Combinatorial and computational approaches in structure-based drug design

Hugo Kubinyi

Address
Combinatorial Chemistry and Molecular Modelling
ZHF/G - A30
BASF AG
D-67056 Ludwigshafen
Germany
Email: HUGO.KUBINYI@MSM.BASF-AG.DE

The increasing number of protein 3D structures and the success of structure-based approaches has led to the development of several experimental and theoretical techniques for the rational design of protein ligands. Combinatorial chemistry significantly speeds up the synthesis of potential new drug candidates. Diversity considerations, as well as the use of 3D structural information of the biological targets, reduce the size of huge libraries to a reasonable number of rationally-designed ligands. New NMR techniques (SAR by NMR) allow the construction of high-affinity ligands from small molecules with much lower affinities. Computer-aided drug design uses building, linking, and/or rigid docking procedures to search for ligands for a certain binding site. Scoring functions provide a rank order of the designed ligands according to their estimated binding affinities. Further developments in computer-aided drug design are automated approaches for the flexible alignment of molecules, the flexible docking of ligands to their binding sites, and the stepwise assembly of synthetically easily accessible ligands from combinatorial libraries of fragments.

Introduction

Structure-based ligand design has adopted a growing importance in pharmaceutical research, especially in the search for new drugs [1,2•,3•,4••,5••]. The application of these techniques is supported by an exponential increase in the number of experimental protein 3D structures [6••]. The design of new ligands is performed in several cycles, most often only by visual inspection and qualitative interpretation of the ligand-binding site interactions. Correspondingly, there is an urgent need for more rational techniques. Several experimental and theoretical approaches that have been developed to aid the design process will be reviewed in this article. Approaches of the greatest importance are the rational design of combinatorial libraries, the SAR by NMR method for the construction of high-affinity ligands, flexible ligand docking, and *de novo* drug design methods.

Combinatorial techniques for structure-based ligand design

Classical drug research depends on a combination of working hypotheses, synthesis, and testing of potential drug candidates, as well as good luck. Combinatorial chemistry and high-throughput screening have added a new dimension to the direction of modern research around the concept of huge libraries of mixtures and the deconvolution of biologically active mixtures to detect new leads. Nowadays, the automated parallel synthesis of specially designed and focused small libraries, made up from single compounds, is at the forefront of research.

Rational design and validation of combinatorial libraries

In addition to synthetic accessibility, diversity is the most important property of combinatorial libraries. Many different descriptor sets have been used to characterize the diversity of combinatorial libraries. There is an ongoing discussion of whether 2D or 3D descriptor sets are superior [11•,12•]. A logical explanation for the observed weakness of 3D descriptors might be that 2D descriptors have undergone much more extensive development. An additional issue is whether diversity considerations should be restricted to the scaffolds and the building blocks or should be applied to the resulting compounds of a library. Diversity profiling was applied to select diverse subsets from structural databases [11•,12•,13•]. HARPick (Rhône-Poulenc Rorer) is a program that selects reagents to build a library on product-based diversity calculations [14]. Combinatorial libraries have also been designed using a genetic algorithm to optimize the distribution of physicochemical or any other properties of a library [15•].

There is, however, no objective definition of diversity. If diversity is understood to be the lack of similarity, one has to be aware that compounds that are closely related chemically might show significantly diverse biological activities [16]. Books [17•,18•] and reviews [19,20a,20b] have been published on molecular diversity considerations in combinatorial chemistry, and can be referred to for further background information.

An interesting approach to the determination of the 'drug-likeness' of series of organic molecules [21] has been pursued by two industrial groups [Ajay, Vertex Pharmaceuticals, personal communication; Sadowski J, BASF AG, personal communication]. Simple structural parameters and scoring values of 0 and 1 were used to train a neural net with sets of chemicals (eg, from the Available Chemicals Directory) and drugs (eg, from the Derwent World Drug Index). The discrimination of the relatively small training sets as well as the predictions for the rest of the huge databases are in the range of 75 to 80%. Surprisingly good results are even obtained if whole series of biologically active compounds (eg, all cardiovascular drugs or all hormones) are eliminated from the training sets. Whilst the 'drug-likeness' assignment of a single compound may be incorrect, the method allows a reasonable ranking within large in-house, external, combinatorial, and virtual libraries.

Structure-based design of combinatorial libraries

The integration of structure-based design into combinatorial chemistry for new pharmaceutical discovery has been reviewed [4••,22] and critically commented upon [23••]. There are many examples of the discovery of enzyme inhibitors and other protein ligands, without considering protein 3D structures, through combinatorial chemistry [24•,25]. Some recent examples of combinatorial libraries that were designed by using information from protein or ligand 3D structures are discussed below.

- Structural variation of the P3 position of a peptidomimetic thrombin inhibitor was performed, at Merck Research Laboratories, USA, by rapid, multiple analog synthesis. Out of > 2,200 commercially and in-house available acid components, 200 were selected and coupled to resin-bound prolyl trans-4-aminocyclohexyl-methyl amide, resulting in the orally available, potent and selective thrombin inhibitor, L-372460 (Merck & Co; K_i thrombin = 1.5 nM, K_i trypsin = 860 nM) [26•]. Novel potent thrombin inhibitors were also discovered by solid-phase synthesis using different, nonbasic P1 building blocks [27].
- Bis-phenylamidine factor Xa inhibitors were designed, at DuPont Merck, USA, by docking and minimizing small fragments in the P1 and P4 binding sites; subsequently, these fragments were connected with a tether, resulting in a potent factor Xa inhibitor (K_i = 34 nM) [28•].
- A library of potential inhibitors of the aspartyl proteinase, cathepsin D, was designed at the University of California, Berkeley, USA, using 3D structural information. Approximately 6 to 7% of the analogs were active at 1 μM concentrations, the most potent analog having a K_i of 73 nM. A second-generation library resulted in the rapid identification of further potent nonpeptide inhibitors (K_i = 9-15 nM) [29].
- The design of matrix metalloproteinase inhibitors, at DuPont Merck, USA, led to combinatorial libraries from which a specific, low molecular weight, MMP-8 inhibitor (MMP-3, K_i = 148 nM; MMP-8, K_i = 1.9 nM) resulted; an unexpected alternative binding mode was observed. Minor structural modification led to a high-affinity MMP-3 inhibitor (K_i = 9 nM) [30].
- A structure-based library design of kinase inhibitors, at Howard Hughes Medical Institute, University of California, Berkeley, USA, produced a 10-fold increase in the inhibitory potency of the natural product, olomoucine [31].
- A library of 4-amino-4H-pyran-6-carbonamides, structurally related to the anti-influenza drug, zanamivir (Monash University, Biota/Glaxo Wellcome), was prepared at Glaxo Wellcome, UK, from a 4-amino-Neu5Ac-2en-derived carboxylic acid and 80 primary and secondary amines; several aliphatic N-dialkylamides and N-phenethyl-N-alkylamides proved to be nanomolar inhibitors of influenza A virus neuraminidase [32,33].
- A targeted library of phosphatase inhibitors was derived at the University of Pittsburgh, USA, from a rational

- Combinatorial libraries for the SH3 domain of *Src* tyrosine kinase were designed at the Howard Hughes Medical Institute, Harvard University, USA, in cycles, by multidimensional NMR spectroscopy investigation of the few highest affinity ligands [35,36].
- A potent, non-peptide GPIIb/IIIa receptor antagonist (collagen-induced platelet aggregation, IC_{50} = 92 nM) was developed at the Life Science Research Center, Nippon Steel Corporation, Japan, from combinatorial libraries based on the Arg-Gly-Asp sequence (the RGD motif) of the natural ligand, fibrinogen [37].
- A selective $\alpha_v\beta_3$ integrin receptor antagonist (IC_{50} = 1.1 nM) was designed at DuPont Merck, USA, as a focused RGD peptidomimetic library, based on an amine or guanidine group to mimic the arginine side chain, a variable linking group, and β-alanine to mimic the aspartate of the RGD motif [38].
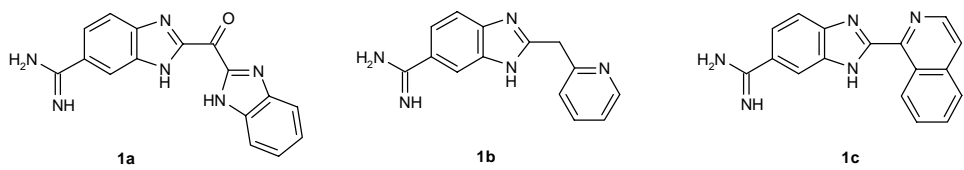
Self-assembly of ligands

In principle, one could imagine that an enzyme could be inhibited by two (or even more) small ligands, binding at different pockets of the protein. The laws of thermo-dynamics are, however, against this concept. Translational and rotational degrees of freedom are lost on binding. Correspondingly, the affinity of a ligand which connects two fragments in an optimal geometry, and which itself does not interfere with the binding, is much higher than the affinity of the two fragments as separate ligands.

Episelection (Arris Pharmaceutical Corporation, USA) is a new strategy in structure-based ligand design. The reaction of various alcohols with a boronic acid trypsin inhibitor produces a series of esters. These are selected either by preferential binding to the protein (epitaxial selection) or assembled at the enzyme surface (epitaxial reaction) [39•].

Huc and Lehn (Université Louis Pasteur, Strasbourg, France) formulated a general concept for the dynamic generation of virtual combinatorial libraries, in which molecular diversity is produced by self-assembly of protein ligands, eg, enzyme inhibitors, from appropriate components [40••]. This approach has been applied to the selective induction of carbonic anhydrase inhibitors by reversible combination of amines and aldehydes; the presence of the enzyme favors the formation of those analogs, which are expected to have high affinities to the protein.

Another example of spontaneous self-assembly was recently observed. Physiological concentrations of zinc ions convert low-affinity, metal-chelating ligands into selective, high-affinity serine proteinase inhibitors [41••]. In the absence of zinc ions, bis(5-amidino-2-benzimidazolyl)methane (BABIM) inhibits human and bovine trypsin with a K_i = 19 •M. The addition of 100 nM of Zn^{2+} increases the affinity for human trypsin to K_i= 90 nM, and for bovine trypsin, by more than four orders of magnitude, to K_i = 5 nM. An even greater effect is observed for keto-BABIM, where the affinity to bovine trypsin increases by a factor of 19,000 to K_i < 1 nM. Further structural variation led to analogs with improved selectivities

Figure 1.

	1a		1b		1c	
	no Zn^{2+}	plus Zn^{2+}	no Zn^{2+}	plus Zn^{2+}	no Zn^{2+}	plus Zn^{2+}
K$_i$ (trypsin) µM	87.5	0.005	> 1,000	136	31.2	22.5
K$_i$ (tryptase) µM	5.7	0.05	358	0.3	8.8	54.5
K$_i$ (thrombin) µM	> 1,000	0.1	> 1,000	10.5	31	0.04
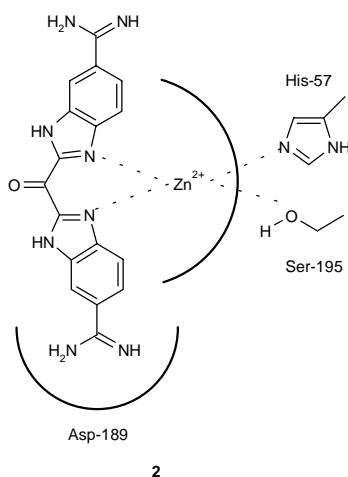
In the presence of zinc ions, the BABIM analog, **1a**, becomes a fairly selective trypsin inhibitor, **1b**, a selective tryptase inhibitor, and the analog, **1c**, a highly selective thrombin inhibitor (all K$_i$ values refer to human enzymes). The lower part of the diagram shows the experimental binding mode of the Zn^{2+}-keto-BABIM complex to bovine trypsin, as determined by protein crystallography. The zinc ion coordinates to the benzimidazole nitrogen atoms of keto-BABIM, **2**, the His-57 nitrogen atom and the Ser-195 oxygen atom [41••].

These results correspond to the activation of GDP complexes of various G-proteins in the presence of aluminum and fluoride ions, which otherwise only takes place in the presence of GTP. Protein crystallography confirmed the hypothesis on the mode of action of this serendipitous discovery, in which the AlF$_4^-$ ion mimics the outer phosphate group of GTP [42-44].

Although the principle of self-assembly of inhibitors in the binding site looks attractive, it is probably too early to decide whether general principles for drug design may result from such single observations.

Experimental methods for combinatorial drug design: SAR by NMR

Ligand design based on the combination of fragments which bind to proximal subsites of a certain protein has already been realized. Stephen Fesik (Abbott Laboratories) has developed an elegant approach for this purpose, ie, the SAR by NMR (structure-activity relationships by nuclear magnetic resonance spectroscopy) method [45••,46•]. In this important

libraries of typically a thousand small molecules are screened against a certain protein. The binding of ligands to a subsite is observed by shifts of the corresponding amide proton signals of the ^{15}N-labeled protein. In the next step, the protein is saturated with the highest affinity ligand for this site and a different library is screened for ligands which bind to another, proximal subsite. If this second step is also successful, both ligands are combined with an appropriate tether. In this manner, high-affinity ligands can be constructed within a short time. The first successful application of the SAR by NMR method was the construction of a high-affinity FK-506 binding protein (FKBP) ligand (K$_d$ = 19 nM), by combining two small molecules (K$_d$ = 2 and 100 µM, respectively) with a linker [45••]. Other applications included the discovery of potent nonpeptide inhibitors of the matrix metalloproteinase, stromelysin (Figure 2) [47•,48•], and of inhibitors which block the DNA binding of a certain *Papillomavirus* protein [49].

Despite the elegance of this approach, SAR by NMR has several limitations:

Figure 2.
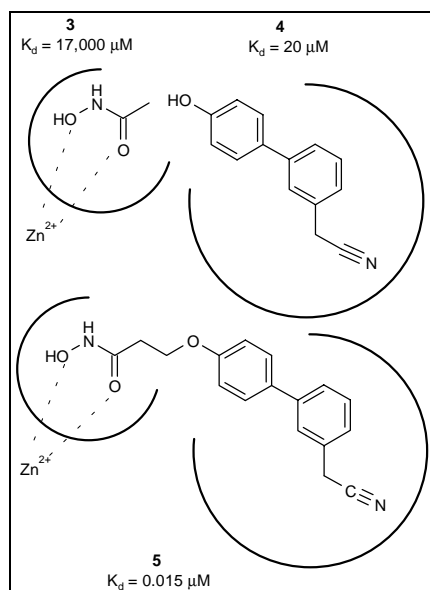
SAR by NMR identifies ligands that bind to proximal subsites of a protein. Acetohydroxamic acid, **3**, and 3-(cyanomethyl)-4'-hydroxybiphenyl, **4**, are low-affinity ligands of the matrix metalloproteinase, stromelysin. Combining them with an appropriate linker produces the high-affinity inhibitor, **5** [47•].

- Large amounts (> 200 mg) of pure ^{15}N-labeled protein are required.
- Sufficient aqueous solubility (~ 2 mM) and stability of the protein, also in the absence of an inhibitor (which is sometimes a problem, especially for proteinases), are preconditions for the NMR measurements.
- The ligands must have sufficient aqueous solubility and stability.
- The assignment of the -NH- signals can take weeks or even months (the 3D structures of the proteins need not be known).
- Ligands for different subsites must be discovered.
- The second subsite should be closely adjacent to the first subsite in order to avoid linkers which are too large.
- A linker which connects the two low-affinity ligands in a relaxed conformation must be designed.
- The linker itself must not have any negative influence on binding affinity.

Alternatives to the SAR by NMR method for large proteins are 1D NMR methods that exploit the changes in relaxation or diffusion rates of small molecules upon binding to unlabeled proteins [46•,50-52]. Different organic solvents have been used to identify specific ligand binding sites on protein surfaces by observing the transfer NOEs to the protein [53].

Another alternative to the SAR by NMR method is the multiple solvent crystal structures (MSCS; Brandeis University, MA, USA) approach [54,55•]. Protein crystals are

protein and the solvent molecule complexes are determined by protein crystallography in order to detect specific binding sites. Although such measurements take only a few days, there is no clear evidence available to suggest that this approach could be as widely applicable as the SAR by NMR method. In addition, no inhibitor has yet resulted from the application of this technique without independent information from other sources.

Computer-aided ligand design

Whereas structure-based design can be regarded as the predominant strategy of the last decade [1,2•,3•,4••,5••], several computer-assisted methods have been developed more recently. If several thousands of candidates, from large structural databases, are to be tested for their suitability as ligands of a certain binding site, molecular modeling [56••] can no longer be performed manually. The design process needs to be automated. The methods of choice for this purpose are computer programs that superimpose molecules by a flexible alignment to derive pharmacophoric patterns and/or quantitative structure-activity relationships, dock molecules to the surface of a protein 3D structure or to a hypothetical pseudoreceptor, or construct new ligands within a predefined binding site [57•,58••].

Automated flexible superposition of molecules

Methods for the alignment of rigid molecules are well established. A simple strategy to perform the alignment of flexible molecules involves the generation of multiple conformations of each compound by a knowledge-based approach (using torsion angle libraries from small-molecule crystal structures), to rank them by an energy function, and to superimpose all of the different pairs of low-energy conformations [59]. Different molecular property fields, such as electrostatic, steric, hydrophobic, hydrogen bond acceptor and donor fields, as well as their weighted combinations, have been used to achieve a fully automated alignment of the molecules. MIMIC (Pharmacia & Upjohn, USA) is a program that matches steric and electrostatic fields to guide the superposition; in a preprocessing step, similar conformations of a molecule are clustered [60]. MIMIC has also been extended to multimolecule alignments [61]. Another approach for the consideration of ligand flexibility starts from conformationally rigid ligands using different template conformations for the superposition of the molecules [62].

The much more demanding flexible superposition of one molecule onto another has been achieved only recently. The GASP program (University of Sheffield, UK) uses a genetic algorithm [63•] to consider conformational flexibility in the optimization of the alignment of a set of molecules [64]. A recent development for time-efficient flexible superposition of pairs of molecules is the computer program, FlexS (German National Research Centre for Information Technology (GMD), Germany), which resulted from a modification of the docking program, FlexX (see the section

protein crystallography) by dissecting the test molecule into rigid fragments, selecting a base fragment to start the alignment, and re-assembling the molecule in a low-energy conformation which fits the template molecule [65••]. The alignment is speeded up by first searching for correspondences of intermolecular interaction centers. A further acceleration comes from the transformation of the Gaussian property functions into Fourier space [66]. FlexS gives reasonable alignments of highly flexible molecules within a few minutes [65••], ie, at least one order of magnitude faster than most other automated programs for flexible alignment.

Docking

Several computer programs for molecular docking have been described within the last years [63•,67•,68••]. The first computer-assisted approach to the discovery of ligands for a given binding site was the program DOCK (UCSF, CA, USA) [69•]. In its original version, DOCK searched in 3D databases for ligands that would fit into a binding cavity based merely on the geometric properties of a certain rigid conformation. Later, the complementarity of other properties was considered. DOCK frequently permits the discovery of micromolar ligands that can serve as lead structures for further development. The latest refinement to DOCK was a significant speed-up of the program [69•]. Molecular docking to ensembles of 3D structures of the same protein allows an indirect consideration of target flexibility [70]. In a recent application, selective micromolar inhibitors of *Pneumocystis carinii* dihydrofolate reductase were derived from a DOCK database search, including > 50,000 molecules from the Fine Chemicals Directory (now Available Chemicals Directory, MDL, CA, USA) [71].

The computer program GRID (University of Oxford, UK) calculates interaction energies between proteins and different probes that are positioned around the surface of the protein [72•]. Porphyrins were superimposed by using a new option of the program that takes into account the flexibility of the propionic acid side chains. Each of the investigated analogs could be correctly placed into the heme binding site of myoglobin [73].

FlexX (GMD, Germany) is an efficient and fast docking program [74••,75••] that starts by dissecting the ligands into rigid fragments. One or several base fragments are selected, either manually [76] or automatically [77], and positioned in favorable orientations within the binding site. Other fragments are added in the next steps, using a tree-search technique for placing the ligand incrementally into the binding site (Figure 3). Only low energy conformations are created, and the different results are ranked according to favorable interaction energies using the scoring function of the *de novo* design program LUDI (see the section on *de novo* ligand design). The program FlexX has been validated by the successful reproduction of the experimental binding modes of 19 ligand-protein complexes [74••,77]. Further extensions will include the combinatorial design of ligands from series of building blocks [Lengauer T, GMD, personal

Following the concepts of a genetic algorithm alignment program [63•] and some strategies of FlexX, the program GOLD (Genetic Optimization for Ligand Docking; Cambridge Crystallographic Data Centre, UK) was developed. For 100 ligand-protein complexes, extracted from the Brookhaven Protein DataBank, GOLD achieved a 71% success rate in identifying the experimental binding mode [78•]. DOCK was also extended to a program that explores ligand flexibility by selecting an anchor fragment of a ligand, positioning it in the binding site, and adding the other parts of the molecule to generate the ligand in a low-energy conformation that fits the binding site [79•].

Different search algorithms for molecular docking have been compared; the results indicate that several different approaches are effective and give satisfactory performance [80a,80b]. An interesting endeavor was discussed during the docking session of the *Second Meeting on the Critical Assessment of Techniques for Protein Structure Prediction* in Asilomar, California, in December 1996. A total of 77 predictions were made by nine groups for the docking of seven small molecules into their binding sites. Overall results were good, with at least one prediction for each target within 3 Å root-mean-square deviation (RMSD), and within 2Å RMSD for over half the targets [81••]. Four groups were invited to describe their experiences in the competition, in separate publications [82-85].

De novo ligand design

De novo design methods have been extensively reviewed [86••,87••,88•,89••,90•]. The first *de novo* design program GROW (Upjohn Laboratories, USA) [89••] started from a simple seed fragment, eg, an amide group that is capable of interacting with the binding site, and continued by adding different amino acids in different conformations, to this fragment. Only the best candidates were selected and the same procedure was repeated several times until a peptide of a certain size had been generated in the binding site.

The *de novo* design program LUDI (BASF, Germany and MSI, CA, USA) [87••] constituted a significant improvement and nowadays, it is the most widely distributed software for computer-aided ligand design. After the definition of a binding site region by the user, the program automatically identifies all of the hydrogen bond donor and acceptor sites, as well as the aliphatic and aromatic hydrophobic areas, of this region of the protein surface. From the program-implemented information on the geometry of interaction of such groups with a ligand, the program creates vectors and regions in space where complementary groups of a ligand should be located. In the next step, LUDI searches databases of 3D structures of small and medium-sized molecules for potential ligands. Each candidate is tested in all possible different orientations and interaction modes. After a rough evaluation by counting the number of interactions and by checking for unfavorable van-der-Waals overlap between the ligand and the protein, the remaining candidates are prioritized by a simple but efficient scoring function which estimates interaction energies on the basis of charged and neutral hydrogen bonding energies, hydrophobic contact areas

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.