Journal of Computer-Aided Molecular Design, **16:** 459–478, 2002.
KLUWER/ESCOM

A validation study on the practical use of automated *de novo* design

Martin Stahl[a,*], Nikolay P. Todorov[b,*], Tim James[b], Harald Mauser[a], Hans-Joachim Boehm[a] & Philip M. Dean[b]
[a]*F.Hoffmann-La Roche Ltd, Basel, Switzerland*
[b]*De Novo Pharmaceuticals Ltd, Compass House, Vision Park, Histon, Cambridge CB49ZR. UK.*

Received 7 May 2002; Accepted in reviewed form 16 August 2002

Key words: de novo design, simulated annealing, drug design, optimization, cyclooxygenase, cyclin dependent kinase, matrix metalloproteinase, estrogen receptor, molecular similarity, maximum common substructure

Summary

The *de novo* design program Skelgen has been used to design inhibitor structures for four targets of pharmaceutical interest. The designed structures are compared to modeled binding modes of known inhibitors (i) visually and (ii) by means of a novel similarity measure considering the size and spatial proximity of the maximum common substructure of two small molecules. It is shown that the Skelgen algorithm generates representatives of many inhibitor classes within a very short time and that the new similarity measure is useful for comparing and clustering designed structures. The results demonstrate the necessity of properly defining search constraints in practical applications of *de novo* design.

Introduction

The massive opportunities for the pharmaceutical industry afforded by the genomics revolution have spurred the development of new *in silico* methods to handle the increased number of targets that will become available. New computational tools such as virtual screening, *de novo* structure-based and ligand-based design methods will play a major role in exploiting the new opportunities. All focus on some aspect of structure of the binding site; even ligand-based methods seek to postulate a model for the site with predicted features at the unknown site surface. This paper draws attention to one of these procedures – *de novo* design – and establishes a methodology for validation of the technique.

De novo design methods are automated computational procedures for the placement and connection of atoms, or molecular fragments, within the 3D structure of a receptor binding site with the aim that the resulting, hypothetical, molecular structures should display high binding affinity to the receptor. Over the past decade, a multitude of software tools for *de novo* design have been developed in parallel to the increasing availability of x-ray structures of pharmaceutically relevant targets [1-3]. These tools differ in many aspects:

- in the sizes and types of molecular fragments and connection rules employed,
- in whether they are built on combinatorial or stochastic algorithmic principles,
- in the strategy used for growing larger scaffolds, e.g. by incremental construction starting from a first fixed fragment, or by linking two or more fragments that were placed into the binding site in previous design steps.

Independent from the details of the method, all *de novo* design approaches suffer from two main deficiencies. First, *de novo* design methods can only employ very general rules about organic chemical synthesis. Therefore, many of the generated structures are difficult to synthesize. Second, all *de novo* design methods must employ some computational estimate of receptor-ligand complementarity or binding affinity; a so-called scoring function [4-6] is used to rank partial solutions relative to each other. The currently available methods for estimating binding affinities are still rela-

tively unreliable. Because experimental and calculated binding affinities can differ vastly, *de novo* design programs are often 'misled' by their scoring functions and produce many false positive structures.

Due to the problem of synthetic feasibility, many structures proposed by *de novo* design programs have not been experimentally examined [7-9] and straight *de novo* design went out of fashion over the last years. Instead, docking methods have been used more widely, which do not assemble novel molecular scaffolds, but determine optimum binding orientations of existing compounds [10-14]. Combinatorial docking programs have been developed [15,16]. They differ from *de novo* design programs not so much in the algorithms used, but in the highly restrictive specification of chemical connection rules, allowing only narrow series of synthetically accessible compounds to be generated. Several successful applications of such tools have been published [17-19].

The problem of inaccurate binding affinity prediction cannot be circumvented so easily. Docking methods and *de novo* design methods both suffer from this problem, but in *de novo* design, the problem can become more apparent, because there are many more possibilities of exploring erroneous local minima of the interaction energy along the optimization trajectory than would be encountered in database docking. It is therefore not surprising that no structure has been published of a protein complex containing a ligand totally designed *de novo* [20]. Where *de novo* design has been successful, it was based not on the structure of an empty binding site alone, but also on knowledge about essential features of a potential ligand [21-23]. For example, the DuPont cyclic urea HIV protease inhibitors were developed from pharmacophore features extracted from x-ray structures of complexes with other HIV protease inhibitors [24]. Likewise, knowledge of the substrate and its interactions in the binding site has facilitated structure-based design of the first neuraminidase inhibitors [25]. More recent publications have shown that structures of neuraminidase inhibitors can considerably deviate from the substrate as long as certain pharmacophore elements are fulfilled [26-28].

Notwithstanding the limitations discussed above (or maybe precisely because they seem to be insurmountable), the true potential of this class of ligand design methods has not yet been exploited. The huge combinatorial problems that limited the application of earlier *de novo* design methods can now be overcome by much improved optimization methods. The chemical diversity available in compound collections for docking is not large enough to cover all relevant regions of chemical space that could be of interest. Most pharmaceutically relevant targets still fall into few structural classes onto which activities of many pharmaceutical companies concentrate. Under these circumstances, the ability to identify novel ligand scaffolds rationally and quickly is a major competitive advantage [29,30], and this is clearly the strength of *de novo* design. *De novo* design tools can cover a large section of the chemical space relevant for a particular binding site with little computational expenditure [31]. However, the experience of the last decade of structure-based design shows that it is necessary to constrain the search by means of pharmacophore features that must be satisfied by the designed compounds. These constraints should be well balanced, as general as possible, but should incorporate all available and relevant information on ligand binding. Lastly, in our judgement the problem of synthetic feasibility can be alleviated by a careful choice of fragments and connection rules.

After the generation of scaffold candidates, essential and variable structural features contained in these scaffolds can be identified to help in the planning of a synthesis strategy, rather than searching for a synthetic route to one particular designed structure. The latter step involves close interactions between molecular design and synthetic chemistry experts and underlines that the critical step in reducing a novel technology to practice is often its integration in a working environment rather than the elimination of deficiencies of the technology itself. For this purpose, procedures for using database searching techniques to find close analogues of designed structures have been proposed [32].

In the present paper, we demonstrate the usefulness of a second-generation *de novo* design program, Skelgen [33,34], by comparing designed structures with those of known ligands. We have chosen four targets representing different classes of proteins that are of central interest to the pharmaceutical industry, cyclin dependent kinase 2 (CDK2), estrogen receptor, cyclooxygenase-2 (COX-2) and stromelysin-1 (MMP3) as a representative of the matrix metalloproteinases (MMPs). For each of these targets, known ligands were selected from the literature. Binding modes for all ligands were manually generated. Interactive, as well as automated procedures were employed for assessing the similarity of these modeled ligands with the structures generated by Skelgen. Overall, we find

that Skelgen is able to comprehensively cover most of the space spanned by the scaffolds and binding modes found with the known inhibitors within only a few hundred designed structures. This retrospective analysis helps to set the direction for prospective applications of Skelgen to generate novel chemotypes and establishes a comparative method for validating *de novo* design tools in general.

Results and discussion

This section starts with a brief description of the algorithmic principles of Skelgen. More details can be found in the Materials and Methods section as well as in a previous publication [33]. Subsequently, we will discuss details of the run-time parameters used in the calculations and the procedure used for the analysis of Skelgen results. This is followed by a discussion of the search constraints and results obtained for each of the four targets. Finally, we will discuss a procedure to automate the comparison of known inhibitors and Skelgen-generated structures.

The Skelgen algorithm

The input to Skelgen consists of a receptor pocket, defined by a protein coordinate file and a rectangular box, a set of constraints that should be fulfilled by each of the generated solutions, and a set of molecular fragments, called *templates*, that are used for the construction process. Skelgen uses a stochastic procedure for the construction of molecular scaffolds. At any given point during the run time of Skelgen, the program always works on only one scaffold, which is modified through a set of *transitions*. Transitions are the rigid-body displacement (translation or rotation) of the whole scaffold, rotations around single bonds, and the addition, deletion or replacement of templates forming the current scaffold. The type of transition is chosen randomly for each modification step. After a modification, the value of an objective function F (or penalty function) is calculated for the modified scaffold. This function is designed such that it becomes zero if all the design constraints are fulfilled, and it returns positive values otherwise. The modified structure is accepted, or rejected, based on the Metropolis criterion $p = \exp(\Delta F/T)$, where ΔF is the difference in the value of the objective function for the scaffold before and after the transition was applied. The control parameter T, analogous to temperature, is slowly lowered during each Skelgen run. Thus, Skelgen searches a combined chemical, orientational and conformational space of potential scaffolds by means of a simulated annealing procedure. A number of search parameters determine the duration of each simulated annealing run. One of them is the length of Markov chains at a particular value of T, and the number of Markov chains to be used as T is decremented. This limits the maximum number of transitions that can be applied before one simulated annealing run is terminated. A single solution structure – a ligand candidate – is generated as output from a single run. The algorithm produces numerous solutions from different random starts. This provides a user with a choice of many different chemotypes from multiple runs. This principle is used here to compare the chemotypes generated with known ligands for the four different sites used in this study.

The validation test set

The four test targets were chosen such that they represent typical design problems in early phases of pharmaceutical projects. The binding sites of two targets, the estrogen receptor and COX-2, are predominantly lipophilic, buried cavities. Promising inhibitor candidates should fit into these cavities such that they form favorable van der Waals interactions with many of the aliphatic or aromatic side chains of the surrounding amino acids. These interactions are short-ranged but do not strongly depend on direction. Thus, for these two targets, the focus is on the overall shape of the designed scaffolds and less on specific interactions. Contrarily, the two other chosen targets, Stromelysin and CDK2, require the fulfillment of specific directed interactions with polar groups, whereas the shape of the whole ligand is of secondary importance. The search constraints used in the Skelgen runs for each of the four targets underline these differences. They are depicted in Figures 1,4,7 and 9, and will be discussed in more detail together with the calculation results for each target.

Skelgen run parameters

Skelgen includes an option for calculating a score value as an estimate of the protein-ligand interaction energy. The user can define a threshold value of this score that must be surpassed by all solutions generated by Skelgen during that run. If this is done, a term is added to the objective function that monitors the difference between the actual and desired score

value in the form of an additional constraint. Calculations for all four targets were run with score threshold values of −20 to −50 in steps of 5 score units. We have generally observed an increase in scaffold size as the score threshold values become more negative, i.e. the calculated binding affinities become stronger. For all targets, scores of −50 units could only be reached with scaffolds significantly larger than known ligands. The size dependence of the score is a general phenomenon with empirical scoring functions [5] and can only be compensated for by additional empirical parameters, such as a restriction of the number of fragments or atoms in a scaffold. In the present case, we have restricted ourselves to the analysis of Skelgen runs with threshold values of −20 to −45 score units. All runs were performed with Markov chain lengths of 500 and 3000. Comparison of the results showed that for the less restrictive score thresholds, differences in chain length have no obvious effect on the output structures, but more negative score values can only be achieved with longer Markov chains. Overall, longer chain lengths tend to lead to more uniform size and more strain-free orientations of scaffolds in the binding pockets. Therefore, only the results of those runs with chain lengths of 3000 were analyzed.

Analysis of Skelgen output

Since the generation of novel scaffolds with Skelgen is a stochastic process aimed at scanning a multidimensional hypersurface containing a multitude of nearly degenerate minima, the question of convergence of the search is a central issue. It is to be expected that when Skelgen is run for a long time with the same parameters, the rate of finding novel solutions will slow down asymptotically [35]. In the present case, although closely related solutions have been found repeatedly for each of the four targets, convergence was certainly not achieved. For each of the six score threshold values, Skelgen was run until 100 scaffolds had been generated. The analysis was thus performed on only 600 structures per target. There are two reasons for this rather low number. First, we wanted to be able to visually compare the generated structures with modeled binding modes of known ligands, which would have been impossible with more than a few hundred solutions. Second, we have observed that although convergence in a strict sense was not achieved, for an observer trained in medicinal chemistry the generated solutions clearly displayed recurring patterns, thus allowing a classification into groups of scaffolds and binding modes. Longer run times would very likely have produced more representatives of these classes (and probably even closer analogues to known inhibitors), but few truly novel solutions. The issue of convergence will be subject to further studies with Skelgen.

Visual assessment of the similarity between structures generated by a *de novo* design program and known inhibitors, as we have performed it in this study, can be very subjective. We have tried to increase the level of objectivity by assembling large and diverse sets of known inhibitors from literature references (Table 1). All compounds were manually modeled into the corresponding binding sites. Details on the modeling procedure can be found in the Materials and Methods section below. General properties of the four sets of known inhibitor structures were identified and it was visually checked whether Skelgen generated solutions belonging to these classes. Known estrogen receptor and COX-2 inhibitors were classified according to their frameworks and binding orientations, ignoring atom types, and searching for closest analogues generated by Skelgen for each of these classes. The chemical structures drawn in Figures 2 and 5 represent such general scaffold patterns. In an analogous way, CDK2 inhibitors were grouped into classes according to their hydrogen-bonding pattern with the receptor using types of donor and acceptor atoms and the number of bonds separating them as criteria. Finally, for stromelysin, the inhibitors extracted from the literature were grouped according to the length and atom types of the atom chain bearing the hydroxamate and S1′ binding moieties. This classification scheme is closely related to an analysis of *de novo* design solutions one would perform for a genuine medicinal chemistry program, and it helps to elucidate the coverage of chemical and orientational space in the solution set.

Results for the four targets

COX-2. The published complex structure of COX-2 bound with the inhibitor SC558 was used for calculations (PDB code 6cox). Three lipophilic pharmacophore elements with radii of 1.5 Å, placed roughly at the corners of an equilateral triangle, were used as search constraints to force Skelgen to leave no part of the active site unoccupied (Figure 1).

Selective inhibitors of COX-2 reflect the angular form and lipophilic nature of their binding site. Most known inhibitors contain three rings connected by sin-

Table 1. Origin of the known ligands of each of the four targets used for comparison purposes.

Target	Existing inhibitor collections	Individual articles
COX-2	[36-40]	[41-44]
Estrogen receptor	[45-48]	[49-53]
CDK2	[54-56]	[57-61]
Stromelysin-1 (MMP3)	[62-65]	

gle bonds or short linkers, or bicyclic systems carrying an additional ring as a substituent. With few exceptions, these ring systems are planar and aromatic or hetero-aromatic. The size of the rings can vary in most classes of inhibitors. It should be noted that the chemical structures in Figure 2 have been generalized as far as possible. Various ring annulation positions have been omitted, aromatic rings have been drawn only where all published inhibitors contain aromatic rings, and single bonds have been drawn where the bond orders can either be single or double. Several classes of inhibitors have been reported that could not be placed in the binding site in a strain-free manner, among them aryl compounds carrying two *tert*.-butyl groups in *meta* position [66]. We assume that these compounds require a change of conformation of the enzyme, although we are not aware of x-ray structures that display a major change in active site size or form. These compounds are omitted from Figure 2.

Figure 2 shows that Skelgen designs representatives of all major classes of known COX-2 inhibitors. Some of the designed structures rearrange considerably upon minimization within MOLOC (e.g. Figure 2b and d). This is due to the fact that Skelgen could satisfy all three spatial constraints only by moving some of the scaffolds away from their minimum energy positions. In several cases, we have observed that Skelgen places ring systems too tightly into subpockets of the binding site than is generally observed with known inhibitors (e.g. Figure 2a), such that repulsive interactions with protein side chains cannot be relieved by force field minimization. On the other hand, Skelgen generated many scaffolds with at least partially saturated ring systems that do fit into the binding pocket in a strain-free manner. This opens the way to many new, interesting scaffolds as potential COX-2 inhibitors. Overall, we are not aware of any *de novo* design methodology that generates a similarly comprehensive set of analogs to known COX-2 inhibitors as Skelgen from its standard set of fragments. Other researchers reproduced individual members of a particular class of inhibitors, but were able to do so only by starting with fragments and connection rules that were generated from exactly this class of COX-2 inhibitors [67].

The majority of COX-2 inhibitors carry a methyl sulfone or sulfonamide substituent pointing towards the residues His 90 and Phe 518. These functional groups do not seem to be essential for inhibitor recognition, since the geometry of their hydrogen bonds with the enzyme is not optimal. The only clearly directed hydrogen bond is formed to the backbone NH of Phe 518. Skelgen does generate a few solutions in which a sulfoxide, instead of a sulfonyl moiety, forms this hydrogen bond. An example is shown in Figure 3.

Finally, we have observed that Skelgen produces many representatives of the 1,2-diaryl heterocycle class of inhibitors (cf. Figure 1). However, Skelgen orients many of these structures in a binding mode that is rotated 120° counterclockwise to the experimentally observed orientation, i.e. with the central ring at the right in Figure 1. We have observed this phenomenon also in several docking experiments and regard it as a relaxed orientation, but are unaware of experimental evidence for it. Moving the sulfonamide substituent to the central ring could test whether this binding mode is indeed possible.

Estrogen receptor. For the estrogen receptor, the complex structure with the antagonist Raloxifen (PDB code 1err) was selected. Two constraints were specified: One ligand atom should form a hydrogen bond to both the Arg 349 and Glu 353 side chains, as indicated for the hydroxyl group of Raloxifen in Figure 4. Essentially, this constraint specifies a hydroxyl group placed between these two side chains, a feature that is found in almost all known agonists and antagonists of the estrogen receptor. Some regions of the estrogen receptor binding-site display considerable flexibility. In particular, the side chain of His 524 can adopt different rotameric positions, such that the imidazole ring either

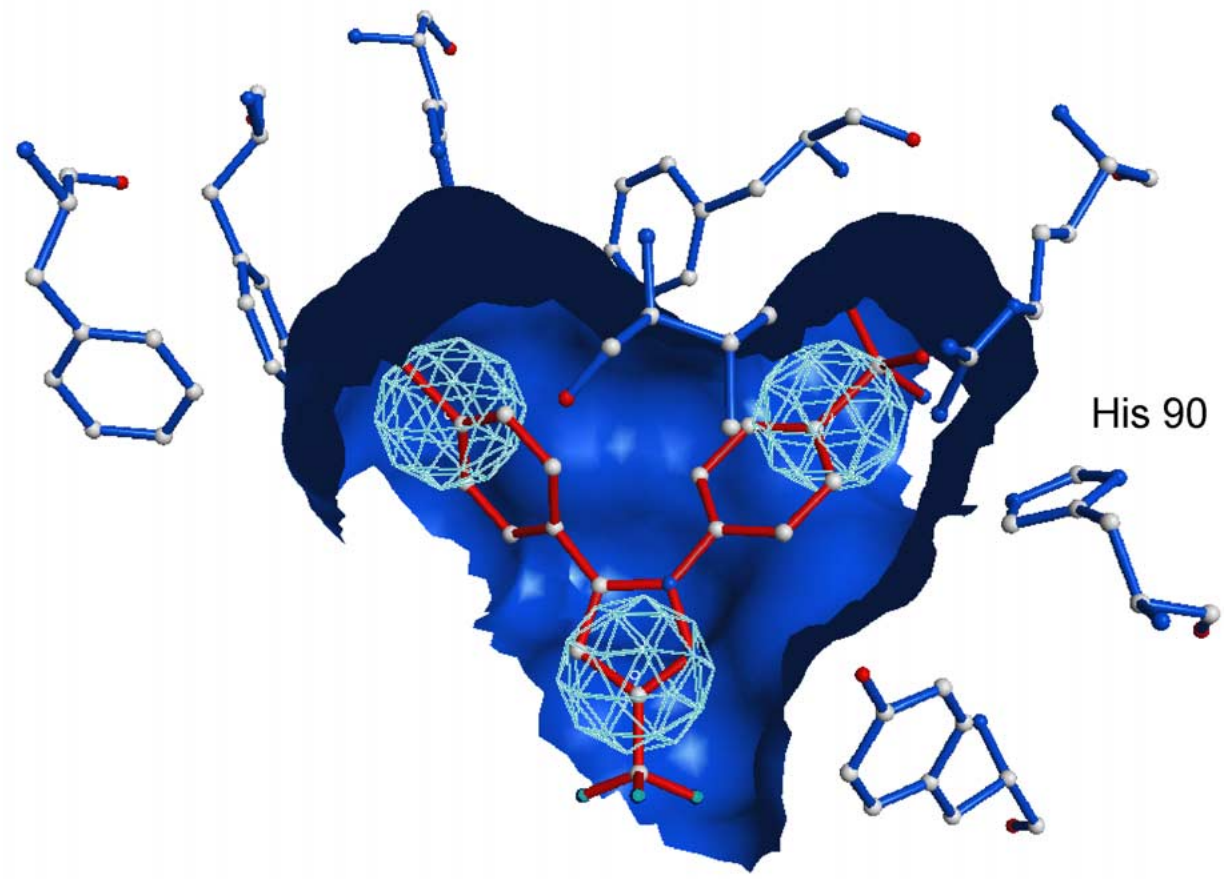

Figure 1. View of the binding site of COX-2 with the ligand SC558 bound (top half of the cavity surface removed). The three spheres correspond to the size and location of the three pharmacophore features used in the Skelgen runs.

forms specific interactions with the ligand or makes room for lipophilic ligand groups. Because of this ambiguity, we have only used a generic second constraint for Skelgen runs that requires any non-hydrogen ligand atom to be situated within one of two large spheres at the distal end of the binding site (Figure 4), ensuring that generated structures would fill the entire steroid binding cavity.

All known ligands of the estrogen receptor – agonists as well as antagonists – possess large, often polycyclic, steroid-size scaffolds. In Figure 5, the classes represented in our inhibitor collection are depicted in their approximate binding orientations (Arg 349 and Glu 353 would be located at the right hand side, His 524 at the left). Note that some of the scaffolds can be oriented in two alternative ways due to their approximate symmetry. X-ray structural information is not available for all of these scaffolds, so it must be left open which orientation is the correct one, as related cases from the literature show that both orientations can be equally likely [68]. It should also be pointed out that we have concentrated on the steroid binding site only. Antagonists such as Raloxifen possess substituents that extend towards and interact with the outer surface of the binding pocket, thereby stabilizing the open, inactive, conformation of the receptor.

Most of the solutions generated by Skelgen superimpose well with modeled binding modes of known ligands. Like most known agonists and antagonists, they are mostly planar, unsaturated ring systems. The majority of solutions consist of a bicyclic ring system connected to another ring via a single bond, as is the case with most known inhibitors. Larger ring systems are not contained in the Skelgen template library, nor does Skelgen close rings by introducing additional bonds or linker fragments. Therefore, larger polycyclic ring systems are not observed in the solution set. Structures containing longer acyclic linkers between rings, such as those depicted in Figure 6, are also not generated. This is probably due to the fact that

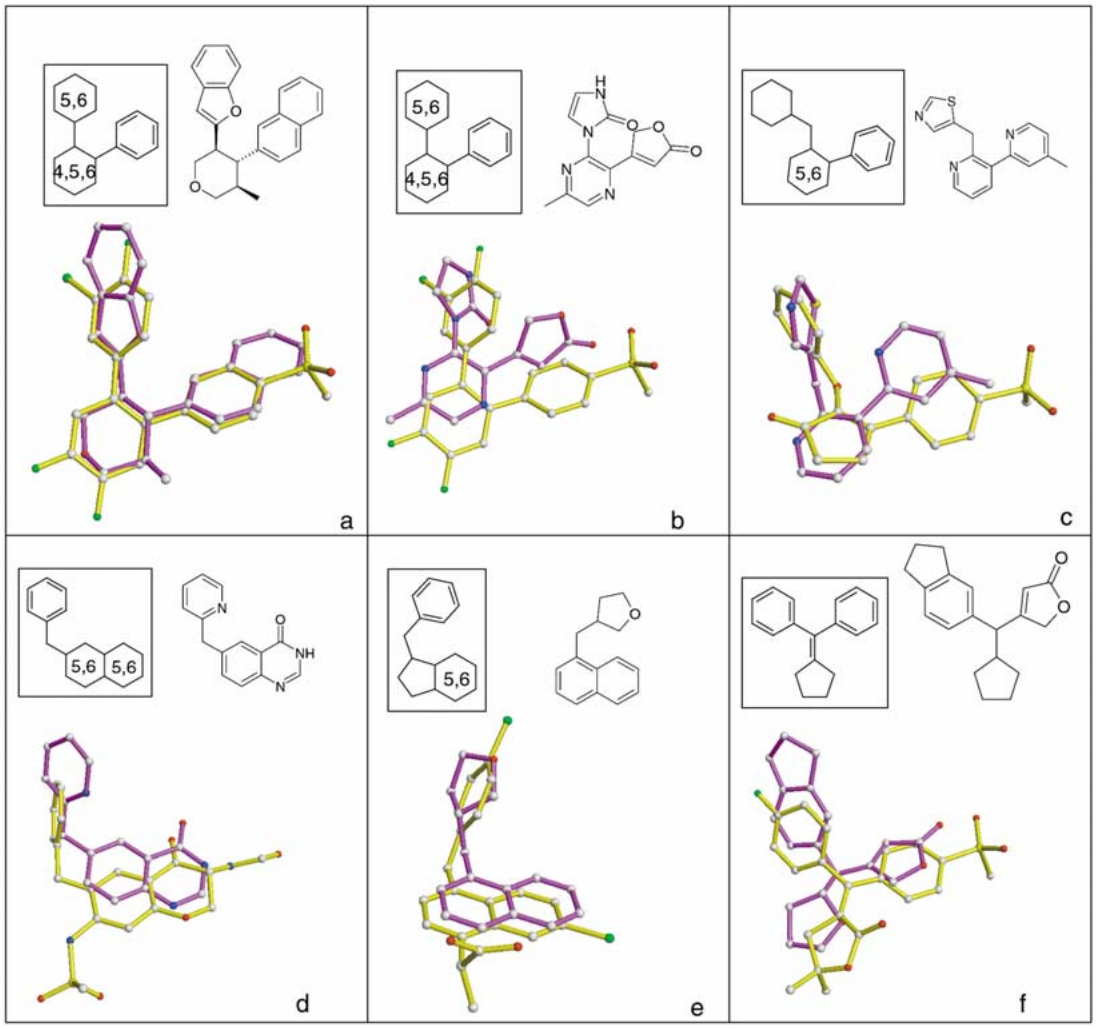

Figure 2. Superimposed pairs of modeled ligands (yellow) and scaffolds designed by Skelgen (magenta). 2D representations of the Skelgen structures are shown next to the general scaffold patterns of known COX-2 inhibitors (in box). For the general inhibitor patterns, numbers inside rings indicate that ligands of alternative ring sizes are experimentally known.

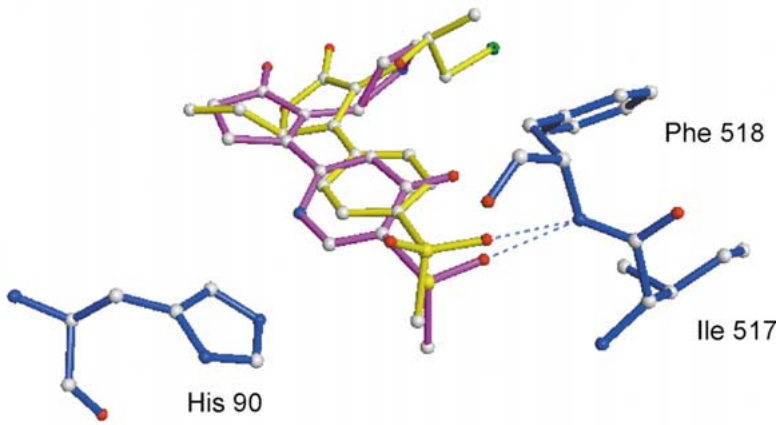

Figure 3. Superimposed structures of a known ligand modeled into the COX-2 active site and a Skelgen structure. Skelgen reproduces the hydrogen bond to the Phe 518 backbone NH group by placing a sulfonyl group in the position of the methylsulfone moiety.

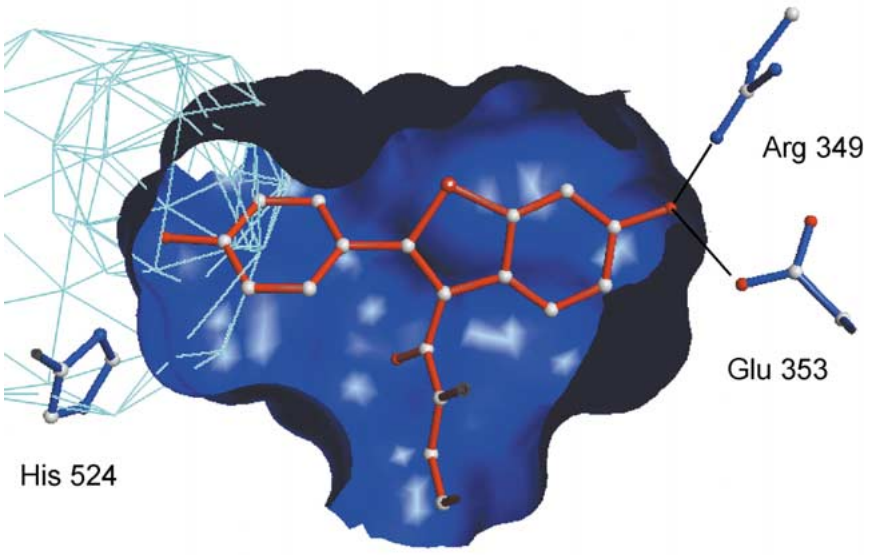

Figure 4. View of the binding site of the estrogen receptor with the ligand Raloxifen bound (PDB code 1err, top half of the cavity surface removed). The search constraints for Skelgen were: (1) placement of a non-hydrogen atom into the area enclosed by the spheres on the left, and (2) formation of a hydrogen bond to the Arg 349 and Glu 353 side chains as indicated by the lines.

Skelgen would have to add two or more one-carbon fragments to a growing structure, and this is a rather unlikely event compared to the addition of a ring system in one step that leads to a much higher increase in calculated binding energy. We have observed the same phenomenon also in the case of COX-2, where far fewer analogs of type d and e (Figure 2) were generated than related structures with the single ring directly attached to the bicyclic ring system. These observations indicate that it might be necessary to fine-tune the probabilities with which rings and acyclic fragments are used for the transition steps in Skelgen. Finally, it should be pointed out that the *cis*-stilbene scaffolds (bottom right in Figure 6) could not be modeled into the 1err conformation of the binding site in a strain-free manner, which makes it even more unlikely for Skelgen to generate analogues.

Because of the restrictive definition of the hydrogen-bonding constraint to Arg349 and Glu353, Skelgen is forced to add a hydroxyl group to every scaffold between these two side chains. In most cases, the hydroxyl group is placed in close vicinity to the crystallographically determined position. However, Skelgen does not check whether alternative tautomeric states exist that could be more stable in aqueous solution. For example, adding a hydroxyl group in alpha position to a pyridine ring nitrogen would produce a pyridone that is more stable in the keto form. Instable chemical structures can also be the result of adding hydroxyl groups (cf. Figure 5a). Additional filter rules are currently being developed to overcome this problem.

CDK2. ATP binding sites of kinases are located between the N-and C-terminal domains of these enyzmes. Inhibitors binding at the ATP binding site all form one or more hydrogen bonds to the so-called hinge strand connecting the two domains. The minimum requirement for an inhibitor is to accept a hydrogen bond from a central backbone NH group, in the case of CDK2 Leu 83 (Figure 7). The majority of known CDK2 inhibitors form an additional hydrogen bond to at least one of the two neighboring carbonyl groups, i.e., that of Glu 81 and Leu 83. Consequently, the search constraints used in Skelgen runs on CDK2 were (i) one hydrogen bond with Leu 83 N and (ii) one hydrogen bond with the backbone carbonyl group of either Glu 81 or Leu 83. In addition, the atom in the ligand structure forming this hydrogen bond had to be different from the acceptor atom interacting with the Leu 83 backbone NH. This additional constraint eliminated the possibility that a ligand hydroxyl group would form both hydrogen bonds. Compounds of this type are known, but to our knowledge have not led to any CDK2 inhibitors of preclinical interest. The PDB complex structure 1di8 was used in the CDK2 calculations.

As detailed above, CDK2 *de novo* design focussed not so much on the identification of structures that

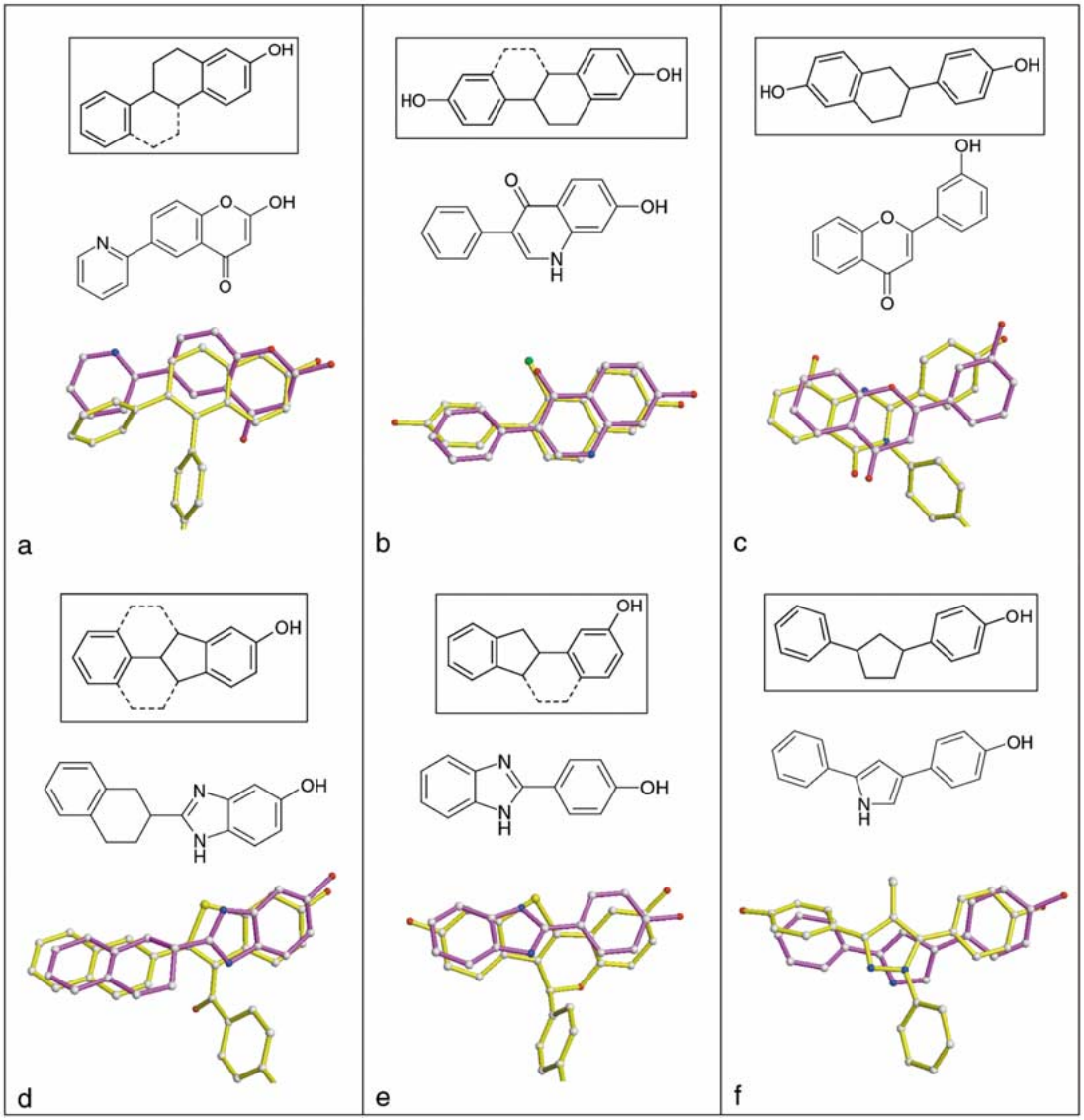

Figure 5. Superimposed pairs of modeled ligands (yellow) and scaffolds designed by Skelgen (magenta) for the estrogen receptor. 2D representations of the Skelgen structures are shown below the general scaffold patterns of known ligands (in box).

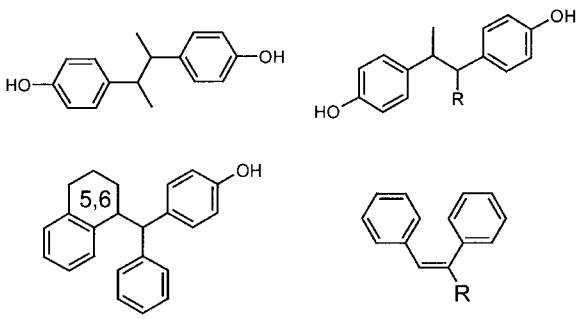

Figure 6. Additional scaffolds of estrogen receptor ligands not re-designed by Skelgen.

would fully occupy the binding site, but rather on finding fragments that would satisfy specific hydrogen bonding interactions with the hinge region. Satisfying these interactions alone does, of course, not lead to high affinity CDK2 inhibitors. Rather, binding affinity is gained through lipophilic interactions between the planar, mostly heterocyclic ring systems carrying the donor and acceptor groups binding to the hinge and surrounding aliphatic peptide side chains. The recurring binding motifs in our CDK2 literature inhibitor collection are depicted in Figure 8 together with analogues generated by Skelgen.

Clearly, all binding motifs are reproduced, but not necessarily incorporated into the same ring systems found in the known inhibitors. The orientation of these binding motifs is generally very close to those in the manually modeled ligands. It should be noted that the CDK2 ATP binding site is a very flexible site that can undergo significant conformational changes. The buried portion of the hinge strand from Phe 80 to the backbone NH of Leu 83 does not move significantly. The Leu 83 carbonyl group, however, can either be oriented into the binding site in a position suitable to form hydrogen bonds, or perpendicular to the adenine binding site (towards the reader in Figures 7 and 8). The x-ray structure selected for this study (1di8) has an intermediate orientation that is not optimally oriented for hydrogen bonding, but also admits placing lipophilic groups alongside the hinge. For this reason, we have observed more, and more optimally oriented, Skelgen solutions for the donor-acceptor pair NH Leu 83 – C=O Glu 81. For example, Skelgen places a related seven-membered ring lactame in almost exact accordance to the modeled orientation of a paullone [57] ring system (Figure 8b).

Since the first step in the design of novel kinase inhibitors is the choice of suitably substituted heterocycle scaffolds, the score constraints do not have to be very restrictive in this particular case. Even without any score constraints, Skelgen produces many interesting scaffolds that could be tested as a first step in a research program aiming at identifying weakly, but specifically binding compounds. This kind of approach has been particularly successful for ATP binding sites [69, Honma, 2001 #833].

MMP3. The complex structure with PDB code 1g49 was selected for calculations on stromelysin. An essential feature of all MMP inhibitors is a functional group forming specific and strongly directed interactions with the catalytic zinc ion and surrounding residues. The majority of existing MMP inhibitors contain either a hydroxamate or a carboxylate moiety forming these interactions and only a limited number of further zinc-binding groups is known [70]. It was not attempted to let Skelgen redesign such zinc-binding functional groups. Instead, a hydroxamate fragment was rigidly placed in the binding site using the ligand coordinates of the 1g49 x-ray structure. This template was incorporated into every structure generated by Skelgen (Figure 9). Two further constraints were used: a hydrogen bond had to be formed to the backbone NH groups of either Leu 664 or Ala 665, and a lipophilic atom had to be placed in the S1′ pocket. Because of the spatial proximity of the constraints, Skelgen primarily has to solve the conformational problem of finding a suitable linker between a lipophilic group in the S1′ pocket and the hydroxamate. Although this is the most focused of the four test scenarios investigated here, it is a realistic one for the application of a *de novo* design tool. Often, research programs in the pharmaceutical industry start with considerable knowledge of the target and its ligand binding requirements, and solving this specific problem has been a major topic of MMP inhibitor research during the 1990s.

Known inhibitor scaffolds for stromelysin are depicted in Figure 10. The known inhibitors differ in the length and atom types of the chain connecting the hydroxamate fragment and the substituent pointing towards the S1′ pocket. The hydrogen bond to the backbone NH groups of Leu 664 and Ala 665 is formed either by a carbonyl group or by a sulfonyl moiety. Visual inspection of the Skelgen results (Figure 10) reveals that all of these binding motifs are redesigned in the appropriate conformation. About half of the 600 Skelgen solutions fall into these classes. Differences in the binding orientations between modeled and designed compounds are mostly due to the fact that the hydroxamate moiety was kept fixed during the Skelgen runs, while the modeled structures are force-field minimized.

Skelgen designs many compounds of class c in Figure 10. A wide variety of analogues are experimentally known, in which the two-atom linker between the hydroxamate and the sulfonyl moiety is substituted or incorporated in various ring systems. Because these ring systems do not form significant direct interactions with the enzyme, Skelgen generates only few examples. One of them is depicted in Figure 10b, where the hydrogen bond to Leu 664 is formed by a carbonyl group instead of a sulfonyl group.

Automated similarity analysis. Only visual inspection through a chemist's eye of the output from a *de novo* design program can lead to a full assessment of potentially interesting binding motifs and scaffolds. Nevertheless, it would be very desirable to have an automated, objective measure to evaluate and cluster the structures for comparison. One possibility could be the use of a simple connectivity-based (2D) similarity measure. However, the similarity of any known inhibitor and analogues generated by Skelgen is rather low when measured in terms of standard 2D molecular similarity measures. For instance, the Tanimoto distance between Daylight fingerprints [71] calculated

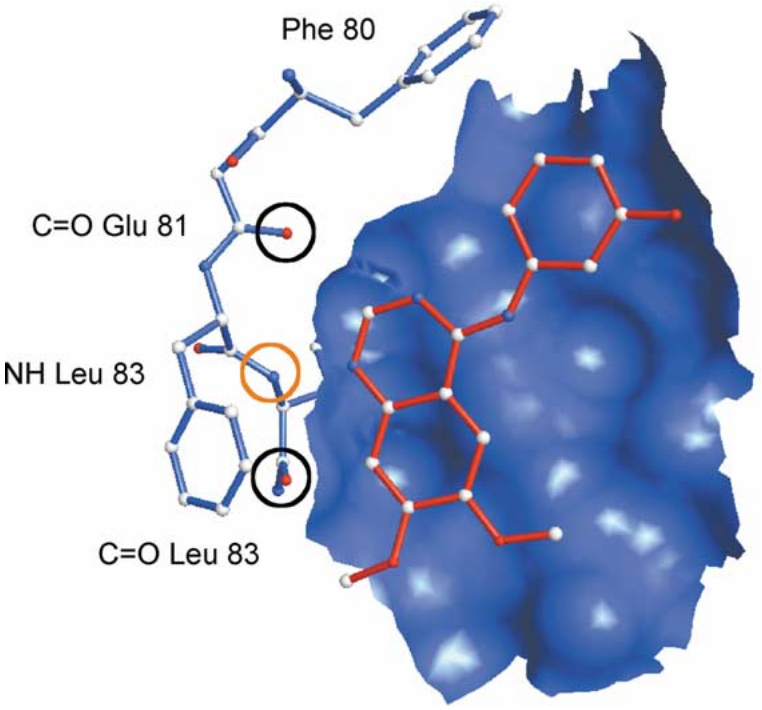

Figure 7. View of the ATP binding site of CDK2 kinase with a ligand bound (PDB code 1di8, top half of the cavity surface removed). The search constraints for Skelgen were: (1) formation of a hydrogen bond with NH Leu 83 and (2) formation of a hydrogen bond to the backbone carbonyl group of either Glu81 or Leu 83.

for a COX-2 ligand and the closest analogue generated by Skelgen is in all cases below 0.8. Furthermore, 2D similarity measures neglect the essential 3D aspect in comparing structures. We have therefore found it necessary to derive a new similarity measure for this purpose. In spite of the low 2D similarity between ligands and Skelgen solutions, they often contain a large common substructure. The size of the common substructure alone, however, is not sufficient as a similarity measure, because it is only relevant if it is oriented in the binding site in the same way. Consequently, we use a weighted sum of the number of atoms in the maximum common substructure (mcs) and the root mean square deviation (rmsd) of these atoms. Hydrogen atoms are disregarded throughout. To calculate the mcs, a program called Macos was written (details see Materials and Methods section) that allows to specify various atom typing criteria. Initially two atoms in different molecules were regarded as identical, i.e. as potential matching partners in the search for the mcs, when they possessed the same element symbol and had the same hybridization (atom typing scheme 1). We used a *proximity score P* that was calculated as

$$P = \omega \cdot rmsd - N_{mcs} \quad (1)$$

where N_{mcs} is the number of atoms in the mcs. A weighting factor ω of 2.0 was empirically found to give the best results. The proximity of two structures increases as the value of P decreases (i.e. becomes more negative). Calculations showed that this score value is useful for comparing structures to single reference compounds, for example for sorting Skelgen-generated structures with respect to their similarity to one known inhibitor. However, the score value is meaningless for comparisons between pairs of compounds that are unrelated to each other (e.g. cannot be used for hierarchical clustering), since it reflects the absolute size of the mcs, and therefore strongly favors matches of large compounds. We therefore introduced the *proximity index p*, which is calculated as -P normalized by the average number of atoms in both molecules:

$$p = \max\left(0, 2\frac{N_{mcs} - \omega \cdot rmsd}{N_1 + N_2}\right) \quad (2)$$

Here, N_1 and N_2 are the numbers of heavy atoms in the two molecules that are compared. The proximity index p was calculated with $\omega = 2.0$. To ensure that values of p are in [0,1], its value is set to zero if

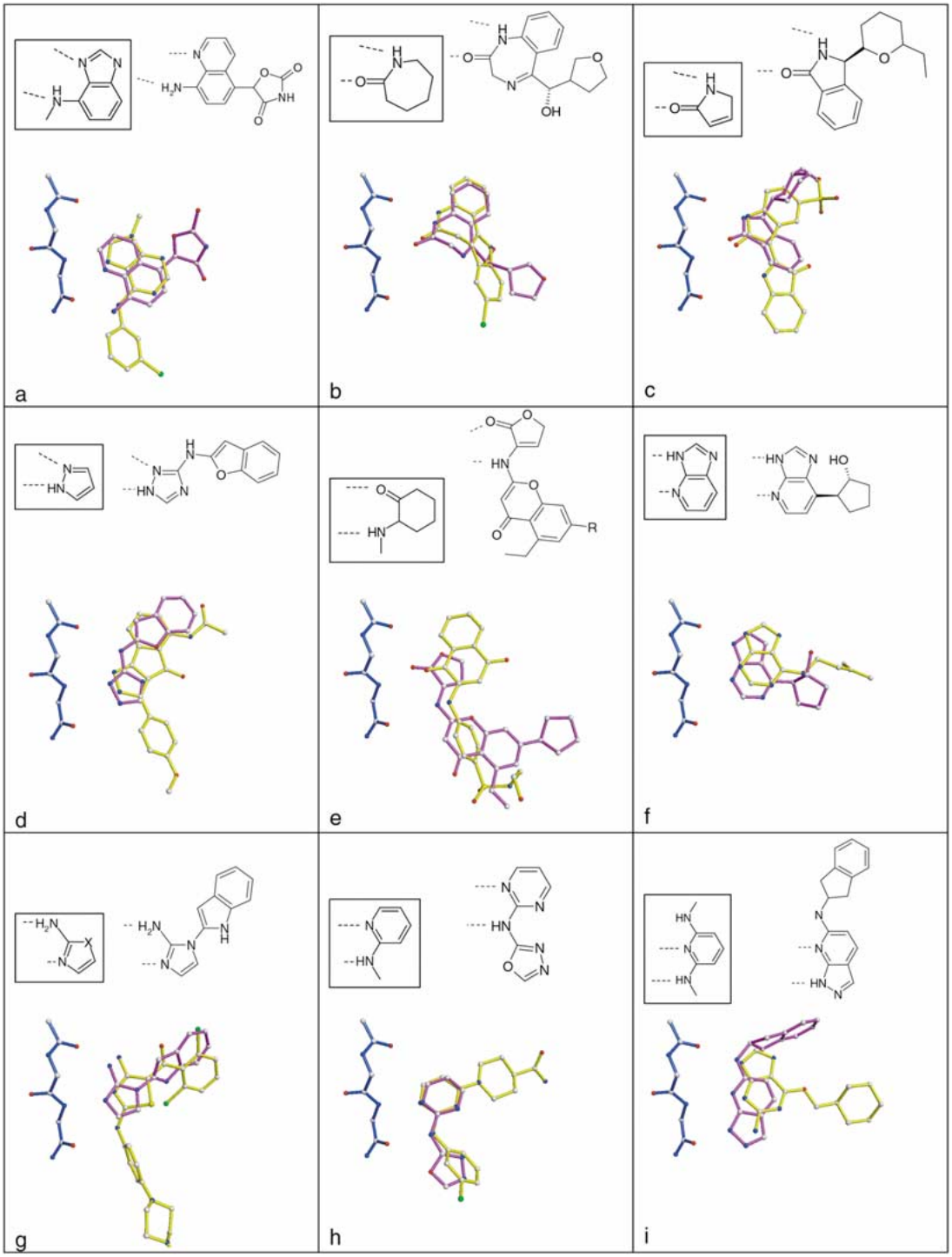

Figure 8. Superimposed pairs of modeled ligands (yellow) and scaffolds designed by Skelgen (magenta) for CDK2. 2D representations of the Skelgen structures are shown next to the general scaffold patterns of known inhibitors (in box).

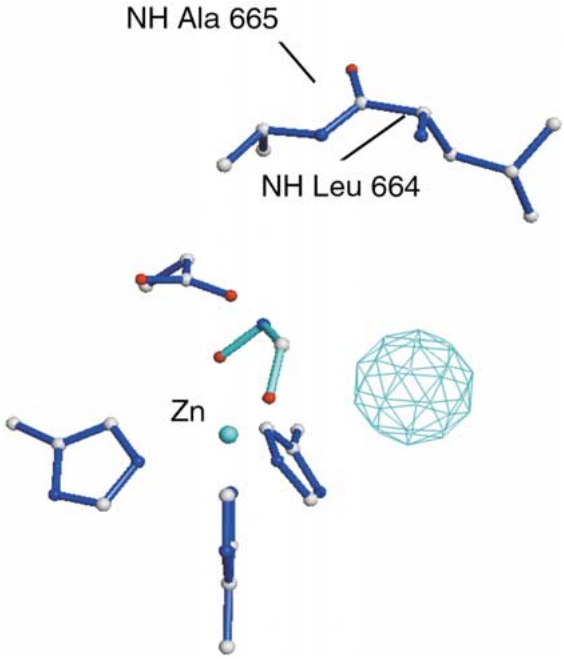

Figure 9. View of the active site of stromelysin-1 (MMP3) with the hydroxamate fragment of the ligand in the crystallographic orientation (PDB code 1g49). The search constraints for Skelgen were: (1) incorporation of the rigid hydroxamate fragment, (2) formation of a hydrogen bond to NH of either Ala 665 or Leu 664, (3) placement of a lipophilic atom within the sphere situated in the S1' pocket.

the normalized expression becomes negative (for large rmsd values or small substructures).

The validity of p was evaluated in the following way. For each of the manually modeled inhibitors, the proximity index p was calculated to each of the 400 structures generated by Skelgen with score threshold values of -20 to -35. The designed structure most similar to each modeled compound was chosen. Figure 11 shows some of the top ranking pairs of structures obtained in this way. In panel A, results for the estrogen receptor are shown. Pairs in the top row were calculated with atom typing scheme 1 described above, i.e. two atoms can be matched onto each other when they have the same element symbol and hybridization. This scheme leads to good pairings of compounds if large portions of two compounds indeed share the same chemical substructure (pairs one and two in the top row of Figure 11) but can also lead to high scores for pairings of completely different or at least differently oriented ring systems (pairs three and four in the top row of Figure 11). We realized that we were generally less interested in whether the two chemical structures share the exact same substructure, but in whether the relative arrangement and topology of ring systems and chains in two scaffolds were identical. Therefore the default atom typing in Macos was modified such that any two atoms could be matched, provided only that the cyclicity of the matched bonds is identical, i.e. two matched bonds must be either both in a ring or in an open chain (atom typing scheme 2). We find that p values calculated in this manner provide robust and general tool for scaffold comparison of modeled inhibitors and *de novo* designed structures. Row two in panel A of Figure 11 shows four examples for the estrogen receptor with the same reference compounds as in row one, and it is clear that the results are meaningful from a medicinal chemistry point of view. The only drawback is that due to the bond cyclicity criterion, large fused-ring systems cannot be easily compared with closely matching open chain analogs (column 2 in Figure 11A) as generated by Skelgen.

The proximity index p yield satisfying results for COX-2 as well. Close analogs of known inhibitors can be identified among the Skelgen-generated structures by means clustering and sorting techniques. Figure 11B shows four of the closest pairs identified in this way. These pairings were again obtained with atom typing scheme 2. Atom typing scheme 1 mostly yields matches between completely different scaffolds (results not shown). Examples for MMP3 are omitted from Figure 11, because in this case substructure topologies of all generated compounds are

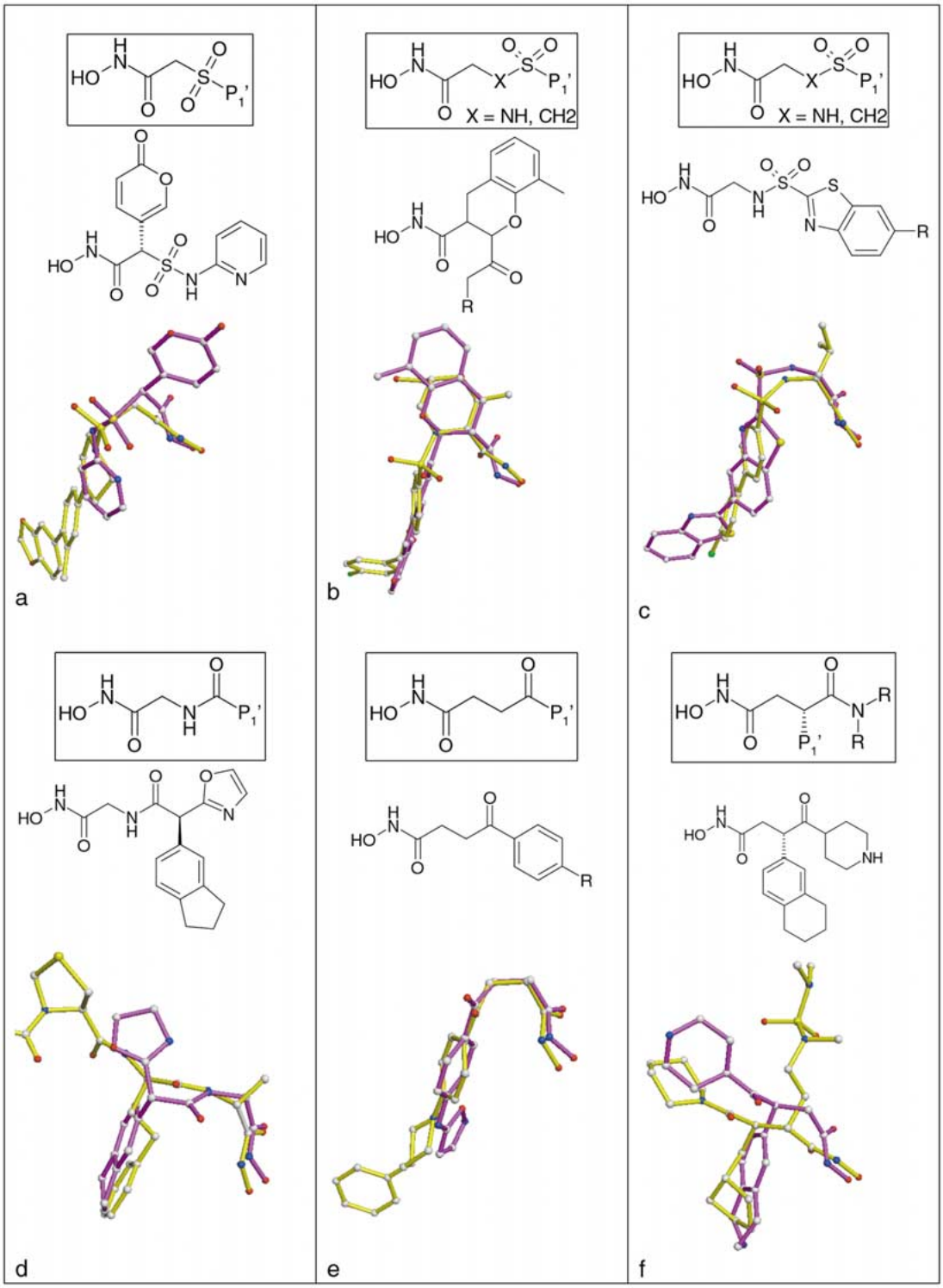

Figure 10. Superimposed pairs of modeled ligands (yellow) and scaffolds designed by Skelgen (magenta) for MMP3. 2D representations of the Skelgen structures are shown next to the general scaffold patterns of known inhibitors (in box).

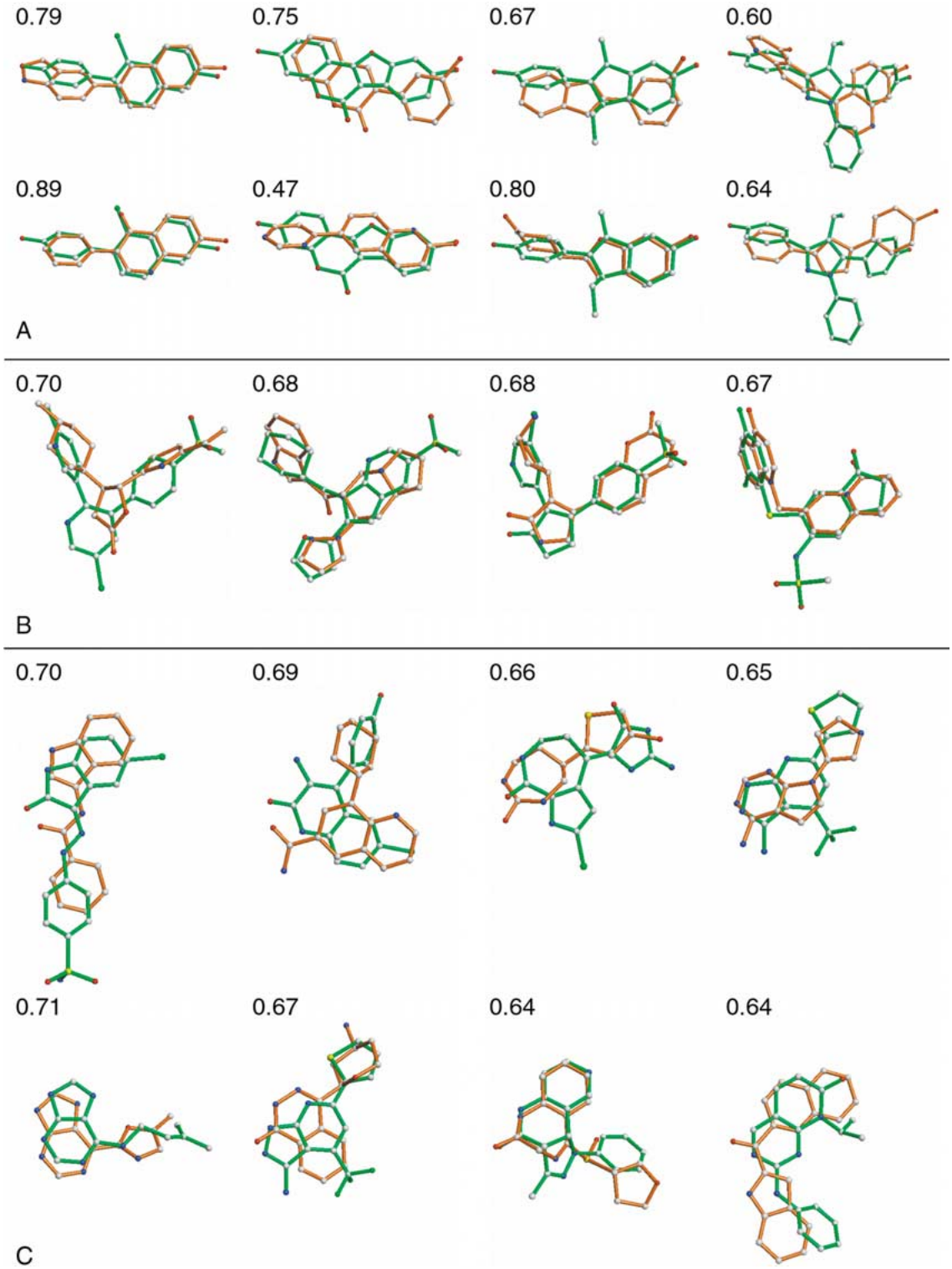

Figure 11. Pairs of known inhibitors (green) and Skelgen designs (orange) having a high similarity according to the proximity index p (details see text). Similarity values are given above each compound pair. Skelgen structures are closest analogs of the shown modeled inhibitors according to p.

very closely related. For CDK2, results are shown in the top row of Figure 11C. Clearly, these pairings do not reflect interaction patterns of compounds. Here, a comparison of the topological properties of structures may be less meaningful, and different means to compare structures could be used.The program Moloc[72] contains the option to calculate a similarity index between two overlayed 3D structures taking into account the volume overlap of lipophilic regions and the degree of matching of directed hydrogen bond donor and acceptor functions. Examples for good matches according to this measure are depicted in Figure 11C, bottom row.

Materials and methods

Modeling known inhibitors

All inhibitors were modeled manually into the binding sites of their respective targets prior to the beginning of the Skelgen calculations. The interactive modelling program MOLOC [72] in conjunction with the MAB force field [73] was used for this purpose. The protein was left rigid in all cases. The united atom version of the MAB force field was used. Inhibitors were minimized starting from manually generated initial conformations and orientations that corresponded best to the experimentally known binding modes of known inhibitors retrieved from the PDB.

The Skelgen program

The Skelgen program allows a drug designer to define design constraints in a flexible way using several mechanisms. Novel ligands are built into a an area of a protein structure defined by a rectangular box by linking predefined fragments together. The composition of the fragment lists and the ways fragments are linked together are under user control. Also, several types of pharmacophore points can be defined to focus the search.

The fragment library used in this study consists of two sublibraries. The first one includes 12 acyclic fragments and the second one 576 ring fragments. The ring fragments were derived from the World Drug Index [74] by extracting ring systems from each structure in the database. Ring systems are defined either isolated rings or more complex (annulated, spiro or bridged) ring systems. Ring systems occurring infrequently in the WDI were not included in the standard fragment library. Also, the attachment points of substituents at each ring system were labeled and only these are used to link other fragments when Skelgen is run. Several chemists have inspected the list and removed unwanted fragments based on their experience.

Structure generation is driven by the Skelgen objective function. Several terms are included in the objective function and they take into consideration different geometric, connectivity and chemical constraints imposed in the design process. When individual requirements are satisfied, the corresponding terms become zero, otherwise they are greater than zero and so is the overall objective function. Skelgen uses simulated annealing to minimize the objective function. A simple atom-repulsion function is used to calculate intramolecular interactions in the ligand:

$$\Sigma_{ij}\Theta(r_{ij} - 0.75r^*_{ij})$$

where Θ is the Heaviside's step function [75], i and j are non-bonded atoms in the ligand, r_{ij} is the distance between them, r^*_{ij} is the sum of the van der Waals radii of atoms i and j. A similar function is used to calculate the repulsive part of the interaction between ligand and receptor atoms. The difference is that interactions between H atoms are not considered and r^*_{ij} is set to 1.5 Å if atoms i and j form a metal - acceptor pair, to 2.5 Å if i and j form a hydrogen bond and to 3 Å otherwise.

A fast empirical scoring function [5] is used to estimate the binding energy of the ligand. It contains terms accounting for hydrogen bonds, interactions with metal ions available in the binding site, hydrophobic interactions and ligand entropic contributions taken into account through the number of rotatable bonds. Skelgen allows the user to define a threshold value for the free energy of interaction and only structures with values lower than the threshold are written out and analyzed further. Several threshold values have been used here from -20 to -50 in intervals of 5 score units. This mechanism allows the designer to sample ligands with different free energies that fulfill all other search constraints. Full details of the derivation and testing of the scoring function have been published elsewhere [33].

Pharmacophore points are a valuable mechanism to focus the ligand search to regions in the active site believed to be important for binding and specificity. These points can be derived from analyses binding modes of known ligands, or from identifying hot spots in the binding site with programs such as GRID [76]. Pharmacophore points can be defined as spheres, or spherical shells, but directionality can

also be included in the definition. Various types of pharmacophores for hydrogen bond donors, acceptors and for lipophilic, polar, or 'any' atom or ring centers are predefined. Distance and angular constraints to hydrogen and neighboring atoms in the ligand as well as the receptor can be flexibly defined. Furthermore, pharmacophore points can be grouped into sets, and a minimum number of pharmacophore points out of each set can be required to be satisfied by each solution structure. For example, a simple query may consist of two pharmacophore spheres at a certain distance and the requirement that any pair of ligand atoms should match. The pharmacophore term in the objective function will become zero if this is the case. Otherwise the penalty term added to the objective function will be the sum of the distances of the nearest ligand atom to the surface of the corresponding sphere. A more complex example could be the definition of five pharmacophore points and the requirement that at least three out of these should be satisfied by ligand atoms. The pharmacophore point constraint will become zero when a ligand is generated that matches any of the three-point pharmacophores. This is easier than defining all 10 possible triplets out of the five pharmacophore points individually and combining search results. Thus Skelgen allows one to define pharmacophore queries at a fine or coarse level as required and to incorporate the chemical intuition of the designer in the process. Multiple examples of pharmacophore definitions are used and have been discussed elsewhere [33].

In addition to using fragments extracted from the WDI and labeling their substitution atoms, adding penalties for predefined unwanted substructures further enhances synthetic feasibility of the ligands. There are 62 such substructures defined. These have been derived by an iterative process. Skelgen has been run repetitively, the output has been examined by chemists, and new substructure rules have been added until no obviously erroneous or undesirable structures were found in the Skelgen output. In this study, every time a new structure is examined inside Skelgen, substructure searches are performed to check if any of the unwanted substructures is present.

Macos algorithm and its implementation

Essential elements of the proximity measures (P and p) used in this study are the maximum common substructure (mcs) of two small molecules and the root mean square deviation (rmsd) of these atoms. A program called Macos performs the mcs calculation. Finding the mcs is an example of an NP complete problem [77]. Only the moderate size of ligand structures encountered in drug design projects allows one to perform the mcs computation within a reasonable amount of CPU time. Within Macos, the common subgraph between two structures, A and B, is found by an exhaustive depth-first search. Common subgraphs are extended one atom at a time, alternating between atoms from A and B. User-defined criteria define whether they can be matched onto each other or not. If, for a newly selected atom, this is not the case, the algorithm backtracks and a different atom from the same molecule is tried. Subgraphs can be extended up to the number of atoms in A or B, whichever is smaller. Proximity measures P or p are calculated for every common subgraph encountered during the search and the best one is kept and updated. The following rules are used at each step of the subgraph construction.

1. An atom from A or B must be included in the subgraph only once.
2. Atoms added to the subgraph must be connected to at least one atom already included in the subgraph.
3. Corresponding atoms and bonds from A and B in the subgraph must match, according to user-defined criteria. Several options can be used. Macos can consider or ignore differences between element types, atom hybridization and cyclicity of atoms or bonds. As a default, hydrogen are ignored in the calculation.
4. The first atom from structure A, included in the common subgraph, should have the smallest number from all atoms of A included in the subgraph, the numbering system being identical to the input order of the atoms.
5. Atoms are added layer by layer around the first atom from A or B included in the subgraph, i.e. atoms which are three bonds away from the first atom in the subgraph should be included only after all atoms which are two bonds away have already been considered.

Rules 4 and 5 ensure that permutationally equivalent subgraphs are excluded from consideration during the construction.

Conclusions and outlook

We have shown that the computer program Skelgen is able to re-discover a broad variety of known ligands for four different targets. To our knowledge, this is the first truly comprehensive evaluation of a *de novo*

design tool. The results show that *de novo* design can be an extremely powerful tool for the generation of potential novel chemotypes. The results also indicate that it should be possible to employ *de novo* design for a comprehensive approach to targets or even target classes – not only for solving *ad hoc* design problems. Furthermore, the analysis has pointed out a number of areas for improvement that would not have become visible without such a thorough analysis: The probabilities for choosing small acyclic linker fragments in a fragment addition step should be increased relative to the probability of choosing a ring system as a default. This should lead to a better mixture of designed structures with ring-ring single bonds and longer linkers between rings. Also, for lipophilic binding pockets the steric repulsive term should be made slightly more stringent and get a higher weight in the objective function.

A key element of Skelgen is the objective function. It consists of a variety of penalty terms, and thus summarizes all deficiencies a designed structure might still have with respect to the chosen target values. It is important to note that this function, and not an empirically calculated interaction energy score alone, drives the generation of new structures. The simultaneous observation of a multitude of constraints leads to a much higher success rate.

Skelgen has a powerful mechanism for the definition of pharmacophore elements and other constraints that should be fulfilled by all designed structures. This allows the inclusion of knowledge on ligand binding that may be available in addition to the 3D structure of the binding site. We regard this as an essential element for successful *de novo* design. Skelgen produces mostly relaxed conformations of scaffolds that require only a final force field minimization step to yield structures suitable for visual inspection and further modeling. While not every designed scaffold will be synthetically feasible by itself, a trained chemist will quickly be able to identify accessible analogues.

The proximity measure that we have developed and employed for assessing the similarity of known ligands and designed scaffolds has proved to be useful for those cases where scaffolds themselves and their binding modes should be compared, rather than specific interactions they form with the receptor. It is a valuable tool for clustering Skelgen results to help molecular design experts as well as chemists to navigate through hundreds of inhibitor candidates. It has also become clear that it will be necessary to develop additional target-specific criteria for filtering and clustering, such as specific interaction patterns or substructures in a particular spatial orientation. The use of the Moloc 3D similarity measure has been shown as an example. It will be worthwhile to invest additional efforts into the analysis of *de novo* design results, because such tools will help to identify structural and interaction patterns that are essential for an understanding of receptor-ligand interactions and for the reduction of *de novo* design to practice. Methods of this type will be invaluable in the post-genomic era for processing the avalanche of target data to be expected soon.

Acknowledgmenents

The authors thank Paul Gerber for providing Moloc and implementing a tailor-made version of the 3D similarity measure, and David Bailey for numerous stimulating discussions.

References

1. Murcko, M. A. In Lipkowitz, K. B. and Boyd, D. B. (Eds.), Reviews in Computational Chemistry 11, Wiley-VCH, New York, 1997, pp. 1–66.
2. Clark, D. E., Murray, C. W. and Li, J. in Lipkowitz, K. B. and Boyd, D. B. (Eds.), Reviews in Computational Chemistry 11, Wiley-VCH, New York, 1997, pp. 67–126.
3. Stahl, M. In Schneider, G. and Boehm, H.-J. (eds.), Virtual Screening for Bioactive Molecules VCH, Weinheim, 2000, pp. 229–264.
4. Tame, J. R. H., J. Comput.-Aided Mol. Design, 13 (1999) 99.
5. Boehm, H.-J. and Stahl, M., Med. Chem. Res., 9 (1999) 445.
6. Muegge, I. and Rarey, M. In Lipkowitz, K. B. and Boyd, D. B. (Eds.), Reviews in Computational Chemistry 17, Wiley-VCH, New York, 1997, pp. 1–60.
7. DeWitte, R. S., Ishchenko, A. V. and Shakhnovich, E. I., J. Am. Chem. Soc., 119 (1997) 4608.
8. Joseph-McCarthy, D., Hogle, J. M. and Karplus, M., Proteins, 29 (1997) 32.
9. Castro, A., Richards, W. G. and Lyne, P. D., Med. Chem. Res., 9 (1999) 98.
10. Burkhard, P., Hommel, U., Sanner, M. and Walkinshaw, M. D., J. Mol. Biol., 287 (1999) 853.
11. Otyepka, M., Krystof, V., Havlicek, L., Siglerova, V., Strnad, M. and Koca, J., J. Med. Chem., 43 (2000) 2506.
12. Schapira, M., Raaka, B. M., Samuels, H. H. and Abagyan, R., Proc. Natl. Acad. Sci. USA, 97 (2000) 1008.
13. Grueneberg, S., Wendt, B. and Klebe, G., Angew. Chem. Int. Ed., 40 (2001) 389.
14. Good, A., Curr. Opin. Drug Disc. Dev., 4 (2001) 301.
15. Sun, Y., Ewing, T. J. A., Skillman, A. G. and Kuntz, I. D., J. Comput.-Aided Mol. Design, 12 (1998) 579.
16. Rarey, M. and Lengauer, T., Persp. Drug Disc. Design, 20 (2000) 63.
17. Boehm, H.-J., Banner, D. W. and Weber, L., J. Comput.-Aided Mol. Design, 13 (1999) 51.

18. Kick, E. K., Roe, D. C., Skillman, A. G., Liu, G., Ewing, T. J. A., Sun, Y., Kuntz, I. D. and Ellman, J. A., Chem. Biol., 4 (1997) 297.
19. Grzybowski, B. A., Ishchenko, A. V., Kim, C.-Y., Topalov, G., Chapman, R., Christianson, D. W., Whitesides, G. M. and Shakhnovich, E. I., Proc. Natl. Acad. Sci. USA, 99 (2002) 1270.
20. Hubbard, R. E., Curr. Opin. Biotechnology, 8 (1997) 696.
21. Kubinyi, H., Curr. Opin. Drug Disc. Dev., 1 (1998) 4.
22. Hicks, S., Assefa, H. and Sindelar, R., Curr. Opin. Drug Disc. Dev., 2 (1999) 223.
23. Murcko, M. A., Caron, P. R. and Charifson, P. S., Annu. Reports Med. Chem., 34 (1999) 297.
24. De Lucca, G. V. and Lam, P. Y. S., Drugs Future, 23 (1998) 987.
25. von Itzstein, M., Wu, W.-Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Phan, T. V., Smythe, M. L., White, H. F., Oliver, S. W., Colmant, P. M., Varghese, J. N., Ryan, D. M., Woods, J. M., Bethell, R. C., Hotham, V. J., Cameron, J. M. and Penn, C. R., Nature, 363 (1993) 418.
26. Babu, Y. S., Chand, P., Bantia, S., Kotian, P., Dehghani, A., El-Kattan, Y., Lin, T.-H., Hutchinson, T. L., Elliott, A. J., Parker, C. D., Ananth, S. L., Horn, L. L., Larver, G. W. and Montgomery, J. A., J. Med. Chem., 43 (2000) 3482.
27. Wang, G. T., Chen, Y., Wang, S., Gentles, R., Sowin, T., Kati, W., Muchmore, S., Giranda, V., Stewart, K., Sham, H., Kempf, D. and Laver, W. G., J. Med. Chem., 44 (2001) 1192.
28. Chand, P., Kotian, P. L., Deghani, A., El-Kattan, Y., Lin, T.-H., Hutchinson, T. L., Babu, Y. S., Bantia, S., Elliott, A. J. and Montgomery, J. A., J. Med. Chem., 44 (2001) 4379.
29. Boyd, D. B., Rational Molecular Design in Drug Research, In Liljefors, T., Jorgensen, F. S. and Krogsgaard-Larsen, P. (Eds.), Alfred Benzon Symposium 42, Munksgaard, Copenhagen, 1998.
30. Bailey, D. and Brown, D., Drug Discovery Today, 6 (2001) 57.
31. Bohacek, R. S. and McMartin, C., J. Am. Chem. Soc., 116 (1994) 5560.
32. Leach, A., Bryce, R. A. and Robinson, A. J., J. Mol. Graphics Mod., 18 (2000) 358.
33. Todorov, N. P. and Dean, P. M., J. Comput.-Aided Mol. Design, 11 (1997) 175.
34. Todorov, N. P. and Dean, P. M., J. Comput.-Aided Mol. Design, 12 (1998) 335.
35. Garland, S. L., New computational tools for the design of peptidomimetics, PhD thesis, University of Cambridge, 1996.
36. Carter, J. S., Exp. Opin. Ther. Patents, 8 (1997) 21.
37. Kalgutkar, A. S., Exp. Opin. Ther. Patents, 9 (1999) 831.
38. Talley, J. J., Prog. Med. Chem., 36 (1999) 201.
39. Carter, J. S., Exp. Opin. Ther. Patents, 10 (2000) 1011.
40. Dannhardt, G. and Laufer, S., Curr. Med. Chem., 7 (2000) 1101.
41. Puig, C., Crespo, M., Godesart, N., Feixas, J., Ibarzo, J., Jimenéz, J.-M., Soca, L., Cardelius, I., Heredia, A., Miralpeix, M., Puig, J., Beleta, J., Huerta, J. M., López, M., Segarra, V., Ryder, H. and Palacios, J. M., J. Med. Chem., 43 (2000) 214.
42. Feixas, J., Jimenéz, J.-M., Godessart, N., Puig, C., Soca, L. and Crespo, M. I., Bioorg. Med. Chem. Lett., 11 (2001) 2687.
43. Bosch, J., Roca, T., Catena, J.-L., Llorens, O., Pérez, J.-J., Lagunas, C., Fernandez, A. G., Miquel, I., Fernandez-Serrat, A. and Farrerons, C., Bioorg. Med. Chem. Lett., 10 (2000) 1745.
44. Khanna, I. K., Yu, Y., Huff, R. M., Weier, R. M., Xu, X., Koszyk, F. J., Collins, P. W., Cogburn, J. N., Isakson, P. C., Koboldt, C. M., Masferrer, J. L., Perkins, W. E., Seibert, K., Veenhuizen, A. W., Yuan, J., Yang, D.-C. and Zhang, Y. Y., J. Med. Chem., 43 (2000) 3168.
45. Magarian, R.vA., Overacre, L. B., Singh, S. and Meyer, K. L., Curr. Med. Chem., 1 (1994) 61.
46. Coghlan, M. J. and Kort, M. E., Exp. Opin. Ther. Patents, 9 (1999) 1523.
47. Gegnas, L. D., Exp. Opin. Ther. Patents, 10 (2000) 833.
48. Sippl, W., J. Comput.-Aided Mol. Design, 14 (2000) 559.
49. Grese, T. A., Pennington, L. D., Sluka, J. P., Adrian, M. D., Cole, H. W., Fuson, T. R., Magee, D. E., Phillips, D. L., Rowley, E. R., Shetler, P. K., Short, L. L., Venugopalan, M., Yang, M. N., Sato, M., Glasebrook, A. L. and Bryant, H. U., J. Med. Chem., 41 (1998) 1272.
50. Fink, B. E., Mortensen, D. S., Stauffer, S. R., Zachary, D. A. and Katzenellenbogen, J. A., Chem. Biol., 6 (1999) 205.
51. Kim, S.-H. and Katzenellenbogen, J. A., Bioorg. Med. Chem., 8 (2000) 785.
52. Stauffer, S. R., Coletta, C. J., Tedesco, R., Nishiguchi, G., Carlson, K., Sun, J., Katzenellenbogen, B. S. and Katzenellenbogen, J. A., J. Med. Chem., 43 (2000) 4934.
53. Tedesco, R., Youngman, M. K., Wilson, S. R. and Katzenellenbogen, J. A., Bioorg. Med. Chem. Lett., 11 (2001) 1281.
54. Rosania, G. R. and Chang, Y.-T., Exp. Opin. Ther. Patents, 10 (2000) 215.
55. Sielecki, T. M., Boylan, J. F., Benfield, P. A. and Trainor, G. L., J. Med. Chem., 43 (2000) 1.
56. Dumas, J., Exp. Opin. Ther. Patents, 11 (2001) 405.
57. Schultz, C., Link, A., Leost, M., Zaharevitz, D. W., Gussio, R., Sausville, E. A., Meijer, L. and Kunick, C., J. Med. Chem., 42 (1999) 2909.
58. Barvian, M., Boschelli, D. H., Cossrow, J., Dobrusin, E., Fattaey, A., Fritsch, A., Fry, D., Harvey, P., Keller, P., Garrett, M., La, F., Leopold, W., McNamara, D., Quin, M., Trump-Kallmeyer, S., Toogood, P., Wu, Z. and Zhang, E., J. Med. Chem., 43 (2000) 4606.
59. Furet, P., Zimmermann, J., Capraro, H.-G., Meyer, T. and Imbach, P., J. Comput.-Aided Mol. Design, 14 (2000) 403.
60. Nugiel, D. A., Etzkorn, A.-M., Vidwans, A., Benfield, P. A., Boisclair, M., Burton, C., Cox, S., Czerniak, P. M., Doleniak, D. and Seitz, S. P., J. Med. Chem., 44 (2001) 1334.
61. Sielecki, T. M., Johnson, T. L., Liu, J., Muckelbauer, J. K., Grafstrom, R. H., Cox, S., Boylan, J., Burton, C. R., Chen, H., Smallwood, A., Chang, C.-H., Boisclair, M., Benfield, P. A., Trainor, G. L. and Seitz, S. P., Bioorg. Med. Chem. Lett., 11 (2001) 1157.
62. Babine, R. E. and Bender, S. L., Chem. Rev., 97 (1997) 1359.
63. Whittaker, M., Curr. Opin. Chem. Biol., 2 (1998) 386.
64. Beckett, R. P. and Whittaker, M., Exp. Opin. Ther. Patents, 8 (1998) 259.
65. Skiles, J. W., Gonnella, N. C. and Jeng, A. Y., Curr. Med. Chem. (2001) 425.
66. Song, Y., Connor, D. T., Sercel, A. D., Sorenson, R. J., Doubleday, R., Unangst, P. C., Roth, B. D., Beylin, V. G., Gilbertsen, R. B., Chan, K., Schrier, D. J., Guglietta, A., Bornemeier, D. A. and Dyer, R. D., J. Med. Chem., 42 (1999) 1161.
67. Zhu, J., Yu, H., Fan, H., Liu, H. and Shi, Y., J. Comput.-Aided Mol. Design, 15 (2001) 447.
68. Meyer, E. F., Botos, I., Scapozza, L. and Zhang, D., Persp. Drug Disc. Design, 3 (1995) 168.
69. Boehm, H.-J., Boehringer, M., Bur, D., Gmuender, H., Huber, W., Klaus, W., Kostrewa, D., Kuehne, H., Luebbers, T., Meunier-Keller, N. and Mueller, F., J. Med. Chem., 43 (2000) 2664.

70. Borkakoti, N., Curr. Opin. Drug Disc. Dev., 2 (1999) 449.
71. Daylight Chemical Information Systems, Santa Fe, New Mexico, USA.
72. Moloc: Gerber Molecular Design, 2002, www.moloc.ch.
73. Gerber, P. R. and Müller, K., J. Comput.-Aided Mol. Design, 9 (1995) 251.
74. World Drug Index, Derwent Information, 1996.
75. Bracewell, R. N. The Fourier transform and its applications, McGraw-Hill, New York, 1999.
76. Goodford, P. J., J. Med. Chem., 28 (1985) 849.
77. Garey, M. and Johnson, D. Computers and Intractability - A guide to the theory of NP-Completeness, Freeman, San Francisco, 1979.