Medicinal Chemistry

International Edition: DOI: 10.1002/anie.201500671
German Edition: DOI: 10.1002/ange.201500671

A Combination of Spin Diffusion Methods for the Determination of Protein–Ligand Complex Structural Ensembles**

*Jens Pilger, Adam Mazur, Peter Monecke, Herman Schreuder, Bettina Elshorst, Stefan Bartoschek, Thomas Langer, Alexander Schiffer, Isabelle Krimm, Melanie Wegstroth, Donghan Lee, Gerhard Hessler, K.-Ulrich Wendt, Stefan Becker, and Christian Griesinger**

Dedicated to Professor Horst Kessler on the occasion of his 75th birthday

Abstract: *Structure-based drug design (SBDD) is a powerful and widely used approach to optimize affinity of drug candidates. With the recently introduced INPHARMA method, the binding mode of small molecules to their protein target can be characterized even if no spectroscopic information about the protein is known. Here, we show that the combination of the spin-diffusion-based NMR methods INPHARMA, trNOE, and STD results in an accurate scoring function for docking modes and therefore determination of protein–ligand complex structures. Applications are shown on the model system protein kinase A and the drug targets glycogen phosphorylase and soluble epoxide hydrolase (sEH). Multiplexing of several ligands improves the reliability of the scoring function further. The new score allows in the case of sEH detecting two binding modes of the ligand in its binding site, which was corroborated by X-ray analysis.*

Structure-based drug design (SBDD) is a powerful and widely used approach to optimize affinity of drug candidates.[1] The method usually requires a high-resolution crystal structure of the receptor–ligand complex. However, for approximately 40% of pharmaceutically relevant protein targets it is extremely hard to obtain crystal structures of sufficient quality for SBDD.[2] As an alternative, structures of protein–ligand complexes can be determined using NMR spectroscopy if the protein can be labeled with stable isotopes such as ^{13}C and ^{15}N. Unfortunately, these target proteins are often non-tractable by conventional NMR spectroscopy because of their large size or because it may not be possible to express them in bacteria, preventing labelling with stable isotopes.

In such cases, INPHARMA,[3] interligand NOEs for pharmacophore mapping, can be employed for gaining structural information for SBDD. Using two ligands that bind competitively to the same binding pocket of a protein, INPHARMA peaks in a NOESY spectrum emerge by the NOE transfer from the protons of one ligand to the protons of the other ligand via the protein protons. Because of the physics of the NOESY experiment the ligands need to dissociate from the protein several times during the mixing time in the NOESY experiment which ranges from around 50 ms to several hundreds of milliseconds. Previously, it has been shown that the ligand conformation and mode in protein–ligand complexes could be unambiguously determined using the INPHARMA approach. We use the term binding mode for NMR-derived complex structures to distinguish it from binding poses directly observed in crystal structures of complexes. For example, for protein kinase A (PKA), the protein–ligand complex that had previously been determined by X-ray crystallography could be reproduced based on the crystal structure of free PKA by ligand-restraint docking and INPHARMA-based restraint cross validation.[4] However, using solely INPHARMA, a large computational effort is necessary to find the correct binding mode, that is, 1000–10000 binding modes per ligand were calculated.[5] In addition, more than one mixing time turned out to be mandatory and so far the ligand was restrained during the docking to the bound conformation known from the crystal structure.

A second ligand-based experiment is saturation transfer difference. In this experiment, saturation of protons of the protein is transferred to a bound ligand by spin diffusion. It has been reported that saturation transfer difference (STD)[6] is feasible to determine the binding paratope of a ligand through group epitope mapping, however, not the binding mode. A third method is the transferred NOE (trNOE). It profits from the property of the NOE to depend on the correlation time of the molecule more or less linearly. Since

[*] Dr. J. Pilger,[+] Dr. A. Mazur,[+] M. Wegstroth, Dr. D. Lee, Dr. S. Becker, Prof. Dr. C. Griesinger
Abteilung für NMR-basierte Strukturbiologie
Max-Planck-Institut für biophysikalische Chemie
Am Fassberg 11, 37077 Göttingen (Germany)
E-mail: cigr@nmr.mpibpc.mpg.de
Dr. P. Monecke, Dr. H. Schreuder, Dr. B. Elshorst, Dr. S. Bartoschek, Dr. T. Langer, Dr. A. Schiffer, Dr. G. Hessler, Dr. K.-U. Wendt
Sanofi-Aventis Deutschland GmbH
Industriepark Hoechst, 65926 Frankfurt a. M. (Germany)
Dr. I. Krimm
Université Claude Bernard Lyon1
Villeurbanne Cedex (France)
[+] These authors contributed equally to this work.

[**] We thank Sandra Schimanski-Breves and Alexander Liesum for crystallization, Dirk Bossemeyer and Michael Gassel for protein production of PKA. This project was performed under the auspices of the BMBF project number 1615870A: "Neue Verfahren der Bio-NMR zur Optimierung und Beschleunigung strukturbasierter Wirkstoff-entwicklung".

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201500671.

a ligand when bound to a protein has a much larger correlation time than in free solution, the conformation of the bound ligand dominates the NOE. This experiment yields valuable restraints to obtain the bound conformation of the ligand, yet the information about the ligand binding mode is very limited.

Here, we report that STI, that is, INPHARMA combined with trNOE and STD, which are all spin diffusion methods results in a straightforward method (see Figure 1) for the

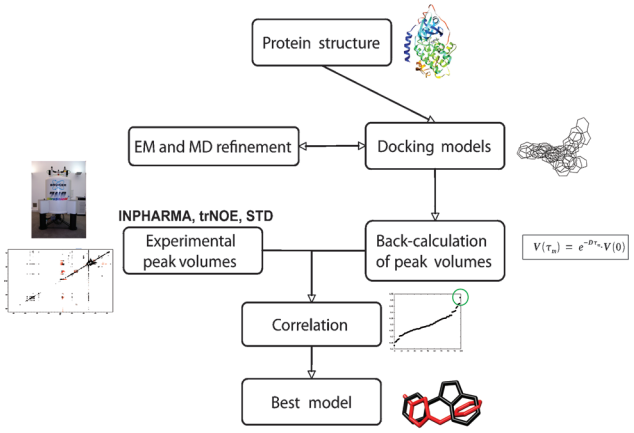


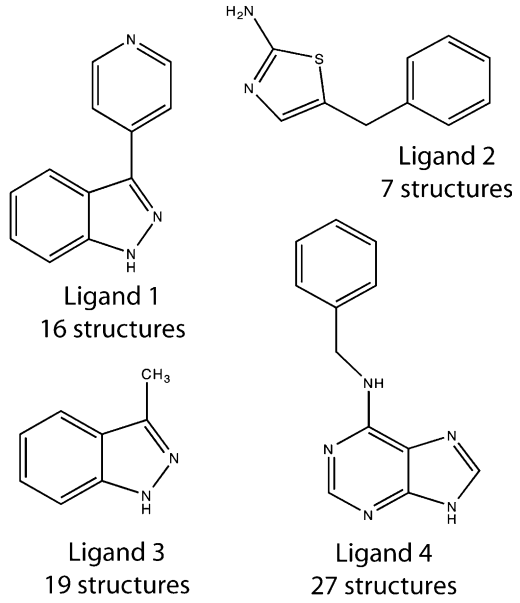
Figure 1. The flowchart gives an overview of the NMR-based binding mode determination methodology (see the Supporting Information for details).

evaluation of the ligand binding mode. Since STD spectra are regularly recorded prior to INPHARMA measurements and tr-NOE peaks are intrinsically present in the NOESY spectra, no additional NMR experiments are required. We would like to emphasize that the focus of this work are the scoring properties of the STI approach, that work on structures created by any docking program.

To evaluate the new approach of combining INPHARMA, tr-NOE and STD restraints, we chose three systems: i) PKA, of which crystal structures in complexes are known, ii) glycogen phosphorylase (GP) and iii) human soluble epoxide hydrolase (sEH). The latter two are pharmaceutically relevant drug targets.

In case of PKA, four heterocyclic ligands were chosen that reversibly target the adenosine triphosphate (ATP)-binding site (Scheme 1). The X-ray structures of these ligands with PKA are known for ligand 1 and 2[4] (pdb accession codes 3DNE and 3DND, respectively) and have been determined here for ligand 3 and 4 (4YXR and 4YXS, respectively). The binding poses in these structures will be compared with the binding modes determined by our approach to combine the spin diffusion methods INPHARMA, trNOE, and STD.

First, restraint-free docking modes using the protein coordinates from PDB entry 3DNE (ligand 1) for the NMR-based refinements as done in previous INPHARMA studies,[4,5] were created for each ligand by the ant colony algorithm-based docking program PLANTS,[7] which is known to perform equally well compared to the established program GOLD.[7] The docking modes showing the best



Scheme 1. Four indazole-based PKA binding ligands. The number of representative structures after docking is shown.

agreement with the experimental data were selected. For these docking modes the INPHARMA, trNOE and STD peak volumes were back-calculated with the software SpINPHARMA (spin diffusion and INPHARMA back-calculation) using the complete-relaxation matrix approach.[8–10] The STD calculation in SpINPHARMA follows the formulation in CORCEMA-ST.[11] Our implementation is faster and, more importantly, it merges with the NOESY calculations seamlessly. These back-calculated volumes were then compared to the experimental NMR results using the Pearson correlation coefficient R to score the fitting. The docking mode with the best fit to the experimental data is preferred. To have an unbiased weight of the different experiments STD, trNOE, and INPHARMA, the correlation coefficients were averaged with the following equation to the correlation coefficient R_{STI} [Eq. (1)]:

$$R_{STI} = \frac{R_{STD} + R_{trNOE} + R_{INPHARMA}}{3} \quad (1)$$

In Figure 2 two examples for ligand pairs (ligands 1 and 2 as well as ligands 3 and 4) are shown. Ranking the docking solutions by R_{STI} identified the structures with closest average RMSD (root-mean-square deviation) to the crystal structures for the two ligands (RMSDs of the single structures can be found in supplement I in the Supporting Information). Separate scoring against INPHARMA and STD or the PLANTS docking scoring function ChemPLP identified structures with far larger RMSDs to the correct structures, that is the crystal structures. It is remarkable, that only the combination of INPHARMA, trNOE, and STD identified the correct binding mode for all ligand combinations, while the docking scoring functions (ChemPLP and GOLDscore), INPHARMA or STD data alone failed to do so.

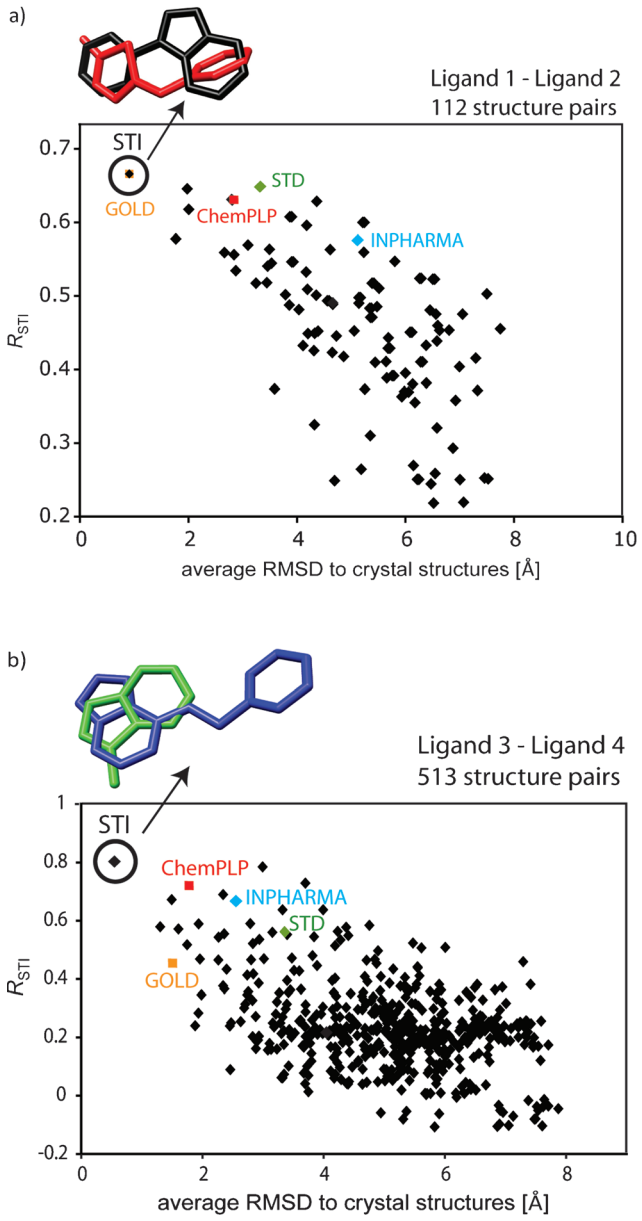


Figure 2. R_{STI} according to Equation (1) plotted against the average RMSD of the two ligand/PKA structures compared to the crystal structures, the best structure according to R_{STI} is indicated with a black circle. The best structures selected by different scoring are represented in blue (INPHARMA) and green (STD) diamonds and red (ChemPLP) and yellow (GOLDScore) squares. For both cases of ligands (1,2) and (3,4), R_{STI} selects the correct binding mode, that is the mode with the smallest RMSD to the average crystal structure. For ligands 1 and 2 this mode is also found by GOLD, for ligands 3 and 4 by no other method. Average $RMSD = (RMSD_{Ligand\ 1} + RMSD_{Ligand\ 2})/2$ (panel a) and average $RMSD$ $(RMSD_{Ligand\ 3} + RMSD_{Ligand\ 4})/2$ (panel b).

The protocol applied in this study is easier to perform than in previous publications[4,5] where on the same system INPHARMA alone reproduced the known X-ray structures. With the combination of INPHARMA, trNOE, and STD, the computational procedure is faster and simpler to implement, because only one mixing time and less docking modes per ligand are sufficient. The docking modes selected with the R_{STI} score can be further refined with MD simulations and cross-validation against the experimental NMR data to obtain structures of higher quality and stability (see supplement I in the Supporting Information).

In lead identification, often many hits are identified, for which determination of the binding mode is valuable to guide selection of a specific compound. Since INPHARMA requires the analysis of pairs of ligands, it is important to use an efficient strategy to minimize the experimental effort, which we describe here. For a set of several ligands binding to the same binding site of the protein, two combinations for each ligand should be measured experimentally: one with a ligand of slightly weaker affinity and one with a ligand of slightly higher affinity (for n ligands n-1 combinations are sufficient). In order to evaluate the quality of the binding modes for the various ligands (i.e. multiplexing), we combined the R_{STI} scores of all ligand pairs to $R_{STI}^{Multiplexing}$, using Equation (2):

$$R_{STI}^{Multiplexing} = \sqrt{\frac{R^2_{STI\_1\&2} + R^2_{STI\_2\&3} + R^2_{STI\_3\&4}}{3}} \quad (2)$$

In the PKA validation study, from the docking modes combining all four ligands, the correct binding mode with the lowest RMSD value is selected with the highest correlation coefficient $R_{STI}^{Multiplexing}$ (see Figure 3). The differentiation is even more robust than in the case of R_{STI} using one pair of ligands only.

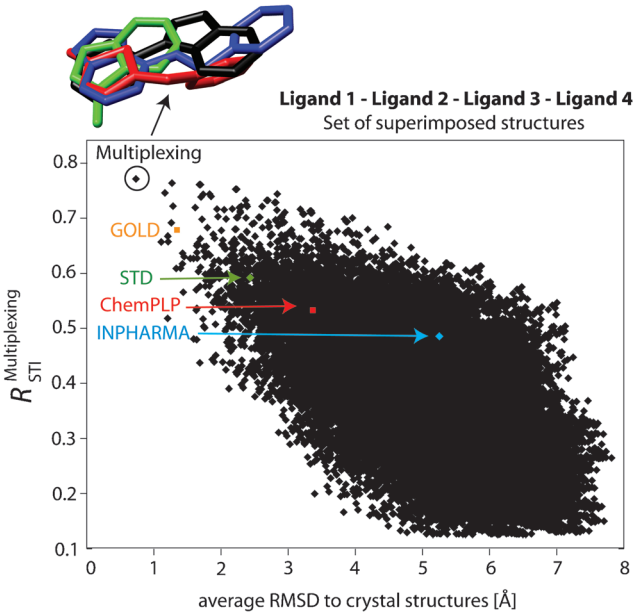


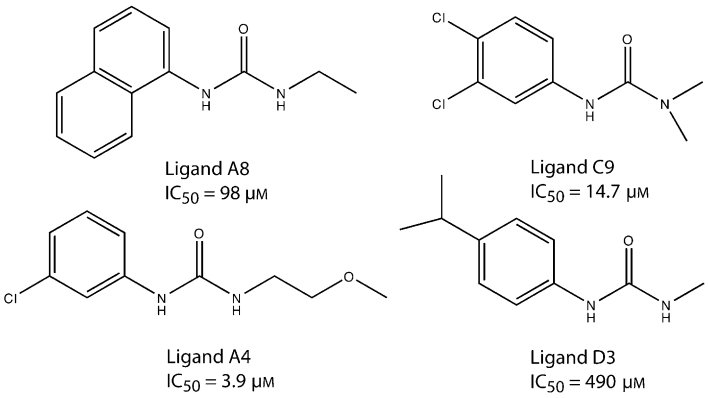
Figure 3. $R_{STI}^{Multiplexing}$ according to Equation (2) plotted against the average RMSD of all four ligand/PKA structures compared to the crystal structures, the best structure according to $R_{STI}^{Multiplexing}$ is indicated with a black circle. The best structures selected by the different scoring are represented in blue (INPHARMA) and green (STD) diamonds and red (ChemPLP) and yellow (GOLDScore) squares. $R_{STI}^{Multiplexing}$ selects the correct binding mode, i.e. the mode with the smallest RMSD to the average crystal structure. Average $RMSD = (RMSD_{Ligand\ 1} + RMSD_{Ligand\ 2} + RMSD_{Ligand\ 3} + RMSD_{Ligand\ 4})/4$.

Encouraged by the result that R_{STI} scoring according to Figure 2 yielded the correct binding mode more reliably than with INPHARMA alone, we applied this approach to two other examples, namely the drug targets glycogen phosphorylase (GP) (see Supplement II in the Supporting Information) and human soluble epoxide hydrolase (sEH).

The next example is sEH, an enzyme that hydrolyses the epoxide moiety formed on unsaturated fatty acids like arachidonic acid during metabolism. Inhibition of this reaction has been shown to have favorable effects on cardiovascular diseases,[12] making sEH an interesting new drug target. An early characterized ligand class consists of a central urea moiety positioned in the middle of the binding site,[13] with one substituent on the left and one on the right side. It is crucial for drug optimization of these ligands to know their orientation in the binding site, which is shaped like a tunnel with entries from both sides. Four ligands (Scheme 2) were chosen and binding to sEH was confirmed with STD spectra.

Subsequently, NOESY spectra of all six pairwise ligand combinations were recorded and INPHARMA peaks were observed for every pairwise ligand combination. For every ligand, five docking modes were created within the crystal structure 1VJ5. All docking modes anchor with the urea moiety and display the aromatic substituents oriented in both directions, respectively. STD, trNOE, and INPHARMA peaks were back-calculated for every structure afterwards using SpINPHARMA. Yet, we did not observe a clear preference for a specific binding mode in contrast to what was found in the case of PKA. Instead, as depicted in Figure 4, the INPHARMA peaks suggest that the two substituents attached to the urea fragment can bind in both directions in the binding pocket such that every ligand populates two distinct binding modes thus calling for an ensemble description. Therefore, population weighting was implemented in SpINPHARMA. For every ligand, two alternate docking modes were chosen, respectively: i) A with the aromatic substituent at the urea moiety pointing towards the Met^{418} side; and ii) B with the same aromatic substituent pointing towards the Met^{337} side (see Figure 5). The populations of binding modes A and B were varied stepwise. Every step was scored by STD, trNOE, INPHARMA and then combined to R_{STI} and $R_{STI}^{Multiplexing}$. A stepsize of 0.1 was used for the populations and scored against the data from all six ligand pairs.

The best overall score $R_{STI}^{Multiplexing}$ for these structure pairs was obtained for the following populations: A4 (pA 0.6, pB: 0.4), A8 (pA: 1.0, pB: 0.0), C9 (pA: 0.8, pB: 0.2) and D3 (pA: 0.4, pB: 0.6). To check these populations that have been obtained from STI measurements for the first time, the ligands A4, A8, and C9 were successfully co-crystallized with sEH and the crystal structures support the observation from NMR. The substituents at the urea moiety for the ligands are always present in both binding modes. The preferred populations according to the crystal structures are: A4 (pA: 0.55, pB: 0.45), A8 (pA: 0.75, pB: 0.25) and C9 (pA: 0.6, pB: 0.4). Populations derived from the NMR experiments agree within the standard deviation of 17% for the calculation of the population by both methods not even taking into account differences between crystal and solution. The standard deviation of the populations was determined in several runs leaving out arbitrarily 10% of the data and calculating the populations, whose standard deviation is the population error. Yet, the preferred binding modes of ligands A4, A8, and C9 are found to be the same in solution and in the crystal. The crystal structures were also applied for the population weighting with NMR data and yielded the populations A4 (pA: 0.8, pB: 0.2), A8 (pA: 0.4, pB: 0.6) and C9 (pA: 1.0, pB: 0.0). It should also be noted that the $R_{STI}^{Multiplexing}$ scores for these structure pairs are in range 0.5–0.6 and thus are lower than



Scheme 2. sEH binding ligands that share the central urea moiety.[13]

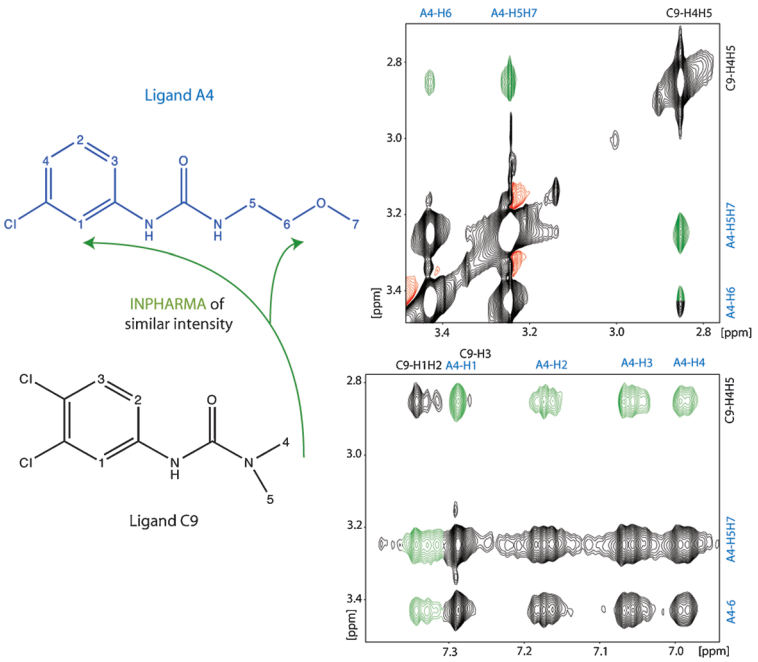


Figure 4. NOESY spectrum of ligands C9 and A4 in the presence of sEH clearly shows trNOE (black) and INPHARMA (green) peaks. Peak intensities between the dimethyl group H4H5 of C9 and the aromatic and aliphatic parts of A4 are very similar, suggesting two flipped binding modes of each ligand.

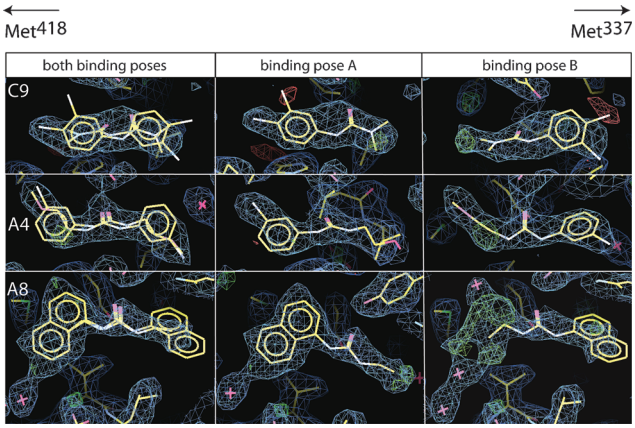


Figure 5. Electron densities of the sEH–ligand complexes confirm the NMR observation, that all ligands are present in two distinct binding modes. In the crystal, an orientation of the aromatic moiety towards the Met^{418} side is preferred (binding pose A), as can be seen by fitting the ligand to the electron densities.

those of PKA and GP, which could reflect the wide conformational space of sEH ligands because of the opposite binding modes and high mobility of the ligands. Yet it is remarkable how well the dual binding poses observed in the crystallographic analysis are reproduced from the STI measurements in solution.

While ligand spectroscopy supported docking was so far limited to single binding modes of each ligand, we have provided compelling evidence that $R_{STI}^{Multiplexing}$ scoring detects ensembles of complex structures, which could then be corroborated by X-ray crystallography.

In conclusion, we have established on the examples of PKA, GP, and sEH a robust method to determine binding modes of small molecule ligands based on scoring of docking modes with INPHARMA, trNOEs, and STD, that is, R_{STI}. This combination provides more robust results than INPHARMA alone, even detecting ensembles of complex structures in solution. The additional experimental data are obtained from NMR experiments that are most of the time done before the INPHARMA measurements: tr-NOESY and STD. Furthermore the inclusion of more than two ligands (if available) can substantially improve the performance of the scoring as shown with the $R_{STI}^{Multiplexing}$ parameter.

Keywords: drug design · glycogen phosphorylase · kinases · NMR spectroscopy · soluble epoxide hydrolase

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 6511–6515
Angew. Chem. **2015**, *127*, 6611–6615

[1] D. C. Rees, M. Congreve, C. W. Murray, R. Carr, *Nat. Rev. Drug Discovery* **2004**, *3*, 660–672.
[2] A. L. Hopkins, C. R. Groom, *Nat. Rev. Drug Discovery* **2002**, *1*, 727–730.
[3] V. M. Sánchez-Pedregal, M. Reese, J. Meiler, M. J. J. Blommers, C. Griesinger, T. Carlomagno, *Angew. Chem. Int. Ed.* **2005**, *44*, 4172–4175; *Angew. Chem.* **2005**, *117*, 4244–4247.
[4] a) J. Orts, J. Tuma, M. Reese, S. K. Grimm, P. Monecke, S. Bartoschek, A. Schiffer, K. U. Wendt, C. Griesinger, T. Carlomagno, *Angew. Chem. Int. Ed.* **2008**, *47*, 7736–7740; *Angew. Chem.* **2008**, *120*, 7850–7854; b) J. Orts, S. K. Grimm, C. Griesinger, K. U. Wendt, S. Bartoschek, T. Carlomagno, *Chem. Eur. J.* **2008**, *14*, 7517–7520.
[5] J. Orts, S. Bartoschek, C. Griesinger, P. Monecke, T. Carlomagno, *J. Biomol. NMR* **2012**, *52*, 23–30.
[6] a) M. Mayer, B. Meyer, *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788; *Angew. Chem.* **1999**, *111*, 1902–1906; b) M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108–6177.
[7] a) O. Korb, T. Stützle, T. E. Exner, *LNCS* **2006**, *4150*, 247–258; b) O. Korb, T. Stützle, T. E. Exner, *J. Chem. Inf. Model.* **2009**, *49*, 84–96; c) O. Korb, H. M. Möller, T. E. Exner, *ChemMedChem* **2010**, *5*, 1001–1006.
[8] R. E. London, *J. Magn. Reson.* **1999**, *141*, 301–311.
[9] M. Reese, V. M. Sanchez-Pedregal, K. Kubicek, J. Meiler, M. J. J. Blommers, C. Griesinger, T. Carlomagno, *Angew. Chem. Int. Ed.* **2007**, *46*, 1864–1868; *Angew. Chem.* **2007**, *119*, 1896–1900.
[10] J. Orts, C. Griesinger, T. Carlomagno, *J. Magn. Reson.* **2009**, *200*, 64–73.
[11] a) V. Jayalakshmi, N. R. Krishna, *J. Magn. Reson.* **2002**, *155*, 106–118; b) V. Jayalakshmi, N. R. Krishna, *J. Magn. Reson.* **2004**, *168*, 36–45.
[12] J. D. Imig, B. D. Hammock, *Nat. Rev. Drug Discovery* **2009**, *8*, 794–805.
[13] N. R. McElroy, P. C. Jurs, *J. Med. Chem.* **2003**, *46*, 1066–1080.

Received: February 9, 2015
Published online: April 15, 2015

