JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Performance of Protein−Ligand Docking with Simulated Chemical Shift Perturbations

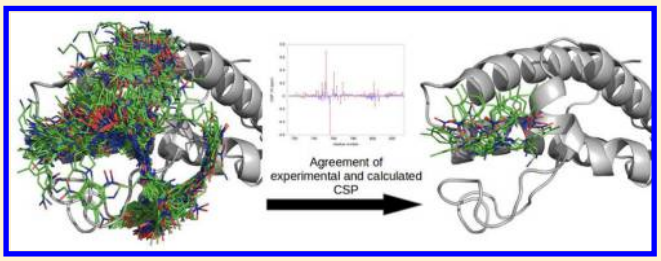
Tim ten Brink,[†] Clémentine Aguirre,[†] Thomas E. Exner,[‡] and Isabelle Krimm*[,†]

[†]Institut des Sciences Analytiques, UMR CNRS 5280, Université Lyon 1, F-69100 Villeurbanne, France

[‡]Theoretical Medicinal Chemistry, Institute of Pharmacy and Biophysics, Eberhard Karls Universität Tübingen, D-72076 Tübingen, Germany

Supporting Information

ABSTRACT: Protein chemical shift perturbations (CSPs) that result from the binding of a ligand to the protein contain structural information about the complex. Therefore, the CSP data, typically obtained during library screening from two-dimensional (2D) nuclear magnetic resonance (NMR) spectra, are often available before attempts to solve the experimental structure of the complex are started, and can be used to solve the complex structure with CSP-based docking. Here, we compare the performance of the post-docking filter and the guided-docking approaches using either amide or α-proton CSPs with 10 protein−ligand complexes. We show that the comparison of experimental CSPs with CSPs simulated for virtual ligand positions can be used to evidence protein conformational change upon binding and possibly improve the CSP-based docking.



INTRODUCTION

The determination of the three-dimensional (3D) structure of a protein−small-molecule complex is a key step in structure-based drug design.[1] X-ray crystallography is usually the experimental method of choice. However, X-ray crystallography experiences some difficulties generating protein−ligand structures for weak affinity ligands such as those used in fragment-based drug design,[2] in particular since a high ligand solubility is required to achieve sufficient occupancy of the binding−site. *In silico* methods, such as computational docking, offer probably the fastest way to solve the binding mode of ligands when bound to a protein. Nevertheless, docking algorithms and scoring functions are far from being perfect,[3,4] especially for fragment docking.[5] To overcome shortcomings of the scoring function, experimental data, such as nuclear magnetic resonance (NMR) data, can be introduced into the scoring function.[6−11] While Nuclear Overhauser effect spectroscopy (NOESY) constraints are commonly used to solve protein complex structures, the method is time-consuming and requires the assignment of the protein proton resonances. Another alternative is to use the protein chemical shifts and measure the change of the chemical shifts of protein nuclei due to the binding of a ligand, the Chemical Shift Perturbations (CSPs), as successfully shown with the well-known structure−activity relationship (SAR) by NMR approach.[12] In the SAR by NMR approach, Hetero-Nuclear-Single-Quantum-Correlation (HSQC) spectra of the free and bound protein are compared, which allows for the identification of ligands and the measurement of the ligand affinity at early stages of molecular design projects. When the protein backbone resonances are assigned, the interaction site of the small molecule on the protein surface can be located.[13]

While the prediction of carbon and nitrogen chemical shifts is possible for proteins,[14−17] there is no appropriate model yet for the calculation of the ligand-induced perturbation of these chemical shifts. By contrast, empirical models exist for the prediction of the perturbation of protein proton chemical shifts upon ligand binding. McCoy and Wyss were the first to use proton CSP data in combination with computational calculations, by comparing experimental CSPs to CSPs predicted with empirical models for different docking poses of the ligand.[18] They used empirical equations based on the ring current effects induced by aromatic rings of ligands. The latter is considered as the major contribution to the proton CSPs in the absence of protein conformational change.[7,18] The binding mode of the ligand corresponds to the ligand position exhibiting the best agreement between experimental CSP data and CSPs calculated with virtual ligand positions. In other published methods, the CSP data are transformed to quite loose distance constraints.[6,19,20]

An important issue for these approaches is the generation of ligand poses. One first method consists of generating an ensemble of possible poses by a docking program and then filtering these according to the CSPs (post-docking filter). This method is straightforward and can be used with any docking program.[18,20] The second method consists of including the CSP data directly into the pose generation step (guided-docking approach). Hybrid methods with pregeneration of poses and

Received: July 23, 2014
Published: October 30, 2014

DOI: 10.1021/ci500446s
J. Chem. Inf. Model. 2015, 55, 275−283

Table 1. Docking Results with the PLANTS ChemPLP Scoring Function, the Post-Docking Filter, and the Guided-Docking Approach[a]

	ChemPLP		Pose Filtering				Guided-Docking			
			αH		Amide H		αH		Amide H	
complex	best	first correct	best	first correct	best	first correct	best	first correct	best	first correct
1J5I	1.64	1.64 (1)	3.16	1.08 (6)	2.67	1.38 (2)	5.50		1.57	1.57 (1)
2L84	1.72	1.72 (1)	3.91	1.73 (3)	6.29	1.86 (5)	1.52	1.52 (1)	2.23	1.53 (15)
2LRR	0.45	0.45 (1)	6.04	1.80 (4)	5.17	1.84 (247)	5.99		6.25	
2LZG	1.07	1.07 (1)	4.69	1.39 (9)	6.67	1.85 (42)	0.90	0.90 (1)	1.03	1.03 (1)
2M5D	1.85	1.85 (1)	4.30	1.94 (2)	2.37	1.84 (8)	2.04	1.97 (8)	3.24	1.79 (104)
1WUG	7.00	0.94 (10)	6.72	1.87 (95)	3.25	1.44 (5)	3.57	1.77 (36)	3.00	1.68 (73)
1WUM	3.65	1.17 (100)	1.89	1.89 (1)	1.74	1.74 (1)	3.09	1.73 (2)	6.46	
2D82	2.85	1.95 (129)	1.28	1.28 (1)	3.63	1.56 (326)	1.76	1.76 (1)	6.21	
2G0L	3.00	0.72 (5)	1.56	1.56 (1)	5.50	1.78 (20)	0.47	0.47 (1)	3.62	1.70 (3)
2L85	6.49	1.75 (94)	n.a.[b]	n.a.[b]	1.34	1.34 (1)	n.a.[b]	n.a.[b]	4.99	1.51 (11)

[a]The RMSD value (in Å) between the experimental and the best ranked pose is given together with the RMSD and the rank of the first correct solution. A missing entry means that the guided-docking approach did not generate any ligand position <2.0 Å to the experimental ligand position. [b]Not available.

local optimization based on the CSP data have also been suggested.[21,22] Amide and Hα protons represent interesting probes for the approach, as their corresponding chemical shifts should be less influenced by amino acid side chain rearrangement of the protein upon ligand binding, by comparison with chemical shifts of side chain protons. Assignment of amide protons is also quite easy and the measurement of CSPs is straightforwardly obtained from $^{1}H-^{15}N$ HSQC spectra.

As already mentioned, some programs that use CSPs for docking are available;[6,7,18−20] however, only HADDOCK,[6] which is principally designed for protein−protein docking, is widely used in the scientific community. Here, our main goal is to compare the performances of the post-docking filter and the guided-docking approaches, using either the amide or the Hα CSPs. We used the docking program PLANTS, which combines an efficient Ant-Colony-Optimization search with the ChemPLP scoring function.[23] The CSP-assisted docking methods are evaluated here with 10 protein−ligand complexes for which the protein apo structure, the actual ligand bound structure, and the NMR data are publicly available. The results indicate that the comparison of the experimental CSPs with the CSPs calculated for a large series of virtual ligand positions is required to ensure that the protein does not undergo a conformational rearrangement that could prevent the method from succeeding, particularly with the post-docking filter. On the other hand, discrepancies between the experimental and simulated CSPs can help identify residues for which an indirect effect of the ligand binding, such as hydrogen bond change or other protein conformational change, strongly contribute to the experimental CSP values. In some of the cases discussed here, removing the CSPs of such residues can significantly improve the CSP-based docking results.

RESULTS

Test Set Assembly and Docking. The protein−ligand complexes used here were assembled from the Protein Data Dank (PDB)[24] with NMR data from the Biological Magnetic Resonance Bank (BMRB).[25] The PDB was searched for complexes with experimental NMR data available. Additional criteria were that NMR data were available for the free protein, that the apo protein structure was recorded in the PDB and that the assignment was nearly complete and contained amide and α-proton assignment for both apo and holo forms. In a third step, the protein and ligand structures were inspected. Only complexes with drug or fragment-like ligands containing at least one aromatic ring were selected, since the CSP simulation is based on the ring current effect of aromatic rings. Complexes were also removed because significant CSPs were observed for almost all protein residues, likely because of their small molecular weights, and no binding sites could be inferred from the NMR data. Finally, 10 complexes could be selected with these criteria. The PDB codes of the protein−ligand complexes are reported in Table 1, and the BMRB codes of the free and ligand bound entries of the complexes, as well as the ligand structures, are listed in Table S1 in the Supporting Information. Three complexes are CBP bromodomains bound to drug-like ligands (PDB codes 2D82, 2L84, and 2L85). Two other complexes are PCAF bromodomains bound to fragment-like inhibitors (PDB codes 1WUG and 1WUM). Two additional complexes are the neocarzinostatin protein bound to flavone (PDB code 2GOL) and to a synthetic chromophore (PDB code 1J5I). The last complexes are the R3H domain of human Sμ bp bound to a desoxy-guanosine-monophosphate (PDB code 2LRR), the P53-binding protein MDM2 in complex with a piperidone ligand (PDB code 2LZG), and the metallo-β-lactamase BCII in complex with *R*-thiomandelic-acid (PDB code 2M5D).

The protein−ligand complexes were docked with PLANTS (see the Methods section) to compare the PLANTS docking results to CSP-based docking results. Five complexes were successfully docked with PLANTS (PDB codes 1J5I, 2L84, 2LRR, 2LZG, and 2M5D), exhibiting a heavy atom RMSD < 2.0 Å to the experimental ligand position. For the five other complexes, PLANTS did not find the correct binding mode, with RMSD > 2.0 Å (2D82 and 2G0L), or did not identify the correct binding site (1WUG, 1WUM, and 2L85).

For each complex, an amide and an α proton CSP set were compiled. Removing CSPs <0.02 ppm from the data was tested. No improvement of the docking results was observed; therefore, all CSPs were retained and only one obviously erroneous value of 11 ppm was removed for the complex 2L84. In addition, the Hα CSPs of complex 2L85 were not used since only two residues exhibited CSPs larger than 0.02 ppm.

DOI: 10.1021/ci500446s
J. Chem. Inf. Model. 2015, 55, 275−283

Post-Docking Filter. For the CSP filtering approach, 1000 ligand positions were generated with PLANTS. For each complex, correct solutions with very low RMSD to the experimental structure (ranging from 0.35 Å to 1.37 Å) were observed, which made the docking results of PLANTS suitable for the pose filtering approach tested here. The CSPs were simulated for the 1000 ligand orientations and compared to the experimental CSPs (see the Methods section). CSP simulation was performed for amide and α protons separately, leading to two independent rankings. The results are listed in Table 1 and Figure 1.

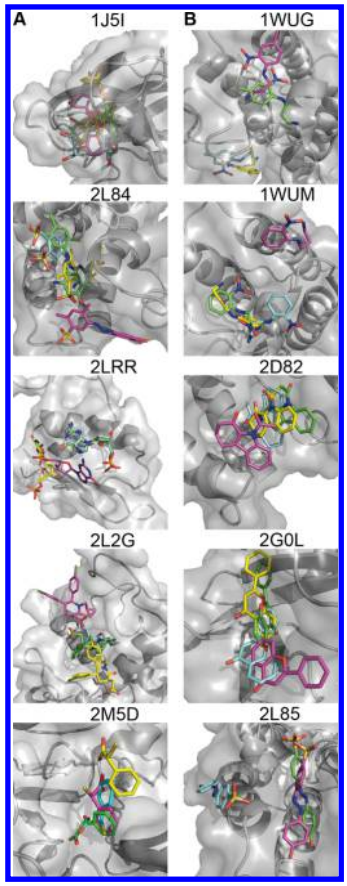


Figure 1. Docking and CSP filter results: (A) docking with PLANTS alone was successful and (B) docking with PLANTS alone failed. The experimental ligand position (green) is compared to the best ranked solution of the PLANTS ChemPLP scoring function (cyan), the best pose selected by the amide CSP filter (magenta) and by the Hα CSP filter (yellow).

For the five complexes successfully docked by PLANTS, the filtering of the poses with the Hα CSPs generated correct solutions in the 10 best-ranked solutions, with RMSD values similar to those obtained with PLANTS alone (Table 1). Nevertheless, the best-ranked solutions displayed an RMSD > 2 Å to the experimental structures for the five complexes. For three of four cases for which the docking with PLANTS alone failed, an improvement over the PLANTS results was obtained (see 1WUM, 2D82, and 2G0L in Table 1 and Figure 1). By contrast, the filtered-docking approach was unsuccessful for complex 1WUG, since the correct position was found at the position 95. No results were obtained for complex 2L85, since the Hα CSPs could not be used.

When the amide proton CSPs were used to filter the docking results, they lead to an improvement in the three cases where PLANTS failed to identify the correct pocket (1WUG, 1WUM, and 2L85). The correct binding site was found for complex 1WUG, and the ligands were docked correctly for complexes 1WUM and 2L85, with RMSD values of 1.74 and 1.34 Å, respectively (see Table 1 and Figure 1). In the two cases where PLANTS found the correct pocket but a wrong binding mode (2D82 and 2G0L), the post-docking filter using amide proton CSPs was not successful. Finally, for the five complexes successfully docked by PLANTS alone, three complexes displayed a correct solution in the first 10 ranked positions (1J5I, 2L84, and 2M5D), while the results including the post-docking CSP filter were clearly worse than the PLANTS scoring function alone for the two complexes 2LRR and 2LZG (Figure 1).

CSP-Guided Docking. The second possibility to use the CSP data for docking is to include the simulation directly into the scoring function, to perform a CSP-guided docking (see the Methods section). Compared to the post-docking filter, where the CSP simulation is applied after the docking and where the ligand poses are selected according to their agreement between simulated and experimental CSPs only, the guided-docking requires a weighting between the normal scoring function and the CSP part of the scoring function. The CSP scoring is based on a Pearson correlation coefficient which ranges from −1 (perfect negative correlation) over 0 (no correlation) to 1 (perfect correlation). To have similar ranges for the CSP and the ChemPLP score, a weighting factor of 100 was chosen. This weight leads to a scoring contribution of ∼50% from the CSP part (see the Methods section).

Concerning the Hα CSPs, the guided-docking approach was successful for complexes 2D82 and 2G0L. The ligand position at rank 2 was a correct solution for complex 1WUM, while the guided-docking failed for 1WUG. For the complexes correctly docked with PLANTS alone, the guided-docking was successful for three complexes (2L84, 2LZG, and 2M5D), while it was not effective for complexes 1J5I and 2LRR (see Table 1).

When the amide proton CSPs were used, none of the five complexes 1WUG, 1WUM, 2D82, 2GOL, and 2L85 are correctly docked with the guided-docking approach (see Figure 2). Nevertheless, if only the aromatic rings are taken into account for the RMSD calculation, complex 2L85 can be considered as successfully docked. With regard to the complexes successfully docked with PLANTS alone, the guided-docking was successful for complexes 1J5I and 2LZG, but failed for complexes 2L84, 2LRR, and 2M5D.

If we compare the performance of the two proton types and the two approaches for the CSP-based docking, the best results were obtained with the post-docking filter using the Hα CSPs, while the guided-docking approach with amide protons yielded very bad results (see Table 1).

The experimental CSPs contain both direct contributions of the ligand (such as the ring current effect) and indirect contributions of the binding event, such as conformational changes and hydrogen bond changes, that are not taken into account in the simulation.[26] If these influences are too large, the selected poses are those that give the best agreement to the experimental CSPs, and therefore show differences in their orientation compared to the experimental structure. These indirect effects are typically larger for the amide proton CSPs than for the Hα CSPs[26] and are therefore particularly

DOI: 10.1021/ci500446s
J. Chem. Inf. Model. 2015, 55, 275−283

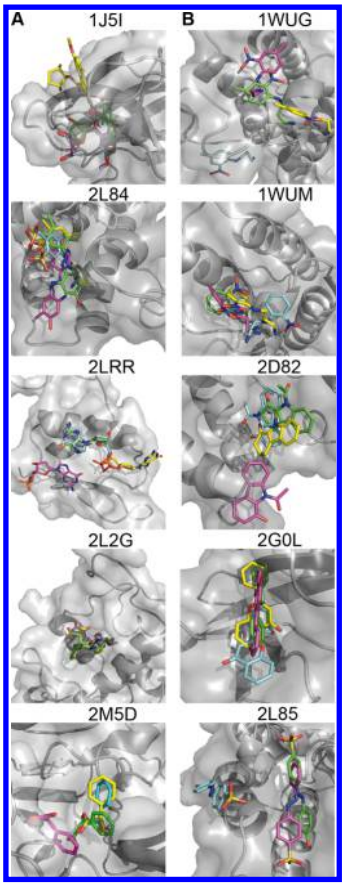


Figure 2. Docking and CSP-guided docking results: (A) docking with PLANTS alone was successful and (B) docking with PLANTS alone failed. The experimental ligand position (green) is compared to the best-ranked solution of the PLANTS ChemPLP scoring function (cyan), the best pose selected by the docking guided by amide proton CSPs (magenta) and guided by Hα CSPs (yellow).

problematic for the post-docking filter and the guided-docking based on amide CSPs.

Comparison of Experimental versus Simulated Amide Proton CSPs. To further analyze the CSP-based docking approaches, the experimental CSPs of amide protons were compared with the CSPs simulated for the 1000 ligand positions. We particularly focused on the complexes where the guided-docking failed to generate correct ligand positions: 2L84, 2LRR, 2M5D, 1WUG, 1WUM, and 2D82. In addition, we analyzed the complex 2LZG that was successfully docked with the guided-docking method but was problematic when the docking results were filtered with the amide CSPs.

For complexes 2LRR and 2M5D, some residues exhibited experimental CSPs larger than the CSP values simulated for the 1000 ligand positions, which is unexpected (see Figure 3 and Figure S1 in the Supporting Information). One would expect the simulated CSPs to have a larger magnitude than the experimental CSPs, since the simulation assumes that the protein is saturated by the ligand. Nevertheless, for both complexes, no clear conclusion could be inferred from the comparison of experimental versus simulated CSPs. For complex 2LZG, the comparison of experimental and simulated CSPs showed that a large number of residues exhibit very large experimental CSPs, compared to the simulated CSPs. We have already shown that such CSP patterns are characteristic of a protein rearrangement upon ligand binding.[27] Even a subtle protein conformational change such as an aromatic side chain movement can induce large CSPs. In the case of the 2LZG protein, aromatic residues Y67, H73, Y100, and Y104 show large side-chain movements upon ligand binding. Because of the large discrepancies between experimental and simulated CSPs, the guided-docking results were very similar to those of the docking alone. By contrast, the post-docking filter failed to generate a correct solution by trying to find a ligand position that has the best agreement with the data.

Conversely, the guided-docking approach failed for complexes 1WUG and 1WUM, while the post-docking filter was successful. The experimental and simulated CSPs of the PCAF bromodomain 1WUG are shown in Figure 3. As can be seen, the residue Y809 displays a much larger experimental CSP value than the values simulated for the 1000 ligand poses. The analysis for the three CBP bromodomains (2D82, 2L84, and 2L85) also highlighted one residue (S1172) that displays a large experimental CSP not explained by any of the ligand poses (for example, see the 2D82 complex data in Figure 3).

The identification of one residue with a large unexplained experimental CSP value allowed us to test whether the post-docking filter and the guided-docking results are improved if the residue is removed from the CSP list. Moreover, even if they do not exhibit large experimental CSPs, the residues in the neighborhood of the identified residue are probably also influenced by the conformational changes. To take these two aspects into account, the post-docking filter and the guided-docking were repeated with (1) removing the CSP of residue Y809 and (2) removing the CSPs of residues E808−A814 from the experimental CSP list, for both complexes 1WUG and 1WUM. As shown in Figure 4, the docking results are clearly improved for complex 1WUG. The RMSD decreases from 3.25 Å to 1.44 Å with the CSP filter, and to 1.32 Å with the CSP-guided docking method when the CSP of residue Y809 are removed. Similar results were obtained if residues E808−E814 are removed, with RMSD = 1.44 and 1.65 Å for the CSP-filter and the guided docking, respectively.

For complex 1WUM, removing residue Y809 from the CSP data improved the RMSD of the top-ranked ligand pose in CSP guided docking by 3.2 Å, but it was still incorrect. For the post-docking filter, the top-ranked solution was correct, as observed when all experimental CSPs are used. When removing the CSPs of residues E808−A814, the CSP guided docking of complex 1WUM improved significantly (RMSD = 1.06 Å). With the post-docking filter, the new top-ranked solution for complex 1WUM was incorrect. Nevertheless, as seen in Figure 4, the higher RMSD is due to the aliphatic chain, for which the CSPs do not provide structural information.

The approach was also tested for the CBP bromodomains (2D82, 2L84, and 2L85). Here, the docking performances were not improved when the CSPs of residue S1172 or the region L1163−L1175 were removed. When comparing the CSP profiles of the CPB bromodomains with the one of complex 1WUG, the CSPs in the binding site region (around R1091 to L1119) of the CBP complexes are comparatively much smaller than those of 1WUG (see Figure 3, as well as Figure S1 in the Supporting Information), which likely explains the results observed.

Complete CSP Set. Finally, the post-docking filter and the guided-docking were tested with the combination of the amide and α-proton CSPs, for the complexes where PLANTS alone failed. In general, the combined Hα and amide proton CSP set lead to results similar to those obtained with the amide CSPs

DOI: 10.1021/ci500446s
J. Chem. Inf. Model. 2015, 55, 275−283

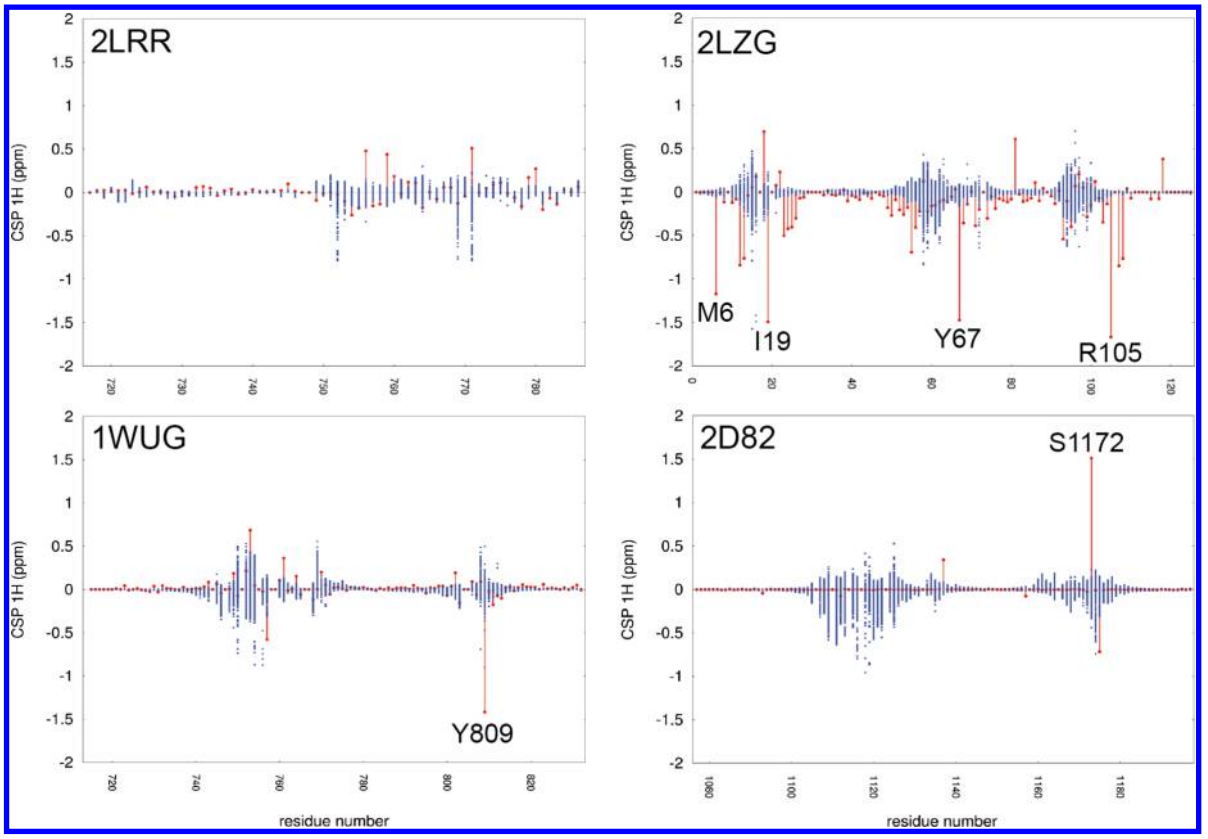


Figure 3. Comparison of experimental and simulated amide CSP profiles. The experimental CSPs (red) are compared to the simulated CSPs of 1000 ligand positions (blue). The residues exhibiting very large unexplained experimental CSPs are labeled. Additional profiles are shown in Figure S1 in the Supporting Information.

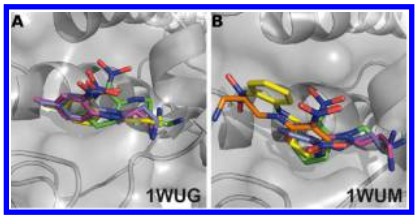


Figure 4. CSP-based docking results for the PCAF bromodomains 1WUG and 1WUG with corrected CSP data. Legend: green structure, experimental structure; blue structure, structure obtained with the post-docking filter when the CSP of residue Y809 is removed; orange structure, structure obtained with the post-docking filter when the CSPs of residues E808−A814 are removed; yellow structure, structure obtained with guided docking when the CSP of residue Y809 is removed; and magenta structure, structure obtained with guided docking when the CSPs of residues E808−A814 are removed.

(see Table 1, as well as Table S1 in the Supporting Information). The guided-docking performed worse in three cases (1WUG, 1WUM, 2D82) when the complete CSP set was used. In the last case (2G0L), correct poses were generated but all had a higher RMSD value than the best-ranked pose obtained with the Hα CSPs (see Table S2 in the Supporting Information).

DISCUSSION

Regarding the ability to identify ligand poses close to the experimental ligand position, both the CSP post-docking filter[20] and the guided-docking method[7] have been shown to give excellent results when artificial CSPs (simulated data for the experimental ligand pose) were used. Both approaches showed lower success rates combined with higher RMSD values, and a lower agreement between simulated and reference CSPs for the best-ranked poses, when experimental CSP data were used.[7,20] This demonstrates that the issue with both approaches can be attributed to the imperfect modeling of the chemical shift changes, which does not simulate all effects that contribute to the CSPs.

The protein CSPs observed upon ligand binding contain direct influence from the ligand, which can be easily simulated (the ring current effect), and additional influences that are difficult to access (protein−ligand hydrogen bonds, for example), as well as indirect influences such as the protein conformational changes and modifications of the protein hydrogen bond network. In case of a subtle conformational change of the protein, Hα CSPs are less affected. By contrast, a subtle protein conformational rearrangement can induce large differences between the experimental and the simulated amide proton CSPs.[27] In addition, amide CSPs can be perturbed by pH modifications and are also sensitive to the solvent effect. Therefore, the CSPs should be corrected from the addition of dimethyl sulfoxide (DMSO) or any other solvent to improve the experimental CSP data.

The comparison of the experimental CSP values with the CSP values calculated for a series of virtual ligand positions can highlight protein residues that contain a large CSP contribution that is not due to the ring current effect of the ligand aromatic rings. This is illustrated here in the case of the protein MDM2 (PDB code 2LZG) (see Figure 3). The comparison of the experimental and simulated CSPs showed that residues M6,

DOI: 10.1021/ci500446s
J. Chem. Inf. Model. 2015, 55, 275−283

Table 2. Summary of the Docking, Post-Docking Filter, and Guided-Docking Results (the PDB Code, a Comment about the Experimental Data Quality, and the Performance of Each Docking Method is Reported)

	exp Data		PLANTS		Filtering				Guided-Docking			
PDB	HN	Hα	best	top 10	Hα best	Hα top 10	HN best	HN top 10	Hα best	Hα top 10	HN best	HN top 10
1J5I	ok	ok	ok	ok	failed	ok	failed	ok	failed	failed	ok	ok
2L84	1 removed	ok	ok	ok	ok	ok	failed	ok	ok	ok	failed	failed
2LRR	ok	ok	ok	ok	failed	ok	failed	failed	failed	failed	failed	failed
2LZG	ok	ok	ok	ok	failed	ok	failed	failed	ok	ok	ok	ok
2M5D	ok	ok	ok	ok	failed	ok	failed	ok	ok	ok	failed	failed
1WUG	ok	ok	failed	failed	failed	failed	ok[a]	ok	failed	failed	ok[a]	ok[a]
1WUM	ok	ok	failed	failed	ok	ok	ok	ok	failed	ok	ok[a]	ok[a]
2D82	ok	ok	failed	failed	ok	ok	failed	failed	ok	ok	failed	failed
2G0L	ok	ok	failed	ok	ok	ok	failed	ok	ok	ok	failed	ok
2L85	ok	not usable	failed	failed	n.a.	n.a.	ok	ok	n.a.	n.a.	failed	failed

[a]CSPs of residues E808−A814 are removed.

I19, Y67, and R105 displayed large experimental CSPs, suggesting that the protein undergoes a structural rearrangement upon ligand binding. In this case, the guided-docking approach gives the same results as the docking alone. By contrast, the post-docking filter is likely to fail because the program will propose wrong ligand orientations that fit as much as possible with the experimental CSPs.

For the bromodomain complexes (1WUG, 1WUM, 2D82, 2L84, and 2L85), the comparison of experimental and simulated CSPs indicated that only one residue exhibited a particularly large experimental CSP. The residues Y809 and S1172 have been identified for the PCAF and CBP bromodomains, respectively. While removing residue S1172 did not improve the docking for the CBP bromodomains, because the CSPs observed for the other regions of the binding site are rather small, a nice improvement was obtained for the PCAF bromodomains (see Figure 4). One important advantage here is that the exact nature of the conformational change that causes the large experimental CSP has not to be known to identify the residues Y809 and S1172.

The guided-docking method offers a clear advantage over the post-docking filter being independent of the initial docking ensemble. If the docking program does not generate native poses or near native poses, the post-docking filter cannot be successful. Here, the PLANTS docking program was able to generate near native poses for all complexes, although they were ranked badly by the scoring function in 5 out of 10 cases (see Table 1). Nevertheless, the sampling of poses can still be insufficient, even with 1000 ligand positions. For complex 1WUM, the post-docking filter using amide proton CSPs selected a correct pose (RMSD = 1.74 Å), but the agreement with the experimental CSPs was poor (P_{score} = 0.0201). By contrast, the guided-docking generated a wrong pose (RMSD = 6.46 Å) with a better agreement to the experimental CSPs (P_{score} = 0.0077). Figure S2 in the Supporting Information shows that few ligand positions were generated by PLANTS in this protein region. Therefore, because of the poor sampling, the CSP filter selected a false positive solution. The observation that rescoring can overestimate the success rate of a scoring function (the CSP scoring in our case) compared to real dockings has already been reported.[28,29] When the CSP data were corrected by removing the CSPs of residues E808−A814, the guided docking generated a correct ligand position (RMSD = 1.06 Å) with a good agreement between simulated and experimental CSPs.

The use of a single CSP set composed of amide and α-proton data was not successful. The main reason is probably the larger magnitude of the amide CSPs, compared to the Hα CSPs. In particular, the guided docking performed poorly when the combined CSP set was used. Clearly, with both approaches tested here, the Hα CSPs performed better than the amide proton CSPs to find the correct binding mode.

CONCLUSION

Here, we have evaluated and compared the post-docking filter and the guided docking with 10 protein−ligand complexes (see Table 2). The CSP guided docking has clear advantages over the CSP post-docking filter, with regard to sampling and pose generation. Nevertheless, the CSP filter is easier to implement and offers the possibility to use different methods for the pose generation, which can be an advantage in some situations. It should be noted that several characteristics of PLANTS help with the pose generation for the post-docking filter. PLANTS writes a user-defined set of clustered poses per docking run, which ensures that a diverse set of input structures for the filtering can be generated.

While all CSP data can be used for the CSP-based docking without any assumption about the location of the binding site, the simulation of amide CSPs for a series of virtual ligand positions can highlight one particular experimental CSP value that contains a large additional contribution to the one of the ring current effect. Removing the residue from the process can significantly improve the CSP-based docking results. Finally, from a performance point of view, the proton choice is important. While the assignment of amide protons is easier, Hα CSPs give better results.

METHODS

Structure and Data Preparation. The structure files were downloaded from the Protein Data Bank (PDB). Seven of the ligand-bound PDB entries contained structure ensembles and three other entries consisted of a single representative structure. For a consistent treatment, only the structure labeled as the best representative structure in the PDB file was used. The ligands were extracted from the complex structures and the protein coordinates were directly used for the dockings. For the ligand structures, the H atoms were removed and readded with the SPORES program.[30] The apo-structures were prepared in a manner similar to that used for the ligand-bound structures. For

DOI: 10.1021/ci500446s
J. Chem. Inf. Model. 2015, 55, 275−283

the CBP bromodomain, the X-ray structure of the apo-form was used. Here, SPORES was used to add the H atoms and to remove water and crystallization buffer molecules.

The free and bound chemical shift data were downloaded from the Biological Magnetic Resonance Bank (BMRB), and amide and α-proton shifts were extracted to calculate the CSP data. The CSPs were defined as the difference between the bound chemical shift and the free chemical shift. In seven cases (1J5I, 1WUG, 1WUM, 2G0L, 2L84, 2LRR, 2LZG), a systematic offset between bound and free values was observed. Here, a constant value was used to correct the zero-line of the CSP profiles. To determine a cutoff value for small CSPs, the method proposed by Schumann[31] was used. In this process, the standard deviation (σ) of the CSPs is calculated iteratively while removing, in each step, CSPs with absolute values larger than 3σ, until no further CSPs are removed. The last σ value obtained is then used as a threshold. The obtained threshold values for the data of the test set were between 0.01 ppm and 0.04 ppm, but no improvement of the docking results was observed when the CSPs below the threshold were set to 0.0. Therefore, all CSPs were retained for the post-docking filter and the guided docking.

CSP Calculation. For the post-docking filter, the CSP simulation included the ring current effect, which is the major contribution to ligand-induced CSPs for ligands that contain aromatic rings,[26,32] and an electric field (ef) term. The ring current effect was calculated with the Haigh−Mallion model:[33]

$$\sigma_{rc} = fB \sum_{ij} S_{ij} \left(\frac{1}{r_i^3} + \frac{1}{r_j^3} \right) \quad (1)$$

Here, f is the ring-specific intensity factor (e.g., 1.00 for benzene type ring), B is the target nucleus factor ($B = 7.06 \times 10^{-6}$ Å for amide and $B = 5.136 \times 10^{-6}$ Å for α-protons[14]). The sum is over the bonds of the ring. The indices i and j mark the atoms forming the bond (e.g., $ij \in \{12, 23, 34, 45, 56, 61\}$ for a six-membered benzene-type ring). r_i and r_j are the distances from the ring atoms i and j to the proton of the protein. S_{ij} is the signed area of the triangle formed by atoms i and j and the proton projected onto the plane of the aromatic ring.

For the additional electric field term, the model from Hunter et al.,[34] without the second-order term, was used:

$$\sigma_{ef} = \varepsilon_1 \left(\sum_i \frac{q_i \cos \theta_i}{r_i^2} \right) \quad (2)$$

The value for ε_1 was set to $\varepsilon_1 = -2.0 \times 10^{-12}$ esu.[34] The sum is over all ligand atoms, r_i is the distance between ligand atom i and the proton, q_i is the partial charge of atom i, and θ_i is the angle between the vector along bond of proton and its bonded heavy atom and the H-ligand atom i vector. The influence of the electric field term for the post-docking filter was shown to be negligible (see Tables S4 and S3 in the Supporting Information).

For the implementation of guided docking, the CSP calculation was modified. The ring current effect calculation was changed to the point-dipole model:[35,36]

$$\sigma_{rc} = fB \left(\frac{1 - 3\cos^2(\theta)}{r^3} \right) \quad (3)$$

f is the ring intensity factor and B is the target nucleus factor analogue to the Haigh−Mallion model ($B = 3.042 \times 10^{-5}$ Å^3).[37,38] r is the distance of the ring center to the proton and θ is the angle between the average ring normal and the vector from the ring center to the proton. Both ring-current models give very similar results.[38] Different CSP values are obtained only for protons located very close to the aromatic ring.[37] The advantage of the point-dipole model is the evaluation time, which is 3−4 times shorter than that observed with the Haigh−Mallion model.

CSP Evaluation and Scoring. The post-docking filter and the guided−docking are based on the comparison of simulated CSPs with the experimental CSPs for a given ligand pose. McCoy et al. suggested the Q_{score} as a measure for the agreement between the two.[18] In the post-docking filter, a normalized version (dubbed pose-score or P_{score}) is used:

$$P_{score} = \frac{1}{N} \sum_i \left(\frac{CSP_{exp}(i)}{CSP_{exp}^{max}} - \frac{CSP_{sim}(i)}{CSP_{sim}^{max}} \right)^2 \quad (4)$$

N is the number of assigned protons (for which a simulation-experiment CSP pair is available). $CSP_{exp}(i)$ is the experimental CSP value for proton i, $CSP_{sim}(i)$ is the calculated CSP value for proton i, and CSP_{exp}^{max} and CSP_{sim}^{max} are the absolute maximal observed and calculated CSP values over all protons, respectively. Low P_{score} indicates that the docking solution is in good agreement with the experimental CSPs. In the case of perfect agreement, the P_{score} value becomes 0. This property renders P_{score} unsuited for guided−docking. The PLANTS scoring function ChemPLP[23] assigns negative score values for favorable ligand poses. Therefore, the Pearson correlation coefficient was chosen as the basis of the CSP scoring in PLANTS. The Pearson correlation coefficient goes from 1.0 (perfect correlation) over 0 (no correlation) to −1.0 (perfect negative correlation). The CSP scoring term was defined as

$$score_{CSP} = -\omega P \quad (5)$$

where P is the Pearson correlation coefficient and ω is a weighting factor that can be set by the user. The negative sign is used to give a negative score to favorable ligand poses and a positive score to unfavorable poses, in accordance with the ChemPLP scoring function. The ChemPLP score of the best-ranked ligand poses for the 10 complexes ranged from −35 score units (for complex 2D82) to −84 score units (for complex 2LZG). An ω of 100 results in an $score_{CSP}$ range between −100 units for perfect agreement and 100 units for perfect negative correlation, which leads to a weighting of ∼1:1 between ChemPLP and $score_{CSP}$ when reasonable correlation for the best-ascored poses is assumed.

Docking. Dockings were conducted with PLANTS[39] with a search algorithm in standard settings and the ChemPLP scoring function.[23] All binding sites were defined as a sphere centered on the center of mass of the experimental ligand structures. The radius was set to the value needed to include all ligand atoms +10.0 Å. For the post-docking filter and the analysis of the PLANTS results, 1000 ligand poses with a clustering RMSD of 0.5 Å were generated. The guided−docking was done with 200 poses with a 1.0 Å cluster RMSD and the CSP term was activated with an weighting factor of 100 (see the CSP Evaluation and Scoring section).

■ ASSOCIATED CONTENT

Supporting Information

Biological Magnetic Resonance Bank (BMRB) entries of the protein−ligand complexes set and ligand structures. Compar-

DOI: 10.1021/ci500446s
J. Chem. Inf. Model. 2015, 55, 275−283

ison of simulated and experimental amide CSP profiles. Influence of the electric field term on filtering with Hα CSPs. Influence of the electric field term on filtering with amide proton CSPs; location of best-ranked solutions in the initial pose ensemble for complex 1WUM. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: isabelle.krimm@univ-lyon1.fr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support was provided by the Agence National de la Recherche (Project No. ANR-11-JS07-0008).

REFERENCES

(1) Böttcher, J.; Jestel, A.; Kiefersauer, R.; Krapp, S.; Nagel, S.; Steinbacher, S.; Steuber, H. Chapter Three: Key Factors for Successful Generation of Protein−Fragment Structures: Requirement on Protein, Crystals, and Technology. In *Fragment-Based Drug Design Tools, Practical Approaches, and Examples*; Kuo, L. C., Ed.; Academic Press: New York, 2011; Vol. *493*, pp 61−89.

(2) Caliandro, R.; Belviso, D. B.; Aresta, B. M.; de Candia, M.; Altomare, C. D. Protein crystallography and fragment-based drug design. *Future Med. Chem.* **2013**, *5*, 1121−1140.

(3) Kubinyi, H. Drug research: myths, hype and reality. *Nat. Rev. Drug Discovery* **2003**, *2*, 665−668.

(4) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; R, C. C. Towards the development of universial, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* **2008**, *153*, 7−26.

(5) Verdonk, M. L.; Giangreco, I.; Hall, R. J.; Korb, O.; Mortenson, P. N.; Murray, C. W. Docking Performance of Fragments and Druglike Compounds. *J. Med. Chem.* **2011**, *54*, 5422−5431.

(6) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731−1737 (PMID: 12580598).

(7) González-Ruiz, D.; Gohlke, H. Steering Protein-Ligand Docking with Quantitative NMR Chemical Shift Perturbations. *J. Chem. Inf. Model.* **2009**, *49*, 2260−2271 (PMID: 19795907).

(8) Korb, O.; Möller, H. M.; Exner, T. E. NMR-Guided Molecular Docking of a Protein−Peptide Complex Based on Ant Colony Optimization. *ChemMedChem.* **2010**, *5*, 1001−1006.

(9) Shah, D. M.; AB, E.; Diercks, T.; Hass, M. A. S.; van Nuland, N. A. J.; Siegal, G. Rapid Protein−Ligand Costructures from Sparse NOE Data. *J. Med. Chem.* **2012**, *55*, 10786−10790.

(10) Orts, J.; Bartoschek, S.; Griesinger, C.; Monecke, P.; Carlomagno, T. An NMR-based scoring function improves the accuracy of binding pose predictions by docking by two orders of magnitude. *J. Biomol. NMR* **2012**, *52*, 23−30.

(11) Skjærven, L.; Codutti, L.; Angelini, A.; Grimaldi, M.; Latek, D.; Monecke, P.; Dreyer, M. K.; Carlomagno, T. Accounting for Conformational Variability in Protein−Ligand Docking with NMR-Guided Rescoring. *J. Am. Chem. Soc.* **2013**, *135*, 5819−5827.

(12) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531−1534.

(13) McCoy, M. A.; Wyss, D. F. Spatial Localization of Ligand Binding Sites from Electron Current Density Surfaces Calculated from NMR Chemical Shift Perturbations. *J. Am. Chem. Soc.* **2002**, *124*, 11758−11763.

(14) Neal, S.; Nip, A.; Zhang, H.; Wishart, D. Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts. *J. Biomol. NMR* **2003**, *26*, 215−240.

(15) Moon, S.; Case, D. A new model for chemical shifts of amide hydrogens in proteins. *J. Biomol. NMR* **2007**, *38*, 139−150 (DOI: 10.1007/s10858-007-9156-8).

(16) Shen, Y.; Bax, A. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR* **2007**, *38*, 289−302.

(17) Han, B.; Liu, Y.; Ginzinger, S.; Wishart, D. SHIFTX2: Significantly improved protein chemical shift prediction. *J. Biomol. NMR* **2011**, *50*, 43−57.

(18) McCoy, M. A.; Wyss, D. F. Alignment of weakly interacting molecules to protein surfaces using simulations of chemical shift perturbations. *J. Biomol. NMR* **2000**, *18*, 189−198.

(19) Schieborr, U.; Vogtherr, M.; Elshorst, B.; Betz, M.; Grimme, S.; Pescatore, B.; Langer, T.; Saxena, K.; Schwalbe, H. How Much NMR Data Is Required To Determine a Protein−Ligand Complex Structure? *ChemBioChem* **2005**, *6*, 1891−1898.

(20) Stark, J.; Powers, R. Rapid Protein−Ligand Costructures Using Chemical Shift Perturbations. *J. Am. Chem. Soc.* **2008**, *130*, 535−545.

(21) Cioffi, M.; Hunter, C. A.; Packer, M. J.; Spitaleri, A. Determination of Protein−Ligand Binding Modes Using Complexation-Induced Changes in ^{1}H NMR Chemical Shift. *J. Med. Chem.* **2008**, *51*, 2512−2517.

(22) Cioffi, M.; Hunter, C. A.; Packer, M. J. Influence of Conformational Flexibility on Complexation-Induced Changes in Chemical Shift in a Neocarzinostatin Protein−Ligand Complex. *J. Med. Chem.* **2008**, *51*, 4488−4495.

(23) Korb, O.; Stützle, T.; Exner, T. E. Empirical scoring functions for advanced protein−ligand docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49*, 84−96.

(24) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−240.

(25) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. BioMagResBank. *Nucleic Acids Res.* **2008**, *36*, D402−D408.

(26) Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1−16.

(27) Aguirre, C.; ten Brink, T.; Walker, O.; Guillière, F.; Davesne, D.; Krimm, I. BcL-xL Conformational Changes upon Fragment Binding Revealed by NMR. *PLoS One* **2013**, *8*, e64400.

(28) O'Boyle, N. M.; Liebeschuetz, J. W.; Cole, J. C. Testing Assumptions and Hypotheses for Rescoring Success in Protein−Ligand Docking. *J. Chem. Inf. Model.* **2009**, *49*, 1871−1878.

(29) Korb, O.; ten Brink, T.; Victor Paul Raj, F.; Keil, M.; Exner, T. Are predefined decoy sets of ligand poses able to quantify scoring function accuracy? *J. Comput.-Aided Mol. Des.* **2012**, *26*, 185−197.

(30) ten Brink, T.; Exner, T. E. Influence of Protonation, Tautomeric, and Stereoisomeric States on Protein−Ligand Docking Results. *J. Chem. Inf. Model.* **2009**, *49*, 1535−1546.

(31) Schumann, F.; Riepl, H.; Maurer, T.; Gronwald, W.; Neidig, K.-P.; Kalbitzer, H. Combined chemical shift changes and amino acid specific chemical shift mapping of protein−protein interactions. *J. Biomol. NMR* **2007**, *39*, 275−289.

(32) Wishart, D. S.; Case, D. A. Use of chemical shifts in macromolecular structure determination. In *Nuclear Magnetic Resonance of Biological Macromolecules Part A*; James, T. L., Dötsch, V., Schmitz, U., Eds. Academic Press: San Diego, CA, 2001; Vol. *338*, pp 3−34.

(33) Haigh, C.; Mallion, R. Ring current theories in nuclear magnetic resonance. *Prog. Nucl. Magn. Reson. Spectrosc.* **1979**, *13*, 303−344.

(34) Hunter, C. A.; Packer, M. J. Complexation-Induced Changes in ^{1}H NMR Chemical Shift for Supramolecular Structure Determination. *Chem.—Eur. J.* **1999**, *5*, 1891−1897.

(35) Pople, J. A. Proton Magnetic Resonance of Hydrocarbons. *J. Chem. Phys.* **1956**, *24*, 1111−1111.

DOI: 10.1021/ci500446s
J. Chem. Inf. Model. 2015, 55, 275−283

(36) Pople, J. Molecular orbital theory of aromatic ring currents. *Mol. Phys.* **1958**, *1*, 175−180.
(37) Moyna, G.; Zauhar, R. J.; Williams, H. J.; Nachman, R. J.; Scott, A. I. Comparison of Ring Current Methods for Use in Molecular Modeling Refinement of NMR Derived Three-Dimensional Structures. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 702−709.
(38) Christensen, A. S.; Sauer, S. P. A.; Jensen, J. H. Definitive Benchmark Study of Ring Current Effects on Amide Proton Chemical Shifts. *J. Chem. Theory Comput.* **2011**, *7*, 2078−2084.
(39) Korb, O.; Stützle, T.; Exner, T. E. An Ant Colony Optimization Approach to Flexible Protein−Ligand Docking. *Swarm Intell.* **2007**, *2*, 115−134.

DOI: 10.1021/ci500446s
J. Chem. Inf. Model. 2015, 55, 275−283