Serendipity meets precision: the integration of structure-based drug design and combinatorial chemistry for efficient drug discovery

F Raymond Salemme, John Spurlino and Roger Bone

Structure-based drug design uses three-dimensional visualization of drug candidates bound to a target receptor to direct structural modifications that increase potency. This widely used approach is limited by the difficulty of accurately predicting drug-binding affinities from three-dimensional structures. The integration of structure-based drug design with combinatorial chemistry can overcome this limitation by providing an empirical understanding of drug-binding energies. This integration allows compound synthesis and evaluation in parallel, and also helps assure that the compounds produced have properties consistent with good bioavailability and safety.

Address: 3-Dimensional Pharmaceuticals, Inc., Eagleview Corporate Center, 665 Stockton Drive, Suite 104 Exton, PA 19341, USA.

Corresponding author.
E-mail: salemme@3dp.com

Electronic identifier: 0969-2126-005-00319

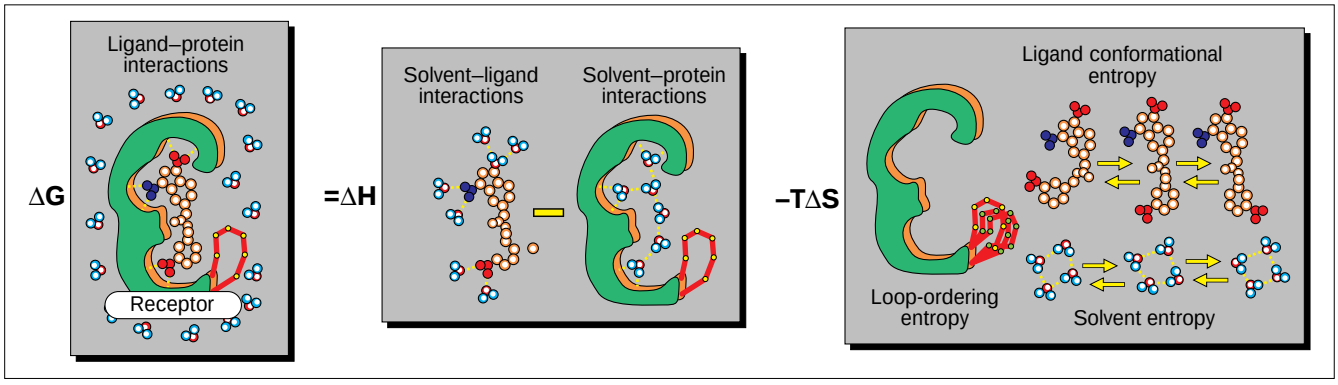
Structure 15 March 1997, 5:319–324

Structure-based drug design uses the three-dimensional structures of complexes made up of a receptor and a 'lead', or candidate drug, as determined by X-ray crystallography or NMR methods, to direct the atom-by-atom modification and improvement of drug leads [1]. Although structure-based design has become widely used in pharmaceutical discovery, fundamental technical issues still limit its efficiency.

Determination of the initial drug-lead structure is just the first step of the drug-optimization process. Optimization is rate limited by the requirement for many rounds of synthesis, activity characterization, and structural analysis in order to map the detailed binding-site energetics for the interactions between the drug and its target protein. It is not possible to simply translate a three-dimensional view of the drug bound to its target protein into new compounds with improved potency, because of the physical complexity underlying the free-energy change that accompanies the binding of a ligand to a macromolecule in water (Fig. 1) [2–4]. Ligand binding involves a multiplicity of factors, including changes in intermolecular interactions between the ligand, the solvent water and the target protein, as well as changes in ligand or receptor polarization, conformation or flexibility. The accurate estimation of how these changes occur and how they affect ligand binding lies beyond the practical capabilities of present computational methods. Consequently, investigators typically carry out iterative synthesis and structural analysis of an extensive series of compounds in order to define the important aspects of ligand-binding energetics empirically, and so to refine the properties of drug candidates.

Furthermore, structure-based design is limited by its focus on the enhancement of drug potency and specificity: it provides little direction for ensuring or improving *in vivo* properties, such as bioavailability or safety. These latter properties, which often determine whether a drug candidate is ultimately developed, are exceedingly difficult to predict quantitatively. Thus far, structure-based drug

Figure 1



A schematic showing the structural correlates of the components of free-energy changes associated with the binding of a drug (ligand) to a macromolecule (receptor) in solvent. Although X-ray crystallography can visualize these effects, accurate estimates of the energetics involved present formidable computational challenges.

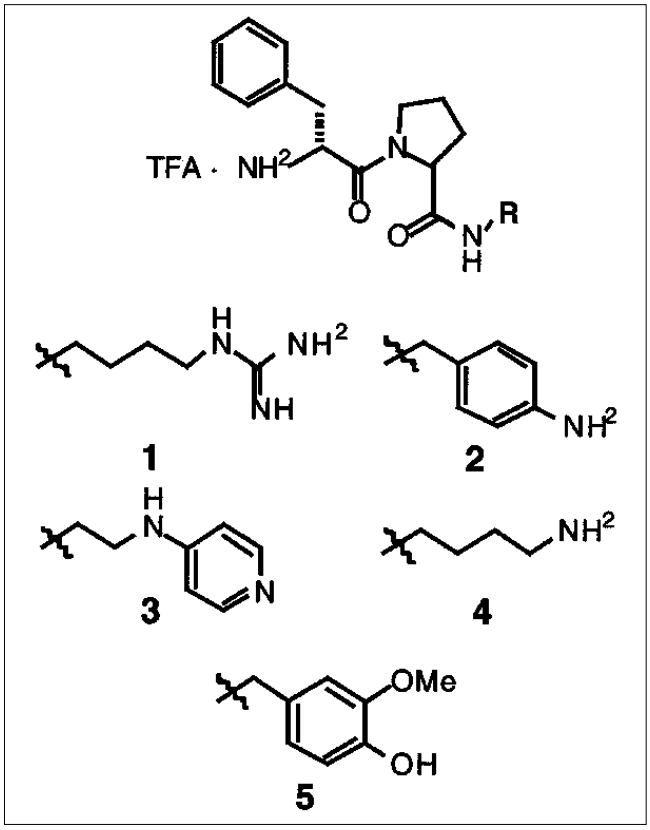
Table 1

K_i values (μM) of thrombin combinatorial library members based on D-Phe–Pro-test scaffold.

	Test group (X)*				
Protease	1	2	3	4	5
Thrombin	0.35	2.6	4.1	2.9	15
Plasmin	(0)†	(0)**	(0)¶	15	(0)††
Urokinase	(0)†	(0)§	(0)#	(0)‡	(0)‡‡
FXa	(0)†	(0)**	(0)¶	31	(0)§§
Chymotrypsin	41	110	(0)	(0)‡	254
Trypsin	4.3	(0)**	(0)#	19	431
Elastase	(0)†	(0)§	(0)#	(0)‡	-

*Test groups (X) are given in Figure 2. †Screened at 6 μM; ‡13 μM; §16 μM; #19 μM; ¶60 μM; **90 μM; ††140 μM; ‡‡1 μM; §§70 μM.

Figure 2



Test groups (X) capable of interacting within the guanidino-binding site (S1 site) of thrombin, when attached to a D-Phe–Pro scaffold.

designers have mainly used structural information to introduce chemical groups which are known from medicinal chemistry experience to aid bioavailability or safety in ways that will not adversely affect binding affinity.

Combinatorial chemistry approaches are based on the development of highly reliable reaction protocols that allow the synthesis of vast numbers of compounds through the exhaustive combination of all possible reactions at two or more sites [5,6]. The development of highly reliable chemical reactions also enables the simultaneous automated synthesis of up to 10^3 individual compounds. The power of combinatorial chemistry to allow the parallel synthesis of compounds forms an effective complement to structure-based design methods. This integration can be accomplished by placing the process under computer control [7], which enables the simultaneous refinement of multiple drug properties. An overall view of how this complementarity works can be illustrated by outlining its application to a long-standing problem in drug discovery, the development of orally active, small-molecule thrombin inhibitors.

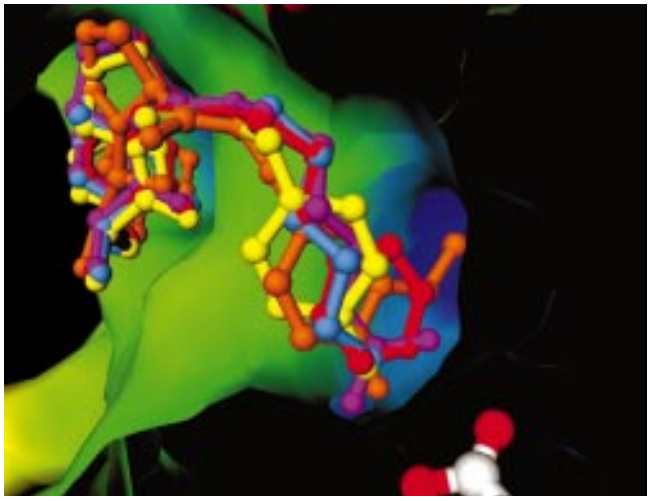
The development of inhibitors of thrombin

Thrombin is a serine protease that plays a central role in thrombosis: it functions as the key inducer of fibrin-clot formation, the most potent activator of platelets, and also amplifies its own production. Despite an extensive knowledge of thrombin structure and function and the availability of inhibitors that act at picomolar concentrations [8–11], the development of safe, orally active agents has proved an elusive objective. Development has been hindered by problems relating to lack of specificity [12] and poor oral bioavailability.

Combinatorial libraries targeted to inhibit enzyme or receptor function can be either based on transition-state analog approaches that exploit the active-site catalytic machinery [13,14], or geometrically directed to make specific interactions with substrate or cofactor subsites. A general drawback of the transition-state analog approach for drug design is that whatever binding energy is gained through active-site interactions, it is often accompanied by a commensurate loss in specificity towards other enzymes that share the same catalytic mechanism. Although approaches to drug development based on transition-state analogs have led to potent inhibitors of serine proteases, including thrombin [8,9,11], lack of specificity [12] and poor bioavailability have thus far prevented the successful development of these compounds as drugs. However, serine proteases typically have well defined binding subsites associated with substrate-recognition specificity, which can potentially provide the basis for potent and specific inhibitors.

The development of novel combinatorial chemistry 'scaffolds' or core structures that can be used to investigate subsite-binding energetics has similar problems to those encountered in designing drugs *de novo*, directly from the three-dimensional protein structures [1,15]. Although the ability to produce many compounds through parallel synthesis greatly increases that rate at which individual compounds can be synthesized and evaluated [7,16], the development of broadly reliable chemistry to allow

Figure 3

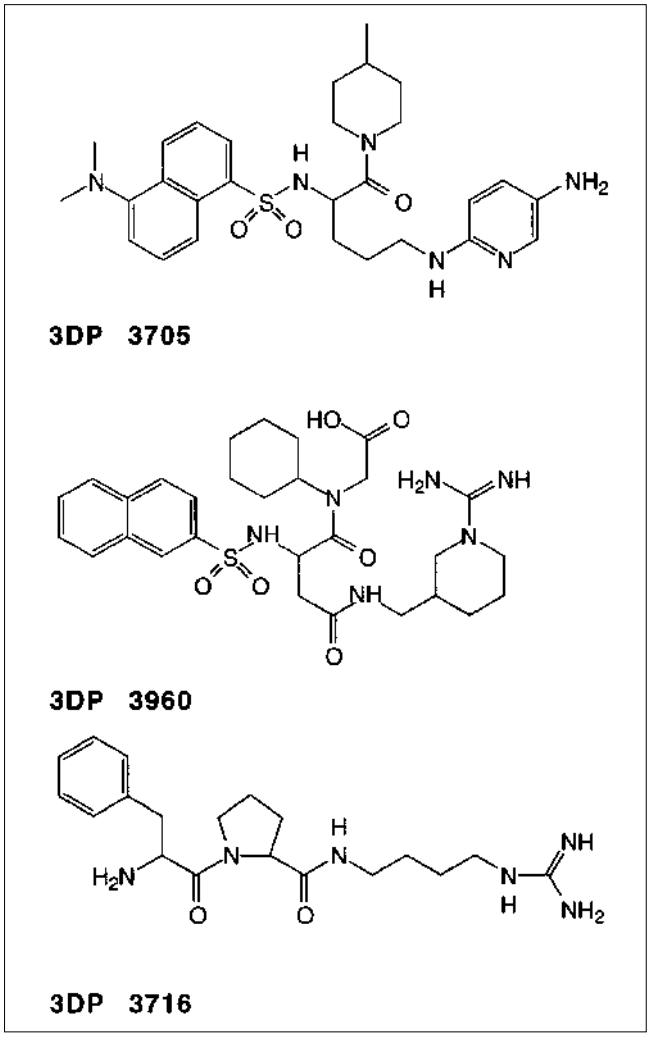


Superposition of crystal structures of thrombin complexes with D-Phe–Pro–X thrombin combinatorial library members based on D-Phe–Pro-test scaffold, on the accessible surface of thrombin. Chemical structures of X groups are given in Figure 2. (Compounds are coded by color: 1, magenta; 2, yellow; 3, red; 4, blue; and 5, orange). (RB and JS, unpublished observations.)

production of large combinatorial libraries is itself a time consuming process. Consequently, 'test scaffolds' are frequently used for initial evaluation of functional groups. These scaffolds allow the rapid development of structure-activity-relationships, even though they are ultimately unsuitable for use as drugs.

Table 1 outlines the results of an experiment using a combinatorial library D-Phe–Pro–(X), based on a D-Phe–Pro-test scaffold. The poor bioavailability of thrombin inhibitors that incorporate a guanidinium group may be related to this group's very basic p*K*a. Consequently, the objective of the experiment was to discover groups (X) that are capable of interacting with the guanidino-binding site (S1 site) of thrombin and which would also provide selectivity for thrombin over other serine proteases with basic S1 specificity (e.g. plasmin). In this experiment, X groups were chosen using a defined computational filter applied to a database of commercially available amines (X groups are given in Fig. 2). The computational filter was designed to identify groups that provided a combination of geometrical, space-filling, and hydrogen-bonding properties which, on the basis of the thrombin structure, could function as guanidine analogs. It is most notable that several of the groups identified, although less potent inhibitors than those containing the guanidinium group, are considerably less basic, so they might be expected to have better bioavailability. Figure 3 shows the superposition of several of these structures determined crystallographically, which generally agree with the expectations of the computational-selection process [16].

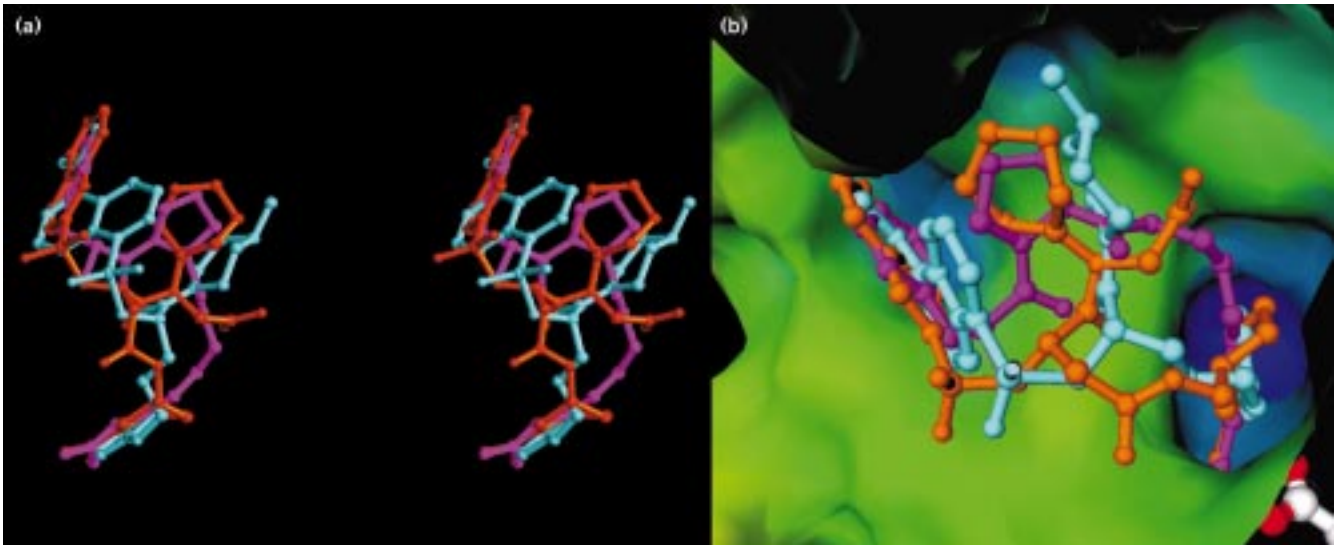
Figure 4



Chemical structures of representative scaffolds for protease-inhibitor design [10,16].

Development of a scaffold that can generate nonpeptidic, drug-like compounds is aided by the structural examination of as large a set of chemically different inhibitors as is readily accessible. In the case of thrombin inhibitors, the list of compounds is extensive and structural data are now available for many of them [10,17–23]. Figure 4 illustrates some representative structural classes of thrombin inhibitors, while Figure 5 shows a superposition of their three dimensional structures determined crystallographically. All of the inhibitors share a basic functional group that binds in the enzyme S1 subsite, and a hydrophobic moiety that binds in the 'aryl subsite', which has been previously identified in a number of thrombin inhibitor studies [10,17–23]. In the current context, the key observation is that the connecting residues play the structural role of scaffolds which differ in the manner in which they project substituents into the aryl and S1 subsites. This

Figure 5



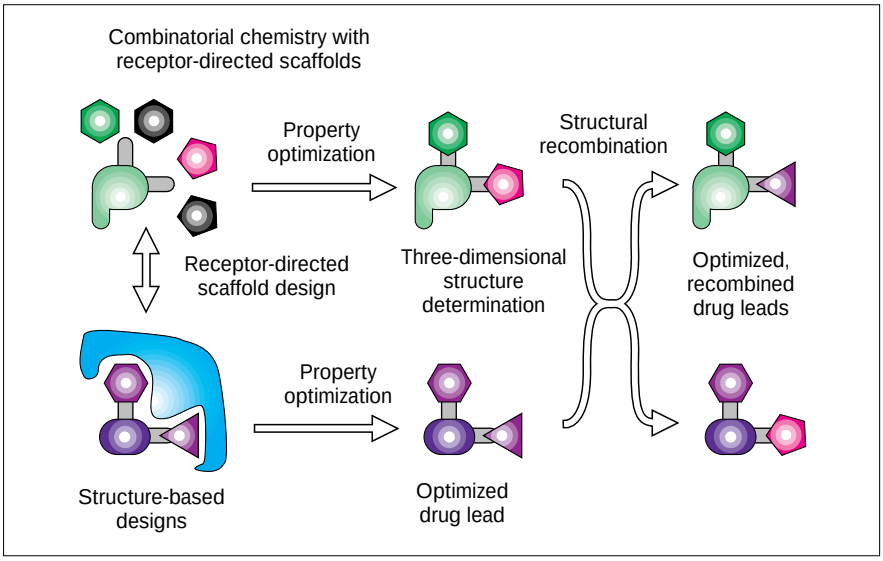
Superposition of representative scaffolds for protease-inhibitor design. Chemical structures are given in Figure 4. (a) Stereo superposition of 3705 (cyan), 3716 (magenta) and 3960 (orange). (b) Superposition of the same three structures on the accessible surface of thrombin.

information is key for two reasons: firstly, it facilitates the design of new scaffolds with more desirable properties as drugs; and secondly, it shows how substituents that have been identified from combinatorial synthesis on a given scaffold can be recombined with other scaffolds to produce improved compounds (Fig. 6).

Virtual libraries

Well developed combinatorial library strategies that are able to explore detailed structure–activity relationships generally require multistep synthetic protocols. These protocols have typically been carried out by modifying scaffolds that are reversibly linked to a solid-phase support, such that reactions are driven to completion and isolation of synthetic intermediates is facilitated. Scaffolds are designed to be elaborated at two or more substitution sites by reaction with custom or commercially available reagents. Development of the structure–activity relationship generally requires the high-fidelity synthesis of individual compounds. Although targeted-combinatorial schemes readily

Figure 6



Schematic of the interaction between structure-based drug design and combinatorial chemistry for drug discovery.

produce libraries of drug-like compounds with 10^6 or more members, current automated equipment is practically limited to the synthesis of 10^2–10^3 compounds for each multistep synthesis cycle. This practical limitation on synthetic throughput motivates the development of effective computer-search strategies that can iteratively define and refine the selection of sub-libraries that will best investigate the structure–property relationships for a given library–target combination.

The most effective way to implement the search of large combinatorial libraries is to precompute the structures and properties of all synthetically accessible compounds within the library and then to apply computational screens to this 'virtual library', in order to select subsets of compounds for iterative rounds of synthesis and testing [7]. Structural features of individual compounds can provide useful criteria for ranking their predicted ability to bind the target receptor and/or to discriminate against binding homologous proteins. Although a great deal of effort has been expended in developing accurate automated-docking algorithms [15], these remain limited by difficulties in estimating the free energy of ligand binding, as outlined in Figure 1. Instead, approaches based on less precise and more rapidly computed molecular properties seem effective in selecting members of receptor-targeted libraries for investigation of structure–activity correlations for a given receptor [24–26].

Selection of compounds for each round of synthesis can be made on the basis of chemical diversity, similarity to previously discovered ligands ('hits'), performance predicted on the basis of structure–activity models, predicted ability to fit the three-dimensional structure of the target, or the incorporation of substructures from known hits [7]. Recently, a computational method based on the use of a genetic algorithms was successfully used to predict which compounds in a 160000-member combinatorial library based on a single pot synthesis (Ugi reaction) would inhibit thrombin [27]. In 20 cycles of 20 Ugi reactions, the effective inhibitory concentration was decreased from around 1 mM to less than 1 μM. Genetic algorithms have been used in a similar manner to identify optimal stromelysin substrates from a hexapeptide library of 64 million members [28].

In addition to selecting molecules to explore specific features suggested from a three-dimensional structural model, computational filters can be designed to simultaneously select molecules with properties that are thought to be important in the ultimate development of a successful drug (e.g. properties relating to bioavailability, toxicology, etc.). Filters can be designed on the basis of computed properties of the structures, such as computed water–octanol partition coefficients (logP values), so that the distribution of logP values of the sub-library selected for synthesis, corresponds to a range of known values for orally acting drugs. In principle, any known property or structure–activity relationship can be introduced as a selection criterion and can be used as part of a multi-parameter optimization scheme for a given combinatorial sub-library. A typical library-design optimization might identify a set of 1000 compounds from a virtual library containing 500000 compounds that fulfill the following parameters: meet geometrical and functional criteria for fitting a target receptor; meet structural criteria aimed at eliminating interactions with homologous targets; maintain a desired range of logP values; have less than four freely rotating bonds; and minimize the cost of reagents.

Structure-based recombination of optimized substituents from independently developed drug-lead series or structure–activity correlations has proved a reliable means of generating nanomolar potent inhibitors of thrombin that show outstanding selectivity [10]. Combinatorial library technology has the ability to collect this information much more rapidly. By simultaneously tracking physical properties of library constituents, it should be possible to maintain a range of physical properties that are perceived to be key in achieving high oral bioavailability [7,16], while optimizing *in vitro* potency. These same recombination principles can be used to create large virtual libraries of protease inhibitors, that can then be systematically investigated to discover potent and specific inhibitors for a variety of proteases of therapeutic interest. The ability to complement the precision approach of structure-based design with the sampling capability of combinatorial chemistry removes many of the limitations of structure-based design and productively focuses the combinatorial selection process. Key to the evolving effectiveness of this integrated technology is the continued development of mathematical approaches that will optimally blend the structural requirements defined by receptor-binding sites with other chemical and physical properties that affect bioavailability and toxicity. Indeed, the most interesting prospects for computer-directed combinatorial chemistry lie in its ability to investigate multiple structure–activity models in parallel, thus overcoming the one-at-a-time throughput limitations of the conventional process of analog synthesis, testing and redesign that characterize both structure-based design and conventional analog chemistry approaches to drug discovery.

The integration of structure-based design and combinatorial chemistry can play a key role in the development of new drugs whose molecular targets are being discovered as a result of genome sequencing efforts. In many cases, targets of potential therapeutic interest will be members of larger protein classes, some of whose members have known three-dimensional structures. The availability of structural information will enable the development of targeted combinatorial libraries, based on pharmacophores that are known to be active for the target class. These libraries can

be searched using the same kinds of strategies as those outlined here for thrombin, in order to produce new pharmaceuticals with the specificity for a receptor or enzyme sub-type that is required for a safe and efficacious drug.

Acknowledgments

The authors acknowledge Richard Soll, Bruce Tomczuk, Tianbao Lu, Nalin Subasinghe, Todd Graybill and Carl Illig for their contributions to work outlined in this review.

References

1. Bugg, C.E., Carson, W.M. & Montgomery, J.A. (1993). Drugs by design. *Sci. Am.* 269, 60–66.
2. Weber, P.C., Wendoloski, J.J., Pantoliano, M.W. & Salemme, F.R. (1992). Crystallographic and thermodynamic comparison of natural and synthetic ligands bound to streptavidin. *J. Amer. Chem. Soc.* 114, 3197–3200.
3. Weber, P.C., Pantoliano, M.W., Simons, D.M. & Salemme, F.R. (1994). Structure-based design of synthetic Azobenzene ligands for streptavidin *J. Amer. Chem. Soc.* 116, 2717–2724
4. Wolfenden, R. (1983). Waterlogged molecules. *Science* 222, 1087–1093.
5. Gordon, E.M., Barrett, R.W., Dower, W.J., Fodor, S.P.A. & Gallop, M.A. (1994). Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies and future directions. *J. Med. Chem.* 37, 1385–1400.
6. Terrett, N.K., Gardner, M., Gordon, D., Kobylecki, R.J. & Steele, J. (1995). Combinatorial synthesis: the design of compound libraries and their application to drug discovery. *Tetrahedron* 51, 8135–8173.
7. Agrafiotis, D., Bone, R.F., Salemme, F.R., & Soll, R.M. (1995). System and method of automatically generating chemical compounds with desired properties, US Patent 5,463,564.
8. Kettner, C., Mersinger, L.M. & Knabb, R. (1990). Selective inhibition of thrombin by peptides of boroarginine. *J. Biol. Chem.* 265, 18289–18297.
9. Costanzo, M.J., *et al.*, & Tulinsky, A. (1996). Potent thrombin inhibitors that probe the S1′ subsite: tripeptide transition state analogues based on a heterocycle-activated carbonyl group. *J. Med. Chem.* 39, 3039–3043.
10. Hilpert, K., *et al.*, & van de Waterbeemd, H. (1994). Design and synthesis of potent and highly selective thrombin inhibitors. *J. Med. Chem.* 37, 3889–3901.
11. Lim, M.S.L., Johnston, E.R. & Kettner, C.A. (1993). The solution conformation of (D)Phe-Pro-containing peptides: implications on the activity of Ac-(D)Phe-Pro-boroArg-OH, a potent thrombin inhibitor. *J. Med. Chem.* 36, 1831–1838.
12. Kettner, C., *et al.*, & Wexler, R. (1996). The preparation of safe thrombin inhibitors by designing compounds selective over Factor I. 212th ACS National Meeting Abstracts, Division of Medicinal Chemistry, No. 112.
13. Campbell, D.A., Bermak, J.C., Burkoth, T.S. & Patel, D.V. (1995). A transition state analog inhibitor combinatorial library. *J. Am. Chem. Soc.* 117, 5381–5382.
14. Kick, E.K. & Ellman, J.A. (1995). Expedient method for the solid-phase synthesis of aspartic acid protease inhibitors directed toward the generation of libraries. *J. Med. Chem.* 38, 1427–1430.
15. Kuntz, I.D., Meng, C.E. & Shoichet, B.K. (1994). Structure-based molecular design. *Accounts Chem. Res.* 27, 117–123
16. Graybill, T.L., *et al.*, & Salemme, F.R. (1996). Enhancing the drug discovery process by integration of high-throughput chemistry and structure-based drug design. Molecular diversity and combinatorial chemistry: libraries for drug discovery. ACS Monograph, American Chemical Society, Washington, 16–27.
17. Engh, R.A., *et al.*, & von der Saal, W. (1996) Enzyme flexibility, solvent and 'weak' interactions characterize thrombin-ligand interactions: implications for drug design. *Structure* 4, 1353–1362.
18. Weber, P.C., Lee, S.-L., Lewandowski, F.A., Schadt, M.C., Chang, C.-H. & Kettner, C.A. (1995). Kinetic and crystallographic studies of thrombin with ac-(D)Phe-Pro-boroArg-OH and its lysine, amidine, homolysine, and ornithine analogs. *Biochemistry* 34, 3750–3757
19. Obst, U., Gramlich, V., Diederich, F., Weber, L. & Banner, D.W. (1995). Design of novel nonpeptidic thrombin inhibitors and structure of a thrombin-inhibitor complex. *Angew. Chem.* 34, 1739–1742.
20. Mathews, I.I. & Tulinsky, A. (1995). Active-site mimetic inhibition of thrombin. *Acta Cryst. D* 51, 550–559.
21. Malley, M.F., *et al.*, & Sack, J.S. (1996). Crystallographic determination of the structures of human α-thrombin complexed with BMS-186282 and BMS-189090. *Protein Sci.* 5, 221–228.
22. Fevig, J.M., Abelman, M.M., Bittrelli, D.R., Kettner, C.A., Knabb, R. & Weber, P.C. (1996). Design and synthesis of ring-constrained boropeptide thrombin inhibitors. *Bioorg. Med. Chem. Lett.* 6, 295–300.
23. Banner, D.W. & Hadvary, P. (1991). Crystallographic analysis at 3.0 Å resolution of the binding to human thrombin of four active site-directed inhibitors. *J. Biol. Chem.* 266, 20085–20093.
24. Brown, R.D. & Martin, Y.C. (1996). Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sci.* 36, 572–584.
25. Martin, E.J., Spellmeyer, D.C., Critchlow, R.E. & Blaney, J.M. (1997). Does combinatorial chemistry obviate computer-aided drug design? *Rev. Comput. Chem.* 10, 75–100.
26. Martin, E.J., Blaney, J.M., Siani, M.A., Spellmeyer, D.C., Wong, A.K. & Moos, W.H. (1995). Measuring diversity: experimental design of combinatorial libraries for drug discovery. *J. Med. Chem.* 38, 1431–1441.
27. Weber, L., Wallbaum, S., Broger, C. & Gubernator, K. (1995). Optimization of the biological activity of combinatorial compound libraries by a genetic algorithm. *Angew. Chem.* 34, 2280–2282.
28. Singh, J., *et al.*, & Treasurywala, A.M. (1996). Application of genetic algorithms to combinatorial synthesis: a computational approach to lead identification and lead optimization. *J. Amer. Chem. Soc.* 118, 1669–1676.