Computer-Aided Design of Thrombin Inhibitors

Amedeo Caflisch, Rudolf Wälchli, and Claus Ehrhardt

Computer-aided ligand design is an active, challenging, and multidisciplinary research field that blends knowledge of biochemistry, physics, and computer sciences. Whenever it is possible to experimentally determine or to model the three-dimensional structure of a pharmacologically relevant enzyme or receptor, computational approaches can be used to design specific high-affinity ligands. This article describes methods, applications, and perspectives of computer-assisted ligand design.

Computer-based approaches are widespread and have applications not only in the efficient administration of already existing data but also in the design and planning of a variety of objects, from cars and airplanes to the exterior and interior of buildings, as well as in the clothing industry and for many projects of the advertisement and entertainment industries. One important application of computer science methods is the computer-aided design of ligand molecules for a given macromolecular target. For about the past 15 years, there has been a significant development and application of computer programs for calculation of optimal conformation(s) of molecules or (macro)molecular assem-

A. Caflisch is in the Dept. of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland; R. Wälchli and C. Ehrhardt are at the Novartis Pharma, Inc., CH-4002 Basel, Switzerland.

 0886-1714/98 5.00 @ 1998 Int. Union Physiol. Sci./Am.Physiol. Soc.

blies, as well as molecular modeling packages for interactive, real-time, three-dimensional computer graphics and chemical database management. These methods have been used and are being used not only by nonprofit research groups but also by many pharmaceutical and biotechnology companies. Computer-aided ligand design is concerned with the prediction of molecules that are expected to bind strongly, i.e., with high affinity, to key regions of pharmacologically relevant enzymes or macromolecular receptors (target macromolecules) so as to inhibit or alter their activity. The ultimate goal is the discovery of a drug molecule to cure a given disease.

In the present context, the name "ligand design" is preferred to "drug design" because the determination of whether or not a ligand will result in a drug involves, besides the affinity of the ligand for its target, a series of pharmacological properties that the current computer-assisted design methods are not able to predict. These properties include transport of the drug to the regions of the organism where the target molecules are located, metabolic stability, selectivity, low toxicity, sufficient half-life, and minor addictive potential. In addition, there is a significant preference for orally available drugs, which limits the molecular mass of the ligands to a maximum of ~600 Da.

In this review, we present our methods for computer-aided ligand design and discuss an application to human thrombin. Thrombin is a serine protease that plays an important role in both hemostasis and thrombosis. In addition to its role in coagulation, thrombin has relevant biological effects on platelets, endothelial and smooth muscle cells, leukocytes, the heart, and neurons (12). Current anticoagulant therapy is limited to three classes of compounds: the heparins, the coumarins (e.g., warfarin), and more recently the low-molecular-weight heparins. Of these drugs, only the coumarins show significant activity when administered orally. All three classes act indirectly to inhibit thrombin. The heparins activate endogenous plasma proteins (notably antithrombin III and heparin cofactor II) that inhibit thrombin and other proteases of the coagulation cascade. The coumarins inhibit the hepatic synthesis of vitamin K-dependent proteins (including thrombin, factors VII, IX, X, and the natural anticoagulants, proteins C and S). These indirect mechanisms largely account for the limitations of these agents as therapeutics. In particular, patient variability necessitates, with both heparins and coumarins, careful dose titration to achieve the optimal therapeutic effect with minimal side effects. This is problematic with the vitamin K antagonists, which show interactions with food and other drugs and require several days for the fully inhibitory effect to manifest itself. The major side effect in clinical use is the hemorrhagic complications, which occur with a frequency as high as 10% and can even be fatal. Hence, it is not surprising that much effort is invested in finding better drugs, including directly acting thrombin inhibitors that offer the potential of a simplified usage and a broader efficacy over existing agents. Furthermore, the detailed understanding of the three-dimensional structure of thrombin and of thrombin-inhibitor complexes, obtained by X-ray crystallography, has motivated in our and other laboratories the computer-aided search and discovery of potent and selective new inhibitor molecules.

Other successful examples of structure-based ligand design include the following protein targets and related therapeutic goals (14): carbonic anhydrase (treatment of glaucoma), renin (treatment of hypertension), dihydrofolate reductase (antibacterial), neuraminidase (antiviral), HIV-1 aspartic proteinase (anti-acquired immunodeficiency syndrome), trypanosomal glyceraldehyde-3-phosphate dehydrogenase (antiparasitic), thymidylate synthase and purine nucleoside phosphorylase (anticancer), elastase (treatment of emphysema), collagenase (rheumatoid and osteoarthritis), phospholipase A_2 (anti-inflammatory), and glycogen phosphorylase (treatment of diabetes mellitus).

"Thrombin . . . plays an important role in both hemostasis and thrombosis."

Structure-based ligand design

Computer-aided ligand design is part of the research strategy called structure-based design, which is an iterative combination of experimental and computer-based approaches (Fig. 1). There are two different entry points in the design cycle: random screening of compound collections and identification of protein target. X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy are used to solve the three-dimensional structure of the protein target. It is important to note that computational techniques are also used during the final phase of experimental structure determination, i.e., the refinement of the data in protein crystallography and the generation of a set of conformations in solution in NMR spectroscopy.

When the structure of a protein-ligand complex is solved, the active site of the protein can be determined; if this is not possible, mutagenesis experiments are needed to help in delineating the binding pocket(s). Knowledge of the protein binding site is used for computer-assisted screening of the corporate three-dimensional database

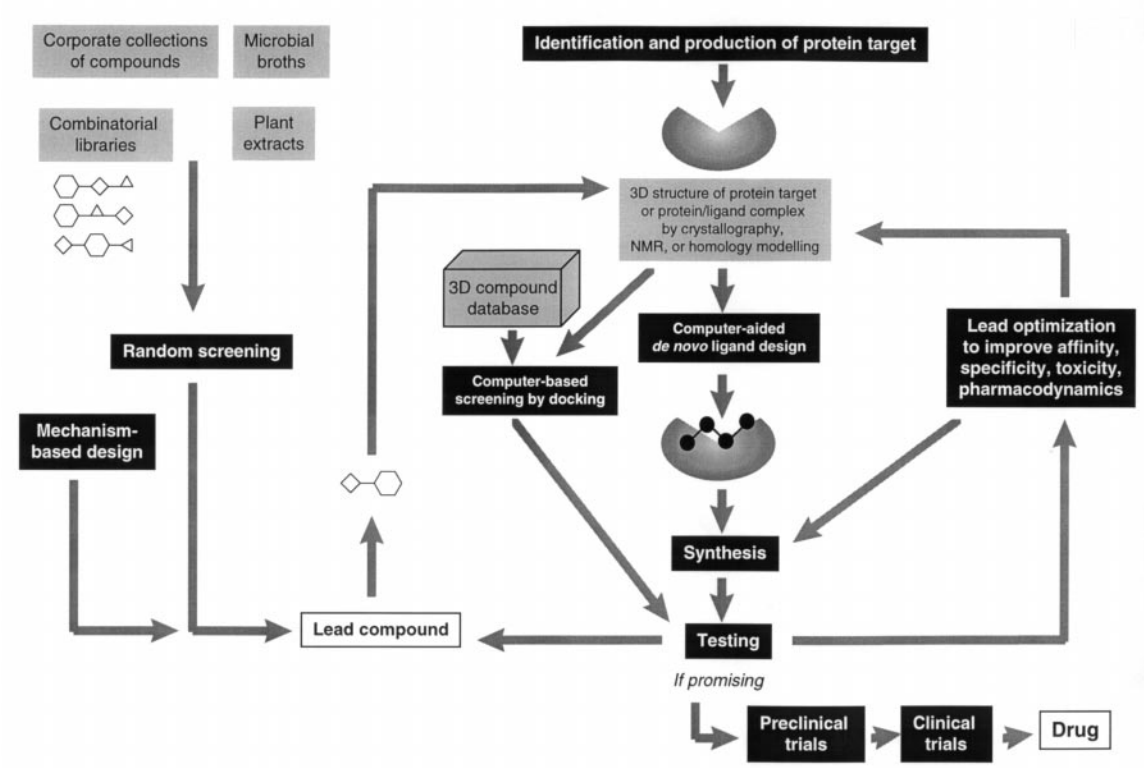

FIGURE 1. Structure-based drug design cycle. A lead compound is a molecule that binds to the protein of interest but often shows weak affinity or is toxic or unstable, yet it forms a starting point for development of better ligands with improved pharmacological properties. It can originate from random screening or from enzyme mechanism-based design or computer-aided design. There are 2 main approaches to structure-based computer-aided design: programs for 3-dimensional (3D) databases searches and procedures for de novo ligand design. 3D database search methods are advantageous because putative ligands retrieved from the database are available and do not have to be synthesized; however, molecule diversity is limited to the database used. De novo methods are virtually unlimited in diversity. They can be divided into 2 categories: programs that link isolated fragments and programs that perform a sequential buildup (8). Structure verification of the protein-lead complex is in many cases essential. New rounds of structure-based design are then performed until a promising ligand reaches preclinical trials. At this stage, the structure is still useful because knowledge of the protein-ligand interactions helps in predicting structural modifications to improve the pharmacodynamic properties without affecting binding potency. There are 3 phases of clinical trials before a compound is approved as a drug: *phase 1*, compound is tested on a small number of healthy volunteers (~100); *phase 2*, compound is administered to a relatively small number of patients (~1,000); *phase 3*, compound is given to a large number of patients (~10,000) in different centers. [Modified from C. Verlinde and W. Hol, *Structure* 2: 577–587, 1994.]

by fast-docking algorithms (6) or for de novo ligand design programs (1, 2, 4–8, 15). Both of these approaches can be used to generate hypothesis and ideas on binding modes for a series of molecules or molecular fragments. If a ligand is already available, either from broad-screening results of synthetic compound or natural product collections or from combinatorial libraries, it is helpful to solve the structure of the protein-ligand complex. Such structural information on the binding mode(s) and conformation of a known ligand is sometimes essential for the discovery of novel compounds with a better pharmacological profile.

Interactive molecular modeling programs are then used to visualize the protein-ligand complex. This is useful for new suggestions and for discussions with medicinal chemists. The novel or modified compounds are synthesized or purchased. Then, their binding properties are determined by biochemical assays and/or biological tests on cell cultures. After the experimental binding data are carefully analyzed, the structure of some interesting ligands in complex with the protein can eventually be solved. The next cycle of ligand design for optimization of binding affinity, physicochemical characteristics (e.g., solubility), and pharmacological properties (e.g., bioavailability) can start. This is again followed by synthesis, testing, and eventually further structural determination. The structure-based approach has the fundamental advantage that the information provided by every additional three-dimensional structure of a protein-ligand complex can be used *1*) to explain the results of the affinity data and *2*) to guide the computer-aided search for an improved ligand.

It is worth noting that there is always a very long time span between the discovery or design of a high-affinity ligand for a pharmacologically relevant protein target and the approval of the corresponding drug by the competent authority. The

vast majority of strong binding ligands are discarded because of one or more of the following problems: toxicity, teratogenicity, too rapid clearance, inability to reach the protein target in sufficient concentration, and instability in solution.

The cloning and sequencing of the human genome promise that an ever-increasing number of proteins will be defined as potential drug targets in the coming years. Advances in X-ray crystallography and NMR spectroscopy (11), the two most used experimental approaches for the determination of protein three-dimensional structure, as well as improvements in computer-based methods for homology modeling (9) will provide the structural information necessary for the design of novel therapeutic agents for the treatment of a variety of diseases. Novel methods, technologies, and approaches are needed to meet the requirements for innovative drugs and to shorten the time from lead discovery, optimization, and development to the clinical trials.

In this review, we focus on our approach to computer-aided ligand design. It is based on the docking of a diverse set of molecular fragments into the active site of a macromolecular target and on the use of a combinatorial strategy to connect them to form candidate ligands. This methodology is illustrated by an application to human thrombin, a trypsinlike serine protease that plays a central role in both hemostasis and thrombosis. The selective and direct inhibition of thrombin is expected to prevent thrombotic diseases without unwanted side effects.

Computer-aided ligand design

The approach we have chosen for computer-aided ligand design consists of three parts (3). First, an efficient method is used to exhaustively search for optimal positions and orientations of a set of diverse molecular fragments (e.g., benzene, indole, benzamidine, *N*-methylacetamide, 2,5-diketopiperazine, and benzodiazepine) on the surface of a macromolecular target (Fig. 2, *top*). For this procedure, the multiple copy simultaneous search (MCSS) procedure was developed (7). Several thousand replicas of each fragment type are randomly distributed inside a sphere with a radius large enough to cover the entire binding site. The replicas are then simultaneously minimized in the force field of the protein. This yields a set of functionality maps of the protein surface, i.e., an exhaustive description of the optimal positions for every fragment type (Fig. 2, *middle*). Such functionality maps are very useful for ligand design, since in the known crystal structures of enzyme-inhibitor complexes most if not all of the functional groups of ligands with high affinity are involved in favorable interactions (electrostatic, van der Waals) with the surrounding protein atoms.

Second, once a set of such positions and orientations for functional groups is found, it is necessary to find optimal connections between these fragments to form putative ligands (Fig. 2, *bottom*). For this purpose, we have developed a program for computational combinatorial ligand design (CCLD). This exhaustive computational approach is based on a series of aspects that have a parallel in the experimental methods of combinatorial chemistry. In a typical CCLD run, all possible ways to link together up to about seven optimally docked MCSS fragments are evaluated, i.e., in the order of 10^{12}–10^{17} compounds from a set of 50–300 MCSS fragment-linker combinations in each pocket of the protein target (2). A series of geometric checks are used to reduce the large number of structures generated by CCLD. Furthermore, the hits are clustered according to spatial and chemical similarity.

"...ways to link together up to about seven optimally docked MCSS fragments are evaluated...."

Evaluating the free energy of binding of the resulting candidates in the third step requires a sophisticated treatment of the interactions as well as a rigorous treatment of solvent (10) and entropic effects. At present, a somewhat crude procedure is used to rank the hits according to the average binding affinity of their fragments. More accurate but still efficient procedures are currently being developed and tested by us and other laboratories.

A stepwise procedure is used because this is more efficient than doing everything at once. It would take an inordinate amount of time to dock hundreds of thousands of molecules into the binding site and use the most accurate theoretical approach to evaluate their binding affinity. By first docking functional groups and then connecting them to form candidate ligands, it is possible to search through a very large number of highly diverse molecules in a relatively short time.

Ligand design programs are being developed at an ever increasing rate, and some are related to various aspects of our ligand design approach. The LEGO software tool is based on a combination of multiple-fragment docking, automatic connection by small linker units (1–4 atom chains), and search of three-dimensional databases for complementary molecules (5). Another related approach is that embodied in the program LUDI (1). This program uses statistical data from small-molecule crystal structures to determine binding sites of molecular fragments, i.e., discrete positions on the binding site surface suitable to form hydrogen bonds and/or fill hydrophobic sites of the receptor. Alternatively, LUDI uses simple rules or the output of the program GRID (15)

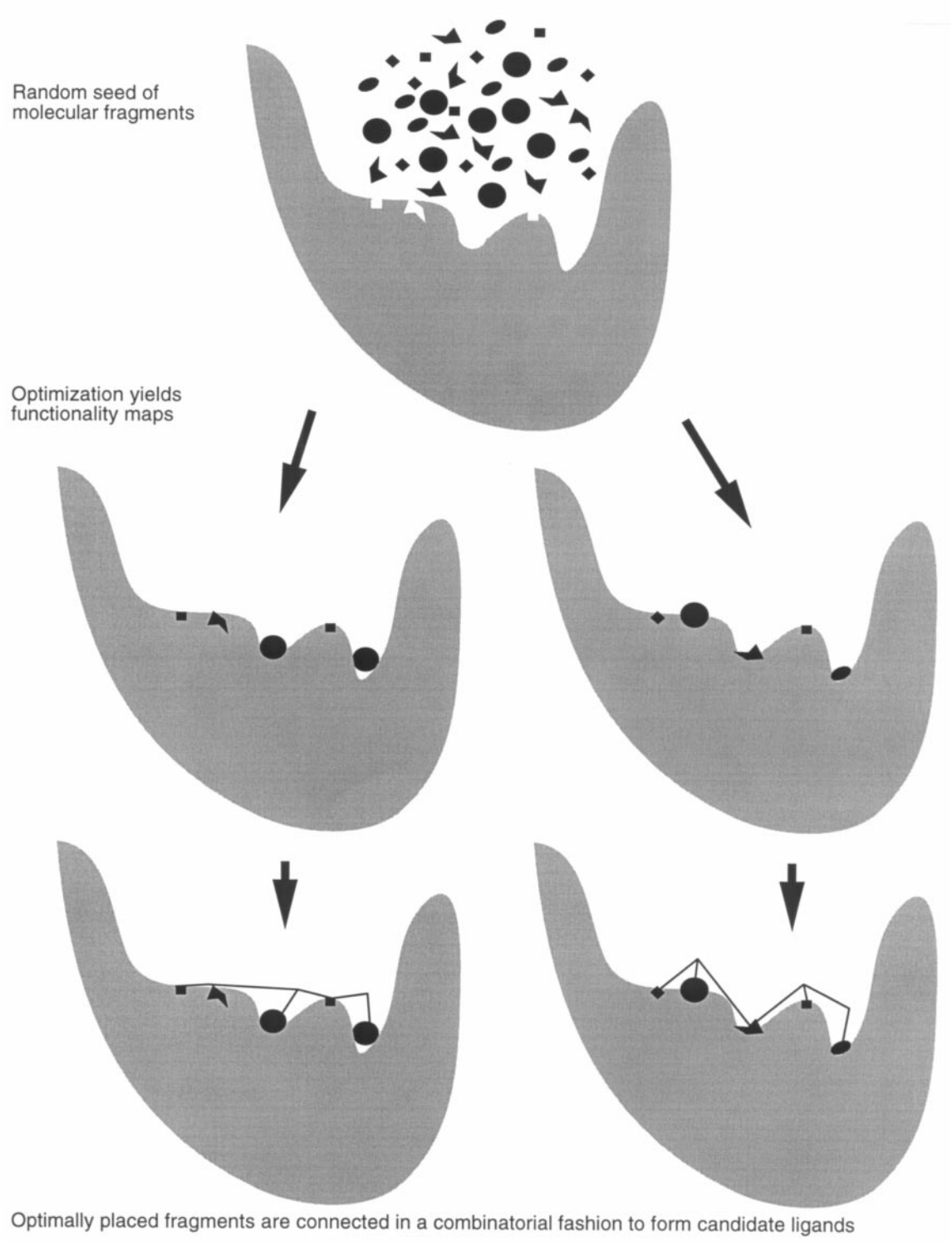

FIGURE 2. Simplified picture of the principles used for multiple copy simultaneous search (MCSS) and for computational combinatorial ligand design (CCLD). The protein target is represented by a gray shape, and molecular fragments are in black. The MCSS method determines energetically favorable positions and orientations (local minima of the potential energy) of functional groups on the surface of a protein or receptor of known 3-dimensional structure (*7*). A diverse set of molecular fragments consisting of charged, polar, aromatic, and aliphatic groups is used to map both the hydrophilic and hydrophobic regions of the protein. First, several thousand replicas of each fragment type are randomly distributed in a sphere centered on the region of interest (e.g., an enzyme active site for the design of inhibitors of catalytic activity). Replicas of the same group are simultaneously minimized in the force field of the protein. During minimization, interactions between the group replicas are omitted. Force on each replica consists of its internal forces and those due to the protein, which has unique conformation and therefore generates a unique field. The following procedures are performed during a regular execution of CCLD (*2*): *1*) MCSS minima are first sorted according to their approximated binding free energies, *2*) a list of bonding fragment pairs and a list of overlapping fragment pairs are then generated, *3*) a combinatorial generation of putative ligands follows, and *4*) ligands are then sorted and clustered.

to generate the interaction sites. Finally, the fragments fitted in the interaction sites can be connected by linker groups. Other fragment-based programs are CONCERTS, GroupBuild (8), and NEWLEAD (13). The fragment-linking strategies for computer-aided structure-based ligand design and other approaches, e.g., procedures that "grow" ligands by a sequential buildup, have been reviewed previously (3, 8).

Thrombin inhibitors

Thrombin is one of the best-characterized enzymes from a structural point of view. It is a trypsinlike serine protease that binds a series of diverse inhibitors without major rearrangements in its conformation, as shown by a number of X-ray crystallography studies. Its S3 and S2 precleavage subpockets have hydrophobic characteristics, whereas at the bottom of the S1 or recognition pocket the carboxy group of Asp-189 is a salt bridge partner for basic side chains. *N*α-[(2-naphthylsulfonyl)glycyl]-DL-*p*-amidinophenylalanylpiperidine (NAPAP; Fig. 3*A*) is the archetypal active site-directed inhibitor of thrombin. It binds reversibly to the thrombin active site by occupying the S3 pocket with its naphthyl group and the S2 pocket with the piperidine ring and by positioning its basic benzamidine moiety into S1 to form a salt bridge with Asp-189.

In continuation of a project aimed at the structure-based design of low-molecular-weight active site-directed inhibitors of human thrombin (12), MCSS was used to generate a series of functionality maps of the thrombin S3 to S1 pockets (2, 3). CCLD was then used to generate a set of candidate ligands. Many of them showed the same interaction patterns as those of known thrombin inhibitors, i.e., hydrophobic moieties in S3 and S2, hydrogen bonds with the polar groups of Gly-216, and basic groups in S1 (2). One of these putative ligands is shown in Fig. 3*B*. It is involved in the same interactions as in the NAPAP-thrombin complex except for the hydrogen bond with the CO of Gly-216. Its cyclohexane ring in S3 is connected to the five-membered ring in S2 by a single methylene linker. This was a novel design, and the candidate ligand appeared to be more rigid than NAPAP, since it has a smaller number of rotatable bonds. Hence, the penalty paid for the loss in entropy on binding should be smaller for this CCLD hit than for NAPAP. This connection represented an interesting new idea that we tried to include in the design of a novel thrombin inhibitor.

As a compromise between a compound that is expected to show reasonable binding affinity and a compound that is not too difficult to synthe-

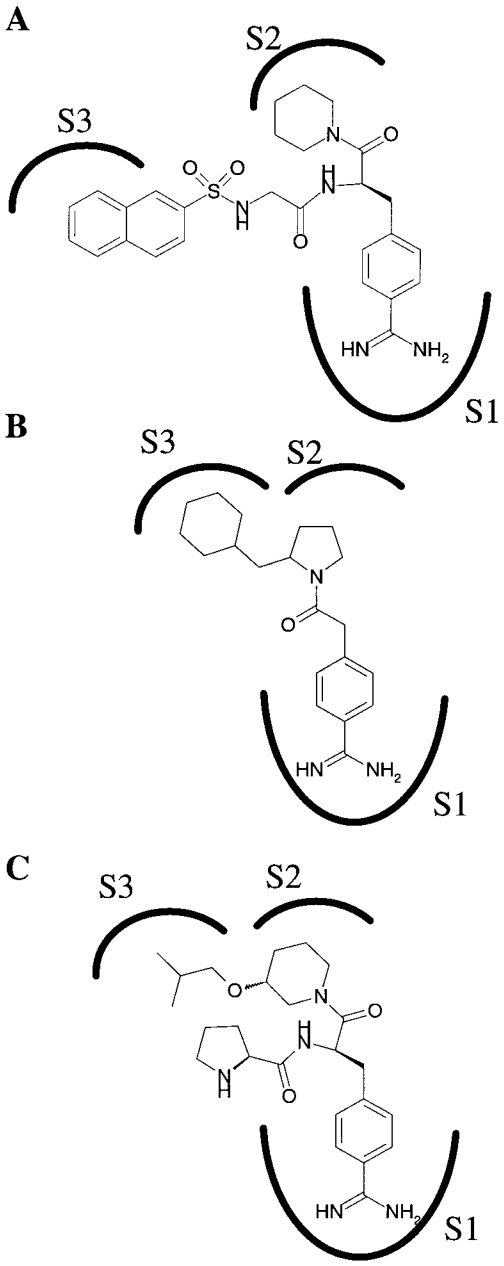

FIGURE 3. *A*: chemical structure of *N*α-[(2-naphthylsulfonyl)glycyl]-DL-*p*-amidinophenylalanylpiperidine (NAPAP). *B*: a candidate thrombin inhibitor obtained by a computational combinatorial ligand design (CCLD) run that used the multiple copy simultaneous search minima of benzamidine, benzene, cyclopentane, and cyclohexane. In few seconds of cpu time of an SGI Indigo2 (R4400 processor), CCLD generated a series of molecules that showed same interaction patterns as those of known thrombin inhibitors, i.e., hydrophobic moieties in S3 and S2, hydrogen bonds with the polar groups of Gly-216, and benzamidine in S1. The compound shown here is involved in same interactions as in NAPAP-thrombin complex except for the hydrogen bond with the CO of Gly-216. CCLD automatically converted one of the sp^3 carbons of the cyclopentane ring into an sp^2 nitrogen to obtain the secondary amide connection between cyclopentane and benzamidine, which should facilitate synthesis. *C*: thrombin inhibitor 1, whose racemic mixture has an inhibition con-

size, we suggested *ligand 1* (Fig. 3*C*), which is more "NAPAP-like" than the original CCLD proposal. In this molecule, we could attach only an isopropyl group to the ring in S2 and therefore D-Pro was added to better fill the S3 pocket. In addition, both hydrogen bonds to the backbone of Gly-216 were kept. *Compound 1* was synthesized and tested in a binding assay and found to be a relatively potent and selective inhibitor of thrombin. Its racemic mixture has an inhibition constant (K_i) value of 1.7 μM for thrombin and is inactive against plasmin. Although *compound 1* is not as potent as NAPAP, which has a K_i value of 9 nM, it can be used for further derivatization.

Conclusion and future developments

"...the prediction of the optimal docking of large and flexible ligands...."

Computational approaches have been used (14) and will contribute to the future success of structure-based ligand design. However, a lot of room for improvement exists in these relatively new technologies. In particular, the prediction of the optimal docking of large and flexible ligands to a protein and an accurate estimation of binding affinity are unresolved problems. These issues are related because the more accurate the description of the binding strength [including solvation (10) and entropic effects] the more time consuming is the sampling of possible solutions. Inversely, the coarser the approximation of the binding affinity the easier it is to sample docked conformations but with the danger that a vast majority of them might be either irrelevant or incorrectly ranked. At the present stage, and probably for the indefinite future, ligand and drug design will work best if it involves a strong and day-by-day collaboration between protein crystallographers (or NMR experts), who determine native and protein-ligand complex structures, theoretical chemists and biophysicists, who develop and apply computational methods to predict new ligands, medicinal chemists, who are willing to synthesize them, and biologists, who can perform the appropriate tests. Some crystallographers are already using computational approaches for ligand design, whereas only a few medicinal chemists like to complement their knowledge and intuition with suggestions originating from computational methods. It is expected that computer programs for (de novo) ligand design will become more efficient and user friendly, which will result in a much larger number of users.

One of the purposes of this review article is to show that structure-based computer-aided ligand design is a fascinating and progressing research field. It is fascinating not only for its ultimate goal, i.e., the discovery of drugs to prevent or cure human diseases, but also because it is based on, and thereby increases, our understanding of molecular interactions and recognition phenomena on an atomic level. Hence, ligand design represents an interesting combination of basic and applied research activities. Another reason is its multidisciplinary character, which requires skills in different branches of science, from theoretical physics and chemistry to computer science and statistics. That structure-based ligand design is a progressing field is evident from the relatively recent increase in the number of structure-based lead discoveries and optimizations (14). The continuous improvement of the currently used methodologies and the development of novel computer-aided approaches as well as the ever-increasing performance of computers will be the basis of further significant progress in the near future.

We thank Dr. Jean-Pierre Evenou for the thrombin-binding assay and J. Apostolakis, D. Arosio, N. Budin, P. Ferrara, N. Majeux, M. Scarsi, and A. Widmer for interesting discussions.

This work was supported by the Swiss National Science Foundation, the Hartmann Müller Foundation for Medical Research, the Olga Mayenfisch Foundation, and Novartis Pharma, Inc.

References

1. Böhm, H. J. Towards the automatic design of synthetically accessible protein ligands—peptides, amides and peptidomimetics. *J. Comput. Aided Mol. Des.* 10: 265–272, 1996.
2. Caflisch, A. Computational combinatorial ligand design: application to human alpha-thrombin. *J. Comput. Aided Mol. Des.* 10: 372–396, 1996.
3. Caflisch, A., and M. Karplus. Computational combinatorial chemistry for de novo ligand design: review and assessment. *Perspect. Drug Discov. Des.* 3: 51–84, 1995.
4. Caflisch, A., A. Miranker, and M. Karplus. Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *J. Med. Chem.* 36: 2142–2167, 1993.
5. Gubernator, K., C. Broger, D. Bur, D. M. Doran, P. R. Gerber, K. Müller, and T. M. Schaumann. Structure-based ligand design. In*: Computer-Aided Drug Design in Industrial Research*, edited by E. C. Hermann and R. Frankle. Berlin: Springer, 1995, p. 61–77.
6. Kuntz, I. D., I. C. Meng, and B. K. Shoichet. Structure-based molecular design. *Acct. Chem. Res.* 27: 117–123, 1994.
7. Miranker, A., and M. Karplus. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins Struct. Funct. Genet.* 11: 29–34, 1991.
8. Rotstein, S. H., and M. Murcko. GroupBuild: a fragment-based method for de novo drug design. *J. Med. Chem.* 36: 1700–1710, 1993.
9. Sali, A., and T. L. Blundell. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* 234: 779–815, 1993.
10. Scarsi, M., J. Apostolakis, and A. Caflisch. Continuum electrostatic energies of macromolecules in aqueous solutions. *J. Phys. Chem.* 101: 8098–8106, 1997.
11. Shuker, S. B., P. J. Hajduk, R. P. Meadows, and S. W. Fesik. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274: 1531–1534, 1996.

12. Tapparelli, C., R. Metternich, C. Ehrhardt, and N. S. Cook. Synthetic low-molecular weight thrombin inhibitors: molecular design and pharmacological profile. *Trends Pharmacol. Sci.* 14: 366–376, 1993.
13. Tschinke, V., and N. C. Cohen. The NEWLEAD program: a new method for the design of candidate structures from pharmacophoric hypotheses. *J. Med. Chem.* 36: 3863–3870, 1993.
14. Veerapandian, P. (Editor). *Structure-Based Drug Design*. New York: Marcel Dekker, 1997.
15. Wade, R. C., and P. J. Goodford. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* 36: 148–156, 1993.