two main approaches have been used: first, fragments are placed in the site and those with favourable interactions connect directly, or *via* bridging fragments [26–28]; second, the structure is kept connected from beginning to end [29,30]. *De novo* design is a combinatorial process that is directly related to the product of the number of fragments that can be combined within various constraints. The chemical nature of the output is controlled by rules that govern the addition, removal or exchange of a fragment on the evolving skeleton. After each modification, the assembled structure is fitted in the site by altering its position and conformation; a scoring function is then used to monitor the predicted free energy of the interaction between ligand and site. To achieve structural diversity, it is necessary to employ optimization strategies, such as genetic algorithms [31] or simulated annealing [32].

Once the output has been examined, promising chemotypes are used to create virtual combinatorial chemical libraries to fully explore the ligand-binding site. This is important for gene family analysis because it offers a way of

designing molecules that are capable of distinguishing between mutational differences in a particular site.

Chemogenomic analysis of human aspartyl proteases

An example of a chemogenomics approach to probe and drug design is the human aspartyl proteases. The protease (peptidase) family of hydrolases is of considerable biological and medical importance, and therefore of interest to academia and the pharmaceutical industry alike [33]. The total complement of human proteases is of the order 500–800, thus providing a major challenge in the selection of relevant targets. It will be important to assess the degree of redundancy between closely related members of protease subfamilies, and this will require the use of specific chemical probes in a similar manner to the kinase inhibitors described previously. To achieve this, individual proteins can be grouped into families according to the primary amino acid sequence homology across the catalytic site. Figure 3 shows an amino acid sequence alignment of the human A1 aspartyl protease family. The 3D structures of several of these proteins have been solved by X-ray crystallography and have a sequence identity to those without solved structures of >40%; this enables the creation of homology models across the A1 family that are good templates for ligand design [34]. Specific chemical probes would be useful for exploring the role of proteases such as napsin or memapsin 1 in human physiology. Proteins such as pepsin however, are of little interest in this regard, whereas memapsin 2 (β-secretase) and renin have major pharmaceutical relevance [35,36].

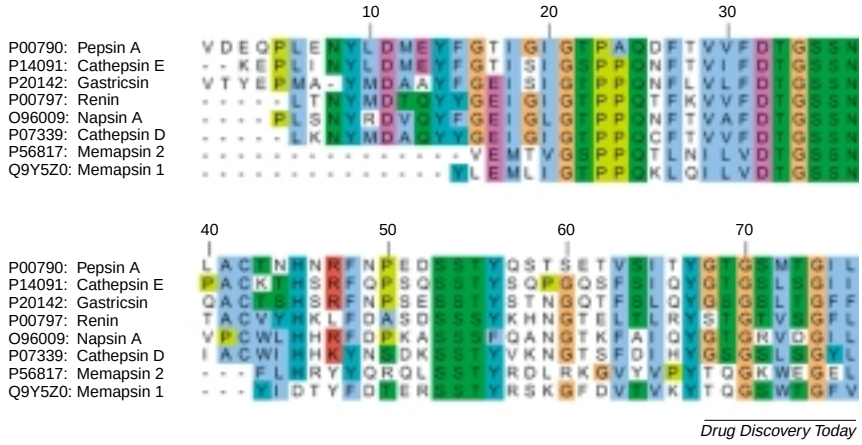

Figure 3. Amino acid alignments of the human A1 aspartyl protease family. Protein sequences of the human aspartyl protease family (A1 in MEROPS classification [33]) were aligned using ClustalW and displayed using JalView (Michelle Clamp: http://circinus.ebi.ac.uk:6543/jalview). The SWISSPROT identifier is shown against the name of each protease.

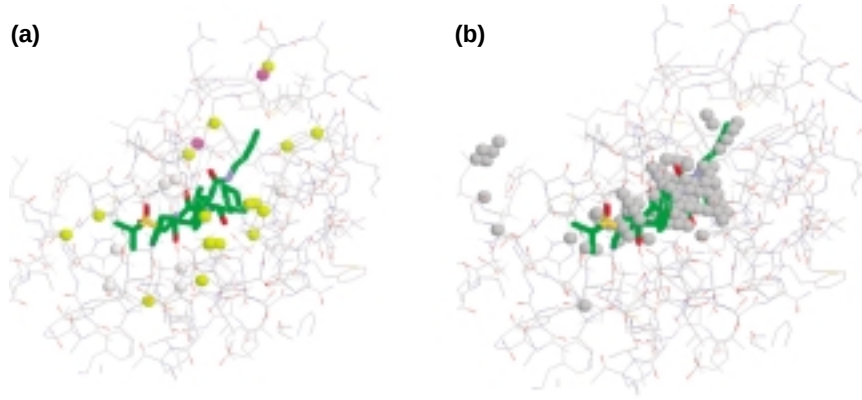

Figure 4. Site analysis of renin aspartyl protease. The three-dimensional structure of renin is shown in a wireframe model with a synthetic inhibitor bound into the active site. **(a)** Hydrogen-binding site points surrounding inhibitor in the active site. Yellow spheres, acceptors; magenta, donors; white, amphiprotic. **(b)** Hydrophobic site points. *De novo* design is performed to subsets of these site points to create virtual chemical scaffolds for further elaboration.

A site analysis of renin is shown in Fig. 4, and illustrates the disposition of hydrogen bonding and hydrophobic site points obtained after analysis of the crystal structure. *De novo* design of small-molecule scaffolds was undertaken by a stochastic assembly process from randomly selected fragments (derived from the World Drug Index) using the Skelgen™ (De Novo Pharmaceuticals, Cambridge, UK) algorithm. The latter uses simulated annealing [32] as an optimization method, and gives rise to a variety of molecular structures, thereby increasing the choice of chemistries available for exemplification (Fig. 5).

Ligand-based design

Structure-based design, by definition, requires a protein structure from which to identify site points. If none is available, as is the case for many receptors of interest, it is possible to employ ligand-based design, provided that suitable ligands are available that demonstrate SARs. Although these ligands would in themselves represent chemical probes of the target, they might have properties that make

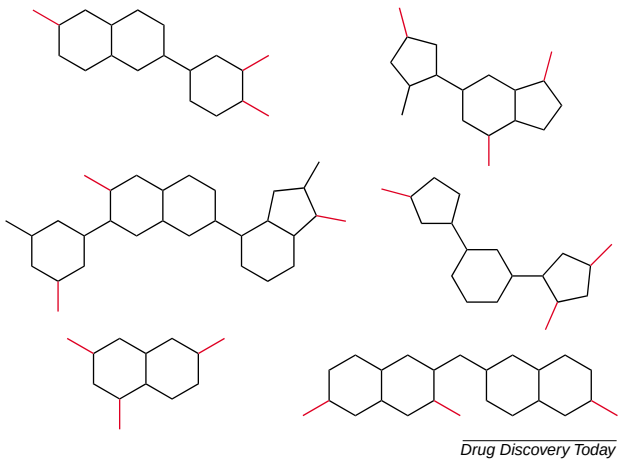

Figure 5. Examples of scaffold shapes produced by Skelgen™ (De Novo Pharmaceuticals, Cambridge, UK) after *de novo* design within the renin site. *De novo* designs are created from pre-selected fragments according to a series of chemical rules that optimize the output of tractable molecules.

them unsuitable for further biological assays in whole cells or *in vivo*. Ligand-based design can provide potent molecules with alternative chemistries that might be more suitable for probing biological function, and indeed appropriate for further development into drugs. These new compounds could also explore the binding sites of closely related proteins (e.g. receptor subclasses), which have yet to be formally identified by pharmacology or molecular biology. Ligand-based design can be applied to any projects where ligands are available, ranging from G-protein coupled receptors to enzymes and other receptor types.

The objective of ligand-based design is to provide a minimal map of the site points within the putative receptor site. Small-molecule design can then be conducted in a way that is analogous to site-directed design. The aim is to infer the geometrical distribution of site points within the protein cavity from the corresponding disposition of ligand points, in effect creating a 'virtual' cavity. The ligand points are then determined from a minimum number of compounds with known activity at the site. The next step in this process is to optimize the molecular similarity of these ligand points using a method that identifies and maximizes partial similarity within a pair of structures. Key features, such as hydrogen-bonding ligand points, hydrophobic regions, areas of electrostatic potential and geometric shape, are prioritized in this search for molecular similarity. At the end of the comparisons, the molecules in the set have to be superposed in an optimal way without preconception and used to assess partial molecular similarity by computing the distance matrix for all the points in each molecule. Subsequently, a difference distance matrix

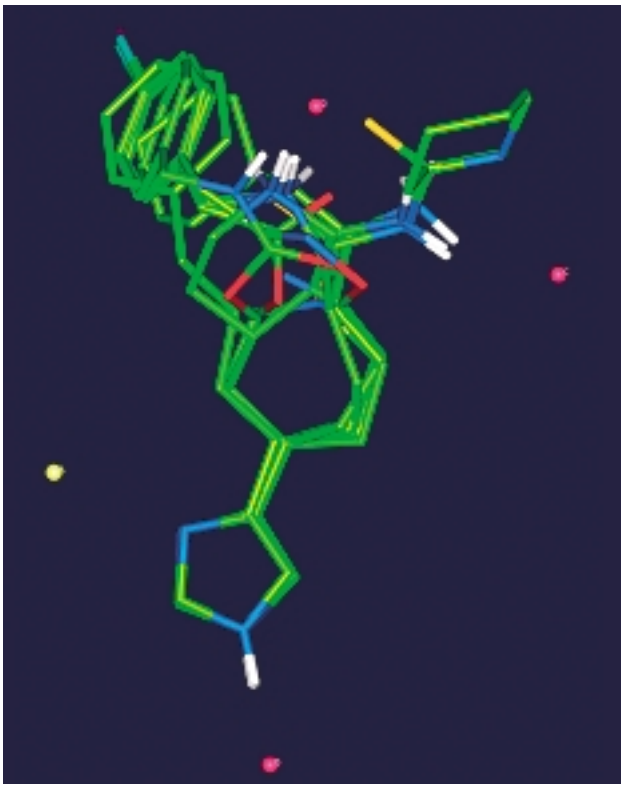

Figure 6. Ligand-based design against the histamine H3 receptor. An overlay of selected ligands for H3 is shown with the site points (small spheres) identified using the SLATE™ algorithm (De Novo Pharmaceuticals, Cambridge, UK) [38]. These points are subsequently exploited in *de novo* design.

is computed by subtraction of the two distance matrices. The partial match is identified from the optimal difference distance matrix using null correspondences. Molecular surfaces are then matched and ranked for different conformations of each molecule. Finally, optimal matches are identified for each molecular pair and the molecular supersurface computed from the partially similar molecular surfaces. This supersurface is then used as the constraining template for novel ligand design.

Ligand-based design to histamine H3 receptor

The method has been used successfully to construct a model for the G-protein-coupled histamine H3 receptor and to design high affinity antagonists [37]. The SLATE™ algorithm (De Novo Pharmaceuticals) [38] was used to optimize the match between the projected site maps of 12 fully flexible H3 antagonists. The resulting superposition of the 12 ligands is shown in Fig. 6. Novel inhibitory molecules with nanomolar to picomolar affinity have been designed and synthesized, based on the molecular supersurface created by this superposed ligand set.

Conclusion

Chemical genomics is emerging as the synthesis of genomics and organic chemistry into a more integrated way of enabling drug discovery. To exploit the vast amount of genomic sequence data that are now available, it is necessary to subdivide and classify the putative targets for small-molecule drugs into families of proteins. This is generally straightforward with targets such as proteolytic enzymes, for example, where key amino acids involved in catalysis are well known, and can be aligned using standard bioinformatics tools. What is more problematical are the large numbers of proteins without known function, and which defy classification (these comprise a significant proportion of the human genome at present; see [39,40]). In both situations, there will be a need to use probes of biological function. This review has discussed the creation and use of small-molecule probes *in silico* to achieve this end, but requires several different supporting conditions; these are the provision of protein structures through structural genomics initiatives, or the availability of ligands for ligand-based design, plus increased computing power and sophisticated algorithms. The rate-limiting step is likely to be in protein structure determination, despite the availability of more than 17,000 structures in the Protein Data Bank (http://www.pdb.org/). The relatively straightforward structures are being solved quite rapidly but there will inevitably be a bias towards those proteins with likely pharmaceutical relevance [41]. Structures for those proteins where no function has been assigned will be longer in coming, but will be necessary if the full potential of chemical genomics is to be realized. It is also these proteins for which ligands are less likely to be found, thus hampering ligand-based design. The obvious solution to this is to create more structures and find more ligands; this is less likely to happen in comparison with pharmaceutically relevant targets, but ways forward can be envisaged. Prediction of tertiary structure from primary sequence is unlikely to be achieved in the foreseeable future, but accurate homology models to known structures could be obtained if the sequence identity is >40%. Ongoing research into the full repertoire of 3D folds in proteins will assist in such modelling with sequences of lower identity, and will increase the structural coverage of different gene families [42].

Finally, increases in the power of computers, such as that obtained through the use of modern computer farms, and the exploration of new algorithms for *in silico* design, will provide the chemical genomic tools needed to understand human biology in health and disease.

References

1 Stockwell, B.R. (2000) Chemical genetics: ligand-based discovery of gene function. *Nat. Rev. Genet.* 1, 116–125

2 MacBeath, G. (2001) Chemical genomics: what it will take and who gets to play? *Genome Biol.* 2, 1–6

3 Ehrlich, P. (1913) Chemotherapeutics: Scientific principles, methods and results. *Lancet* ii, 445–451

4 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964

5 Bailey, D. and Brown, D. (2001) High-throughput chemistry and structure-based design: survival of the smartest. *Drug Discov. Today* 6, 57–59

6 Croston, G.E. (2002) Functional cell-based uHTS in chemical genomic drug discovery. *Trends Biotechnol.* 20, 110–115

7 Kino, T. *et al.* (1987) FK506: A novel immunosuppressant isolated from a *Streptomyces*. I. Fermentation, isolation and physico-chemical and biological characteristics. *J. Antibiot.* XL, 1249–1255

8 Liu, J. *et al.* (1991) Calcineurin is a common target of cyclophilin–cyclosporin A and FKBP–FK506 complexes. *Cell* 66, 807–815

9 Scherf, U. *et al.* (2000) A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.* 24, 236–244

10 Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126

11 Chan, T-F. *et al.* (2000) A chemical genomics approach toward understanding the global functions of the target of rapamycin protein (TOR). *Proc. Natl. Acad. Sci. U. S. A.* 97, 13227–13232

12 Peterson, R.T. *et al.* (2000) Small-molecule developmental screens reveal the logic and timing of vertebrate development. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12965–12969

13 Badger, A.M. *et al.* (1996) Pharmacological profile of SB 203580, a selective inhibitor of cytokine suppressive binding protein/p38 kinase, in animal models of arthritis, bone resorption, endotoxin shock and immune function. *J. Pharmacol. Exp. Ther.* 279, 1453–1461

14 Sebolt-Leopold, J.S. *et al.* (1999) Blockade of the MAP kinase pathway suppresses growth of colon tumors *in vivo*. *Nat. Med.* 5, 810–816

15 Ui, M. *et al.* (1995) Wortmannin as a unique probe for an intracellular signalling protein, phosphoinositide 3-kinase. *Trends Biochem. Sci.* 20, 303–337

16 Schreiber, S.L. (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* 287, 1964–1969

17 Klebe G. (2000) Virtual screening: an alternative or complement to high throughput screening. *Perspect. Drug Des. Discov.* 20, Preface, 9

18 Stevens, R.C. *et al.* (2001) Global efforts in structural genomics. *Science* 294, 89–92

19 Pieper, U. *et al.* (2002) MODBASE, a database of annotated comparative protein structure models. *Nucleic Acids Res.* 30, 255–259

20 Falquet, L. *et al.* (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res.* 30, 235–238

21 Thornton, J.M. *et al.* (2000) From structure to function: approaches and limitations. *Nat. Struct. Biol.* 7, 991–994

22 Hendlich, M. (1998) Databases for protein–ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* 54, 1178–1182

23 Danziger, D.J. and Dean, P.M. (1989) Automated site-directed drug design: the prediction and observation of ligand point positions at hydrogen-bonding regions on protein surfaces. *Proc. R. Soc. London B Biol. Sci.* 236, 115–124

24 Danziger, D.J. and Dean, P.M. (1989) Automated site-directed drug design: a general algorithm for knowledge acquisition about hydrogen-bonding regions at protein surfaces. *Proc. R. Soc. London B Biol. Sci.* 236, 101–113

25 Dean, P.M. (1995) Defining and using molecular similarity or complementarity for drug design. In *Molecular Similarity in Drug Design*. (Dean, P.M., ed.), pp. 1–23, Blackie Academic and Professional

26 Boehm, H-J. (1996) Computational tools for structure-based ligand design. *Prog. Biophys. Mol. Biol.* 66, 197–210

27 Lewis, R.A. and Dean, P.M. (1989) Automated site-directed drug design: the formation of molecular templates in primary structure generation. *Proc. R. Soc. London B Biol. Sci.* 236, 141–62

28 Pearlman, D.A. and Murcko, M.A. (1996) CONCERTS: dynamic connection of fragments as an approach to *de novo* ligand design. *J. Med. Chem*. 39, 1651–1663

29 Liu, H. *et al*. (1999) Structure-based ligand design by dynamically assembling molecular building blocks at binding site. *Proteins* 36, 462–470

30 Rotstein, S.H. and Murcko, M.A. (1993) GroupBuild: a fragment-based method for *de novo* drug design. *J. Med. Chem*. 36, 1700–1710

31 Goldberg, D.E. (1989) *Genetic Algorithms, Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA, USA

32 Van Laarhoven, P.J.M. and Aarts, E.H.L. (1987) *Simulated Annealing: Theory and Applications*. Reidel, Dordrecht, The Netherlands

33 Barrett, A.J. *et al*., eds (1998) *Handbook of Proteolytic Enzymes*, Academic Press

34 Sanchez, R. and Sali, A. (2000) Comparative protein structure modeling. Introduction and practical examples with modeller. *Methods Mol. Biol*. 143, 97–129

35 Lin, X. *et al*. (2000) Human aspartic protease memapsin 2 cleaves the β-secretase site of β-amyloid precursor protein. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1456–1460

36 Rahuel, J. *et al*. (2000) Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. *Chem. Biol.* 7, 493–504

37 De Esch, I.J. *et al*. (2001) Development of a pharmacophore model for histamine H3 receptor antagonists, using the newly developed molecular modeling program SLATE. *J. Med. Chem*. 44, 1666–1674

38 Mills, J.E. *et al*. (2001) SLATE: a method for the superposition of flexible ligands. *J. Comput.-Aided Mol. Design* 15, 81–96

39 International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921

40 Venter, J.C. *et al*. (2001) The sequence of the human genome. *Science* 291, 1304–1351

41 Harris, T. (2001) The commercial use of structural genomics. *Drug Discov. Today* 15, 1148

42 Vitkup, D. *et al*. (2001) Completeness in structural genomics. *Nat. Struct Biol*. 8, 559–566