J Biomol NMR (2014) 60:147–156
DOI 10.1007/s10858-014-9864-9

ARTICLE

Protein–ligand structure guided by backbone and side-chain proton chemical shift perturbations

Clémentine Aguirre · Tim ten Brink · Olivier Cala · Jean-François Guichou · Isabelle Krimm

Received: 1 August 2014 / Accepted: 19 September 2014 / Published online: 26 September 2014

Abstract The fragment-based drug design approach consists of screening libraries of fragment-like ligands, to identify hits that typically bind the protein target with weak affinity (100 μM–5 mM). The determination of the protein–fragment complex 3D structure constitutes a crucial step for uncovering the key interactions responsible for the protein–ligand recognition, and for growing the initial fragment into potent active compounds. The vast majority of fragments are aromatic compounds that induce chemical shift perturbations (CSP) on protein NMR spectra. These experimental CSPs can be quantitatively used to guide the ligand docking, through the comparison between experimental CSPs and CSP back-calculation based on the ring current effect. Here we implemented the CSP back-calculation into the scoring function of the program PLANTS. We compare the results obtained with CSPs measured either on amide or aliphatic protons of the human peroxiredoxin 5. We show that the different kinds of protons lead to different results for resolving the 3D structures of protein–fragment complexes, with the best results obtained with the H_α protons.

Keywords Chemical shift perturbation · Fragment · Protein–ligand complex · Ring current effect · Ligand

Clémentine Aguirre and Tim ten Brink contributed equally to this work.

C. Aguirre · T. ten Brink · O. Cala · I. Krimm (✉)
UMR5280 CNRS, Institut des Sciences Analytiques, Ecole Normale Supérieure de Lyon, Université Lyon 1, Villeurbanne, France
e-mail: isabelle.krimm@univ-lyon1.fr

J.-F. Guichou
Centre de Biochimie Strucurale, INSERM U1054/ UMR5048, Université de Montpellier 1 et 2, Montpellier, France

Introduction

Fragment-based drug design (FBDD) allows screening larger parts of chemical space with smaller compound libraries than traditional high throughput screening (HTS) of drug- or lead-like compounds (Harner et al. 2013; Kuo 2011). This has lead FBBD to become an accepted alternative to other lead discovery techniques (Harner et al. 2013; Kuo 2011). However, fragments also lead to new challenges and difficulties especially for the hit verification and the determination of the protein–fragment complex 3D structure. Fragment-like molecules are by definition smaller and possess less possibilities to form directed interactions with the target protein, e.g. by hydrogen bonds. Combined with the aromatic character of the fragments core, these properties lead to the main difficulties in FBDD, namely the weak affinity and a relatively high rate of binding through unspecific interactions like hydrophobic contacts (Bissantz et al. 2010). X-ray crystallography is the method of choice to solve the complex 3D structure but here the low affinity of fragments, which necessitates high compound solubility to ensure a sufficient ligand occupancy, can lead to unsuccessful crystallisation assays (Caliandro et al. 2013).

On early stages of FBDD projects, during the hit identification phase, an idea of the fragment binding modes is particularly helpful, especially to discriminate between binding mostly through hydrophobic contacts or through directed interactions. NMR methods like SAR by NMR (Shuker et al. 1996) or sparse NOE (Shah et al. 2012) approaches have been developed especially for cases where X-ray crystallography fails. The SAR by NMR approach utilises chemical shift perturbations (CSPs) that occur in the protein's NMR spectra upon ligand binding to identify two ligands that bind two

different adjacent pockets on the protein surface (Shuker et al. 1996). This method has been further developed using different approaches in data analysis for the comparison of the binding modes of a series of analogous ligands (Medek et al. 2000; Riedinger et al. 2008). CSPs have been also used quantitatively in combination with computational methods. These computational methods, especially molecular docking approaches, offer possibly the fastest way to solve the binding mode of a small molecule in complex with a target protein. However, being predictions, based on mostly empirical models, docking results still need experimental verification. CSPs have been shown to be able to find the ligand orientation in the protein binding site (McCoy and Wyss 2000) and to refine docking results (Cioffi et al. 2008; Stark and Powers 2008; Cioffi et al. 2009). The experimental data can also be included directly into the docking process to give docking results with experimental confidence (Dominguez et al. 2003; Schieborr et al. 2005; González-Ruiz and Gohlke 2009; Korb et al. 2010).

We have recently shown that we could predict the binding modes of a series of analogous fragments bound to the human peroxiredoxin 5 (PRDX5) by a combination of docking and rescoring with back-calculated CSPs from amide protons (Aguirre et al. 2014). We consider that the aromatic rings of fragments represent valuable probes for the calculation of fragment-induced CSPs based on the ring current effect of aromatic rings. Amide protons are typically the nuclei of choice for CSP-based docking approaches because of the easier assignment of the ^{15}N–^{1}H HSQC spectrum as compared to the ^{13}C–^{1}H HSQC spectrum. However, our results also showed that the CSPs from amide protons contained additional influences from hydrogen bonding and conformational changes that were not included in the back-calculation due to the lack of appropriate models (González-Ruiz and Gohlke 2009; Williamson 2013; Aguirre et al. 2013). CSPs measured for aliphatic protons can offer an alternative as they do not form hydrogen bonds and they are often located closer to the ligand than the backbone amide protons. Here, we investigate whether the additional experimental work and additional NMR measurements needed for the assignment of the aliphatic protons would be justified because they result in more accurate binding mode predictions. As a test example we use the PRDX5 protein target with a series of three fragments, to compare the amide proton CSP-derived binding modes with the aliphatic CSP-derived ones. Additionally we exchanged the rescoring approach by a more sophisticated guided docking approach where the CSP back-calculation is directly implemented in the scoring function of the PLANTS program (Korb et al. 2007, 2009).

Materials and methods

Protein production and purification

Protein production and purification was performed at the *Platform of IBCP-Lyon "Bioengineering of proteins"*. Human peroxiredoxin 5 (PRDX5) was expressed as a 6×His-tagged protein in *Escherichia coli* strain M15 using the pQE-30 expression vector. Cells were grown at 37 °C in M9 minimal medium supplemented with thiamine and containing $^{15}NH_4Cl$ as the sole nitrogen source and [$^{13}C_6$]-D-glucose as the sole carbon source to produce uniformly $^{15}N/^{13}C$-labelled protein. Expression was induced with isopropyl β-D-1-thiogalactopyranoside for 4 h. Cells were then lysed in 10 mM imidazole, 50 mM sodium phosphate, 300 mM NaCl (pH = 8) supplemented with lysozyme and DNase by sonification and clarified by centrifugation. The 6×His-tagged protein contained in the supernatant was purified using a Ni_2^+-affinity chromatography. The protein was eluted with 400 mM imidazole, 50 mM sodium phosphate and 300 mM NaCl (pH = 7.4). Eluted protein was then dialysed (10 kDa cutoff) against PBS buffer (pH = 7.4, NaCl 137 mM, KCl 2.8 mM, Na_2HPO_4 10 mM, KH_2PO_4 1.8 mM).

Resonance assignment

Spectra were acquired with a Varian Inova 600 MHz spectrometer, equipped with a standard 5 mm triple resonance inverse probe with a *z*-axis field gradient and with a Bruker 950 MHz TGIR spectrometer, equipped with cryo TCI, 5 mm probe with a *z*-axis field gradient.

The PRDX5 ^{15}N-HSQC spectrum has been assigned as previously published (Barelier et al. 2010) using NMR samples that contained 500 µM of uniformly $^{15}N/^{13}C/50\,\%^{2}H$ labelled reduced PRDX5 and 5 mM DTT (Dithiothreitol). 3D HNCA, HN(CO)CA, HNCACB and CACB(CO)NH experiments were recorded at 28 °C on a Varian Inova 600 MHz NMR spectrometer. 94 % of the amide resonances could be assigned.

The PRDX5 ^{13}C-HSQC spectrum has been assigned using NMR samples that contained 500 µM of uniformly $^{15}N/^{13}C$ labelled reduced PRDX5 with 5 mM DTT. $^{1}H_m$–$^{13}C_m$ resonances (where *m* denotes methyl) were assigned using (H)CC$_m$H$_m$-TOCSY and H(C)C$_m$H$_m$-TOCSY experiments (Yang et al. 2004; Permi et al. 2004; Würtz et al. 2006) on a Varian Inova 600 MHz NMR spectrometer. The 3D (H)CC$_m$H$_m$-TOCSY data set consisting of 110 × 50 × 1,024 complex points with spectral widths of 80, 40 and 8 ppm in F1, F2 and F3 dimensions was acquired using eight scans. The 3D H(C)C$_m$H$_m$-TOCSY data set consisting of 256 × 40 × 1,024 complex

points with spectral widths of 9, 25 and 9 ppm in F1, F2 and F3 dimensions was acquired using eight scans. Other ^{1}H–^{13}C resonances were assigned using classical (H)CCH-TOCSY and H(C)CH-TOCSY experiments recorded on a Bruker 950 MHz NMR spectrometer. The 3D (H)CCH-TOCSY data set consisting of $65 \times 65 \times 1{,}024$ complex points with spectral widths of 75, 75 and 14 ppm in F1, F2 and F3 dimensions was acquired using eight scans. The 3D H(C)CH-TOCSY data set consisting of $65 \times 65 \times 1{,}024$ complex points with spectral widths of 75, 12 and 14 ppm in F1, F2 and F3 dimensions was acquired using eight scans. The assignment was checked using ^{15}N-NOESY-HSQC with a mixing-time of 150 ms recorded on a Varian Inova 600 MHz NMR spectrometer.

All NMR spectra were processed using NMRPipe (Delaglio et al. 1995) and analysed using NMRView (Johnson and Blevins 1994) and Sparky (Goddard and Kneller 2004).

CSP measurements

Chemical shift perturbations were measured using ^{15}N-HSQC and ^{13}C-HSQC spectra recorded on a Varian Inova 600 MHz and a Bruker 950 MHz NMR spectrometer, respectively. NMR samples contained 500 µM uniformly $^{15}N/^{13}C$ labelled protein, 5 mM DTT, and ligand concentration varied from 0 to 5 mM. 2D ^{15}N-HSQC and ^{13}C-HSQC spectra were acquired at 28 °C, using 64 and 128 t_1 increments, respectively. A control 1D ^{1}H spectrum was recorded prior to each HSQC experiment to assess the purity and stability of the fragments. Solutions at maximal fragment concentration were checked for alteration of the sample pH to prevent confounding sources of CSP. All NMR spectra were processed using NMRPipe (Delaglio et al. 1995) and analysed using NMRView (Johnson and Blevins 1994) and Sparky (Goddard and Kneller 2004).

For a given ^{15}N-HSQC or ^{13}C-HSQC cross peak, the proton CSPs (CSP_H) induced by fragment-binding were defined as the difference between the corresponding chemical shifts in the bound (δ^{1H}_{bound}) and the free (δ^{1H}_{free}) states:

$$CSP_H = \delta^{1H}_{bound} - \delta^{1H}_{free} \quad (1)$$

Docking

All dockings were conducted with PLANTS (Korb et al. 2007, 2009) with the search algorithm in standard settings. The binding site included all residues exhibiting CSPs larger than 0.02 ppm. The ChemPLP scoring function (f_{score}) (Korb et al. 2009) was used along the CSP scoring term. For scoring the agreement between the experimental CSPs and the simulated CSPs, a term of the form

$$f_{CSP} = -\omega P \quad (2)$$

was used. Here, P is the Pearson correlation coefficient calculated between the simulated and experimental values and ω is a weighing factor to generate scores in the range of the ChemPLP values for the docked poses. The minus sign is needed because the ChemPLP assigns negative value to favourable ligand poses. A weighting factor of 100 was chosen for all dockings. This leads to f_{CSP} values between −100 for very good poses and 100 for poses with extremely bad agreement between the simulated and experimental CSPs. The ChemPLP scores for docking of the fragments without CSP ranged from −53 scoring units (fragment **1**) to −73 scoring units (fragment **2**). A weighting factor of 100 therefore ensures enough contribution of the CSP scoring while the ChemPLP still has influence on the substituent placement and directed interactions like hydrogen bonds. When CSP values were different for the two CH_3 of a valine or a leucine, the ambiguous CSPs were tested in the CSP-guided docking. Here we have generated 10 positions for each ligand. In all dockings the cluster RMSD of PLANTS (Korb et al. 2007) was set to 0.5 Å.

CSP simulation

The CSP simulation uses as only contribution the ring current effect of the aromatic rings in the ligand. For the implementation in PLANTS the point-dipole model proposed by Pople (Pople 1956, 1958) was chosen:

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

Here, i is a ring-specific intensity factor (e.g. 1.00 for benzene type ring), B is a factor for the target nuclei (3.042×10^{-5} Å^3 for protons), r is the distance between the ring center of the ligand and the protein's proton and θ is the angle between the average ring normal and the vector from the ring center to the proton. The electric field effect (Hunter and Packer 1999) was tested but was negligeable in our case, so only the ring current effect was used.

Results

In a previous study, we have shown that the fragments **1** (4-methylcatechol), **2** (4-*tert*-butyl-catechol) and **3** (1-1′-biphenyl-3,4-diol) bind to the PRDX5 with a common binding mode (Aguirre et al. 2014), with the aromatic ring and the hydroxyl functions superimposed in the binding pocket of the protein. We have also reported the X-ray structures (Pdb codes 4K7N, 4K7O and 4MMM for the complexes formed between the PRDX5 and fragments **1**, **2**

and **3**, respectively) as well as the affinities of the complexes ($K_D = 330 \pm 40\,\mu M$, $K_D = 50 \pm 20\,\mu M$ and $K_D = 150 \pm 20\,\mu M$ for fragments **1**, **2** and **3**, respectively). In this previous study, the binding modes of the fragments were investigated by filtering virtual fragment positions obtained by docking with AutoDock4 (Morris et al. 2009) according to their agreement between simulated and experimental amide proton CSPs. The CSPs alone were not able to identify a unique binding mode, while the combination of CSPs and STD data with docking generated a unique protein–fragment structure for the three fragments. The results obtained with the NMR data were in good agreement with the X-ray structures obtained by soaking or co-crystallisation. Nevertheless, we noticed that the aromatic rings of the ligands were tilted as compared to the X-ray structures, in particular for small fragments (fragment **1** for example). Such deviations from the experimental structure could likely be related to the influence of additional contributions in the experimental CSPs, such as subtle protein rearrangement and perturbation of the hydrogen bond network. Such contributions are not included in the calculation (Ösapay and Case 1991; Wishart and Case 2001; Parker et al. 2006; Moon and Case 2007; González-Ruiz and Gohlke 2009). Here, we compare the results obtained with either amide (HN) or aliphatic proton (Hα, Hβ, Hγ, ...) CSPs. McCoy and Wyss showed that CSPs of Hα protons could be used, but no comparison with amide protons were reported. They used amino acids to probe the influence of the ligands on the protein chemical shifts, using the SHIFTS program for CSP calculation. In addition, they used CSPs as a filter to select the best docked positions (McCoy and Wyss 2000). Here, we directly introduced the CSPs into the docking process of PLANTS (Korb et al. 2009). Our main intention is to compare the results obtained with amide proton CSPs and aliphatic proton CSPs and discuss their performances toward the determination of protein–ligand complex structures.

Experimental CSP

CSPs were measured on ^{15}N-HSQC and ^{13}C-HSQC spectra of the free and the fragment-bound protein. Amide proton CSPs have been previously reported and discussed in details (Aguirre et al. 2014). An extract of the ^{13}C-HSQC spectra of the free protein and the protein bound to the fragments **1** to **3** is presented in Fig. 1. The maximum absolute CSP values measured for amide protons reached 0.08, 0.15 and 0.16 ppm for fragments **1**, **2** and **3** respectively, while the maximum absolute CSP values measured for aliphatic protons reached 0.63, 0.71 and 0.70 ppm for the same fragments. We observed 6, 11 and 17 strong CSPs ($|CSP| > 0.1$ ppm) as well as 6, 6 and 9 medium CSPs ($0.05 < |CSP| \leq 0.1$ ppm) for fragments **1**, **2** and **3**, respectively. Additionally, 10, 16 and 28 weak CSPs ($0.02 < |CSP| \leq 0.05$ ppm) were observed on the ^{13}C-HSQC spectra for the three fragments.

^{1}HN resonance assignment has been previously published (Barelier et al. 2010). For the assignement of aliphatic protons, we used the H(C)C$_m$H$_m$-TOCSY, (H)CC$_m$H$_m$-TOCSY, H(C)CH-TOCSY and (H)CCH-TOCSY experiments, as well as the HNCA, HN(CO)CA, HNCACB and CBCA (CO)NH experiments. Methyl-containing residue resonances were assigned through the HCC$_m$H$_m$-TOCSY experiment (Yang et al. 2004) (where *m* denotes methyl groups that are selected for observation) using uniformly labelled protein U-[^{15}N/^{13}C] (Fig. 2). The ^{13}C-HSQC resonances were assigned using the (H)CC$_m$H$_m$-TOCSY experiment with the previously assigned Cα and Cβ resonances (Fig. 2a). Side chain protons were assigned using the H(C)C$_m$H$_m$- TOCSY experiment (Fig. 2b). For the other non-methyl containing residues (Phe, Glu, ...), the ^{1}H and ^{13}C edited HCCH-TOCSY experiments were used.

We assigned 100 % of the resonances displaying strong CSPs (larger than 0.1 ppm) upon ligand addition. 100 % of the resonances displaying medium CSPs ($0.05 < |CSP| \leq 0.1$ ppm) could be assigned in the cases of fragments **1** and **2**, but only 4 out the 9 medium CSPs could be assigned for fragment **3**. For the weak CSPs ($0.02 < |CSP| \leq 0.05$ ppm), we assigned 6 out of 10 (fragment **1**), 9 out of 16 (fragment **2**) and 7 out of 28 (fragment **3**) resonances. The repartition of the measured CSPs for the amide protons HN and aliphatic protons H$_{ali}$ are displayed in Fig. 3. As previously observed by González-Ruiz and Gohlke in 2009 (González-Ruiz and Gohlke 2009), in the case of sparse experimental CSP data, it is the distribution of the CSPs rather than the amount of data that determine the success of the method. If the CSPs are uniformly distributed around the binding site, as it is the case for the PRDX5 assigned protons (Fig. 3), they offer better restraints for the ring orientation than CSPs distributed unevenly in the binding site (González-Ruiz and Gohlke 2009). Figure 3 also shows the difference of the proximity of the protein aliphatic protons to the ligand as compared to the protein amide protons. Due to this proximity, aliphatic protons are more perturbed by the ring current induced by the fragment-binding and the resulting experimental CSPs are more intense than those of the amide protons.

CSP-guided docking

By contrast with our previous approach where the ligand positions provided by the docking program were sorted according to their agreement between experimental and simulated CSPs, here the CSPs are included into the docking process. With this approach, the results do not

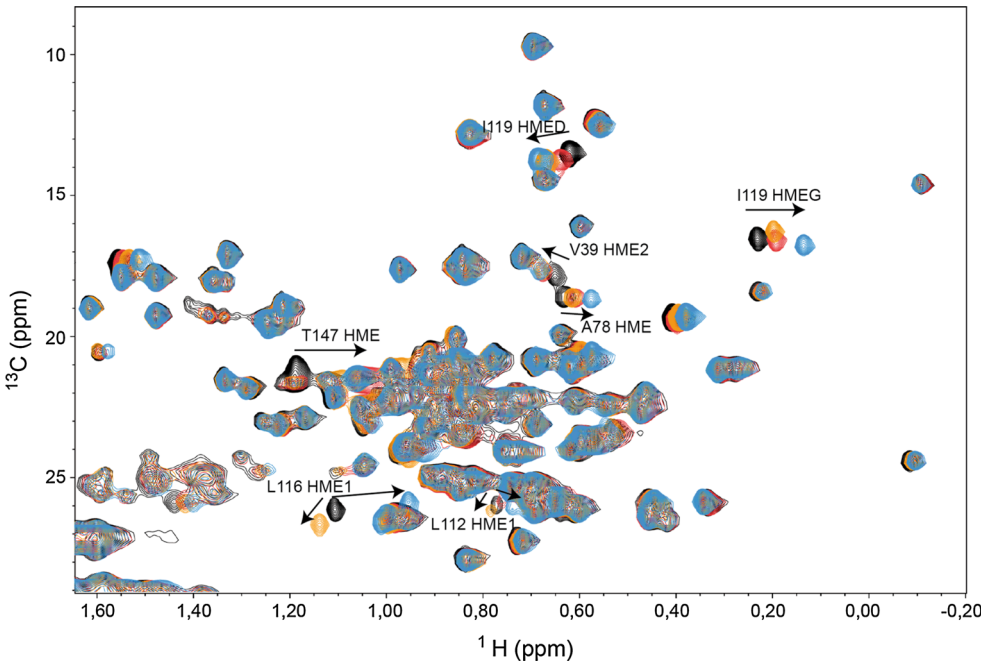

Fig. 1 Superimposition of ^{13}C-HSQC spectra of the free protein PRDX5 (*black*) and the protein bound to the fragments **1** (*red*), **2** (*orange*) and **3** (*blue*). Only the part of the spectra that mainly contain the methyl proton signals is displayed. Resonances that are perturbed by the fragment binding are labelled

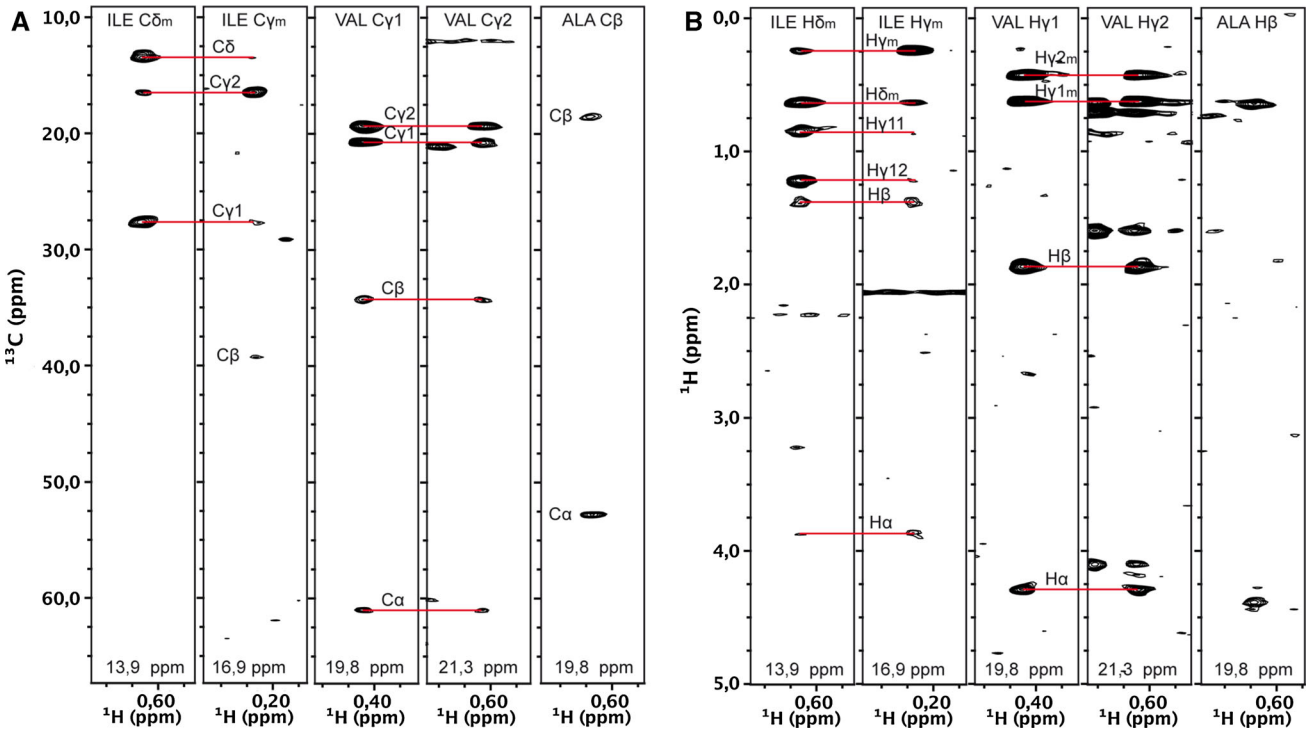

Fig. 2 Representative F1–F3 slices from (H)CC$_m$H$_m$-TOCSY (**a**) and H(C)C$_m$H$_m$-TOCSY (**b**) spectra of the PRDX5 protein. Two identical slices are observed in the case of two methyl-containing residues (Ile, Val, Leu)

depend on the starting positions generated by the docking program, and it is more likely that the generated fragment positions are in agreement with the CSP data (Fig. 4). Here, we used the PLANTS program (Korb et al. 2009), which was modified to integrate the simulation and the comparison to the experimental values. This new procedure requires a weighting between the normal scoring function (f_{score}) and the CSP part of the scoring function (f_{CSP}). The CSP scoring function f_{CSP} is based on a Pearson correlation coefficient which ranges from −1 (perfect negative

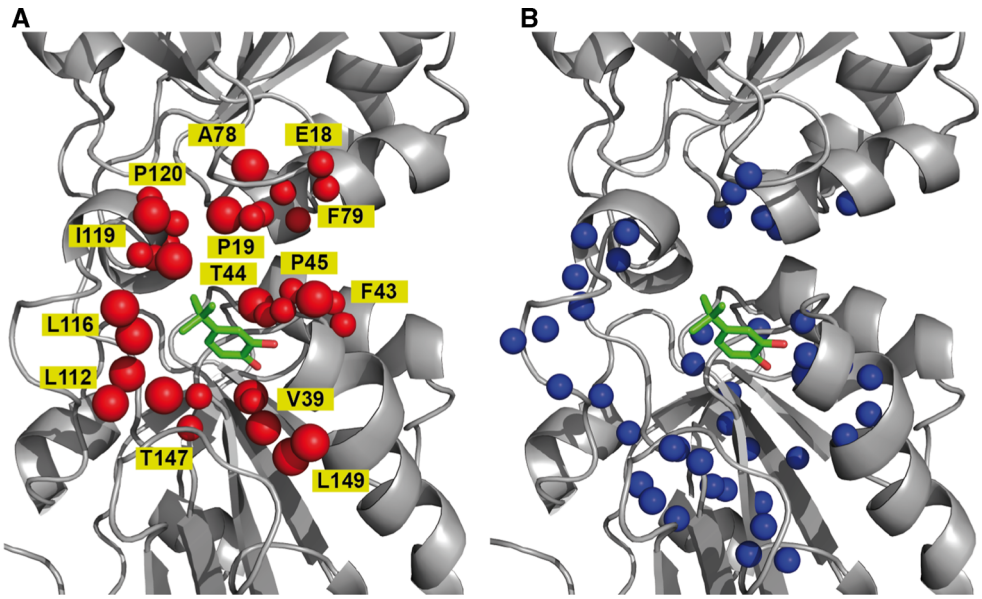

Fig. 3 Distribution of assigned protons on the 3D protein PRDX5 structure. Aliphatic protons are displayed in *red* with larger sphere radius for methyl protons (3 protons for one sphere) (**a**), whereas amide protons are displayed as *blue* spheres (**b**). Only protons that are less than 10 Å away from the fragment aromatic ring center (X-ray structure 4K7O with the fragment **2** in *green* is represented) and with NMR signal that are not overlapping on the protein HSQC spectra, are displayed

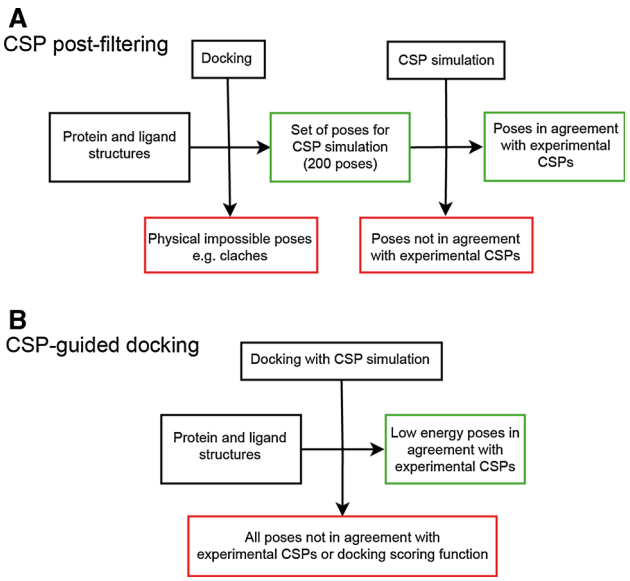

Fig. 4 Comparison scheme of the CSP post-filtering and the CSP-guided docking approach. **a** In the CSP post-filtering, a docking program is used to pre-generate poses that are then rescored by the simulated CSPs. To improve the coverage of poses for the filtering, more poses (200–1,000) than in a normal docking are generated. The final pose selection only includes the CSPs and high energy poses can be selected. **b** In the CSP-guided docking approach, the poses (10–20 poses) are simultaneously optimised for docking score and CSP agreement, which leads in an ideal case only to low energy poses that are in agreement with the experimental CSPs

correlation) over 0 (no correlation) to 1 (perfect correlation). A weighting factor of 100 was chosen to ensure a similar range for the CSP and the normal scoring function part. This weighting leads to a scoring contribution of about 50 % from the CSP part (see "Materials and methods" section). Here, we want to compare the results obtained with this CSP-guided docking from different kind of protons (Hα, H_m, HN, ...).

Figure 5 shows the fragment positions generated with the CSP-guided docking, using CSPs measured for different kind of protons. We can compare the results obtained with amide protons HN and aliphatic protons H_{ali} (Fig. 5a, b). It is clear from the figure that the positions provided by the NMR-guided docking are closer to the X-ray structures when using H_{ali} CSPs than HN CSPs for the three ligands. This is also shown by the RMSD values of the 10 best poses that range in the case of amide protons from 0.7 to 1.6 Å for fragment **1**, from 1.8 to 2.2 Å for fragment **2** and from 1.0 to 1.6 Å for fragment **3**, whereas in the case of aliphatic protons the RMSD range from 0.3 to 1.3 Å, from 0.6 to 1.5 Å and from 0.4 to 1.4 Å for the fragments **1**, **2** and **3**, respectively (Fig. 6). This is very likely related to the fact that HN protons are more sensitive to changes in the protein environment (secondary structure, hydrogen bonds, ...) than the aliphatic protons (Wishart 2011). It appears that amide proton CSPs contain, in addition to the ring current effect, supplementary effects that lead to a CSP-guided docking bias, and that it is in general more appropriate to use CSPs measured on aliphatic protons to determine the ligand binding mode on a target protein.

To go further into the CSP analysis of aliphatic protons, we focused on the methyl protons H_m that represent probes closer to the ligand and that can be easily assigned thanks to the HCC_mH_m-TOCSY experiment. The virtual positions of the fragment obtained with the CSPs of H_m protons are closer to the X-ray structures than those obtained with the CSPs of amide protons (Fig. 5c). Nevertheless, if only the Hα are used for the CSP-guided docking, better results are obtained (Fig. 5d). Hα CSPs are less influenced by a

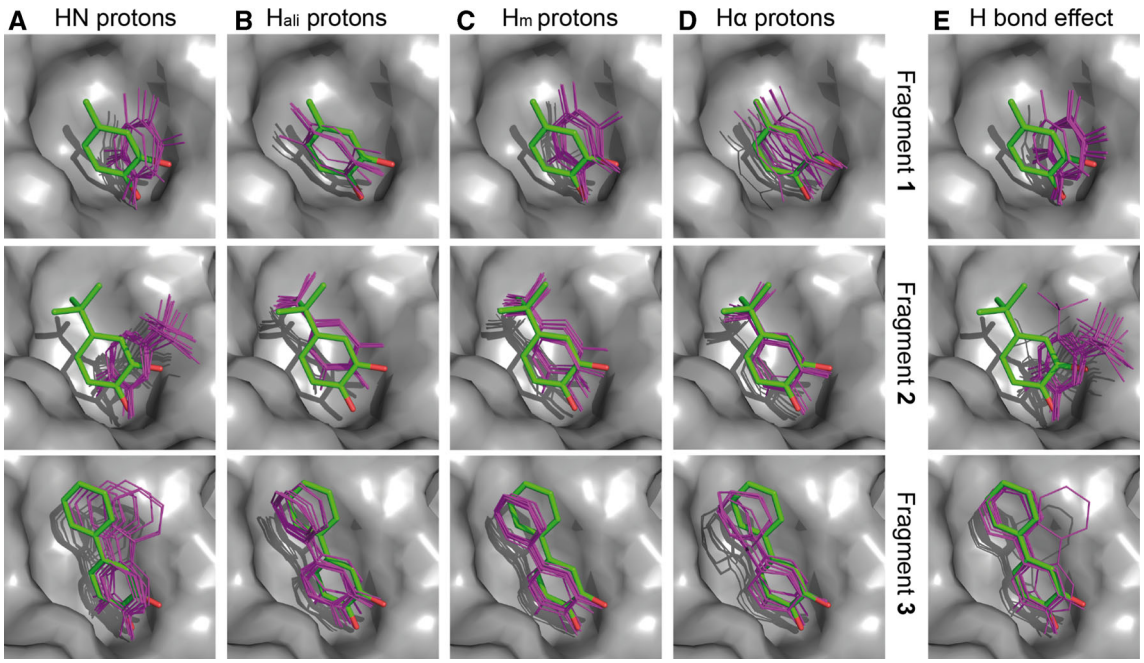

Fig. 5 Virtual fragment positions obtained from the CSP-guided docking. The X-ray structure (*green*) is compared to the 10 best positions generated by the CSP-guided docking with PLANTS (*purple*). Results are given for the fragments **1** to **3** (from *top* to *bottom*). CSP were measured for (**a**) amide protons (HN), (**b**) aliphatic protons (H_{ali}), (**c**) methyl protons (H_m) and (**d**) α protons (Hα). (**e**) Results obtained using HN CSPs by removing values for amide protons involved in hydrogen bonds (residues G46 and C47) from the set of CSPs

possible rearrangement of protein side chains, by comparison with the other aliphatic protons. These results highlight the potency of the Hα CSPs to determine the binding mode of fragments on the PRDX5 protein. The average RMSD values over the 10 first positions are 0.45 Å for fragment **1**, 0.75 Å for fragment **2** and 0.79 Å for fragment **3** in the case of Hα protons; 0.94, 0.93 and 0.78 Å for methyl protons although they are 0.66, 1.20 and 0.79 Å for all aliphatic protons and 1.36, 1.98 and 1.27 Å for amide protons (Fig. 6).

Discussion

In this study, we used ligand-induced CSPs to guide the docking of fragment-like compounds into the protein 3D structure. NMR should play an important role in FBDD, since X-ray crystallography can fail due to the weak affinity of the ligands and the high compound solubility required. All the reports that used the comparison between calculated and experimental ligand-induced CSPs to determine the protein–ligand 3D structure are based on the backbone chemical shifts and only one of them used CSPs from Hα (Wyss et al. 2004; Gorczynski et al. 2007; Cioffi et al. 2008, 2009; González-Ruiz and Gohlke 2009). Here, we used not only the amide and the Hα CSPs, but also the CSPs from the protein side chain protons ($H\beta$, $H\gamma$, H_m, ...) to compare and evaluate the different results. In addition, we have implemented the CSP scoring into the docking program PLANTS.

As illustrated in Fig. 5, the results of the CSP-guided docking depend on the kind of protons used. Nevertheless, the approach generates fragment positions close to the X-ray structures for all kind of protons.

Empirical models for hydrogen bond effects have already been used for the calculation of protein chemical shifts (Neal et al. 2003; Shen and Bax 2007; Nielsen et al. 2012). However, the prediction of the influence of an hydrogen bond formed between a protein and a ligand is much more complicated (Ösapay and Case 1991; Wishart and Case 2001; Parker et al. 2006; Moon and Case 2007; González-Ruiz and Gohlke 2009). The main problem is that the ligand binding can displace solvent molecules and therefore the exact change in the hydrogen bond network is difficult to predict. Nevertheless, as hydrogen bonds typically affect few protons from the total number of protein protons affected by the ligand interaction, González-Ruiz and Gohlke suggested to remove experimental CSPs for the amide protons involved in hydrogen bonds, in order to improve the CSP-guided docking results (González-Ruiz and Gohlke 2009). For the PRDX5 complexes, we have identified the residues involved in hydrogen bonds with the ligands using the Ligplot+ software (Laskowski and Swindells 2011) and the X-ray structures (4K7N, 4K7O and 4MMM). Both the amide protons of residues G46 and

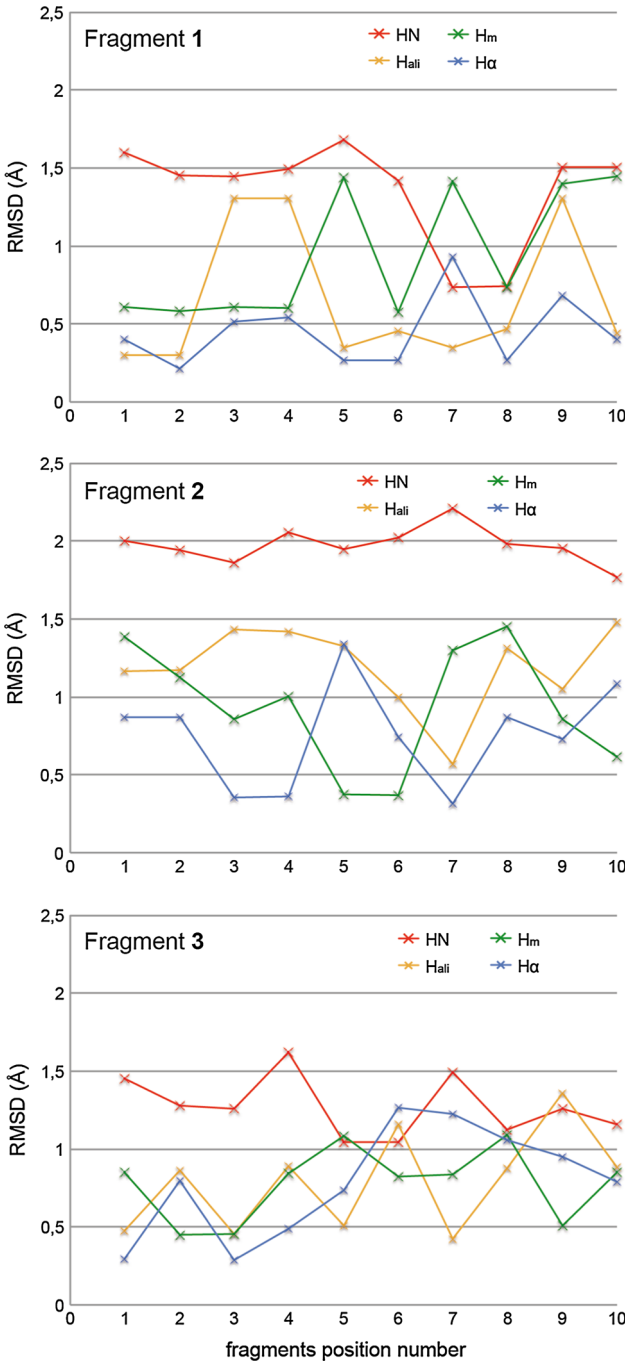

Fig. 6 RMSD values for fragment positions generated by the CSP-guided docking to the X-ray structure. Values are given only for the 10 best scored positions of the fragments **1** to **3** (from *top* to *bottom*)

C47 are involved in hydrogen bond with the fragments. To check if hydrogen bonds could be responsible for the bias observed between the ligand positions generated by the CSP-guided docking and the X-ray structures, the CSPs of G46 and C47 were removed from the docking process. As illustrated in Fig. 5e, no noteworthy improvements are observed for fragments **1** and **2** (compared to Fig. 5a), while a small improvement is observed in the case of fragment **3**. In our case, the hydrogen bonds involve amide protons that are very close to the aromatic ring of the ligands. Therefore, the ring current effect is large for both protons, and compensates the hydrogen bond effect. As a consequence, the experimental CSPs observed for both residues are rather weak when compared to the calculated CSP values. This likely explains why removing the CSPs for those residues have not a significant influence. By contrast, when hydrogen bonds concern residues for which the ring current contribution is weak, the observed CSPs can be very large (due to the hydrogen bond effect) and removing the data will have a significant impact in this case. As previously proposed, HN protons involved in hydrogen bonds should be removed from the set of experimental data, when they exhibit a large experimental CSP that is not correlated with the ring current effect back-calculated using the ligand virtual positions (González-Ruiz and Gohlke 2009).

From the experimental point of view, protein backbone resonance assignment is easily achievable compared to the side chain resonance assignment. In addition, the amount of correlation peaks is larger for protein ^{13}C-HSQC spectra than for ^{15}N-HSQC spectra with more overlapped peaks difficult to assign. However, methyl protons signals can be easily observed using HCC$_m$H$_m$-TOCSY experiment recorded for uniformly labelled protein (U-[^{15}N $-^{13}$ C]), which reduces significantly the ^{13}C-HSQC spectrum complexity. Some ambiguities still appear in the case of the methyl groups of valine and leucine residues (see Fig. 2). Stereospecific assignment of the CH_3 *pro-R* and *pro-S* groups is usually obtained using fractionally ^{13}C labelled proteins (Neri et al. 1989). Protein growing on minimal media containing 10 % [$^{13}C_6$]-glucose results in a doublet for the ^{13}C resonance of the *pro-R* (γ^1 in Val and δ^1 in Leu) and a singlet for the ^{13}C resonance *pro-S* (γ^2 in Val and δ^2 in Leu). More recently, a new labelling procedure was proposed for which only methyl groups for *pro-S* in valine and leucine are protonated using acetolactate 2S [^{1}H, ^{13} C] in expression media (Gans et al. 2010; Plevin et al. 2011). Nevertheless, when the assignment ambiguities are not resolved, and if the CSPs are different for the two groups, the approach can be tested while taking into account these ambiguities. When stereospecific assignment is not mandatory, using CSPs from methyl groups represents a quick and easy way to generate protein–ligand complexes for fragments, even with large proteins > 30 kDa (as it is the case for the PRDX5, 36 kDa).

Chemical shifts of protons from protein side chains are mainly affected by the ring current effect from the aromatic ligand and weakly by the hydrogen bonds. By contrast, they are very sensitive to protein conformational

rearrangement upon ligand interaction. While CSPs measured for side chain protons are very sensitive to the ligand-induced electronic environment (they are closer to the ligand than amide protons), their use can be problematic in case of protein structural rearrangement. Even a very subtle rearrangement of a side chain upon fragment binding can have a significant impact for the CSP calculation and prevent the determination of the correct binding mode. These rearrangements are difficult to model and anticipate. Similar issues appear if the protein structure used for the CSP calculation is different from the protein structure in solution. This could explain why, among the aliphatic protons, Hα appear as the best probes, even if the measured CSPs are weak and if only a small set of protons display significant CSPs. These Hα protons are not sensitive to side chain rearrangement, and not involved in hydrogen bonds with the ligands. They are nevertheless sensitive to backbone conformational change, but this rearrangement is less likely observed with weak affinity ligands.

As we have mentioned, not all the aliphatic proton resonances for which CSPs have been observed could be assigned. As shown in Fig. 5b, the results obtained using these experimental data are clearly sufficient to obtain complex structures close to the X-ray structures. Moreover, the CSP-guided docking based on the Hα protons (4, 5 and 7 CSPs for fragments **1**, **2** and **3**) or the H_m protons (6, 7 and 12 CSPs) generated highly satisfactory results, demonstrating that few experimental CSPs can lead to the determination of the protein–fragment complex structure. We have also tested if removing all the weak CSPs $(0.02 < |\mathrm{CSP}| \leq 0.05\ \mathrm{ppm})$ would have an impact on the results. For fragments **2** and **3**, no influence are noticed, and the RMSD of the 10 best positions are similar to those obtained when using the full set of experimental CSPs. For fragment **1** that displays in overall weak CSPs, the results are slightly different and the results are in better agreement with the X-ray structure when all CSPs are included in the process. In overall, these results show that a complete assignment of aliphatic protons is not necessary for the CSP-guided docking approach.

In conclusion, in the test case reported here, the CSPs measured on Hα protons lead to the best results for the determination of protein–fragment complex structures by CSP-guided docking. The complex structures obtained with the amide proton CSPs displayed the highest RMSD to the X-ray structures, leading nevertheless to correct results. Their easy assignment make therefore the CSPs from amide protons relevant probes to assess the binding mode of fragment hits at the early stage of the FBDD process. Finally, CSPs from methyl groups lead to better results than those obtained with the CSPs of amide protons. For large proteins, it can be particularly convenient to measure CSPs on methyl groups. Nevertheless, one must ensure that no conformational change occurs upon fragment-binding. In the future, the CSP-guided docking approach will need further improvement, especially by taking into account the protein rearrangement upon ligand binding.

Acknowledgments Financial support from the TGIR-RMN-THC Fr3050 CNRS for conducting the research is gratefully acknowledged. The authors want to thank the ANR (Agence Nationale de la Recherche), ANR-11-JS07-0008, for financial support.

References

Aguirre C, ten Brink T, Walker O, Guillière F, Davesne D, Krimm I (2013) BcL-xL conformational changes upon fragment binding revealed by NMR. PLoS One 8(5):e64,400

Aguirre C, ten Brink T, Guichou JF, Cala O, Krimm I (2014) Comparing binding modes of analogous fragments using NMR in fragment-based drug design: application to PRDX5. PLoS One 9(7):e102,300

Barelier S, Linard D, Pons J, Clippe A, Knoops B, Lancelin JM, Krimm I (2010) Discovery of fragment molecules that bind the human peroxiredoxin 5 active site. PLoS One 5(3):e9744

Bissantz C, Kuhn B, Stahl M (2010) A medicinal chemists guide to molecular interactions. J Med Chem 53(14):5061–5084

Caliandro R, Belviso DB, Aresta BM, de Candia M, Altomare CD (2013) Protein crystallography and fragment-based drug design. Future Med Chem 5(10):1121–1140

Cioffi M, Hunter CA, Packer MJ, Spitaleri A (2008) Determination of protein–ligand binding modes using complexation-induced changes in (1)H NMR chemical shift. J Med Chem 51(8):2512–2517

Cioffi M, Hunter CA, Packer MJ, Pandya MJ, Williamson MP (2009) Use of quantitative (1)H NMR chemical shift changes for ligand docking into barnase. J Biomol NMR 43(1):11–19

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6(3):277–293

Dominguez C, Boelens R, Bonvin AMJJ (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125(7):1731–1737

Gans P, Hamelin O, Sounier R, Ayala I, Durá MA, Amero CD, Noirclerc-Savoye M, Franzetti B, Plevin MJ, Boisbouvier J (2010) Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. Angew Chem Int Ed Engl 49(11):1958–1962

Goddard TD, Kneller DG (2004) Sparky 3. University of California, San Fransisco, CA

González-Ruiz D, Gohlke H (2009) Steering protein–ligand docking with quantitative NMR chemical shift perturbations. J Chem Inf Model 49(10):2260–2271

Gorczynski MJ, Grembecka J, Zhou Y, Kong Y, Roudaia L, Douvas MG, Newman M, Bielnicka I, Baber G, Corpora T, Shi J, Sridharan M, Lilien R, Donald BR, Speck NA, Brown ML, Bushweller JH (2007) Allosteric inhibition of the protein–protein interaction between the leukemia-associated proteins Runx1 and CBFβ. Chem Biol 14(10):1186–1197

Harner MJ, Frank AO, Fesik SW (2013) Fragment-based drug discovery using NMR spectroscopy. J Biomol NMR 56(2):65–75

Hunter CA, Packer MJ (1999) Complexation-induced changes in ^{1}H NMR chemical shift for supramolecular structure determination. Chem Eur J 5(6):1891–1897

Johnson BA, Blevins RA (1994) NMR view: a computer program for the visualization and analysis of NMR data. J Biomol NMR 4(5):603–614

Korb O, Stützle T, Exner TE (2007) An ant colony optimization approach to flexible protein–ligand docking. Swarm Intell 2(1):115–134

Korb O, Stützle T, Exner TE (2009) Empirical scoring functions for advanced protein–ligand docking with PLANTS. J Chem Inf Model 49(1):84–96

Korb O, Möller HM, Exner TE (2010) NMR-guided molecular docking of a protein–peptide complex based on ant colony optimization. ChemMedChem 5(7):1001–1006

Kuo LC (2011) Fragment-based drug design: tools, practical approaches, and exemples. Academic Press, San Diego 591 p

Laskowski RA, Swindells MB (2011) Ligplot+: multiple ligand–protein interaction diagrams for drug discovery. J Chem Inf Model 51(10):2778–2786

McCoy MA, Wyss DF (2000) Alignment of weakly interacting molecules to protein surfaces using simulations of chemical shift perturbations. J Biomol NMR 18(3):189–198

Medek A, Hajduk PJ, Mack J, Fesik SW (2000) The use of differential chemical shifts for determining the binding site location and orientation of protein-bound ligands. J Am Chem Soc 122(6):1241–1242

Moon S, Case DA (2007) A new model for chemical shifts of amide hydrogens in proteins. J Biomol NMR 38(2):139–150

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30(16):2785–2791

Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts. J Biomol NMR 26(3):215–240

Neri D, Szyperski T, Otting G, Senn H, Wüthrich K (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling. Biochemistry 28(19):7510–7516

Nielsen JT, Eghbalnia HR, Nielsen NC (2012) Chemical shift prediction for protein structure calculation and quality assessment using an optimally parameterized force field. Prog Nucl Magn Reson Spectrosc 60:1–28

Ösapay K, Case DA (1991) A new analysis of proton chemical shifts in proteins. J Am Chem Soc 113(25):9436–9444

Parker LL, Houk AR, Jensen JH (2006) Cooperative hydrogen bonding effects are key determinants of backbone amide proton chemical shifts in proteins. J Am Chem Soc 128(30):9863–9872

Permi P, Tossavainen H, Hellman M (2004) Efficient assignment of methyl resonances: enhanced sensitivity by gradient selection in a DE-MQ-(H)CC(m)H(m)-TOCSY experiment. J Biomol NMR 30(3):275–282

Plevin MJ, Hamelin O, Boisbouvier J, Gans P (2011) A simple biosynthetic method for stereospecific resonance assignment of prochiral methyl groups in proteins. J Biomol NMR 49(2):61–67

Pople J (1958) Molecular orbital theory of aromatic ring currents. Mol Phys 1(2):175–180

Pople JA (1956) Proton magnetic resonance of hydrocarbons. J Chem Phys 24(5):1111–1111

Riedinger C, Endicott JA, Kemp SJ, Smyth LA, Watson A, Valeur E, Golding BT, Griffin RJ, Hardcastle IR, Noble ME, McDonnell JM (2008) Analysis of chemical shift changes reveals the binding modes of isoindolinone inhibitors of the MDM2-p53 interaction. J Am Chem Soc 130(47):16,038–16,044

Schieborr U, Vogtherr M, Elshorst B, Betz M, Grimme S, Pescatore B, Langer T, Saxena K, Schwalbe H (2005) How much NMR data is required to determine a protein–ligand complex structure? ChemBioChem 6(10):1891–1898

Shah DM, AB E, Diercks T, Hass MAS, van Nuland NAJ, Siegal G (2012) Rapid protein–ligand costructures from sparse NOE data. J Med Chem 55(23):10,786–10,790

Shen Y, Bax A (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR 38(4):289–302

Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274(5292):1531–1534

Stark J, Powers R (2008) Rapid protein–ligand costructures using chemical shift perturbations. J Am Chem Soc 130(2):535–545

Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc 73:1–16

Wishart DS (2011) Interpreting protein chemical shift data. Prog Nucl Magn Reson Spectrosc 58:62–87

Wishart DS, Case DA (2001) Use of chemical shifts in macromolecular structure determination. Methods Enzymol 338:3–34

Würtz P, Hellman M, Tossavainen H, Permi P (2006) Towards unambiguous assignment of methyl-containing residues by double and triple sensitivity-enhanced HCCmHm-TOCSY experiments. J Biomol NMR 36(1):13–26

Wyss DF, Arasappan A, Senior MM, Wang YS, Beyer BM, Njoroge FG, McCoy MA (2004) Non-peptidic small-molecule inhibitors of the single-chain hepatitis C virus NS3 protease/NS4A cofactor complex discovered by structure-based NMR screening. J Med Chem 47(10):2486–2498

Yang D, Zheng Y, Liu D, Wyss DF (2004) Sequence-specific assignments of methyl groups in high-molecular weight proteins. J Am Chem Soc 126(12):3710–3711