

NMR-Guided Molecular Docking of a Protein–Peptide Complex Based on Ant Colony Optimization

Oliver Korb,^[a, b] Heiko M. Möller,^[a] and Thomas E. Exner^{*,[a]}

Standard docking approaches used for the prediction of protein–ligand complexes in the drug development process have problems identifying the correct binding mode of large flexible ligands. Herein we show how additional experimental data from NMR experiments can be used to predict the binding mode of a mucin 1 (MUC-1) pentapeptide recognized by the breast-cancer-selective monoclonal antibody SM3. Distance constraints derived from trNOE and saturation transfer difference NMR experiments are combined with the docking approach PLANTS. The resulting complex structures show excel-

lent agreement with the NMR data and with a published X-ray crystal structure. The method was then further tested on two complexes in order to demonstrate its more general applicability: T-antigen disaccharide bound to *Maclura pomifera* agglutinin, and the inhibitor SBI279 bound to S100B protein. Our new approach has the advantages of being fully automatic, rapid, and unbiased; moreover, it is based on relatively easily obtainable experimental data and can greatly increase the reliability of the generated structures.

Introduction

Theoretical methods such as molecular docking have become important tools for predicting protein–ligand complexes, especially in the early stages of the drug development process. In evaluation studies on large test sets, up to 80% of the complexes with small-to-medium-sized ligands can be predicted with an accuracy $< 2 \text{ \AA}$ heavy-atom RMSD to the crystal structure using state-of-the-art docking tools.^[1–11] However, with increasing ligand size and flexibility, the success rates drop dramatically because of insufficient sampling and scoring function problems in particular. For that reason, peptides and carbohydrates cause many problems due to their flexibility and the large number of potentially favorable interactions with protein functional groups. For reasons of efficiency, docking tools must use very simple scoring functions to evaluate the energetic contributions, in most cases approximated by the sum of pairwise atom–atom interactions. For each type of interacting atom, an average contribution is added. Thus, similar groups, like the backbone of each amino acid of a given peptide or the hydroxy groups of carbohydrates, contribute almost the same score regardless of the local surroundings. In reality, these can have a substantial influence on the energetics. For example, hydrogen bond energies, which are normally $\sim 5 \text{ kcal mol}^{-1}$, may be $> 10 \text{ kcal mol}^{-1}$ in “strong” or “low-barrier” hydrogen bonds.^[12] These and other approximations of the scoring functions (no explicit solvent, inexact treatment of electrostatic and steric interactions, and/or neglect of entropic contributions) are more severe for larger ligands and preclude the identification of the correct relative placement and conformation of the ligand in the binding sites (correct pose) out of the immense number of other energetically favorable—at least with respect to the scoring function used—but incorrect structures (decoys). The development of more sophisticated scoring functions, which at the same time can be evaluated effectively

so that they can be used in docking applications, is not straightforward. Even with the ongoing worldwide research in this area, a major breakthrough is not expected in the foreseeable future. To circumvent this dilemma, the incorporation of experimental data could guide the search algorithm toward structures in accordance with these data and therefore toward more reasonable regions of the search space. In this regard, NMR spectroscopy is especially well suited to offer a number of experiments. Combinations of docking with NMR measurements on free and complexed protein^[13,14] or intermolecular NMR constraints^[15–19] have already been proposed. The HADDOCK^[15,16] program in particular has often been applied successfully. The disadvantage of these previous methods is that they require specific labeling of the protein; this is time consuming, expensive, and limited to small protein domains. In contrast, NMR information for the ligand in its bound conformation taken from STD,^[20,21] trNOE, and INPHARMA^[22,23] experiments, for example, can be obtained relatively easily for a large set of weakly bound protein–ligand systems that are in fast-to-intermediate exchange between the free and bound

[a] Dr. O. Korb, Dr. H. M. Möller, Dr. T. E. Exner
Department of Chemistry and Zukunftskolleg
University of Konstanz, 78457 Konstanz (Germany)
Fax: (+ 49) 7531-88-3587
E-mail: thomas.exner@uni-konstanz.de

[b] Dr. O. Korb
Present address: The Cambridge Crystallographic Data Centre
Cambridge (UK)

states.^[24,25] Therefore, such ligand-based information (STD and trNOE) is exclusively used in this work.

As a challenging test case, we show that the complex of a mucin 1 (MUC-1) peptide with the monoclonal antibody SM3 can be predicted correctly with additional constraints derived from NMR data. The MUC-1 glycoprotein is overexpressed and aberrantly glycosylated in breast cancer cells.^[26-28] Due to the highly truncated oligosaccharide side chains, the peptide core of cancerous MUC-1 is more exposed, so that the development of cancer-selective antibodies is possible.^[28] One of these, the SM3 monoclonal antibody, is highly specific, and the corresponding epitope was identified as Pro1-Asp2-Thr3-Arg4-Pro5 by using a library of octapeptides in a pepsin assay with ELISA detection.^[29] The binding pose of this epitope in a tridecapeptide was determined by X-ray crystallography and is available from the Protein Data Bank (PDB)^[30] as entry 1SM3.^[31] Möller et al. studied the complex of SM3 with the five-residue-long epitope in solution by STD NMR and trNOE-based conformational analysis.^[32] Saturation transfer difference (STD) NMR exploits magnetization transfer from the protein to the ligand in order to differentiate between strongly bound and water-exposed parts of the ligand.^[20,21] Protons of Pro1 and Asp2 show the highest intensities in the STD spectra, which can be explained by more and tighter contacts with the antibody surface. Medium and low intensities are observed for Thr3 and Arg4/Pro5, respectively. Additionally, 13 intra-ligand distance constraints were obtained from trNOESY spectra. Based on these data, the bound conformation and the binding mode were determined in a time-consuming, largely manual way by combining constrained molecular dynamics of the ligand, docking guided by the X-ray structure, and finally molecular dynamics of the complex. In contrast, our new approach is fully automated, fast, and unbiased. To demonstrate the general applicability of the docking protocol developed for the SM3-antigen complex, it was also applied to determine the complex of *Maclura pomifera* agglutinin with the T-antigen disaccharide (β -D-Gal(1 \rightarrow 3)- α -D-GalNAc-(1 \rightarrow O)-Me)^[33] and calcium-bound S100B protein in complex with inhibitor SBI279.^[34] Weimar et al. studied the first complex with trNOE and STD NMR so that the same information as for the SM3 complex is available.^[33] For S100B, only STD spectra were measured.^[34]

Results and Discussion

SM3 antibody-peptide complex

The starting point of this study is our docking program PLANTS.^[35-37] Relative to other state-of-the-art docking approaches, PLANTS has demonstrated excellent performance on a large number of protein-ligand complexes, with success rates of up to 80% for sets of ligands with 10 rotatable bonds or fewer, using the newly developed ChemPLP scoring function.^[37] Taking these published results into account, we expected that the quality of the results obtained on the system studied here would be at least as good as those obtainable by other methods. As input, the 3D structure of the active site of the protein and reasonable values for bond lengths, bond

angles, and ring conformations for the ligand must be provided, as these are not changed during the optimization. The antibody structure was extracted from entry 1SM3^[31] of the Protein Data Bank.^[30] In this, the crystallographic atomic coordinates of SM3 in complex with a 13 amino acid epitope are stored (Thr18'-Ser19'-Ala20'-Pro1-Asp2-Thr3-Arg4-Pro5-Ala6-Pro7-Gly8-Ser9-Thr10; herein we follow the convention reported by Gendler et al.,^[38] who defined the peptide start at PDTRP; residues resulting from elongation at the N terminus are denoted by an apostrophe). Hydrogen atoms were added automatically to the antibody by the SPORES setup procedure.^[39] The same pentapeptide as used by Möller et al.^[32] served as the input ligand structure. Because the torsion angles are fully randomized prior to docking, only bond lengths, bond angles, and ring conformations are conserved.

The docking results using the standard docking procedure are shown in Figure 1. Figure 1a shows an overlay of the best-ranked structure of 20 independent docking runs. Two clusters of conformations can be distinguished, out of which the green cluster probably corresponds to the global optimum, as the structure with the best score across all docking runs can be found in this cluster. The appearance of the second cluster (with less favorable scores than in the first cluster) reveals that there are some remaining sampling problems in a single docking run. But these are not severe, as in 16 of the 20 docking runs, structures close to the "global" optimum belonging to cluster 1 were found, so that sampling problems can be excluded here by the additional sampling of multiple docking runs. Comparing cluster 1 with the crystal structure (red), the backbone of the peptide appears to be oriented more or less correctly. However, upon closer inspection of the pose, it is evident that all side chains except Pro1 are located completely differently in the crystal structure and the docking poses. This is clearly noticeable for the best-ranked structure of the first run, representatively shown in Figure 1b.

As described in the Experimental Section below in greater detail, PLANTS is designed to generate a diverse set of 10 high-ranked complex structures (including the best-ranked structure) in each run. Taking all these structures into account in follow-up studies, performed manually by visual inspection or with a more sophisticated computational approach, has been shown to increase the chances of finding the correct structure.^[35-37] Notably, this is not the case for the system studied here. All 10 diverse poses are very different from the crystal structure (shown exemplarily with poses 2 and 4 in Figure 1c and 1d, respectively), although the major binding motif is identified (Pro1-Asp2-Thr3 in docking pose 4, Figure 1d). One way to explain the differences is to propose highly diverging structures in the crystal and in solution. In this case, however, the predicted structures should fulfill the constraints derived by the NMR experiments even when these are not imposed during the docking calculation. In checking the compatibility of the best-scored pose with the NMR constraints as determined by Möller et al.,^[32] we found that ten trNOE and eight STD constraints are strongly violated by up to 4 and 1.5 Å, respectively. Thus, the docking results are neither in accordance with the X-ray structure nor with the NMR data. PLANTS, using

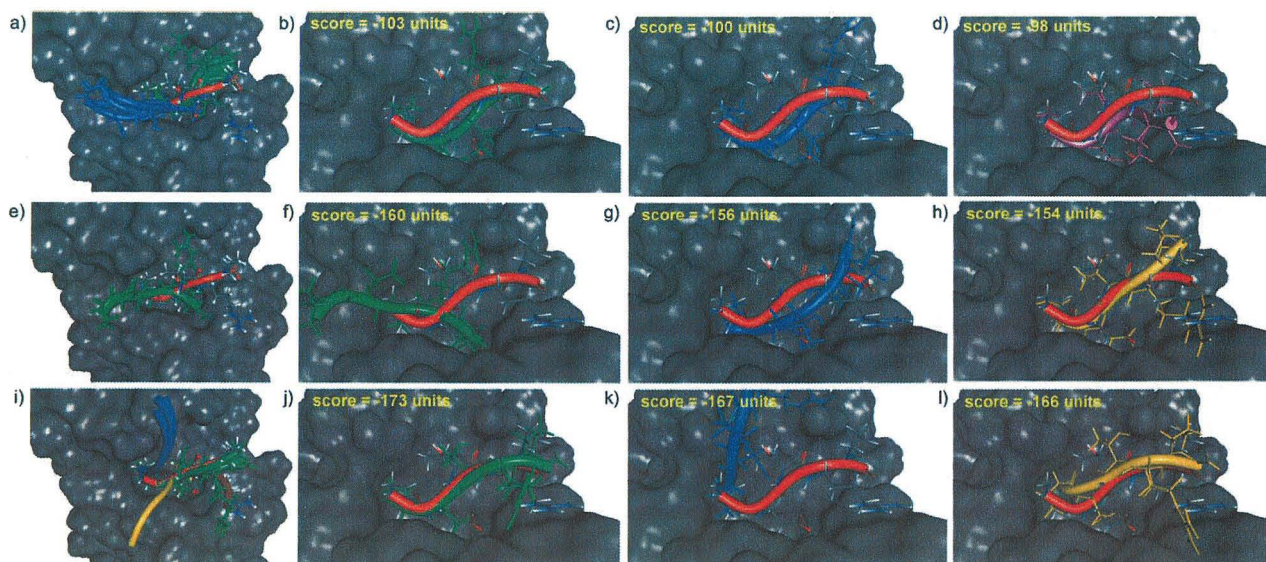


Figure 1. Docking results for the complex of mucin 1 (MUC-1) peptide with the monoclonal antibody SM3. a)–d) ChemPLP scoring function only; a) overlay of the best-ranked poses of 20 independent docking runs and the ligand poses on rank: b) 1 (green), c) 2 (blue), and d) 4 (magenta) of run 1, according to the ChemPLP scoring function. One representative structure is shown in ball-and-stick rendering for the two distinguishable clusters in green (16 structures) and blue (four structures). For comparison, the five corresponding amino acids of the crystal structure are also shown in atom-type coloring with red ribbon representation. Pose 4 is the most similar to the crystal structure of all generated structures of the diverse set and was therefore also chosen for visualization. e)–h) ChemPLP/trNOE scoring function; e) overlay of the best-ranked poses of 20 independent docking runs and the ligand poses on rank: f) 1 (green), g) 2 (blue), and h) 3 (yellow) of run 1, according to the ChemPLP scoring function including trNOE distance constraints. All best-ranked structures belong to only one cluster, which is shown in green. For comparison, the five corresponding amino acids of the crystal structure are also shown in atom-type coloring with red ribbon representation. i)–l) ChemPLP/trNOE/STD scoring function; i) overlay of the best-ranked poses of 20 independent docking runs and the ligand poses on rank: j) 1 (green), k) 2 (blue), and l) 3 (yellow) of run 1, according to the ChemPLP scoring function as well as trNOE and STD distance constraints. Besides the one large cluster that probably represents the global optimum shown in green, a smaller cluster (three structures in blue) and two single structures (yellow and brown) can be seen. These deviating poses are caused by sampling problems (less favorable scores relative to the large cluster) and could probably be avoided by optimization of the search parameters. For comparison, the five corresponding amino acids of the crystal structure are also shown in atom-type coloring with red ribbon representation.

the ChemPLP scoring function on its own, is not able to correctly predict the complex of the pentapeptide bound with SM3. Because sampling problems have already been excluded, this must be fully attributed to the scoring function. As described above in the Introduction, it is known that large highly flexible ligands, to which the pentapeptide studied herein with its 22 degrees of freedom clearly belongs, are very difficult to predict with conventional scoring functions.

As also described above, there is no clear-cut way to improve the scoring functions without losing the efficiency required for docking approaches. In a normal docking run, the scoring function must be evaluated $\sim 10^5$ – 10^6 times to reach exhaustive sampling of the search space. We therefore took a different approach and included the available information as determined by NMR in solution. In the study reported by Möller et al.,^[32] upper and lower limits of 13 trNOE distance constraints were determined. These data can be used to evaluate the poses generated by PLANTS (as done above for the best-ranked pose) and to identify the single pose that best obeys the experimental constraints. Because none of the ten diverse poses given as output of PLANTS is close to the correct structure, such an approach is guaranteed to fail for this complex. Thus, the experimental data have to be an integral part of the docking algorithm to guide it to the correct region of the search space. To do so, one scoring function term per con-

straint according to Equation (1) was added to the ChemPLP scoring function.

$$f_{\text{NOE}}(r, w_{\text{NOE}}, r_{\text{min}}, r_{\text{max}}) = \begin{cases} w_{\text{NOE}} + w_{\text{penalty,NOE}} \cdot (r_{\text{min}}^2 - r^2) & \text{if } r < r_{\text{min}} \\ w_{\text{NOE}} & \text{if } r_{\text{min}} \leq r \leq r_{\text{max}} \\ w_{\text{NOE}} + w_{\text{penalty,NOE}} \cdot (r^2 - r_{\text{max}}^2) & \text{if } r > r_{\text{max}} \end{cases} \quad (1)$$

The weight w_{NOE} was empirically set to -15 units, and constraint violations were heavily penalized by multiplying the deviation from the upper/lower limit by $w_{\text{penalty,NOE}} = 10$. w_{NOE} is used to ensure that structures with small violations still get favorable negative scores, which is a prerequisite for the pheromone update mechanism employed in PLANTS. For details of the ant colony optimization algorithm used in PLANTS, see references [35–37]. More important is $w_{\text{penalty,NOE}}$, as it determines the relative importance of the experimental data relative to the other scoring function terms. The chosen value resulted in small violations with reasonable scores for ChemPLP. The best-ranked poses obtained in 20 independent docking runs with PLANTS, ChemPLP, and trNOE constraints are shown in Figure 1e. It is clear that only one cluster with a new binding motif occurs, which is completely different from the crystal structure. Thus, it appears the NOE constraints would have

made things worse. Here, however, the benefit of the feature of PLANTS to generate a diverse set of highly ranked structures becomes evident. Especially the third-ranked pose of this set shows a striking similarity to the X-ray structure (Figure 1 h). The agreement of the backbone and side chain orientations of residues 1–3 is remarkable. Arg4 and Pro5 are slightly displaced, but for these, only very low STD effects are observed, so that these are exposed to the solvent and their positions are less well defined. Furthermore, Arg4 and Pro5 are affected by crystal packing: they come into contact with an adjacent antibody in the crystal structure, which could also explain the different conformations of these amino acids in solution (see figure S1 in the Supporting Information).

The docking runs including NMR constraints demonstrate that the use of additional experimental data can help guide the docking algorithm toward the correct docking pose. However, the combination of ChemPLP and trNOE constraints still has problems to correctly rank the different structures. As further experimental support, we tried to also incorporate the information from STD NMR. In these experiments, protons of the ligand that come into close contact with the surface of the protein show the highest degree of saturation and thus the largest signal in the STD NMR spectrum, whereas protons located far away from the receptor are saturated to a much lesser extent, and therefore show weak signals in the spectrum. To obtain distance constraints from this information, all protons observable in the spectrum were classified into strong, medium, and weak, as summarized in Table 1. Distance ranges,

Table 1. Protons showing STD effects and their classification. ^[a]	
Classification (STD intensities)	Protons
Strong	Pro1-ββ', Pro1-δδ'
Medium	Asp2-ββ', Thr3-α
Weak	Thr3-ββ', Thr3-γ, Arg4-ββ', Arg4-γγ', Arg4-δδ', Pro5-ββ', Pro5-γγ', Pro5-δδ'

[a] STD effects of individual protons were quantified by analyzing original spectra,^[32] strongly overlapping signals were excluded.

where the constraints are considered to be fulfilled, are set to 1–3 Å, 1–5 Å, and 4–∞ Å for ligand protons showing a strong, medium, and weak STD effect, respectively. To calculate the distance, the closest protein atom to the ligand proton is taken. The STD scoring function term used [Eq. (2)] is analogous to the NOE term, but the violations are less penalized ($w_{penalty,STD}=5$) because the distance ranges are less well defined.

$$f_{STD}(r, w_{STD}, r_{min}, r_{max}) = \begin{cases} w_{STD} + w_{penalty,STD} \cdot (r_{min}^2 - r^2) & \text{if } r < r_{min} \\ w_{STD} & \text{if } r_{min} \leq r \leq r_{max} \\ w_{STD} + w_{penalty,STD} \cdot (r^2 - r_{max}^2) & \text{if } r > r_{max} \end{cases} \quad (2)$$

w_{STD} was set to –3 units. The total score is then defined as the sum of the three contributions [Eq. (3)]:

$$f_{total} = f_{ChemPLP} + \sum f_{NOE} + \sum f_{STD} \quad (3)$$

When this new scoring function is used, the correct pose is identified as the best-ranked pose in 15 of 20 docking runs (Figure 1 i and 1 j). The best-ranked structures of the other five runs originate from sampling problems, as their scores are much less favorable (–164 to –168 units, compared with –173 units for the large cluster). From the diverse set, pose 3 also shows a very similar binding pattern (Figure 1 l). In looking at the scores and the constraint violations, it is evident that the poses are a well-balanced compromise between optimizing the score with respect to ChemPLP and minimizing the constraint violations (Supporting Information). On the one hand, the ChemPLP score increases from –103 units to –36 units in the corresponding best-ranked pose of the docking runs without any and with trNOE and STD constraints, respectively. On the other hand, all STD constraints are fulfilled and only five out of 13 trNOE constraints are slightly (≤ 0.4 Å) violated. This is even more remarkable when considering the incompatibility of the structures of the ChemPLP-only docking run with the NMR spectroscopic information (see above).

Maclura pomifera agglutinin in complex with T-antigen

The core 1 disaccharide (β-D-Gal(1→3)-α-D-GalNAc), also called Thomson–Friedenreich or T-antigen, is one of the best-characterized carbohydrate tumor markers. Besides others, *M. pomifera* agglutinin (MPA) shows high specificity toward the T-antigen structure and has potential as a diagnostic probe. Weimar et al.^[33] analyzed the conformation of the complex between MPA and the T-antigen disaccharide by NMR spectroscopy. From their study, we extracted three trNOE and 14 STD distance constraints. The classification of the STD signals into weak, medium, and strong as well as the upper and lower limits for the NOE constraints are given in the Supporting Information. As starting point, the X-ray structure of (β-D-Gal(1→3)-α-D-GalNAc-(1→O)-H; PDB entry 1JOT) was used.^[40] Hydrogen atoms were added with the SPORES program,^[39] the ligand was extracted, and the hydrogen was manually modified into a methyl group to end up with the β-D-Gal(1→3)-α-D-GalNAc-(1→O)-Me disaccharide of the NMR study. PLANTS was used with the same ChemPLP/trNOE/STD scoring function described above. Only the lower limits of the STD constraints were adapted slightly. Tyr122 was treated as flexible in the docking, as it is stated in the original publication that this flexibility is needed to accommodate the methyl group.^[33] Sampling is no problem here, and the same structure is obtained in all 20 runs, which is shown in Figure 2 a. A perfect match of GalNAc is obtained in the top-ranked pose. The rotation of Tyr122 can also be observed as described in the original publication.^[33] Only Gal is slightly displaced relative to the crystal structure. This is not unexpected since the STD data clearly show that this portion of the disaccharide is more solvent exposed and less important for binding. Relative to the SM3 antibody–peptide complex described above, the experimental data are less important for determining the correct pose of the

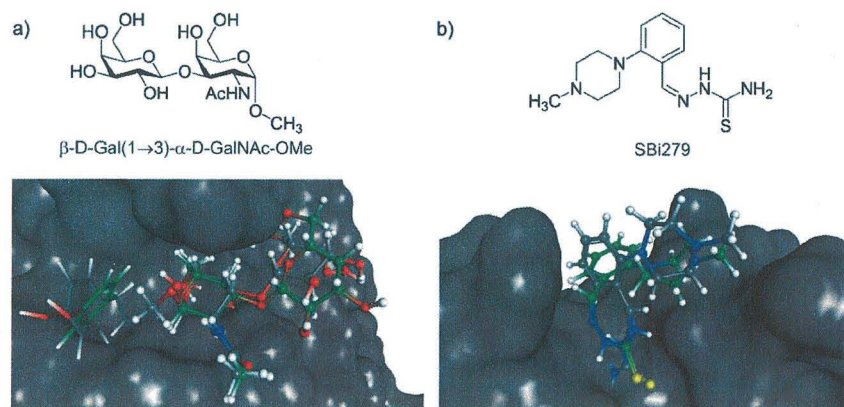


Figure 2. Comparison of the X-ray structure (carbon atoms in green) and the NMR docking poses (carbons in gray): a) *M. pomifera* agglutinin in complex with T-antigen disaccharide (flexible side chain Tyr122 in capped-sticks representation); b) S100B protein in complex with SBi279.

MPA–disaccharide complex, as ChemPLP on its own already results in a very similar pose (see figure S2 in the Supporting Information). However, the additional experimental information gives more confidence in the docking pose regarding the strong binding of the GalNAc moiety and the weakly bound Gal residue.

S100B protein–SBi279 complex

Charpentier et al.^[34] described a number of small molecules complexed to unique sites of the calcium-bound S100B protein by NMR and X-ray crystallography. S100B binds directly to the p53 tumor suppressor protein and in this way inhibits its function; thus, the inhibition of S100B is of great interest in cancer therapy. From the small-molecule inhibitors described,^[34] we picked SBi279 as an example. The complex was prepared by starting from PDB entry 3GK2,^[34] and the docking was performed in the same way as for the MPA–T-antigen complex. The results are shown in Figure 2b, which were obtained using STD constraints only, as no trNOE data are available. The classification of the STD signals is again given in the Supporting Information. The location of the sulfur atom is reproduced perfectly, and the two ring systems are in locations very similar to those in the X-ray structure. However, in the proposed docking pose, the rings are more solvent exposed, which is in better agreement with the STD intensities, as no significant STD signals were obtained for the piperazine ring, and also the phenyl ring shows only relatively weak signals relative to the methyl group and the olefinic proton. Thus, the differences could indeed represent the differences between the solution and the crystal structures. Additionally, we note that the use of ChemPLP alone results in complete failure, with a completely different docking pose that disagrees with the NMR and X-ray experimental data (cf. figure S3 in the Supporting Information).

Conclusions

Herein we demonstrate that the combination of experimental data obtained from NMR experiments with docking tools can

greatly increase the reliability of predicted docking poses. For the complex of the SM3 antibody with its epitope, the ChemPLP scoring function complemented with trNOE and STD distance constraints is able to correctly predict the complex structure as the best-ranked docking pose. Importantly, all three contributions—ChemPLP, trNOE, and particularly STD—are required to obtain a successful prediction. Moreover, for the other two examples used to show the transferability of this new approach, good agreement with the complex structures ob-

tained by X-ray crystallography was also observed. For all three complexes, especially for the solvent-exposed regions, much better agreement was observed than for the docking guided by the ChemPLP scoring function only. The advantage of this approach is that the experimental data used are relatively easy to obtain (relative to a full structure determination of the complex by NMR or X-ray crystallography), as STD NMR and trNOE experiments do not rely on isotope labeling and have no size limitation for the receptor protein. The only experimental limitation is for the dissociation rate constant k_{off} to be large enough so that the saturated ligand can accumulate in solution, which is the case for many ligands with affinities in the micromolar to high nanomolar range. In combination with the docking approach, the need for human intervention can be minimized, which should speed up the determination of complex structures by NMR spectroscopy while minimizing human bias.

Additional work is underway to verify and optimize the weighting factors for the NOE and STD scoring function terms, which were heuristically defined in this work, so that the optimized and fully automatic version of our new approach can become a valuable tool for NMR spectroscopists. The next publicly available release of the PLANTS software includes this new feature, which is available at <http://www.tcd.uni-konstanz.de>. Additionally, the ligand poses of our NMR-guided PLANTS docking might serve as ideal starting points for a more sophisticated refinement. The CORCEMA^[41,42] and CORCEMA-ST^[43,44] approach based on the full calculation of the relaxation and conformational exchange matrix could be used for trNOE and STD data, respectively. However, owing to the large amount of computer time needed for these calculations, these would be applied in a post-processing step for rescoreing the generated poses.

Experimental Section

As mentioned above, the docking studies were performed with our docking program PLANTS.^[35–37] PLANTS uses Ant Colony Optimization (ACO),^[45] a state-of-the-art global optimization algorithm

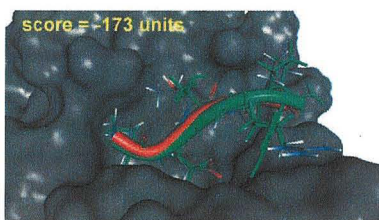
to find minima of a scoring function representing favorable complex structures. ACO belongs to the swarm intelligence approaches, and is inspired by the foraging behavior of real ants, which are able to find shortest paths between their nest and a food source via indirect communication using pheromone trails.^[46] In PLANTS, MAX-MIN Ant System,^[46] one of the best-performing ACO algorithms, is used. Artificial ants laying artificial pheromone trails optimize the relative location and orientation of the ligand in the active site as well as the torsional degrees of freedom of the ligand and of hydrogen bond donor groups in the protein. For a more in-depth description, see references [35–37].

To guide the search, the recently described ChemPLP scoring function^[37] was employed to score protein–ligand interactions as well as intra-ligand clash terms. Additionally, NMR constraints were used as described above. Standard settings for all parameters were used for the scoring function as well as the optimization algorithm (search speed setting: “speed1”).^[37] Only σ , which determines the number of iterations performed by the ACO, was set from 1.25 to 5 to increase the sampling. To take the stochastic nature of the ACO algorithm into account, 20 independent docking runs were performed for each of the three experiments (ChemPLP, ChemPLP/trNOE, ChemPLP/trNOE/STD). All generated structures of one run were sorted according to their scoring function value. Starting from the overall best-scoring structure (cluster structure 1), all other ligand conformations in the sorted list that have a heavy-atom RMSD < 2 Å to cluster structure 1 are eliminated. The clustering process then continues with the next non-eliminated structure (cluster structure 2) in the sorted list, again eliminating all structures within a heavy-atom RMSD of 2 Å to cluster structure 2. This process is repeated until 10 clusters are identified or all structures are eliminated. In this way, a set of diverse complex structures is generated that can be additionally analyzed manually or with more sophisticated approaches to increase the chances of finding the correct pose.

Keywords: drug design · NMR spectroscopy · protein–ligand docking · saturation transfer difference · trNOE

- [1] E. Kellenberger, J. Rodrigo, P. Muller, D. Rognan, *Proteins Struct. Funct. Bioinf.* **2004**, *57*, 225–242.
- [2] E. M. Krovat, T. Steindl, T. Langer, *Curr. Comput.-Aided Drug Design* **2005**, *1*, 93–102.
- [3] E. Perola, W. P. Walters, P. S. Charifson, *Proteins Struct. Funct. Bioinf.* **2004**, *56*, 235–249.
- [4] M. von Korff, J. Freyss, T. Sander, *J. Chem. Inf. Model.* **2009**, *49*, 209–231.
- [5] J. B. Cross, D. C. Thompson, B. K. Rai, J. C. Baber, K. Y. Fan, Y. Hu, C. Humblet, *J. Chem. Inf. Model.* **2009**, *49*, 1455–1474.
- [6] N. Huang, B. K. Shoichet, J. J. Irwin, *J. Med. Chem.* **2006**, *49*, 6789–6801.
- [7] K. E. Hevener, W. Zhao, D. M. Ball, K. Babaoglu, J. Qi, S. W. White, R. E. Lee, *J. Chem. Inf. Model.* **2009**, *49*, 444–460.
- [8] T. Cheng, X. Li, Y. Li, Z. Liu, R. Wang, *J. Chem. Inf. Model.* **2009**, *49*, 1079–1093.
- [9] G. L. Warren, C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, M. S. Head, *J. Med. Chem.* **2006**, *49*, 5912–5931.
- [10] P. Englebienne, N. Moitessier, *J. Chem. Inf. Model.* **2009**, *49*, 1568–1580.
- [11] C. R. Corbeil, N. Moitessier, *J. Chem. Inf. Model.* **2009**, *49*, 997–1009.
- [12] C. L. Perrin, J. B. Nielson, *Annu. Rev. Phys. Chem.* **1997**, *48*, 511–544.
- [13] B. Wang, L. M. Westerhoff, K. M. Merz, *J. Med. Chem.* **2007**, *50*, 5128–5134.
- [14] D. Gonzalez-Ruiz, H. Gohlke, *J. Chem. Inf. Model.* **2009**, *49*, 2260–2271.
- [15] C. Dominguez, R. Boelens, A. M. J. J. Bonvin, *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737.
- [16] S. J. de Vries, A. D. J. van Dijk, M. L. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, A. M. J. J. Bonvin, *Proteins Struct. Funct. Bioinf.* **2007**, *69*, 726–733.
- [17] X. Morelli, A. Dolla, M. Czjzek, P. N. Palma, F. Blasco, L. Krippahl, J. J. G. Moura, F. Guerlesquin, *Biochemistry* **2000**, *39*, 2530–2537.
- [18] M. J. Gradwell, J. Feeney, *J. Biomol. NMR* **1996**, *7*, 48–58.
- [19] A. P. R. Zabell, C. B. Post, *Proteins* **2002**, *46*, 295–307.
- [20] M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108–6117.
- [21] M. Mayer, B. Meyer, *Angew. Chem.* **1999**, *111*, 1902–1906; *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788.
- [22] J. Orts, C. Griesinger, T. Carlomagno, *J. Magn. Reson.* **2009**, *200*, 64–73.
- [23] V. M. Sánchez-Pedregal, M. Reese, J. Meiler, M. J. J. Blommers, C. Griesinger, T. Carlomagno, *Angew. Chem.* **2005**, *117*, 4244–4247; *Angew. Chem. Int. Ed.* **2005**, *44*, 4172–4175.
- [24] M. Pellecchia, I. Bertini, D. Cowburn, C. Dalvit, E. Giralt, W. Jahnke, T. L. James, S. W. Homans, H. Kessler, C. Luchinat, B. Meyer, H. Oschkinat, J. Peng, H. Schwalbe, G. Siegal, *Nat. Rev. Drug. Discov.* **2008**, *7*, 738–745.
- [25] B. Meyer, T. Peters, *Angew. Chem.* **2003**, *115*, 890–918; *Angew. Chem. Int. Ed.* **2003**, *42*, 864–890.
- [26] F. G. Hanisch, G. Uhlenbruck, J. Peter-Katalinic, H. Egge, J. Dabrowski, U. Dabrowski, *J. Biol. Chem.* **1989**, *264*, 872–883.
- [27] S. R. Hull, A. Bright, K. L. Carraway, M. Abe, D. F. Hayes, D. W. Kufe, *Cancer Commun.* **1989**, *1*, 261–267.
- [28] K. O. Lloyd, J. Burchell, V. Kudryashov, B. W. Yin, J. Taylor-Papadimitriou, *J. Biol. Chem.* **1996**, *271*, 33325–33334.
- [29] J. Burchell, J. Taylor-Papadimitriou, M. Boshell, S. Gendler, T. Duhig, *Int. J. Cancer* **1989**, *44*, 691–696.
- [30] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.
- [31] P. Dokurno, P. A. Bates, H. A. Band, L. M. D. Stewart, J. M. Lally, J. M. Burchell, J. Taylor-Papadimitriou, D. Snary, M. J. E. Sternberg, P. S. Freemont, *J. Mol. Biol.* **1998**, *284*, 713–728.
- [32] H. Möller, N. Serttas, H. Paulsen, J. M. Burchell, J. Taylor-Papadimitriou, B. Meyer, *Eur. J. Biochem.* **2002**, *269*, 1444–1455.
- [33] T. Weimar, R. Bukowski, N. M. Young, *J. Biol. Chem.* **2000**, *275*, 37006–37010.
- [34] T. H. Charpentier, P. T. Wilder, M. A. Liriano, K. M. Varney, S. Zhong, A. Coop, E. Pozharski, A. D. MacKerell, E. A. Toth, D. J. Weber, *Biochemistry* **2009**, *48*, 6202–6212.
- [35] O. Korb, T. Stützle, T. E. Exner in *Lecture Notes in Computer Science, Vol. 4150* (Eds.: M. Dorigo, L. M. Gambardella, M. Birattari, A. Martinoli, R. Poli, T. Stützle), Springer, Berlin, **2006**, pp. 247–258.
- [36] O. Korb, T. Stützle, T. E. Exner, *Swarm Intell.* **2007**, *1*, 115–134.
- [37] O. Korb, T. Stützle, T. E. Exner, *J. Chem. Inf. Model.* **2009**, *49*, 84–96.
- [38] S. J. Gendler, J. M. Burchell, T. Duhig, D. Lampion, R. White, M. Parker, J. Taylor-Papadimitriou, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6060–6064.
- [39] T. ten Brink, T. E. Exner, *J. Chem. Inf. Model.* **2009**, *49*, 1535–1546.
- [40] X. Lee, A. Thompson, Z. Zhang, H. Ton-that, J. Biesterfeldt, C. Ogata, L. Xu, R. A. Z. Johnston, N. M. Young, *J. Biol. Chem.* **1998**, *273*, 6312–6318.
- [41] H. N. B. Moseley, E. V. Curto, N. R. Krishna, *J. Magn. Reson. Ser. B* **1995**, *108*, 243–261.
- [42] E. V. Curto, H. N. B. Moseley, N. R. Krishna, *J. Comput.-Aided Mol. Des.* **1996**, *10*, 361–371.
- [43] V. Jayalakshmi, N. R. Krishna, *J. Magn. Reson.* **2002**, *155*, 106–118.
- [44] V. Jayalakshmi, N. R. Krishna, *J. Magn. Reson.* **2004**, *168*, 36–45.
- [45] M. Dorigo, T. Stützle, *Ant Colony Optimization*, MIT Press, Cambridge, MA (USA), **2004**.
- [46] T. Stützle, H. H. Hoos, *Fut. Gen. Comput. Syst.* **2000**, *16*, 889–914.

The combination of NMR experimental data and docking tools can greatly increase the reliability of predicted docking poses. For the complex of the antibody SM3 with its epitope, the PLANTS docking program and the ChemPLP scoring function complemented with intra-ligand trNOE and STD distance constraints are able to correctly predict the complex structure as the best-ranked docking pose.



*O. Korb, H. M. Möller, T. E. Exner**

NMR-Guided Molecular Docking of a Protein–Peptide Complex Based on Ant Colony Optimization