Journal of Computer-Aided Molecular Design, **13:** 51–56, 1999.
KLUWER/ESCOM

Combinatorial docking and combinatorial chemistry: Design of potent non-peptide thrombin inhibitors

Hans-Joachim Böhm, David W. Banner & Lutz Weber
Hoffmann-La Roche Ltd, Pharmaceuticals Division, CH-4070 Basel, Switzerland

Received 23 July 1998; Accepted 4 August 1998

Key words: combinatorial, de novo design, docking, scoring functions, thrombin

Summary

A computational algorithm was used to design automatically novel thrombin inhibitors that are available from a single-step chemical reaction. The compounds do not contain amide bonds, are achiral and have a molecular weight below 400. Of the 10 compounds that were synthesized, five bind to thrombin with a K_i in the nanomolar range. Subsequent X-ray structure determination of the thrombin-inhibitor complex for the best compound ($K_i = 95$ nM) confirms the predicted binding mode. The novel algorithm is applicable to a broad range of chemical reactions.

Introduction

Structure-based design of enzyme inhibitors has been successfully applied to a large number of targets resulting in several clinical candidates and even in marketed drugs [1–3]. In the past, the 3D structure of the molecular target was usually displayed with graphics systems, and human chemical intuition was then used to derive ideas about novel ligands [4]. About 5 years ago, the first algorithms for *de novo* ligand design were disclosed [5]. These programs attempt to automatically design novel ligands to a given 3D protein target structure. Again, several successful applications have been reported [6, 7]. The experience with the structure-based approaches points to several challenges in the further improvement of the approach.

First, structure-based design can be used as a guide to possible routes for modifications of a given lead, e.g. by filling unoccupied lipophilic pockets, using a new chemical structure to satisfy key interactions with the protein, or by forming additional hydrogen bonds. However, it has turned out to be very difficult to identify the optimal chemical structure to achieve maximum binding affinity. The problem relates in part to the absence of reliable methods to predict binding affinities. Furthermore, small conformational changes of the protein structure lead to subtle changes of the size and shape of binding pockets which are difficult to predict at present. One possible way to approach the problem is to explore protein binding pockets by a series of structurally closely related compounds, e.g. by the design and synthesis of combinatorial libraries.

Second, in our opinion, synthetic accessibility was not sufficiently taken into account in many structure-based design efforts in the past – resulting sometimes in the suggestion of molecules that are very difficult to synthesize. While this has again led to many interesting new protein ligands, further optimization turned out to be very time-consuming afterwards [8, 9] due to the lengthy synthesis. One approach to this problem is to focus on chemical reactions that are amenable to parallel or combinatorial synthesis.

In the present communication we report the application of a novel algorithm to structure-based design that addresses the problems discussed above. The design of novel thrombin inhibitors was chosen to test this approach, which tries to combine the strengths of both computational ligand design and combinatorial chemistry. Within a small set of 10 designed compounds that were subsequently synthesized, we were able to identify several potent non-peptide low molecular weight inhibitors of thrombin. The predicted binding mode was confirmed by 3D structure determination of the protein-ligand complex using X-ray analysis. To our knowledge, the present work is the first example of an enzyme inhibitor with nanomo-

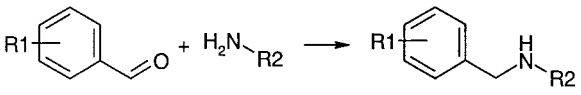


Figure 1. Reductive amination.

lar binding affinity that was designed using automatic docking tools.

Methods

Computational algorithm

The combinatorial docking algorithm is built to a large extent on the previously published program LUDI [10]. Briefly, the concept of interaction sites [11] is used to generate suitable positions for functional groups in the binding site of the possible ligand to form hydrogen bonds or to fill a hydrophobic pocket. These interaction sites are derived from a statistical analysis of nonbonded contacts in the crystal packing of small organic molecules. As the next step, 3D coordinates of fragments are read from a library and the structures are docked into the binding site by fitting them onto the interaction sites. Fragments which can be fitted onto the interaction sites with an rms deviation smaller than a certain value (0.7 Å in the present example) and do not overlap with the protein are considered as hits. Finally, a simple empirical scoring function [12] is used to rank the hits. An important new feature of the algorithm is the ability to connect structures in a chemically and structurally sensible manner. Building blocks are linked by using 'link sites' [13]. The program differentiates between chemically different types of link sites (-NH_2, -CHO, -OH, -SO_2Cl, -CH_2Br, -COCl, etc.). A simple set of rules determines which types of link sites can be connected together. The program fits a building block onto the interaction sites and simultaneously links it to a previously docked building block. In the present paper, we report the design of compounds that can be obtained by the reductive amination reaction starting from various aromatic benzaldehydes and primary amines (Figure 1).

Suitable building blocks were taken from the Available Chemicals Directory (ACD) [14]. Using several exclusion criteria (molecular weight below 300, no dialdehydes, no compounds with aliphatic chains longer than C_6) we arrived at 5300 primary amines and 540 benzaldehydes. The virtual size of the combinatorial library of possible reductive amination products was therefore $5300 \times 540 = 2\,862\,000$. The program CORINA [15] was used to generate a 3D structure for every building block. In order to facilitate the chemically correct linking between the aldehyde Ar-CHO and the amine R2-NH_2 into the final structure Ar-CH_2-NH-R2, the aldehyde moiety was converted into a methyl group and this modified fragment was then used to link it to the amine. The link was performed by superimposing the C-H bond of the methyl group with the H-N bond of the amine. The dihedral angle at the link C_{aro}-CH_2-NH-C was targeted to be $\pm 90°$, in accordance with a statistical analysis of small molecule crystal structures.

Synthesis and biological evaluation

All compounds were synthesized according to [16] by coupling 1 mmol of the corresponding aldehyde with 1 mmol of p-amino-benzamidine dissolved in ethanol using sodium cyanoborohydride and catalytic amounts of acetic acid over a reaction time of 24 h. The crude products were purified by solid-phase extraction with an RP-18 filled cartridge, by washing with water and obtaining the products with a water/methanol 1/1 solvent. All compounds gave satisfactory MS and NMR analysis. The compounds were tested for their thrombin activity with a chromogenic assay as described [17].

X-ray structure determination

The crystal structure was determined as follows: Recombinant human thrombin [18] was crystallized in the C2 form using the hanging drop method at 4 °C. The reservoir buffer was 25% PEG 3350, 100 mM NaCl, 100 mM phosphate pH 7.4. Crystals were soaked overnight in a 1 mM solution of the inhibitor in this buffer, switched briefly to the same buffer with 30% glycerol added, flash-frozen in liquid nitrogen and data collected at 120 K (Oxford Cryostream) on an FR591 X-ray generator (Nonius) with double mirrors (Supper) and a 30 cm image plate (marresearch). Larger crystals cracked on soaking or on freezing, and the inhibitor also crystallized out. Eventually a small crystal ($0.1 \times 0.1 \times 0.2$ mm) was successfully frozen from fresh inhibitor solution. Images of 1° in 300s were measured over 155° to a maximum resolution of 2.4 Å. Data were processed with the XDS program [19], and the unit cell found to be: a=70.02 Å, b=71.59 Å, c=71.93 Å, β=99.53°. The R-factor on intensities was 4.1%, and data was accepted to a resolution of 2.8 Å, with, in the outermost shell intensity

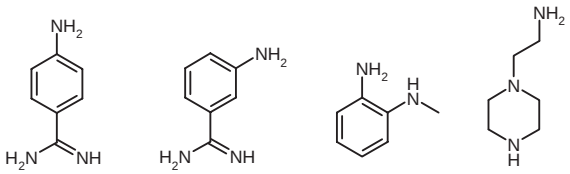
R-factor=22.3%; I/σI=3.6; 60% I>3σI; 99% completeness. Positional and constrained B-value refinement with X-PLOR [20], starting from other inhibitor structures (work in progress) converged to an R-factor (on F) of 15.4% (R-free = 25.6%) for 2581 atoms (including 167 waters) for 8672 reflections. Geometry was good, with rms bond errors of 0.012 Å and rms angle errors of 1.93°. The inhibitor density is well defined and the model was refined with tight planarity but no constraints on the three dihedral angles mentioned in the text. B-values are ∼30 Å^2 for the benzamidine moiety, ∼65 Å^2 for the central phenyl ring and ∼55 Å^2 for the terminal phenyl ring. Perhaps this is indicative of some thermal motion of the inhibitor (but at this limited resolution it is difficult to be certain).

Results

The combinatorial docking run was carried out in a two-step process. First, all amines were docked into the thrombin binding site, using the 3D protein structure from [17]. As a constraint, all docked molecules were required to form at least one hydrogen bond with the side chain of Asp 189 located at the bottom of the thrombin P1 specificity pocket. Furthermore, only hits with a predicted K_i smaller than 1 mM were accepted. This first step produced 65 hits and took 8 min on an SGI R5000 workstation. The 4 top scoring amines are shown in Table 1. The top-scoring hit is p-amino-benzamidine **1** which is a known thrombin inhibitor with a K_i of 34 μM. For comparison: unsubstituted benzamidin binds to thrombin with a K_i of 250 μM [17]. The top scoring p-amino-benzamidine was then used to attach the various substituted benzaldehydes onto the amine group. This second docking calculation took 20 min and produced 98 hits. Based on the predicted score and the availability from a major vendor, 10 benzaldehydes were selected for synthesis. The results of the biological screening of the resulting 10 reductive amination products are summarized in Table 2.

For all synthesized compounds, the inclusion of additional side chains leads to an improved binding affinity by at least a factor of 10 compared to unsubstituted p-amino-benzamidine. Some of the compounds have only small substituents at the aromatic ring originating from the benzaldehyde (e.g. compounds **5** and **6**). We believe that the comparatively weak binding of these compounds is due to the lack of a suitable

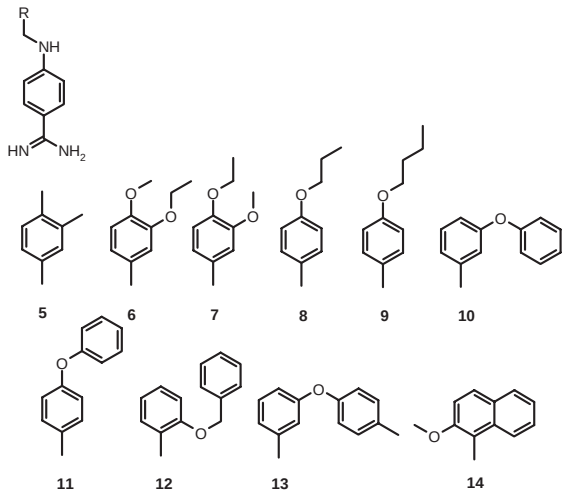
Table 1. Binding affinities of the top scoring hits from the computational docking of amines into the P1 pocket of thrombin



1 **2** **3** **4**

Compound No.	K_i Thrombin [μM]	K_i Trypsin [μM]
1	34	5.7
2	170	18
3	>100	>100
4	>100	>100

Table 2. Binding affinities of compounds **5**–**14** for thrombin and trypsin



5 **6** **7** **8** **9** **10** **11** **12** **13** **14**

Compound No.	K_i Thrombin [μM]	K_i Trypsin [μM]
5	0.91	2.8
6	1.3	2.0
7	1.5	3.4
8	0.27	1.2
9	0.13	0.55
10	1.1	7.4
11	0.095	0.52
12	1.9	2.2
13	0.99	6.9
14	2.9	10.0

lipophilic group to occupy the P3 pocket of thrombin. This occupation is, however, possible by substituents in the para position in compounds **8**, **9** and **11** that indeed showed nanomolar binding affinity. The most potent thrombin inhibitor found in the present study is compound **11** with a K_i of 95 nM. This compound binds 5 times more strongly to thrombin than to trypsin. This is even more remarkable since the anchor fragment **1** binds 5 times less strongly to thrombin than to trypsin. Thus, the inclusion of the lipophilic side chain improves K_i of **11** by a factor of 350 over the unsubstituted compound **1** and it reverses the selectivity from trypsin-selective towards thrombin-selective.

In order to confirm the predicted binding mode of compound **11**, the 3D structure of the complex of thrombin with this inhibitor was determined by X-ray structure analysis. The experimentally determined 3D structure of **11** bound to thrombin is compared with the conformation obtained from combinatorial docking in Figure 2. The X-ray structure confirms the predicted binding mode. The benzamidine group occupies the P1-pocket and the terminal phenyl substituent occupies the P3 pocket. The deviations between the atomic positions of experimental and the docked structure are smaller than 1 Å for these two end parts of the inhibitor. Significant deviations are, however, observed for the central phenyl ring, and to a lesser extent for the terminal phenyl ring. In the experimentally observed structure, the central phenyl ring is rotated by $\sim 45°$ as compared to the corresponding part in the docked structure. This difference is due to the use of predefined torsion angles at the linkage between the amine and aldehyde fragments in the combinatorial docking calculation. According to a statistical analysis of crystal structures containing the molecular fragment Ph-CH_2-NH-X, the preferred dihedral for the rotation around C_{aro}-C_{aro}-CH_2-NH is 90°. This value was used in the docking calculations. In the experimentally observed structure this dihedral refines to a value of $-45°$. The dihedrals around the ether linkage were predicted as ($-27°$; $-27°$, generated using the program CORINA [15]) and observed as (35°; 74°). As depicted in Figure 2, the difference between modelled and observed structures is principally that in the modelled structure the inhibitor bends inwards to best occupy the hydrophobic P2 pocket, whereas in the observed structure residues Tyr60A and Trp60D move down by ~ 1 Å to reduce the size of the P2 pocket so as to optimize the binding. The possibility of such adaptation of the protein to the inhibitor has long been known for thrombin [21] but was not taken into consideration in the present example.

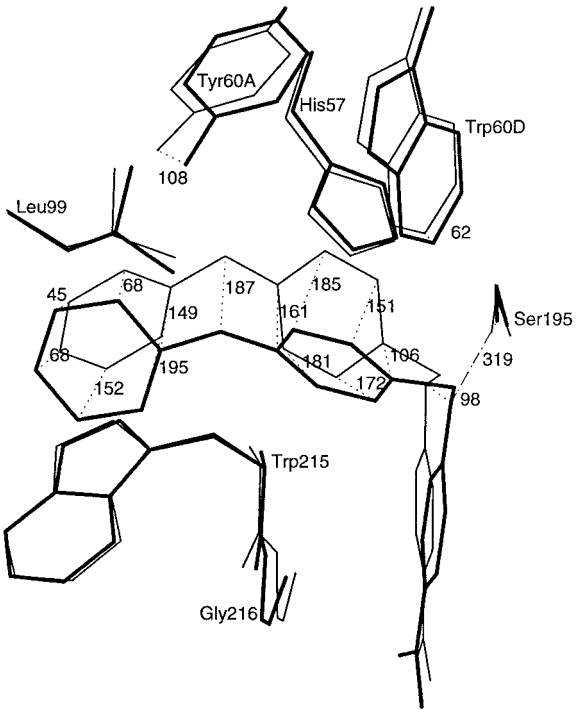


Figure 2. Comparison of the observed X-ray structure of compound **11** bound to thrombin (bold lines) with the predicted structure from the combinatorial docking procedure. Distances between selected atom pairs are given in units of 0.01 Å. The benzamidine moiety (lower right) fits into the P1 (recognition) pocket as predicted (differences between equivalent atoms <1 Å). The terminal phenyl group, at the left, also fits as predicted in the hydrophobic P3, or distal, pocket (only Trp215 and Leu99 are shown here). The central part of the molecule is observed to lie up to 2 Å away from the predicted position. As discussed in the text, the hydrophobic P2, or proximal, pocket (formed by Tyr60A, Trp60D, His57, and Leu99) is reduced in size by movements of ~ 1 Å in Tyr60A and Trp60D, compatible with a preferred conformation of the inhibitor.

Discussion and conclusions

In the present paper, we report the computational *de novo* design, synthesis and biological evaluation of novel potent thrombin inhibitors. All compounds can be made by reductive amination. The predicted binding mode for the most potent compound **11** was confirmed by subsequent X-ray structure determination.

A number of very potent and selective thrombin inhibitors have been published with binding affinities in the single digit nanomolar or even subnanomolar range [17, 22]. However, these are all compounds

with a significantly higher molecular weight. Moreover, they are only available by a multi-step synthesis. In contrast, the present compounds are obtained by a single-step chemical reaction and have a low molecular weight of 300–350. To our knowledge, the present paper is the first disclosure of a noncovalent thrombin inhibitor with MW<400 and K_i<100 nM.

A paradigm held by many researchers in the field is that inhibition of trypsin-like serine proteases requires the formation of at least one hydrogen bond with Gly 216. Indeed, such hydrogen bonds are observed in crystal structures of many disclosed nanomolar thrombin inhibitors so far. However, the present compounds do not contain any suitably positioned polar functional groups that could form such hydrogen bonds. Therefore, the present data provide evidence that these hydrogen bonds are not required for high affinity binding to thrombin. This is also in agreement with a recent study by Obst et al. [9] who observed only a fourfold loss in binding affinity by replacing a C=O group hydrogen bonding to Gly 216 NH by -CH_2-.

It should be noted that the chosen system is advantageous for our approach for a number of reasons. First, thrombin contains a clearly defined P1-pocket that is deeply buried into the protein structure and therefore ideally suited to bind small molecules similar to benzamidine. If such a 'needle-pocket' does not exist, it may be more difficult to find very small ligands with measurable binding affinity which could then be further improved by attaching side chains using combinatorial docking. Second, a highly reliable binding assay was available to us allowing for the determination of binding affinities in the 100 μM range. If such a reliable and robust assay does not exist, there is of course little hope to find small 'needle-type' molecules with binding affinities in the high micromolar range. An alternative approach to the optimization of the biological activity of combinatorial libraries using a genetic algorithm has been described by Weber et al. [23].

An interesting aspect of the present work relates to the performance of the scoring function. Figure 3 shows a plot of the predicted binding affinities versus the measured data. The scoring performs very well in the prediction of the binding affinities of compounds **1** and **2**. This result is not too surprising since the thrombin-benzamidine complex was part of the calibration data set used to derive the scoring function. It is also capable to predict the binding affinities of compounds **5**–**14** with an error of less than 1.6 log units. We believe that **5**–**14** bind to the same lipophilic

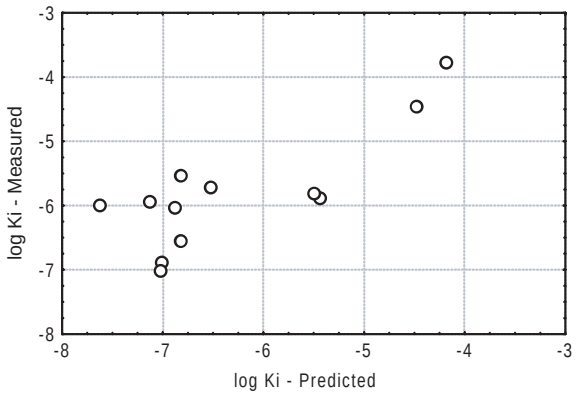


Figure 3. Plot of the predicted K_i of compounds **1**–**14** versus the experimentally determined K_i.

pocket and differences in K_i are due to differences in lipophilic contacts. In our experience, this non-bonded interaction is described well by the scoring function. However, it should be noted that the scoring function did not predict compound **11** as the most potent compound within the present series of compounds. The highest predicted binding affinity was obtained for compound **13**. One limitation of the scoring function is clearly the neglect of conformational changes of the protein. As described previously, we observe a slight movement of the residues Tyr60a and Trp60D. This conformational change increases the lipophilic contact surface for **11** by roughly 10 $Å^2$ (as calculated by the grid-based algorithm used in the scoring function [12]) which amounts to 2 kJ/mol in binding affinity. Another difference between the predicted and the experimentally determined complex structure is that in the experimental structure the hydrogen bonds have geometries that are slightly closer to the ideal values used in the scoring function. Therefore, the score for the hydrogen bonds is higher for the experimental structure. Using the experimental structure for thrombin complexed with **11**, the calculated K_i is 20 nM ($\log K_i = -7.7$).

In summary, in the present communication we have shown that a combination of a novel automatic structure-based design tool used together with combinatorial chemistry can be successfully used to discover novel and potent enzyme inhibitors that are also accessible by a facile synthetic strategy.

Acknowledgements

We would like to thank Klaus Gubernator, Fritz Winkler, Gérard Schmid and Paul Hadváry for their continuous support of our work.

References

1. Babine, R.E. and Bender, S.L., Chem. Rev., 97 (1997) 1359.
2. Kaldor, S.W., Kalish, V.J., Davies, J.F., Shetty, B.V., Fritz, J.E., Appelt, K., Burgess, J.A., Campanale, K.M., Chirgadze, N.Y., Clawson, D.K., Dressmann, B.A., Hatch, S.D., Khalil, D.A., Kosa, M.B., Lubbehusen, P.P., Muesing, M.A., Patick, A.K., Reich, S.H., Su, K.S. and Tatlock, J.H., J. Med. Chem., 40 (1997) 3979.
3. Hubbard, R.E., Curr. Opin. Biotech., 8 (1997) 696.
4. Cohen, N.C., Blaney, J.M., Humblet, C., Gund, P. and Barry, D.C., J. Med. Chem., 33 (1990) 883.
5. Lewis, R.A. and Leach, A.R., J. Comput.-Aided Mol. Design, 8 (1994) 467.
6. Böhm, H.J., Prog. Biophys. Mol. Biol., 66 (1996) 197.
7. Rich, D.H., Bohacek, R.S., Dales, N.A., Glunz, P. and Ripka, A.S., Chimia, 51 (1997) 45.
8. Mack, H., Pfeiffer, T., Hornberger, W., Böhm, H.J. and Höffken, W., J. Enz. Inh., 9 (1995) 73.
9. Obst, U., Banner, D.W., Weber, L. and Diederich, F., Chem. Biol., 4 (1997) 287.
10. Böhm, H.J., In van de Waterbeemd, H., Testa, B. and Folkers, G. (Eds), Computer-assisted Lead Finding and Optimization, Wiley-VCH, Weinheim, 1997, p. 125.
11. Böhm, H.J., J. Comput.-Aided Mol. Design, 6 (1992) 61, 593.
12. Böhm, H.J., J. Comput.-Aided Mol. Design, 8 (1994) 243.
13. Böhm, H.J., J. Comput.-Aided Mol. Design, 10 (1996) 265.
14. The available chemicals directory can be obtained from MDL Information Systems, Inc., San Leandro, CA.
15. Sadowski, J., Rudolph, C. and Gasteiger, J., Tetrahedron Comput. Method., 3 (1990) 537.
16. Yan, I.Y., Slusher, R.M., Broom, A.D., Ueda, T. and Cheng, Y.C., J. Med. Chem., 31 (1988) 2137.
17. Hilpert, K., Ackermann, J., Banner, D.W., Gast, A., Gubernator, K., Hadvary, P., Labler, L., Müller, K., Schmid, G., Tschopp, T.B. and van de Waterbeemd, H., J. Med. Chem., 37 (1994) 3889.
18. Russo, G., Gast, A., Schlaeger, E.J., Angiolillo, A. and Pietropaolo, C., Protein Expr. Purif., 10 (1997) 214.
19. Kabsch, W., J. Appl. Crystallogr., 26 (1993) 795.
20. Brünger, A., 1992, X-PLOR v3.1 Manual (Yale University Press, New Haven, CT).
21. Banner, D.W. and Hadváry, P., J. Biol. Chem., 266 (1991) 20085.
22. Tucker, T.J., Lumma, W.C., Lewis, S.D., Gardell, S.J., Lucas, B.J., Baskin, E.P., Woltmann, R., Lynch, J.J., Lyle, E.A., Appleby, S.D., Chen, I., Dancheck, K.B. and Vacca, J.P., J. Med. Chem., 40 (1997) 1565.
23. Weber, L., Wallbaum, S., Broger, C. and Gubernator, K., Angew. Chem. Int. Ed. Engl., 34 (1995) 2280.