Flexible Docking of Ligands into Synthetic Receptors Using a Two-Sided Incremental Construction Algorithm

Andreas Steffen,* Andreas Kämper, and Thomas Lengauer

Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, D-66123 Saarbrücken, Germany

Received March 4, 2006

We present a new algorithm for the fast and reliable structure prediction of synthetic receptor−ligand complexes. Our method is based on the protein−ligand docking program FlexX and extends our recently introduced docking technique for synthetic receptors, which has been implemented in the program FlexR. To handle the flexibility of the relevant molecules, we apply a novel docking strategy that uses an adaptive two-sided incremental construction algorithm which incorporates the structural flexibility of both the ligand and synthetic receptor. We follow an adaptive strategy, in which one molecule is expanded by attaching its next fragment in all possible torsion angles, whereas the other (partially assembled) molecule serves as a rigid binding partner. Then the roles of the molecules are exchanged. Geometric filters are used to discard partial conformations that cannot realize a targeted interaction pattern derived in a graph-based precomputation phase. The process is repeated until the entire complex is built up. Our algorithm produces promising results on a test data set comprising 10 complexes of synthetic receptors and ligands. The method generated near-native solutions compared to crystal structures in all but one case. It is able to generate solutions within a couple of minutes and has the potential of being used as a virtual screening tool for searching for suitable guest molecules for a given synthetic receptor in large databases of guests and vice versa.

INTRODUCTION

During the development of a synthetic receptor the major goal is to achieve sufficient binding affinity. In some cases it is even more important to secure high selectivity of the receptor to a particular ligand. Thus, often additional efforts have to be expended until the receptor is specific enough to discriminate between similar ligands. Efficient computational approaches for the design of synthetic receptors are of great interest in current supramolecular chemistry research since synthetic receptors have proved to be of high demand in various fields: In analytical chemistry synthetic receptors can serve as separators that selectively bind guest molecules. If the binding process is accompanied by a change in the chromophore of a synthetic receptor the complex formation is directly visible.[1] The structural simplicity of synthetic receptors in contrast to the complex nature of proteins affords a focused view on the processes involved in molecular recognition. For example, Schrader et al. developed selective synthetic receptors for catecholamines.[2] Cyclodextrins serve as solubilizers for poorly soluble drugs such as diclofenac sodium[3] and are components in several drug formulations on the market.[4,5] The selective recognition of proteins by synthetic receptors is an emerging topic that will gain increasing importance in the future.[6]

Due to the industrial relevance of tailored synthetic receptors the need for methods that allow for faster, more rational development increases. This need mirrors a recent trend in medicinal chemistry. There, in earlier times, the discovery of drugs was often a question of serendipity.[7] Nowadays, a wide range of diverse methods and tools are available that assist in the fast and rational development of novel drugs. Whereas in drug design the optimal ligand for a given protein target is searched, in supramolecular chemistry often the inverse scenario is given: For an existing ligand an optimal synthetic receptor has to be developed.

One of the methods which a medicinal chemist has at his disposal employs so-called docking tools.[8−10] Here, the conformational space of the ligand is explored within the defined binding site of a protein. Scoring functions estimate the binding affinity. With docking methods so-called virtual screenings can be performed, in which typically large data sets of molecular compounds are tested for their binding affinity to the protein in silico. The top-scoring molecules are supposed to be the most promising candidates for the further optimization steps and are selected for subsequent refinement. Whereas current state-of-the-art tools for docking are able to handle the flexibility of the ligand, the efficient and reliable modeling of the protein's flexibility still remains a challenging task. In most attempts the assumption is made that the protein stays rigid during the binding process.[11−13] Only some docking tools allow for limited flexibility, e.g. side-chain movement,[14] movement of hydroxyl groups,[15] or use ensembles of different protein conformations.[16−18] Despite the fairly rough assumption of a rigid receptor, numerous success stories in which docking tools have led to novel drug candidates have been published.[19,20] This supports the applicability of the approach.

However, for a number of synthetic receptors the assumption of a rigid receptor cannot be carried over, as they can exhibit a high degree of flexibility similar to ligands. In many cases not only the ligand but also the receptor adapts its conformation during complex formation. This process is often referred to as induced fit.[21]

* Corresponding author phone: +49-681-9325-328; fax: +49-681-9325-399; e-mail: asteffen@mpi-inf.mpg.de.

10.1021/ci060072v CCC: $33.50
Published on Web 06/03/2006

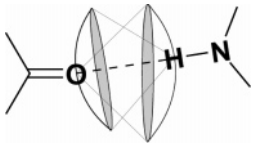


Figure 1. The FlexX interaction scheme. Two interaction groups enter into an interaction if their surfaces mutually lie on the center of the counter group.

In recent work some of us presented a docking program specifically tailored toward synthetic receptors.[22] This program demonstrated the transferability of the protein−ligand docking algorithm of FlexX[11] to the field of synthetic receptors. Two different docking strategies were applied, forward and inverse docking. Whereas in forward docking the ligand is flexibly docked into a comprehensive sample of rigid receptor conformations, in inverse docking the roles of the two molecules are exchanged, i.e., the receptor is constructed around the ligand. This approach successfully predicted the structures from 9 out of 10 experimentally determined complexes of the validation set. Nevertheless a limitation of the method became apparent for two cases in which both molecules of the complex exhibited a high degree of flexibility. In one case, a selective receptor for dicarboxylic acids,[23] the docking time was unjustifiably long. In another case, a receptor for tricarboxylic acids,[24] forward docking could not be performed at all, due to the immense number of discrete conformations of the receptor.

In this paper we present a novel docking approach which circumvents these limitations. This technique handles the conformational space of ligand and receptor more efficiently. The conformational space of both molecules is explored simultaneously using information from the respective countermolecule. In the following sections we introduce our method and demonstrate its capabilities on a test set of 10 complexes involving synthetic receptors and their ligands. For all of the test complexes crystal structures are available.

METHODOLOGY

The new docking strategy for synthetic receptors extends our previously described method.[22] The overall principle of the algorithm relies on the incremental construction of the complex from fragments.[11] To model the molecular flexibility a discrete set of preferred torsional angles is used for each rotable bond.[25] These torsional angles have been derived from a statistical analysis of the Cambridge Structural Database (CSD).[26]

The interaction model used in FlexX and FlexR has been adapted from LUDI.[28] Each interacting group in the molecules is described by an interaction center and interaction surface. Two groups of different molecules enter into an interaction if their surfaces mutually lie on the center of the counter group (Figure 1). For computational reasons the surfaces are represented by discrete point sets (Figure 2).[11]

In contrast to our previous work[22] the structures of both molecules—ligand and receptor—are built up incrementally. Since the conformations of both molecules are unknown initially, we cannot use the conformation of one molecule to direct the generation of the conformation of the other molecule. Due to this fact a *precomputation phase* is introduced that determines putative interaction patterns between the two interacting molecules. These interaction

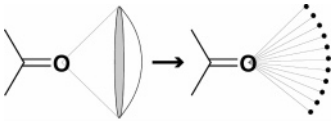


Figure 2. The interaction surfaces are represented by discrete point sets.

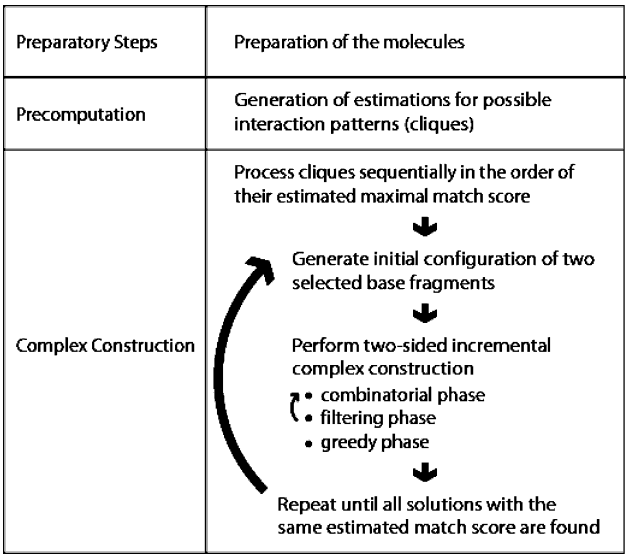


Figure 3. Flowchart of the algorithm.

patterns (cliques) direct the subsequent phase of *complex construction*. A schematic flow diagram of the algorithm is shown in Figure 3 and detailed below.

Phase 1: Preparatory Steps. Initially, the molecules are severed at each acyclic single bond and molecular fragments are obtained. Torsion angles at double bonds, bond lengths, and bond angles are taken from the input structure. The conformations of small ring systems up to a ring size of 10 atoms are computed with the program CORINA.[27] Depending on the type and the neighborhood of an interacting group an appropriate interaction surface and an interaction center is assigned.

Phase 2: Precomputation Phase. In the precomputation phase we collect putative interactions between ligand and receptor which can be realized simultaneously. The search is performed by first identifying putative interaction pairs between the molecules, then generating a docking graph which comprises all identified putative interaction pairs, and finally executing a clique search for finding maximal sets of interactions that can be realized simultaneously.

To illustrate the procedure consider the complex **1** (Figure 4A). From the molecular structures, first molecular graphs are derived in which atoms are represented by nodes, and edges denote covalent bonds between atoms (Figure 4B). We define *centers of directional short-range interactions* as atoms that can form hydrogen bonds or salt bridges. In the case of hydrogen-donor interaction groups we regard the hydrogen atoms themselves as interaction centers. Centers of directional short-range interactions in the two molecules are identified, labeled, and colored according to their functionality (Figure 4B).

From the molecular graphs of the receptor and the ligand a *docking graph* (Figure 5) is derived similarly to the approach used by the program Dock.[29] A node represents

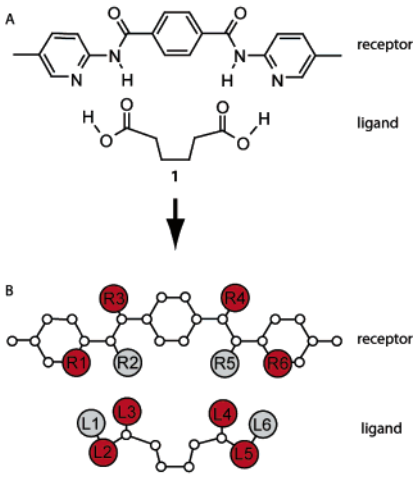


Figure 4. (A) Structural formula of complex **1**. Only polar hydrogen atoms are shown for clarity. (B) Molecular graph representation of complex **1**. Nodes (circles) represent atoms; edges represent chemical bonds between two atoms. Centers of directional short-range interactions are labeled with unique identifiers. Atoms that do not exhibit directional interactions are depicted in white. Hydrogen atoms at hydrogen donor sites and hydrogen-bond acceptors are shown in gray and red, respectively. Apolar hydrogen atoms are not shown for clarity.

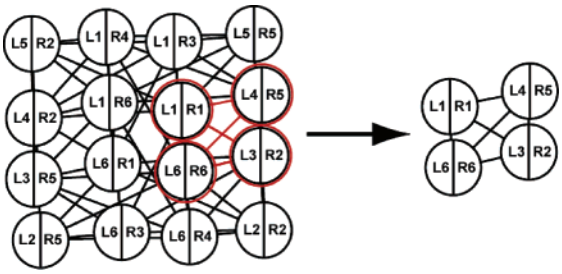


Figure 5. Generated docking graph of complex **1**. Nodes are generated for each possible interaction pair. Two nodes share an edge if they are compatible, i.e., the corresponding interactions can be realized simultaneously. The clique exemplifies one possible complex interaction pattern.

each possible directional short-range interaction that can be formed between an atom of the receptor and an atom of the ligand. The node is labeled with the identifiers of both atoms (Figure 5A). Edges between two nodes are inserted only if two interactions can be realized simultaneously. Two interactions can be realized simultaneously if the two corresponding interaction centers in the receptor are at about the same distance as the two corresponding interaction surfaces in the ligand. In this case, the two nodes representing both interactions are connected with an edge in the docking graph (Figure 6).

The respective distance property is checked by computing bounds on the maximal and minimal distances that can occur between two centers of directional short-range interactions and their corresponding interaction surfaces, respectively, within the conformational space of the molecules (see Supporting Information—Distance-Range Estimation). The resulting graph is then submitted to a clique search using the Bron-Kerbosch algorithm.[30] In a clique all nodes are mutually connected, thus a clique represents sets of interactions that can be realized simultaneously within the same complex structure (Figure 5B). However, it is important to

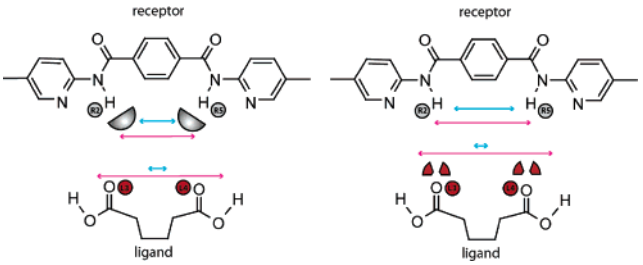


Figure 6. The distance range between the interaction surfaces of the host atoms R2 and R5 exhibits an overlap with the distance range between the interaction centers L3 and L4 of the guest molecule (left). The same applies for the interaction centers R2 and R5 of the host molecule and the interaction surfaces L3 and L4 of the guest molecule (right). This results in an edge between the corresponding nodes in the docking graph.

note that the generated cliques have to be further validated in the complex construction step because, at this point, only simple distance constraints are considered. Additional constraints, such as the exclusion of possible atom overlaps are not taken into account. In cases where the molecules exhibit a high degree of flexibility degenerate cliques might occur. A degenerate clique is a clique comprising a subclique of interactions that can be realized simultaneously plus several interactions that cannot be realized. To deal with these cases not only the maximal cliques that are contained in the docking graph but also their subcliques are added to the clique list.

Whether the generated cliques can yield a valid complex structure is assessed in the complex construction step. So far, in docking algorithms based on the incremental construction principle,[11,22] only one molecule is constructed incrementally. The alternative docking strategy presented here builds up both molecules incrementally, guided by the calculated cliques.

Phase 3: Complex Construction. The precomputation phase results in a set of cliques each of which represents a putative interaction pattern for the molecular complex between receptor and ligand. We denote the interactions represented by a single clique as *targeted interactions*. Complex construction essentially amounts to a heuristic optimization of the complex configuration with respect to a specific score. Scores are heuristic estimates of binding energies that mostly account for contributions by directional interactions. Thus scores are minimized during the optimization. At first, we determine for each clique a simple lower bound on the best attainable score. This lower bound is calculated by simply adding the optimal scores for all participating interactions. In the complex construction phase the cliques are processed sequentially in increasing order of this score. For complex construction first, an initial configuration is generated (see Phase 3.1: Generation of Initial Configurations). In the second step the adaptive two-sided incremental complex construction algorithm (see Phase 3.2: Adaptive Two-Sided Incremental Complex Construction) is applied. The algorithm terminates if a solution is found that fulfills all targeted interactions of a clique. If other cliques exist that exhibit the same estimated maximal interaction energy, then the algorithm proceeds until all cliques with the same estimated score are processed. From each of the cliques that lead to a valid complex structure the 10 best scoring solutions are included in the solution set.

Phase 3.1: Generation of Initial Configurations. The docking process starts with the generation of initial configurations of two selected fragments, one fragment from each molecule. The precomputation of cliques facilitates the targeted generation of an initial configuration. For each single clique of the clique list at first the two fragments are selected that accomplish as many of the targeted interactions as possible. In the following these two fragments are called *base fragments*. In the case that only one directional short-range interaction between the selected base fragments is targeted the *one-point base fragment placement algorithm* is used as described in Kämper et al.[22] In this algorithm a number of discrete placements are generated by just using geometrical information of the two participating interaction surfaces. If the two selected base fragments exhibit more than a single interaction, one of the possible interactions is selected for generating different sterically possible configurations with the one-point placement algorithm. Here, only those configurations are retained for the next step in which all targeted interactions between the base fragments are realized, i.e., the interaction criterion is fulfilled (Figure 1) and no atom overlaps between the fragments exist. All complexes are assessed by means of the scoring function (see Scoring function). To reduce the number of highly similar structures for all generated complexes a clustering procedure is applied (see Clustering). The remaining initial configurations are then submitted to the subsequent *adaptive two-sided incremental construction algorithm*. All of them exhibit the same directive short-range interaction pattern, namely the targeted interactions between the two base fragments.

Phase 3.2: Adaptive Two-Sided Incremental Complex Construction. In this phase both molecules are constructed, starting with the initial complexes, to complete the whole complex structure. Here, we apply an iterative procedure which consists of three repetitive steps (Figure 3): A *combinatorial* step, a *filtering* step, and a *greedy* step.

At the beginning of the two-sided incremental construction the following information has already been computed: (1) the initial base fragment configurations; (2) the list of the remaining targeted interactions that have to be realized in the final complex; (3) the order in which the remaining fragments have to be added to each of the two molecules (see Supporting Information—Determining the fragment order); and (4) the distance ranges from any *outgoing* atom to its respective next targeted interaction atom (see Supporting Information—Distance range estimation method).

For each of the remaining interactions the following procedure is iteratively repeated. In the *combinatorial* step the molecule which requires fewer fragments to reach the next targeted interaction is expanded combinatorially in torsion space until the next targeted interaction group is reached. Partial solutions with inter- or intramolecular clashes are discarded. The scoring function is applied for obtaining an estimation of the energy of each partial solution. To reduce the number of highly similar structures a clustering is performed (see Clustering). If the molecule has reached the next targeted interaction group, then the algorithm proceeds to the *filtering* step, and the other molecule is incrementally built up in torsion space until the targeted interaction is realized. Here, two kinds of filters are applied (see Applied Filters). If none of the (partial) solutions corresponds to the constraints of the currently processed clique, then this particular clique is skipped as the interaction pattern represented by the clique does not lead to a valid complex structure. In this case, the algorithm proceeds to the next clique. All (partial) solutions that fulfill the mandatory targeted interactions are submitted to the clustering procedure again. After all interactions of the processed clique are realized and still fragments remain, the algorithm proceeds to the greedy step. In this step the molecules are expanded alternately. Here, maximally the 100 best scoring partial solutions are taken into the next construction iteration regardless of the absolute value of the scores. This cycle is repeated until the entire complex is completely built up. Finally all solutions are ranked by their scores.

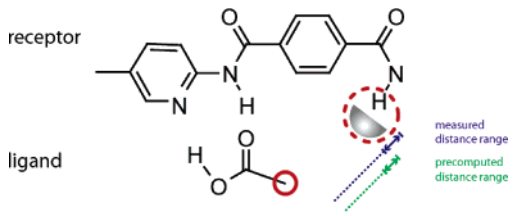


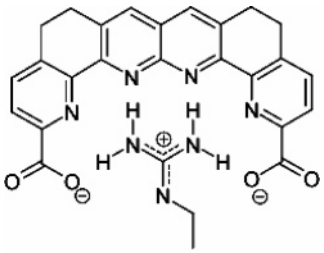
Figure 7. Distance filter. The distance from the current outgoing atom (red circle) to the targeted interaction (dashed red circle) is calculated (blue). If this distance exhibits an overlap with the precomputed distance range (green), the partial solution exhibits the potential to lead to the targeted interaction pattern prescribed by the clique.

Applied Filters in Phase 3.2. In the filtering step of the complex construction two different types of filters are applied in order to direct the complex construction to solutions that exhibit the targeted interaction pattern of the currently processed clique. Two types of filters are implemented, a distance filter and an interaction pattern filter. The distance filter (Figure 7) is applied whenever the position of only one interaction group of the next targeted interaction is known, whereas the counter group of this particular interaction has not been placed yet. In such a case the algorithm compares the current distances between the *outgoing* atom (see Supporting Information—Distance range estimation method) and the discrete surface points of the counter group with the corresponding precomputed distance range. If the two distance ranges do not overlap, the solution is discarded. Otherwise the current construction state of the complex exhibits the potential to fulfill all targeted interactions and is retained.

The interaction pattern filter verifies whether in a given (partial) solution all targeted directional short-range interactions considered so far are realized at this particular construction step. The filter is applied only when a fragment is placed that comprises an interaction group which should form an interaction with its already placed counter group from the clique. A (partial) solution that does not meet all requirements is discarded.

Scoring Function. A scoring function based on the work of Böhm is used for fast energy evaluation throughout the algorithm.[11,31] Five terms are included that describe neutral H-bonds, ionic interactions, aromatic interactions, lipophilic contributions, and entropic costs. Penalty functions are used that penalize deviations from ideal interaction distances and for directional interactions also deviations from ideal angles. In this scoring scheme interactions are considered as being independent from each other, such that the scores of the single interactions within a complex are summed up to obtain

Chart 1. Structural Formula of Complex **10**[a]



10

[a] Apolar hydrogen atoms are omitted for clarity.

the score of the entire complex. This allows for scoring partial solutions.

Clustering. Highly similar (partial) solutions can be clustered to one representative solution. This reduces the number of (partial) solutions while ensuring that no important structural information is lost. After each construction step a complete-linkage clustering algorithm is performed.[11] Here, the distance between two (partial) solutions is measured by means of the root-mean-square deviation (RMSD) for all atoms that are placed in the current state. The RMSD threshold has been set to 0.8 Å. For the purpose of the new algorithm an additional clustering criterion is implemented. If only one interaction group of a targeted interaction has been placed, then the two partial solutions are not clustered if the distance between the two atoms exceeds a threshold. This guarantees that partial solutions with different properties regarding the targeted interactions are retained. For this distance the threshold has been set to 0.4 Å.

Test Data Set. To evaluate our docking strategy we assembled a set of 10 experimentally determined crystal structures of synthetic receptors and their comprised ligands from the Cambridge Structural Database (CSD).[26] We started with the test data as described by Kämper et al.[22] We discarded the macrocyclic receptor by Hamilton and van Engen,[21] since currently the method is unable to handle macrocyclic ring systems. Instead we included a guanidinium receptor (Chart 1) by Bell et al.[32] which was prepared using the protocol which was described by Kämper et al.[22]

RESULTS

The complexes of the validation set exhibit various degrees of flexibility and challenge our method in different ways. A common test for docking tools is the redocking experiment in which the ability of a docking tool to reproduce crystal structures is assessed. Here, FlexR only received the single mol2-files of the molecules as input. No information on the crystal structure was used. All generated solutions were scored by means of the scoring function and sorted in increasing order of their scores. The best scoring solution, i.e., the one with the lowest score, is on rank 1.

Table 2. Docking Results[a]

ID	RMSD [Å] of best scoring solution	min. RMSD [Å] within first 10 best scoring solutions	CPU time MM:SS
1	0.93	0.76	03:30
2	1.17	1.12	01:27
3	3.86	1.04	49:01
4	1.07	0.96	00:31
5	1.08	0.94	00:06
6	0.60	0.60	00:01
7	0.56	0.56	00:01
8	0.73	0.73	01:17
9	0.70	0.70	00:19
10	0.63	0.30	00:03

[a] For each complex, we list the root-mean-square deviation (RMSD) of the best scoring solution, the best RMSD within the 10 best scoring solutions, and the CPU time. CPU times are obtained on an Intel P4 Xeon 3.06 GHz. We use the symmetry corrected all-atom RMSD of the docked complex structure superimposed to the experimentally determined complex structure. (For now symmetry correction is done manually; code for automating this procedure is in the advanced stages of development.)

The results obtained in the redocking experiment are summarized in Table 2. In protein−ligand docking a commonly used criterion for the evaluation of docking results is the RMSD value of a predicted complex structure compared to an experimentally determined crystal structure. Here, typically an RMSD of below 2 Å is considered as a successful docking result. However, the RMSD of the best scoring solution is not necessarily the best criterion for an assessment of docking solutions as it is highly dependent on the applied scoring function. Sometimes only slight deviations between the scores of two solutions prevent a solution with low RMSD at high rank (i.e. rank with a low rank number). For this reason we also report the lowest RMSD within the 10 best scoring solutions. Although we can demonstrate that our scoring function directed the docking algorithm to reasonable solutions, the particular scores are not tabulated as it is a known fact that they do not provide a reliable estimate of the binding affinity in most cases. The scores for the solutions of a particular clique do not vary much as all of them exhibit the same interaction pattern. The number of all solutions found for the complexes varies from 2 for complex **6** to 417 for complex **8**.

Following the defined RMSD criterion, for all of the complexes within our test set our new docking method predicts a reasonable structure at one of the first 10 ranks (Figures 8−19). Furthermore, for complexes **1**, **2**, **4**−**7**, **9**, and **10** near-native solutions were found at rank 1. Consider

Table 1. Test Data Set of Receptor−Ligand Complexes with Known Crystal Structure

ID	short description	CSD code	lit.
1	glutaric acid receptor	JEWNUU	Garcia-Tellado et al.[23]
2	ammonium ion receptor	CUKTUX	Chin et al.[33]
3	tricarboxylic acid receptor	QIJPEE	Ballester et al.[24]
4	two-point binding receptor	POLFUR	Pascal and Ho[34]
5	barbiturate receptor	DAQVAS	Berl et al.[35]
6	creatinin receptor	ZESFEI	Bell et al.[1]
7	bis(guanidinium) receptor for phenyl phosphate	HASWUT	Kneeland et al.[36]
8	bis(guanidinium) receptor for sulfate	QAFVAV	Grossel et al.[37]
9	caffeine receptor	WEWTEX	Waldvogel et al.[38]
10	receptor for guanidinium derivates	DAMQUD	Bell et al.[32]

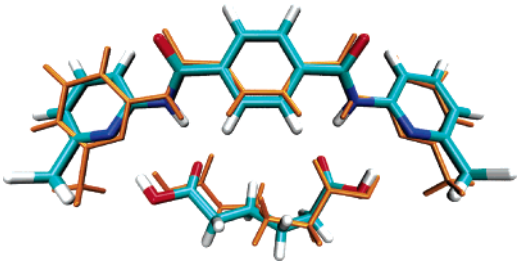


Figure 8. Docking results of complex **1** (rank 1, atom coloring) superimposed on the X-ray structure (orange).

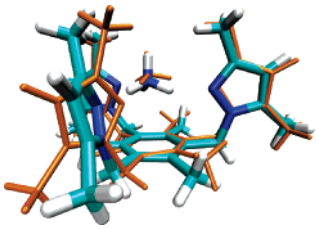


Figure 9. Docking results of complex **2** (rank 1, atom coloring) superimposed on the X-ray structure (orange).

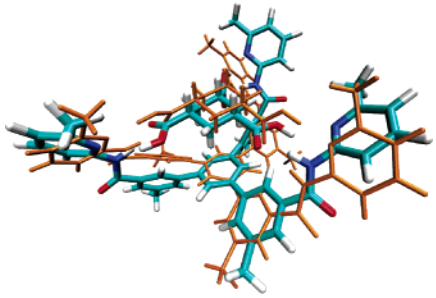


Figure 10. Docking results of complex **3** (rank 1, atom coloring) superimposed on the X-ray structure (orange).

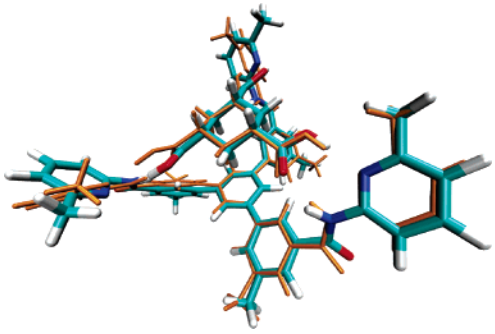


Figure 11. Docking results of complex **3** (rank 7, atom coloring) superimposed on the X-ray structure (orange).

the complexes **2**, **5**, **6**, **9**, and **10**, where an almost perfect prediction has been obtained (Figures 9, 13, 14, 18, and 19). Complexes **4** and **7** exhibit slight deviations in the orientation of aromatic rings, whereas in complex **1** mainly the alkyl part of ligand differs from the crystal structure. The solution found at rank 1 of complex **3** exhibits an RMSD that exceeds the defined threshold of 2 Å, although the exact hydrogen bond pattern of the crystal structure was reproduced (Figure 10). Nevertheless, the complex at rank 7 of complex **3** has an RMSD of 1.0 Å (Figure 11) and was scored only slightly worse than the solution at rank 1. The best scoring solution generated for complex **8** falls within the defined threshold and could thus be considered as near-native. However, regarding the generated H-bond pattern the solution differs

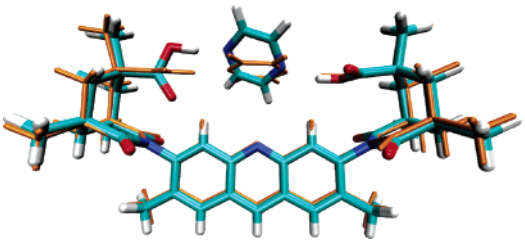


Figure 12. Docking results of complex **4** (rank 1, atom coloring) superimposed on the X-ray structure (orange).

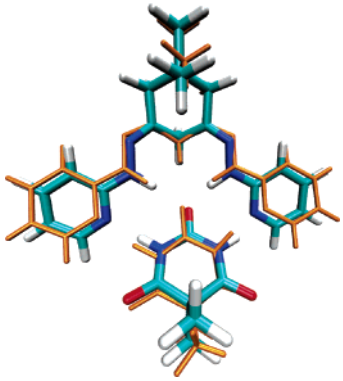


Figure 13. Docking results of complex **5** (rank 1, atom coloring) superimposed on the X-ray structure (orange).

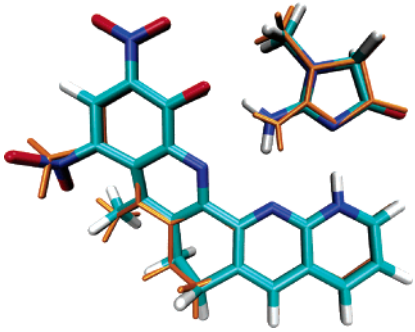


Figure 14. Docking results of complex **6** (rank 1, atom coloring) superimposed on the X-ray structure (orange).

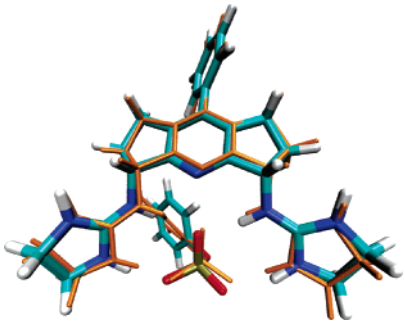


Figure 15. Docking results of complex **7** (rank 1, atom coloring) superimposed on the X-ray structure (orange).

from the crystal structure. In the experimentally determined structure six salt bridges are found of which four are bifurcate. None of the 10 best scoring solutions exhibits this interaction pattern. Considering all generated docking structures of complex **8,** a solution was found at a low rank (minimal RMSD of 0.48 Å observed on rank 263 as shown in Figure 17) that offers the same H-bond pattern as observed in the crystal structure. This supports the ability of our algorithm to generate a near-native structure for this test case

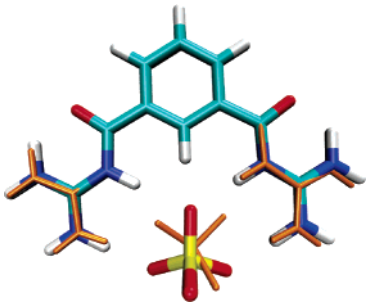


Figure 16. Docking results of complex **8** (rank 1, atom coloring) superimposed on the X-ray structure (orange).

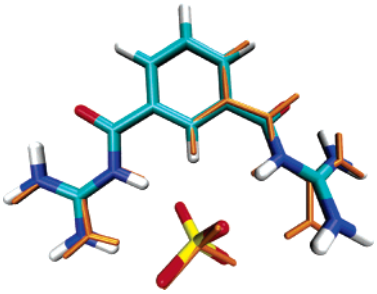


Figure 17. Docking results of complex **8** (solution with minimal RMSD, atom coloring) superimposed on the X-ray structure (orange).

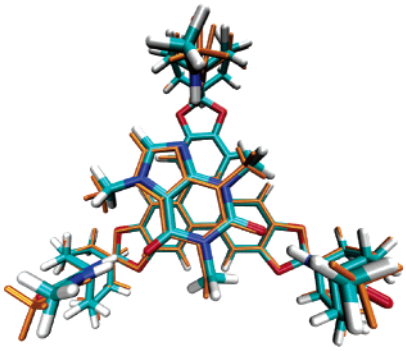


Figure 18. Docking results of complex **9** (rank 1, atom coloring) superimposed on the X-ray structure (orange).

but at the same time reveals some problems of our scoring function in assessing them adequately.

Regarding the computation time the results can be classified into three groups. Complexes **4**−**7**, **9**, and **10** were generated in a couple of seconds. For complexes **1**, **2**, and **8** a couple of minutes were needed. Only the highly flexible complex **3** caused a longer computation time of about 49 min. In comparison to our previously presented work,[22] most importantly, a significant acceleration has been achieved for the two highly flexible complexes **1** and **3**. Here, the computation time could be reduced by about a factor of 80 in the case of complex **1** and to about one fourth in the case of complex **3**. To summarize, for five cases (complexes **1**, **3**, **5**−**7**) the new docking algorithm is faster than the previous one, and for four cases it is slower (complexes **2**, **4**, **8**, **9**).

DISCUSSION

The redocking experiment showed that, in general, our docking strategy produces reliable predictions for complexes between synthetic receptors and ligands. Our approach to tackle the flexibility of two molecules simultaneously successfully predicted all examples of our test set with respect to an RMSD of below 2 Å. In a previous study we already showed the general transferability of the FlexX concepts to the docking of synthetic receptor−ligand systems. Here, additionally we focused on the more efficient handling of systems in which both molecules exhibit a high degree of flexibility. One limitation of our previous approach[22] was observed for docking of complex **1**, where the docking times for forward and inverse docking exceeded several hours of CPU time. The second limitation was observed for complex **3** where no forward docking was possible at all due to the large conformational space of the receptor. The inverse docking strategy, however, worked but was comparably slow. Our new method predicted these structures significantly faster, and, at the same time, near-native complexes were obtained. However, considering the complexes **2**, **4**, **8**, and **9** our new method was noticeably slower than the recently presented method,[22] although the computation time was still in the range of seconds to minutes.

The following parameters significantly influence the computation time used for a complex: (i) the number of possible interactions between the two molecules, (ii) the flexibility of the molecules, and (iii) the symmetry of the system. The more interactions are possible the larger the docking graph gets. Flexible molecules can cause unspecifically wide distance ranges between centers of directional short-range interactions, and thus many interaction pairs are compatible. This results in many edges in the docking graph. Consequently, the search for cliques in the docking graph slows down. However, it has to be noted that in none of the examples presented the precomputation phase lasted longer than a few seconds. Thus, the precomputation phase is not the rate-limiting step of the algorithm. However, as the number of interactions and the flexibility of the molecules rises further, this might be a limiting factor for the method. So far our method does not consider symmetry information of the molecules. Due to this fact many symmetrical cliques are generated if the molecules of the complex exhibit symmetry. Currently, all of them are processed although no additional information is gained, and, thus, the computation time for one complex rises with dependence to the inherent symmetry of the complex. Consider complex **3** where each of the molecules has 3-fold symmetry. By manually removing symmetric cliques we can show that the computation can be accelerated by a factor of 10 without losing any information. The automatic detection of symmetries in synthetic receptor−ligand complexes[10] thus promises to speed up the computation of such complexes. In our tests, the number of cliques for the complexes has varied from 2 for complex **4** to 360 for complex **3**.

In highly flexible complexes, additionally, many degenerate cliques are generated that do not represent valid complex structures. There are several reasons why this occurs. At first, in the precomputation phase only distances between the centers of the particular directional short-range interaction are considered. Clashes of the remaining atoms are not taken into account at this stage. Furthermore the estimated distance ranges are treated as being independent from each other, although this assumption is not valid in every case (Figure 20). We see some potential in incorporating some additional geometric criteria that help to discard degenerate cliques at an early stage.

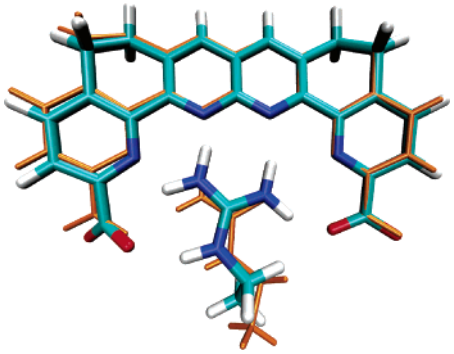


Figure 19. Docking results of complex **10** (rank 1, atom coloring) superimposed on the X-ray structure (orange).

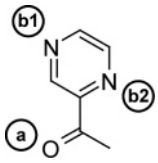


Figure 20. Dependence of distance ranges. In this example, the maximal distance between atom a and atom b2 can only be realized if the distance between atom a and atom b1 is minimal. This shows the dependency of distance ranges which is disregarded in the docking graph generation step. Here, the distance ranges are treated as being independent from each other.

In our approach both molecules—synthetic receptor and ligand—are incrementally constructed from fragments. As stated above, in the beginning no conformation is known, and thus the guiding role of one molecule for docking the other one is missing. In the precomputation phase our method needs the presence of directional short-range interactions such as hydrogen bonds or salt bridges which exhibit spatially much more constrained interaction geometries as lipophilic interactions. The consideration of the latter in this step would lead to infeasibly large docking graphs as generally many lipophilic interaction combinations are possible in common synthetic receptor−ligand complexes. Furthermore, their geometrically ambiguous nature would not allow for applying strict distance filters. However, it is important to note that lipophilic interactions are assessed during the complex construction as the scoring function considers them. The molecules in synthetic receptor−ligand systems that are solely based on lipophilic interactions are generally less flexible due to the weak nature of the lipophilic interactions that would not outweigh entropic costs of binding. Thus, at least one molecule is commonly rigid, and a slightly modified version of our recently proposed docking strategy can be applied.[39]

So far, supramolecular complexes between ligands and macrocyclic synthetic receptors cannot be generated using our incremental build-up principle. The efficient conformational sampling of large rings is still a challenging research topic. Work along this line is in progress in a project within our group.[40]

The effect of the applied scoring function on the prediction of synthetic receptor−ligand complexes has been discussed by Kämper et al.[22] The test set used in our recent paper consisted of complexes that have been crystallized from aprotic solvents. Here, a complex has been integrated into the test set that was crystallized from aqueous solution. Although the structural influence of water, which is present in the crystal structure, is not tackled explicitly, near-native structures were generated. So far solvation is considered only implicitly in the scoring function which has been parametrized on experimentally derived protein−ligand complexes that have been crystallized from water. A scoring function for different solvents and with a more specific description of the intermolecular interaction terms is currently under development.[41]

Besides the forward and inverse docking strategy[22] our new docking algorithm is the second approach that transfers the concepts of the Flex* program suite to the synthetic receptor−ligand system and predicts near-native results for all test cases. This underlines the reliability of the whole concept. Which of the two methods is applied best for a given synthetic receptor−ligand complex depends on the flexibility of the molecules. In the case that only one molecule has to be treated as flexible the approach introduced by Kämper et al.[22] is the method of choice. If both molecules are flexible, then our new algorithm should be applied.

CONCLUSIONS

A fast and fully automated method for predicting the structure of binary complexes between synthetic receptors and their ligands has been developed. Our new approach tackles the flexibility of both molecules simultaneously. For all but one of the selected test cases, we observed an excellent agreement between predicted and experimental structures.

The high efficiency of our adaptive two-sided incremental build-up approach originates from the fact that the conformational space of both molecules is searched with respect to the counter molecule, and thus the search space is significantly reduced. In comparison to our recently presented work[22] a significant acceleration has been achieved for complexes that consist of highly flexible molecules, while the quality of the results has been maintained. Together with some future modifications our tool will open new scenarios for the computer-assisted design of novel synthetic receptor−ligand complexes. In fact, even in its present state the speed of the approach allows for large-scale computation such as virtual screenings for an optimal ligand for a given synthetic receptor and vice versa.

ACKNOWLEDGMENT

We thank C. M. Marian and J. Apostolakis for many helpful suggestions on improving the manuscript, L. Kunert and C. Hartmann for helpful discussions regarding this work, and J. Büch for technical support. Financial support from the Deutsche Forschungsgemeinschaft through project KA 1804/1-1 is gratefully acknowledged.

Supporting Information Available: The algorithmical details of the distance range estimation, the fragment order computation, and the optimization of matches. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Bell, T. W.; Hou, Z.; Luo, Y.; Drew, M. G.; Chapoteau, E.; Czech, B. P.; Kumar, A., Detection of Creatinine by a Designed Receptor. *Science* **1995**, *269*, 671−674.

(2) Maue, M.; Schrader, T. A Color Sensor for Catecholamines. *Angew. Chem., Int. Ed.* **2005**, *44*, 2265−2270.

(3) Davis, M. E.; Brewster, M. E. Cyclodextrin-based Pharmaceutics: Past, Present and Future. *Nat. Rev. Drug Discovery* **2004**, *3*, 1023−1035.
(4) Stuerzebecher, C.-S.; Witt, W.; Raduechel, B.; Skuballa, W.; Vorbrueggen, H. Prostacyclins, their Analogs or Prostaglandins and Thromboxane Antagonists for Treatment of Thrombotic and Thromboembolic Syndromes. U.S. Pat. 5,523,321, 1996.
(5) Stanton, J.; Vincent, P. Tumor Necrosis Factor Receptor 2. U.S. Pat. 6,673,908, 2001.
(6) *Functional Synthetic Receptors*; Schrader, T., Hamilton, A. D., Eds.; Wiley-VCH: Weinheim, 2005.
(7) Kubinyi, H. Chance Favors the Prepared Mind − From Serendipity to Rational Drug Design. *J. Recept. Signal Transduction Res.* **1999**, *19*, 15−39.
(8) Lyne, P. D., Structure-based Virtual Screening: an Overview. *Drug Discovery Today* **2002**, *7*, 1047−1055.
(9) Brooijmans, N.; Kuntz, I. D., Molecular Recognition and Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335−373.
(10) Chen, H.; Lyne, P. D.; Giordanetto, F.; Lovell, T.; Li, J. On Evaluating Molecular-docking Methods for Pose Prediction and Enrichment Factors. *J. Chem. Inf. Model.* **2006**, *46*, 401−415.
(11) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.
(12) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.
(13) Shoichet, B.; Bodian, D.; Kuntz, I. Molecular Docking Using Shape Descriptors. *J. Comput. Chem.* **1992**, *13*, 380−397.
(14) Leach, A. R.; Lemon, A. P. Exploring the Conformational Space of Protein Side Chains Using Dead-end Elimination and the A* Algorithm. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 227−239.
(15) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(16) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient Molecular Docking Considering Protein Structure Variations. *J. Mol. Biol.* **2001**, *308*, 377−395.
(17) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated Docking to Multiple Target Structures: Incorporation of Protein Mobility and Structural Water Heterogeneity in AutoDock. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 34−40.
(18) Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. Testing a Flexible-receptor Docking Algorithm in a Model Binding Site. *J. Mol. Biol.* **2004**, *337*, 1161−1182.
(19) Kämper, A.; Rognan, D.; Lengauer, T. Lead Identification by Virtual Screening. In *Bioinformatics − from Genome to Therapy*; Lengauer, T., Ed.; VCH−Wiley: Weinheim, to appear.
(20) Klebe, G. Virtual screening: Scope and Limitations. In *Virtual Screening in Drug Discovery*; Alvarez, J., Shoichet, B., Eds.; Taylor & Francis: Boca Raton, 2005.
(21) Hamilton, A. D.; van Engen, D. Induced Fit in Synthetic Receptors: Nucleotide Base Recognition by a "Molecular Hinge". *J. Am. Chem. Soc.* **1987**, *109*, 5035−5036.
(22) Kämper, A.; Apostolakis, J.; Rarey, M.; Marian, C. M.; Lengauer, T. Fully Automated Flexible Docking of Ligands into Flexible Synthetic Receptors Using Forward and Inverse Docking Strategies. *J. Chem. Inf. Model.* **2006**, *46*, 903−911.
(23) Garcia-Tellado, F.; Goswami, S.; Chang, S.-K.; Geib, S. J.; Hamilton, A. D. Molecular Recognition: A Remarkably Simple Receptor for the Selective Complexation of Dicarboxylic Acids. *J. Am. Chem. Soc.* **1990**, *112*, 7393−7394.
(24) Ballester, P.; Cápo, M.; Costa, A.; Deyà, P. M.; Gomila, R.; Decken, A.; Deslongchamps, G. Selective Binding of cis-1,3,5-Cyclohexane Tricarboxylic Acid vs its Epimeric trans Isomer by a Tripodal Amidopyridine Receptor; Crystal Structures of the 1:1 Complexes. *Org. Lett.* **2001**, *3*, 267−270.
(25) Klebe, G.; Mietzner, T. A Fast and Efficient Method to Generate Biologically Relevant Conformations. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 583−606.
(26) Allen, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr., Sect. B.: Struct. Sci.* **2002**, *58*, 380−388.
(27) Sadowski, J.; Gasteiger, J. From Atoms and Bonds to Three-dimensional Atomic Coordinates: Automatic Model Builders. *Chem. Rev.* **1993**, *93*, 2567−2581.
(28) Böhm, H.-J., LUDI: Rule-based Automatic Design of new Substituents for Enzyme−inhibitor Leads. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593−606.
(29) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E., A Geometric Approach to Macromolecule-ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.
(30) Bron, C.; Kerbosch, J., Algorithm 457: Finding all Cliques of an Undirected Graph. *Commun. ACM* **1973**, *16*, 575−577.
(31) Böhm, H.-J., The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein Ligand Complex of known Three-dimensional Structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.
(32) Bell, T. W.; Khasanov, A. B.; Drew, M. G., Role of Pyridine Hydrogen-bonding Sites in Recognition of Basic Amino Acid Side Chains. *J. Am. Chem. Soc.* **2002**, *124*, 14092−14103.
(33) Chin, J.; Walsdorff, C.; Stranix, B.; Oh, J.; Chung, H. J.; Park, S.-M.; Kim, K. A Rational Approach to Selective Recognition of $NH4^+$ over K^+. *Angew. Chem., Int. Ed.* **1999**, *38*, 2756−2759.
(34) Pascal, R. A.; Ho, D. M. Molecular Structures of Host−guest Complexes with Rebek's Diacid. *Tetrahedron* **1994**, *50*, 8559−8568.
(35) Berl, V.; Huc, I.; Lehn, J.-M.; DeCian, A.; Fischer, J. Induced Fit Selection of a Barbiturate Receptor from a Dynamic Structural and Conformational/Configurational Library. *Eur. J. Org. Chem.* **1999**, *11*, 3089−3094.
(36) Kneeland, D. M.; Ariga, K.; Lunch, V. M.; Huang, C.-Y.; Anslyn, E. V., Bis(alkylguanidinium) Receptors for Phosphodiesters: Effect of Counterions, Solvent Mixtures, and Cavity Flexibility on Complexation. *J. Am. Chem. Soc.* **1993**, *115*, 10042−10055.
(37) Grossel, M. C.; Merckel, D. A. S.; Hutchings, M. G. The Effect of Preorganisation on the Solid State Behaviour of Simple 'Aromatic-Cored' Bis(guanidinium) Sulfates. *CrystEngComm* **2003**, *5*, 77−81.
(38) Waldvogel, S. R.; Fröhlich, R.; Schalley, C. A. First Artificial Receptor for Caffeine: A New Concept for the Complexation of Alkylated Oxopurines. *Angew. Chem., Int. Ed.* **2000**, *39*, 2472−2475.
(39) Kämper, A.; Lengauer, T., work in progress.
(40) Kämper, A.; Kaspar, M.; Lengauer, T. Efficient Conformational Analysis of Synthetic Receptors with Macrocyclic and Fused Ring Systems. In *Synthetic Receptors 2005, Proceedings of the Second World Congress on Synthetic Receptors*; Elsevier Publishers: Oxford, 2005; p O4.
(41) Raub, S.; Marian, C., work in progress.

CI060072V