

The description of protein internal motions aids selection of ligand binding poses by the INPHARMA method

Benjamin Stauch · Julien Orts · Teresa Carlomagno

Received: 24 February 2012 / Accepted: 13 August 2012 / Published online: 22 September 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Protein internal motions influence observables of NMR experiments. The effect of internal motions occurring at the sub-nanosecond timescale can be described by NMR order parameters. Here, we report that the use of order parameters derived from Molecular Dynamics (MD) simulations of two *holo*-structures of Protein Kinase A increase the discrimination power of INPHARMA, an NMR based methodology that selects docked ligand orientations by maximizing the correlation of back-calculated to experimental data. By including internal motion in the back-calculation of the INPHARMA transfer, we obtain a more realistic description of the system, which better represents the experimental data. Furthermore, we propose a set of generic order parameters, derived from MD simulations of globular proteins, which can be used in the back-calculation of INPHARMA NOEs for any protein–ligand complex, thus by-passing the need of obtaining system-specific order parameters for new protein–ligand complexes.

Keywords INPHARMA · Protein dynamics · Ligand binding · Order parameters

Introduction

Protein surfaces are not static but plastic boundaries, interacting with and adapting to ligands. Besides steric and electrostatic interactions, dynamic features of proteins and protein–ligand interactions have been shown to be functionally relevant (Karplus and Kuriyan 2005; Kay et al. 1998). Protein dynamics can be probed experimentally and computationally (Hub and de Groot 2009; Kohn et al. 2010), with NMR spectroscopy standing out as an especially well-suited experimental tool to study the dynamics of complexes (Mittermaier and Kay 2006) in a close-to-native liquid environment.

NMR spectroscopy is particularly powerful in the investigation of transient protein–ligand complexes (Carlomagno 2005). Ligand binding epitopes can be mapped using STD techniques (Jayalakshmi and Krishna 2002; Mayer and Meyer 2001), protein residues contacting the ligand can be identified using chemical shift perturbation experiments (McCoy and Wyss 2002), and transferred-NOEs or transferred-CCR (cross-correlated relaxation) rates allow for the determination of the bioactive conformation of the ligand (Blommers et al. 1999; Ni 1994; Carlomagno et al. 1999).

Recently, we have developed the INPHARMA method (Orts et al. 2008; Sanchez-Pedregal et al. 2005) to determine the relative binding mode of two competitive, transiently bound ligands. INPHARMA relies on interligand, protein-mediated, transferred-NOE signals between two ligands L_A and L_B , binding competitively and weakly to a receptor T . The efficiency of the INPHARMA transfer at each ligand site depends on the relative binding mode of

Electronic supplementary material The online version of this article (doi:10.1007/s10858-012-9662-1) contains supplementary material, which is available to authorized users.

B. Stauch · J. Orts · T. Carlomagno (✉)
Structural and Computational Biology Unit, European Molecular
Biology Laboratory (EMBL), Heidelberg, Germany
e-mail: carlomagno@embl.de; teresa.carlomagno@embl.de

Present Address:

B. Stauch
European Bioinformatics Institute (EBI), Hinxton, UK

Present Address:

J. Orts
Eidgenössische Technische Hochschule (ETH), Zurich,
Switzerland

the ligands to the protein; the transfer is generally more efficient between two protons of the two ligands that are close to the same protein protons in the binding pocket (Supplementary Figure S1). Because of this dependency, a quantitative analysis of the INPHARMA NOEs allows for the determination of the relative binding mode of L_A and L_B to the target protein. In favorable cases, the absolute orientation of the ligands within the protein binding pocket can be determined as well from INPHARMA data (Orts et al. 2008). In agreement with standard structural based drug-design workflows, the INPHARMA data are used to select the correct binding modes from a pool of pairs of complex structures generated by molecular docking. The agreement between experimental and back-calculated INPHARMA data for each complexes pair is used as selection criterion (Reese et al. 2007).

For Protein Kinase A (PKA) and two competitive ligands L_A and L_B , the INPHARMA NOEs allowed selection of the correct ligand binding poses from a pool of pairs of PKA/ L_A and PKA/ L_B structures representing combinations of very different orientations of the ligands (Orts et al. 2008, 2012). A high correlation coefficient is found between experimental and back-calculated INPHARMA NOEs for the complexes' pair representing the correct ligands binding poses, that is the crystal structures of the PKA/ L_A and PKA/ L_B complexes. In this favorable case, the INPHARMA data allowed a clear selection not only of the relative, but also of the absolute binding mode of both L_A and L_B .

Despite the high correlation coefficient ($R = 0.82$) of the experimental INPHARMA NOEs with the data back-calculated for the correct structures of the PKA/ L_A and PKA/ L_B complexes, the experimental data were consistently lower than the theoretical ones (slope of the linear fit = 0.33), indicating overestimation of the magnetization transfer efficiency (Orts et al. 2008, 2009). To explain this effect, we suggested the influence of protein internal motion on the INPHARMA NOEs, as order parameters smaller than 1 would reduce the efficiency of the magnetization transfer. Since the interligand NOEs observed in an INPHARMA experiment are mediated by protein protons through spin-diffusion, their value depends on protein internal motions. This is in contrast to intraligand transferred-NOEs, which are dominated by the direct dipolar-dipolar interaction between protons of the ligand and are only mildly affected by the protein protons.

In this work, we set out to include NMR order parameters in the INPHARMA calculations. To demonstrate the usefulness of order parameters in improving the quality of the fit of experimental to theoretical data and in increasing the selective power of the INPHARMA NOEs, we first estimate a set of order parameters for the PKA/ L_A and PKA/ L_B complexes from Molecular Dynamics simulations; next, we

modify the implementation of the full relaxation matrix approach (Nilges et al. 1991), used to back-calculate the INPHARMA data, to allow for incorporation of order parameters in the spectral density function. Lastly, we introduce a set of generic order parameters to be used in INPHARMA calculations, thus by-passing the need of performing MD simulations on each protein/ligand complex of interest; we show that even the use of generic, non-tailored order parameters increases the discrimination power of the INPHARMA method.

Theory

Protein internal motions

Dipolar cross-relaxation rates σ_{kl}^{NOE} determine magnetization transfer between spins k and l through space and depend on the spectral density functions (Ernst et al. 1987):

$$J_{kl}(\omega) = \int_{-\infty}^{\infty} C_{kl}(t) \exp(-i\omega t) dt \quad (1)$$

which are the Fourier transforms of the dipolar correlation functions:

$$C_{kl}(t) = 4\pi \left\langle \frac{Y_{20}(\theta_{kl}^{\text{lab}}(t_0 + t)) Y_{20}^*(\theta_{kl}^{\text{lab}}(t_0))}{r_{kl}^3(t_0 + t) r_{kl}^3(t_0)} \right\rangle \quad (2)$$

with θ_{kl}^{lab} being the angle between the inter-nuclear vector r_{kl} and the external magnetic field, Y_{2m} being the rank 2 spherical harmonics of order m and the angled brackets denoting a Boltzmann ensemble average. Assuming that the overall tumbling motion of the molecule is much slower than fast internal motion, the two kinds of motions can be treated independently of each other (Wallach 1967): for isotropic diffusional tumbling, the correlation function of the overall motion is an exponential $C^{\text{tumbling}}(t) = e^{-|t|/\tau_c}$ with τ_c being the correlation time of the molecule. The contribution of the internal motions to the dipolar correlation function has the form:

$$C_{kl}^{\text{internal}}(t) = \frac{4\pi}{5} \sum_{m=-2}^2 \left\langle \frac{Y_{2m}(\theta_{kl}^{\text{mol}}(t_0 + t), \phi_{kl}^{\text{mol}}(t_0 + t)) Y_{2m}^*(\theta_{kl}^{\text{mol}}(t_0), \phi_{kl}^{\text{mol}}(t_0))}{r_{kl}^3(t_0 + t) r_{kl}^3(t_0)} \right\rangle \quad (3)$$

with r_{kl}^3 , θ_{kl}^{mol} , and ϕ_{kl}^{mol} being the spherical coordinates in a molecular fixed frame.

For $t \rightarrow \infty$ the internal correlation function $C^{\text{internal}}(t)$ assumes a plateau value S^2 , which is called the NMR order parameter (Lipari and Szabo 1982).

The dipolar spectral density functions can then be rewritten as (Brueschweiler et al. 1992):

$$J_{kl}(\omega) = \left\langle \frac{1}{r_{kl}^6} \right\rangle S_{kl}^2 \frac{2\tau_c}{1 + \omega^2\tau_c^2} + \left\langle \frac{1}{r_{kl}^6} \right\rangle (1 - S_{kl}^2) \frac{2\tau_{tot}}{1 + \omega^2\tau_{tot}^2} \quad (4)$$

where

$$S_{kl}^2 = \frac{4\pi}{5} \langle r_{kl}^{-6} \rangle^{-1} \sum_{m=-2}^2 \left| \left\langle \frac{Y_{2m}(\theta_{kl}^{\text{mol}}, \phi_{kl}^{\text{mol}})}{r_{kl}^3} \right\rangle \right|^2 \quad (5)$$

and

$$\frac{1}{\tau_{tot}} = \frac{1}{\tau_c} + \frac{1}{\tau_{kl}} \quad (6)$$

and τ_{kl} being the internal correlation time. The second term of Eq. (4) can be neglected if $\tau_{kl} \ll \tau_c$.

Assuming that the angular and the radial fluctuations are uncorrelated, the order parameter of Eq. (5) can be factorized as:

$$S_{kl}^2 \approx S_{r,kl}^2 \cdot S_{\Omega,kl}^2 \quad (7)$$

where

$$S_{r,kl}^2 = \langle r_{kl}^{-3} \rangle^2 / \langle r_{kl}^{-6} \rangle \quad (8)$$

and

$$S_{\Omega,kl}^2 = \frac{4\pi}{5} \sum_{m=-2}^2 \left| \langle Y_{2m}(\theta_{kl}^{\text{mol}}, \phi_{kl}^{\text{mol}}) \rangle \right|^2 \quad (9)$$

are the radial and angular contributions, respectively. As previously observed (Brueschweiler et al. 1992), for PKA this was found to be true to a good approximation (data not shown).

Results

INPHARMA calculations using order parameters

In previous work (Orts et al. 2008) conducted on PKA in complex with two ligands L_A and L_B (Supplementary Figure S2), we had reported a high correlation coefficient for the linear fit between experimental INPHARMA NOEs and INPHARMA NOEs calculated from crystal structure distances in the complexes PKA/ L_A and PKA/ L_B (Pearson correlation coefficient $R = 0.82$). In this work, internal motions of both the protein and the ligands had been neglected ($S_{kl}^2 = 1$ for all pairs of protons (k, l)). Despite the high correlation coefficient of the linear fit, the slope of only 0.33 indicated the systematic overestimation of the magnetization transfer (Orts et al. 2008) by a factor of ~ 3 .

In order to explain the deviation of the slope of the linear fit from 1, we explored the impact of including

internal motions in the INPHARMA calculations. Protein internal motions are expected to have an impact on the values of interligand INPHARMA NOEs, as these NOEs are mediated by the protons of the protein via spin diffusion. Unlike transferred-NOEs, which have been shown to depend mostly on direct interactions between protons of the ligand, INPHARMA NOEs strictly depend on the interaction of the ligand(s) with the protein protons.

In order to provide a general understanding of the influence of internal motions on INPHARMA NOEs, we simulated the effect of uniform order parameters $S^2 < 1$ of different size for both the protein and the ligands in both the free and bound states (Fig. 1). The data were calculated for the system consisting of the PKA/ L_A and PKA/ L_B complexes, for which the correlation time is varied artificially between 0 and 1,000 ns to simulate the effect of receptor size. For a medium-sized protein ($\tau_c = 15$ –20 ns), the intensities of the INPHARMA NOEs are very sensitive to internal motion; on the other hand, for large receptors lower order parameters are tolerated before observing a considerable effect on the slope of the fit. This is due to a compensatory effect in large receptors, where $S^2 < 1$ reduces the efficiency of the INPHARMA transfer at the protein–ligand interface (as it does for smaller receptors) but at the same time reduces the loss of ligand magnetization in the protein core due to spin-diffusion. In the absence of internal motions, the loss of ligand magnetization due to spin-diffusion in the protein core is more prominent for large receptors than for small ones; consequently, larger receptors benefit more from slowing down this process. Our results emphasize the importance of considering internal motions for small and medium-sized systems. At the same time, the discriminatory power of the INPHARMA calculations proves remarkably robust with respect to variations of the order parameters for both small and large receptors, as correlation coefficients stay high over a wide range of order parameter values, especially when using full build-up information.

Next we asked the question about the dynamics of which group of proton pairs has the strongest effect on the magnitude of magnetization transfer. We systematically varied the order parameters for the inter-molecular and intra-molecular NOEs for a fixed receptor size ($\tau_c = 17$ ns) and monitored the intensity of the INPHARMA NOEs (data not shown). As expected, the reduction of the order parameters for the inter-molecular NOEs has the largest effect and reduces the efficiency of the protein-mediated magnetization transfer between the two ligands. This reduction can be compensated by the presence of $S^2 < 1$ for the intra-protein NOEs, which, as explained before, reduces the loss of magnetization in the protein core. Intra-ligand order parameters $S^2 < 1$ contribute the least.

The calculations of Fig. 1 and Supplementary Figure S3 predict that a uniform order parameter $S^2 < 1$ of about 0.5 is

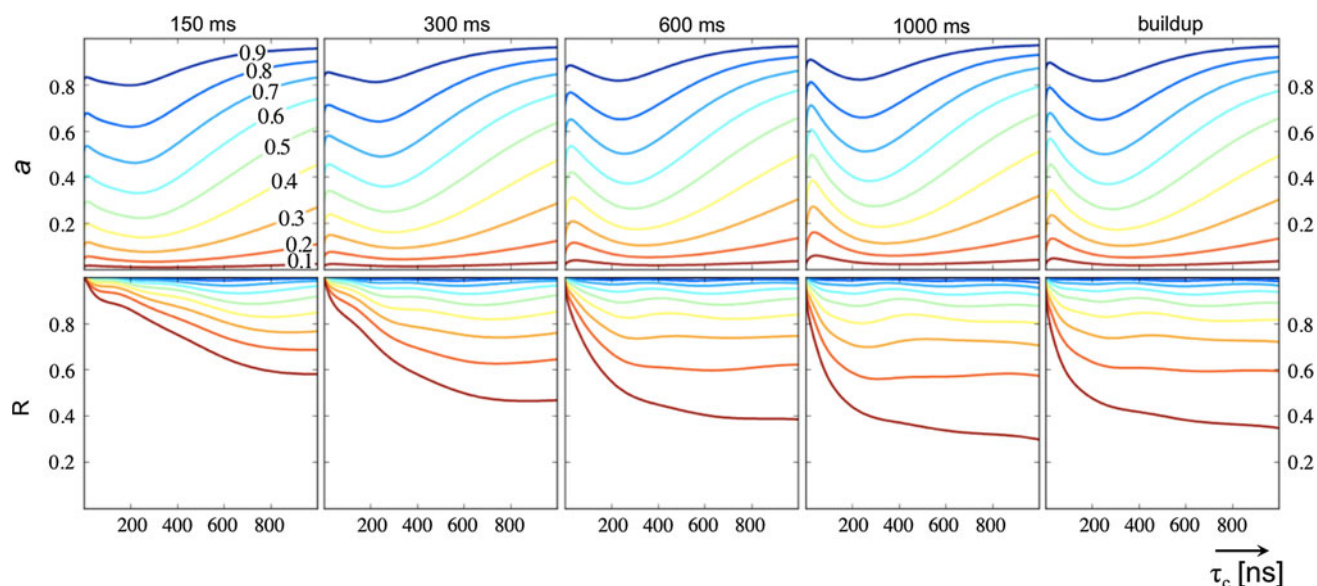


Fig. 1 Plots for the parameters of linear fits of INPHARMA NOEs calculated for the PKA/L_A and PKA/L_B complexes with uniform order parameters $S^2 < 1$ versus reference INPHARMA NOEs calculated for the rigid case with $S^2 = 1$. Slopes (*upper panel*) and Pearson correlation coefficients (*lower panel*) of best fit lines are shown in dependence of complex size (x axis, $\tau_c = 1$ –1,000 ns) and order parameter S^2 (*contour lines*) for different mixing times (*left to right*)

necessary to cause a three-fold reduction of the INPHARMA NOEs for the PKA system. To verify whether a more realistic representation of the protein and ligand dynamics would be able to explain the observed slope of 0.33 in the fit between the experimental and the back-calculated INPHARMA NOEs for the PKA/L_A and PKA/L_B complexes, we obtained estimation of order parameters in the complexes from trajectories of Molecular Dynamics (MD) simulations (Brueschweiler et al. 1992). We performed 30 ns MD simulations of the PKA/L_A and PKA/L_B complexes and extracted a set of order parameters for all proton pairs within 10 Å of the ligand binding pocket that have less than 6 Å mutual inter-nuclear distance. Due to these cutoffs, proton pairs that are not included in the calculations would either have mutual distances beyond NOE detection, or be too distant from the binding pocket for spin-diffusion mediated magnetization transfer. To remove the overall tumbling motion of the molecule, each frame was superimposed to the crystal structure as common reference frame. The order parameter was factorized in the radial and angular part according to Eq. (7), which resulted to be a good approximation for our system (correlation coefficient $R > 0.996$ between S_{kl}^2 and product $S_{r,kl}^2 \cdot S_{\Omega,kl}^2$). The order parameters extracted from the MD simulations have an average of 0.62 ± 0.22 , in good agreement with order parameters derived from MD simulations in another study (Schneider et al. 1999). Spectral density functions containing the first term of Eq. (4) were used in the full relaxation matrix to

and for the full build-up consisting of combined data from all four mixing times. *Contour lines* at values [0.1;0.9] are shown *color-coded* (red to green to blue). All combinations of INPHARMA NOEs between the groups of protons of Fig. S2 have been calculated; data are normalized to diagonal peak intensities in a NOESY spectrum at 150 ms mixing time

calculate the INPHARMA NOEs for the PKA/L_A and PKA/L_B complexes with the correct ligands orientations.

To our disappointment, the incorporation of internal motions in the INPHARMA calculations considerably deteriorated the quality of the fit between experimental and back-calculated data, yielding $R = 0.66$. However, the slope increased to 0.71, suggesting that internal motions can indeed explain the over-estimation of the INPHARMA NOEs in back-calculation performed for rigid complexes.

The decrease in the value of the theoretical INPHARMA NOEs of more than two-fold upon inclusion of internal fluctuations indicates that the internal motions primarily affecting the INPHARMA NOEs are of angular nature. A strong effect of radial fluctuations would in fact increase the rate of the NOE-transfer, according to $\langle r_{kl}^{-3} \rangle^2 \geq \langle r \rangle^{-6}$, and therefore result in a decrease of the slope of the correlation between experimental and theoretical data.

The poor fit obtained with the order parameters and distance averaging from the MD runs indicates that Molecular Dynamics is not able to reproduce the motional features of the complexes at a high level of accuracy. This is in line with the notion that obtaining accurate quantitative predictions of the NMR relaxation parameters from MD simulations is a challenging task (Trbovic et al. 2008; Markwick et al. 2008; Case 2002), despite recent improvements in force-field parameterization (Showalter et al. 2007). In general, MD simulations are able to reproduce order parameters for backbone NH and methyl group CH bond vectors measured by NMR (Showalter

and Brueschweiler 2007; Ming and Brueschweiler 2004), which contain only the angular part of Eq. (7). Also in our hands, quantitative estimation of backbone NH and methyl group order parameters works reasonably well on a benchmark set of 4 globular proteins (see below). It is therefore reasonable to assume that in this case the discrepancy between the experimental and the back-calculated data can be mainly attributed to the failure of MD simulations to reproduce the distance average $\langle r_{kl}^{-3} \rangle^2$ or $\langle r_{kl}^{-6} \rangle$. A similar conclusion is reported also in (Vogeli et al. 2009); in our case the situation is aggravated by the intermolecular ligand–protein distances, which are even more challenging to reproduce theoretically due to the much worse definition of ligand force fields with respect to protein force fields.

To verify the ability of the MD simulation to reproduce the correct distance distribution, we compared the average distances $\langle r_{kl} \rangle$ from the MD simulations with the distances extracted from the crystal structure, which can be considered a good approximation of the average state in solution and the most accurate distance information available. This comparison reveals that the correlation between $r_{kl,cryst}$ and $\langle r_{kl} \rangle$ is of poor quality for interproton distances below 6 Å ($R = 0.65$), with the MD-derived distances being consistently larger than the statistic distances in the crystal structures. This is in agreement with a recent study on perdeuterated ubiquitin (Vogeli et al. 2009), which shows that inaccuracies in order parameter estimations from MD simulations can be attributed to distance effects and MD derived distances exhibit a poor correlation to NMR distances derived from cross-relaxation measurements. In our case, if only the intermolecular distances between the protein and the ligand are considered, the quality of the correlation between $r_{kl,cryst}$ and $\langle r_{kl} \rangle$ drops even further ($R = 0.18$ for the PKA/L_A complex), which underlines the inability of MD simulations to correctly reproduce the motions of the ligand in the binding pocket.

Inaccuracies in the MD simulations, especially in the short distance range, which dominates the average, are expected to have a large effect on $\langle r_{kl}^{-6} \rangle$, while the effect on $S_{r,kl}^2 = \langle r_{kl}^{-3} \rangle^2 / \langle r_{kl}^{-6} \rangle$ is expected to be less (for example, 10 % error on the internuclear distances translates to 10 % error on the radial order parameter $S_{r,kl}^2$, but 25 % error on $\langle r_{kl}^{-6} \rangle$, as observed in numerical simulations). Therefore, while the MD-derived averaged distances $\langle r_{kl}^{-6} \rangle$ are substantially wrong, the MD-derived radial order parameters $S_{r,kl}^2 = \langle r_{kl}^{-3} \rangle^2 / \langle r_{kl}^{-6} \rangle$ are expected to be closer to the correct ones. We therefore decided to use the $S_{r,kl}^2$ order parameters derived from MD simulations, while turning to alternatives for the estimation of $\langle r_{kl}^{-6} \rangle$.

In the absence of an accurate estimation for $\langle r_{kl}^{-6} \rangle$ in Eq (4) from MD simulations, we attempted substituting $\langle r_{kl}^{-6} \rangle$

with $r_{kl,cryst}^{-6}$. This choice was made following previous work, which reported on a good correspondence between the effective distances extracted from NOE data and distances from crystal structures. In particular, (Vogeli et al. 2009) showed that for backbone NH interproton distances up to 5 Å in perdeuterated ubiquitin, the crystal structure distances are generally within 5 % of effective averaged distances extracted from NOESY NMR experiments. This result suggests that for most distances $r_{kl,cryst}^{-6}$ might be a good surrogate of $\langle r_{kl}^{-6} \rangle$, as expected for internal motions of moderate amplitude.

Using $r_{kl,cryst}^{-6}$ as a surrogate for $\langle r_{kl}^{-6} \rangle$ we are able to improve the correlation of the back-calculated INPHARMA NOEs to the experimental data ($R = 0.86$ vs. $R = 0.82$ in the static case) (Fig. 2); the slope of the linear fit reaches 0.86, which indicates that internal motions on a fast time-scale are responsible for most of the over-estimation of the INPHARMA NOEs in back-calculations using static complexes. Interestingly, setting a uniform order parameter of 0.62 (equal to the average of all order parameters extracted by the MD runs) would only result in a slope of 0.6, indicating that our set of order parameters captures specific characteristics of the interaction.

In any case, in a routine application of INPHARMA, $\langle r_{kl}^{-6} \rangle$ would not be available from MD simulations for any protein and ligand binding pose of interest, due to the expensiveness of the calculation. A feasible approach could consist in back-calculating the INPHARMA NOEs using static distances from complex structural models in combination with generic order parameters that can be applied to any protein. In view of these limitations, the fact that the experimental data for the PKA/L_A and PKA/L_B system can be best reproduced by using the crystal structure static distances and the order parameters extracted by MD simulations is encouraging. In the next paragraph we explore the possibility of defining generic order parameters that can be used for any protein/ligand complex in the INPHARMA back-calculations.

At this point it should be noted that the usage of an incorrect correlation time for the free ligands can also alter the slope of the correlation between the back-calculated and the experimental INPHARMA NOEs (Zheng and Post 1993). Therefore it is essential to optimize the parameters of the system, such as the correlation time of the complex and the correlation time of the free ligands, by fitting the intraligand transferred-NOEs before calculating the INPHARMA NOEs.

Generic order parameters

MD simulations are computationally costly and it is unfeasible to run them for every application of INPHARMA. In

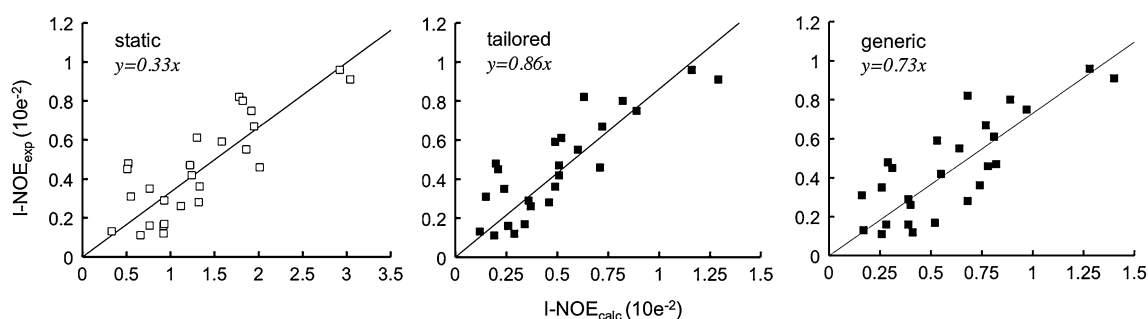


Fig. 2 Linear regression of experimental INPHARMA-NOEs ($I\text{-NOE}_{\text{exp}}$) at mixing times 300, 450, 600, and 750 ms versus simulated data ($I\text{-NOE}_{\text{calc}}$) ignoring (*left panel*) and considering (*central and right panels*) internal motions. In the *central panel*, tailored order parameters, derived from MD-simulations for the PKA, L_A , L_B system, are used; in the *right panel* generic order parameters

are used. INPHARMA cross-peak intensities are normalized to diagonal peak intensities of L_A in a NOESY spectrum at mixing time of 150 ms. Best-fit lines ($y = ax$, *black*) are plotted after performing a linear regression (*left*, $R = 0.82$, $a = 0.33$; *centre*, $R = 0.86$, $a = 0.86$, *right*, $R = 0.81$, $a = 0.73$)

addition, current force fields are designed for proteins, while simulations including ligands require manual adjustment and extension of the force field. To by-pass the demanding task of running MD simulations, it would be desirable to obtain a set of order parameters that can be easily transferred between systems.

To generate such *generic* order parameters, containing both angular and radial contributions according to Eq. (5), we follow the following approach: from MD simulations of different globular proteins, we derive order parameters for pairs of non-exchanging protons; we decompose our order parameters into contributions of individual protons (*vide infra*) and assign them to different chemical groups; last, we propose to use group-wise averaged values as *generic* order parameters that represent the motional behavior of equivalent groups in any system of interest.

In general, order parameters S_{kl}^2 are defined for pairs of nuclei (k , l). To derive generic order parameters for each proton type in a protein, the first required step is the decomposition of S_{kl}^2 in proton-specific contributions S_k and S_l such that $S_{kl}^2 \approx S_k \cdot S_l$. To find the optimal values S_k for all protons k , we minimize the l^2 -norm (Euclidean norm) of the difference $\mathbf{A} - \mathbf{x}^T \mathbf{x}$, with $A_{kl} = S_{kl}^2$, to obtain a vector \mathbf{x} with $x_k = S_k$. In other words, we decompose the pair-wise order parameter S_{kl}^2 into contributions from the single protons k and l ; by minimizing the target function for all protons at the same time, all available dynamic information about the protein is considered, i.e. for a given nucleus k , S_k contains information about all S_{kl}^2 for $k \neq l$.

Non-exchanging protons are classified according to the heavy atom they are covalently bound to, yielding 9 groups: $C\text{-}\alpha$, CH_3 , $\text{CH}_2\text{-}\beta$, $\text{CH}_2\text{-}\gamma$, $\text{CH}_2\text{-}\delta$, $\text{CH}_2\text{-}\epsilon$, CH_2 -proline, CH_1 , and aromatic protons. Each of the groups is split in two sub-groups, depending on whether the proton k belonging to that group is involved in a dipolar interaction with a proton l belonging to the same or to another residue. On average, we find that every proton has 18 neighbors

within the defined distance cutoff (5 \AA), 6 of which belong to the same residue, and the remaining 12 are in other residues (11 in non-neighboring residues). The average S -factor, calculated over the S_k for all protons inside one of the eighteen defined chemical classes, represents the *generic* S -factor for that class.

To transfer this information and assign expected motional behavior to an unknown system, we assign to each proton k of the system an atom-type, according to the grouping scheme described above, along with the corresponding *generic* S -factor S_k . The generic order parameter S_{kl}^2 defined for the dipolar interaction between k and any other proton l is obtained by multiplying the respective S -factors S_k and S_l .

We obtain the set of expected S -factors from 25 ns long MD simulations of 4 globular proteins, for which a high-resolution crystal structure is available in the Protein Data Bank (Berman et al. 2000) (human ubiquitin, PDB identifier 1ubq (Vijay-Kumar et al. 1987); human FYN tyrosine kinase SH3 domain, PDB identifier 1shf (Noble et al. 1993); the murine adipocyte lipid binding protein, PDB identifier 1lib (Xu et al. 1993); fibronectin type III domain from human tenascin, PDB identifier 1ten (Leahy et al. 1992). For the internal correlation function $C(t)^{\text{internal}}$ to be considered to converge to S^2 , we require it to be constant within a range around the plateau value for an extended time. An extensive description of how order parameters are extracted is given in the [Methods](#) section. The dynamics of each of these proteins has been investigated experimentally by NMR spectroscopy (Best et al. 2004; Constantine et al. 1998; Lee et al. 1999; Mittermaier et al. 2003) and can thus be used to benchmark our ability to estimate order parameters from MD simulations. For human ubiquitin, experimental order parameters can be reproduced well by our MD simulation ($R = 0.87$, root mean square error, $\text{RMSE} = 0.15$ for methyl groups and $R = 0.75$, $\text{RMSE} = 0.07$ for the backbone); for the human FYN tyrosine kinase SH3 domain, methyl axis order parameters are in excellent agreement with experiments ($R = 0.76$,

RMSE = 0.13), while backbone order parameters are not available; for both the murine adipocyte lipid binding protein and the fibronectin type III domain from human tenascin, methyl axis order parameters are well reproduced ($R = 0.84$, RMSE = 0.14 and $R = 0.71$, RMSE = 0.21, respectively), while the reproduction of backbone order parameters is less accurate ($R = 0.53$, RMSE = 0.14 and $R = 0.62$, RMSE = 0.05, respectively). For the murine adipocyte lipid binding proteins, the lower quality of the fit of backbone order parameters can be explained by the fact that the dynamics studies were conducted on the human protein, while the crystal structure of the human protein was not available at the time when we performed the simulations. The human and the murine proteins differ in 11 residues, which can explain the differences in the dynamics. Similarly, for the fibronectin type III domain from human tenascin, the dynamic studies were performed on a 2 amino acids longer construct (aa 1–92), while the crystal structure is on a truncated construct (aa 1–90). Hamill et al. report that the C-terminal extension has a stabilizing effect on the protein and alters its dynamics. Interestingly our studies can detect differences in the dynamic behavior of different constructs (Hamill et al. 1998).

Encouraged by the apparent good quality of the MD simulations in reproducing (angular) fluctuations, we used the sets of theoretical order parameters from the MD simulations of all four proteins to derive generalized order parameters. After decomposition of the order parameters S^2 to obtain S -factors, these S -factors are averaged for each proton class for all four proteins (Table 1). To judge the quality of the decomposition, we compute linear correlation coefficients between S_{kl}^2 and the product $S_k \cdot S_l$ for all pairs of nuclei (k, l). For the four test systems, we obtain average correlation coefficients R of 0.979 (intra-residue, RMSE 0.045–0.055) and 0.935 (inter-residue, RMSE 0.076–0.085), respectively.

The values we obtain reflect the expected dynamic behavior of a protein: C- α protons are largely static, while methyl group rotation results in low S -factors for CH_3 protons; for the different CH_2 subgroups, an increasing mobility (as reflected in decreasing S -factor) can be appreciated when moving away from the main chain (from β to ϵ). It should be noted that the angular contribution to the S -factor dominates the intra-residue values (that is, the order parameter between closer protons), while the radial contribution becomes more important for the inter-residue values (that is the order parameter between protons at longer distance) (data not shown).

Performance of generic order parameters in INPHARMA calculations

Next, we used the set of generic order parameters in the INPHARMA calculations for the PKA system. Protein

Table 1 S -factors for “inter-” and “intra-residue” proton pairs averaged over the four globular proteins (Table S1)

	Inter-residue (SD)	Intra-residue (SD)	N (inter/intra)
C- α	0.96 (0.01)	0.93 (0.03)	372/297
CH_3	0.62 (0.03)	0.56 (0.03)	611/587
CH_2 - β	0.87 (0.02)	0.81 (0.04)	366/307
CH_2 - γ	0.78 (0.04)	0.73 (0.04)	154/160
CH_2 - δ	0.69 (0.04)	0.60 (0.02)	52/67
CH_2 - ϵ	0.44 (0.04)	0.39 (0.05)	35/48
CH_2 -proline	0.82 (0.03)	0.77 (0.04)	62/64
CH_1	0.92 (0.03)	0.79 (0.06)	98/99
Aromatic	0.78 (0.02)	0.88 (0.03)	121/124

Both angular and radial fluctuations are contained in these values

N number of protons used to derive the corresponding S -factor; SD standard deviation of each value

hydrogens of PKA were assigned to the chemical groups as defined above; the ligand hydrogens were assigned to equivalent groups, as if they belonged to the protein, i.e. the “aromatic” or “ CH_2 ” group. This is an approximation; however, the lack of reliable force fields for organic ligands precludes a reliable calculation of ligand order parameters from MD simulations, leaving no other choice than using this approximation. S -factors of individual protons were re-multiplied to retrieve generic order parameters, and used in the INPHARMA calculations. Similar to what observed with the set of PKA-specific order parameters, generic order parameters result in an increased slope of the best fit line of 0.73; the correlation coefficient R of 0.81 is only slightly worse than that obtained with PKA-specific order parameters and similar to that obtained for the rigid case (Fig. 2). This result confirms the usefulness of the generic order parameter for the INPHARMA back-calculations.

Validation of calculated order parameters

To further validate our set of order parameters, we investigate its performance with respect to randomization (Fig. 3). We create different sets of random order parameters and evaluate how well these sets reproduce the experimental data in the INPHARMA calculations in comparison to the performance of the original set. Each test is iterated 1,000 times, each time using a different random set of order parameters. We perform three tests: (1) we shuffle the original set, i.e. to each pair of nuclei, we randomly assign the order parameter originally derived for another pair of nuclei; (2) we assign random order parameters drawn from a Gaussian distribution centered at 0.62 and with standard deviation 0.22; (3) we assign random order parameters drawn uniformly from [0;1]. Figure 3 shows that both the shuffled and the Gaussian set of order parameters cluster in

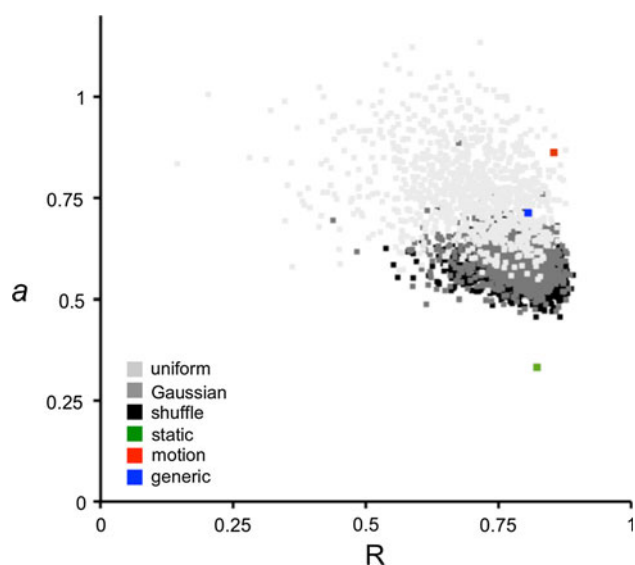


Fig. 3 Effect of randomization of the order parameters set. The INPHARMA NOEs, which are back-calculated using different sets of (randomized) order parameters are linearly fit to the experimental data. The values of the slopes a of the respective best-fit line (y axis) is plotted against the Pearson correlation coefficient R of the fit (x axis). Each dot represents a set of order parameters. The *color code* is as follows: *red*, the tailored order parameters extracted from MD simulation; *blue*, the generic set of order parameters; *green*, the *rigid case*; *black*, sets of order parameters randomized by shuffling proton pairs and order parameters; *dark gray*, sets of order parameters drawn from a Gaussian distribution with mean 0.62 and SD 0.22 to resemble the order parameter dataset extracted from MD simulation; *light gray*, a set of order parameters drawn uniformly from [0;1]

the same region, while the uniform dataset samples a wider range of quality space. Our sets of tailored and generic order parameters, however, dominate most solutions according to the Pareto criterion of multidimensional optimization. This demonstrates that the set of order parameters extracted from the MD simulation is appropriate to describe differential dynamics in the protein.

Impact of order parameters on the discriminatory power of INPHARMA

The generic order parameters can be used in INPHARMA calculations of any complex of interest. A relevant question is whether the representation of internal motions through the generic order parameters can improve the selection of the correct ligand binding modes, for example by providing an improved clear-cut discrimination in the linear fit of the experimental data to the correct or wrong binding poses.

To answer this question, we use the experimental data for the complexes PKA/L_A and PKA/L_B to select among the 16 pairs of complexes analyzed in (Orts et al. 2008). These complex pairs result from the combination of four different binding modes of L_A and four different binding modes of L_B,

which all differ from each other by 180° rotations around three orthogonal axes. Figure 4a shows the correlation coefficient and the slope of the linear fits of the experimental data versus the back-calculated data for all 16 pairs with (filled symbols) and without (empty symbols) including the generic order parameters in the INPHARMA calculations. If internal motions are ignored, four models show a correlation coefficient higher than 0.8 and therefore pass the INPHARMA selection. As explained in (Orts et al. 2008), further discrimination between the binding poses is obtained by additional criteria, such as the systematic deviation of INPHARMA peaks stemming from different structural moieties of the ligands and the semi-quantitative use of further weak INPHARMA peaks. However, when using the generic order parameters to describe the internal motions, a much better discrimination of the binding modes is achieved. Both the high correlation coefficient and the slope point to the PKA/L_A and PKA/L_B complexes pair indicated with a triangle as the one best reproducing the experimental data; these complexes correspond to the correct binding pose for both ligands, as seen in the crystal structures.

In alternative to evaluating the correlation coefficient and the slope separately, a composite quality factor of the type $[m(1 - R)^2 + n(1 - a)^2]^{-1}$ can be applied to select the pairs of binding modes that is best in agreement with the experimental data. Figure 4b shows this composite quality factor for the sixteen combinations of PKA/L_A and PKA/L_B models, both considering (y axis) and ignoring (x axis) internal motions. The better discrimination between poses achieved when including the description of internal motions through the generic order parameters in the INPHARMA back-calculations is striking (Fig. 4b). As a note of caution, we point out that the impact of the generic order parameters on INPHARMA calculations has been tested for one experimental system only (PKA/L_A and PKA/L_B complexes). A more extensive validation through experimental data for other complex systems, including proteins of different sizes, is in progress in our laboratory. In general, we expect a larger impact of order parameters in the discrimination potential of INPHARMA NOEs for proteins of smaller size (Fig. 1).

Conclusions

In this work, we present an extension of the INPHARMA method by incorporating protein plasticity in the calculations. By improving the realism of the underlying physical model, i.e. by incorporating NMR order parameters representing protein internal motions into the spectral density function, we achieve an improvement of the correlation coefficient between simulated and experimental data (0.86 as opposed to 0.82 for the rigid case) and of the slope of the

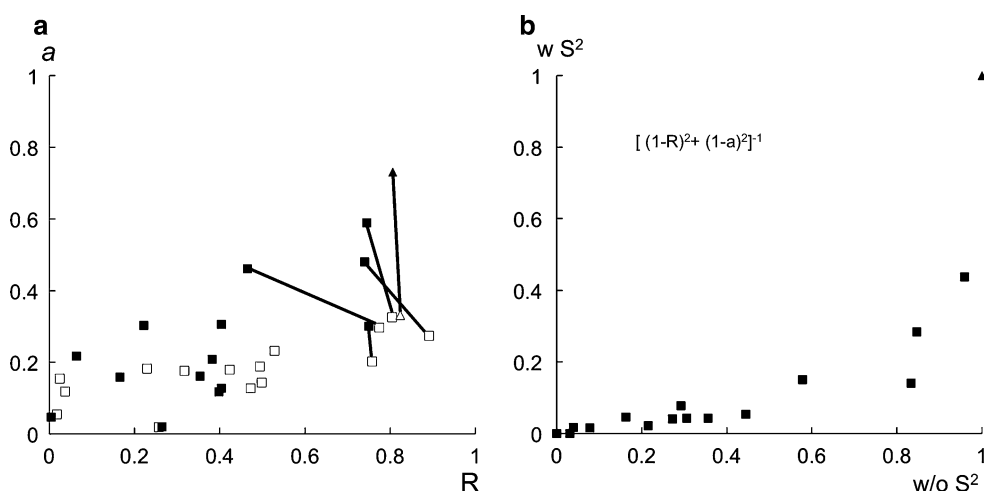


Fig. 4 Comparison of the selectivity of INPHARMA, when including (*solid symbols*) or excluding (*empty symbols*) internal motions, for the PKA/L_A and PKA/L_B complexes represented by a test set of four binding poses per ligand (yielding 16 ligands combinations). The correct pair of ligands binding poses, as seen in the crystal structures of PKA/L_A and PKA/L_B (PDB IDs 3dne and 3dnd, respectively), is indicated as *triangle*; other, incorrect, solutions as *squares*. **a** Slope of

linear regression line (0.86 as opposed to 0.33 for the rigid case). This confirms that the systematic over-estimation of magnetization transfer observed when treating the protein as rigid is largely accounted for by the use of order parameters. Importantly, we suggest generic order parameters that can be used for any experimental system irrespective of atomic coordinates. For the PKA system, we find that the use of generic order parameters to represent internal motions improves the efficacy of INPHARMA in selecting between different ligands binding poses.

Materials and methods

MD simulations and force field parameterization

Proteins and protein–ligand complexes were simulated using NAMD 2.6 (Phillips et al. 2005) and the CHARMM22 force field (MacKerell et al. 1998) in a periodic cubic box of explicit TIP3 water with a side length of the maximum internuclear distance of the respective protein atoms (~ 45 Å for 1lib, ~ 50 Å for 1shf, ~ 47 Å for 1ten, and ~ 45 Å for 1ubq), plus a padding of 25 Å to avoid mutual interaction of the protein images. Crystal structure coordinates, with hydrogens added with REDUCE (Word et al. 1999), were used as starting points. After 10^4 steps of initial energy minimization, systems were heated stepwise from 0 to 298 K with a temperature increment of 3 K per 1 ps, followed by an equilibration phase of 5 ns and the production runs. The time step of the integrator was set to 2.0 fs and

best fit line a plotted against Pearson correlation coefficient R for rigid and motional models. Equivalent solutions with $R > 0.70$ for the rigid model are connected by *black lines*. **b** Combined quality factor of motional against rigid model; Pearson correlation coefficients R and slopes of best fit line a are combined according to the formula $[m(1 - R)^2 + n(1 - a)^2]^{-1}$ with $m = n = 1$, and resulting values are normalized to the interval [0;1]

a Langevin thermostat was applied. Force field parameters of ligands (equilibrium bond lengths, angles, dihedral angles, and non-bonded interactions) were assigned analogous to known compounds as described previously (Vanommeslaeghe et al. 2010). Analogous substructures were extracted from thiazole, thiophene, histidine, indole, and aminopyridine moieties, and corresponding parameters assigned to unknown ligand parameters. After MD simulation, each frame of the trajectory was superimposed to the crystal structure conformation, minimizing protein heavy atom RMSD. Water coordinates were deleted and one snapshot per 1 ps was subjected to further analysis.

Extraction of order parameters

Distance-dependent NMR order parameters (Lipari and Szabo 1982) were calculated directly from the MD trajectories utilizing VMD (Humphrey et al. 1996) according to Eq 5 (Brueschweiler et al. 1992). Purely radial and purely angular order parameters were calculated as in Eqs. (8) and (9). (Brueschweiler et al. 1992).

Criterion for convergence of correlation functions

The existence of an order parameter S_{kl}^2 requires the internal correlation function $C_{kl}(t)$ to converge to a plateau value. After removing overall tumbling motion by superimposing each snapshot to the initial structure as a reference, normalized internal correlation functions were computed directly from the MD trajectory as:

$$C_{kl}(t_i) = c_1 \frac{1}{N-i} \sum_{j=1}^{N-i} \frac{P_2(\theta_{kl}^{lab}(t_j) \cdot \theta_{kl}^{lab}(t_{i+j}))}{r_{kl}^3(t_j) \cdot r_{kl}^3(t_{i+j})} \quad (10)$$

with $C_{kl}(t)$ containing both angular and radial fluctuations and $P_2(x) = \frac{1}{2}(3x^2 - 1)$ the second Legendre polynomial, $\theta_{kl}^{lab}/r_{kl}(t_j)$ the unit vector orientation/distance between nuclei k and l at time t_j , respectively, N the finite length of the MD trajectory and $c_1^{-1} = \frac{1}{N-i} \sum_{j=1}^{N-i} r_{kl}^{-6}(t_j)$. It should be noted that the choice of the normalization constant is arbitrary and was set to correspond to the definition of the order parameter (Eq. 5) for convenience.

For rapid internal motions, C decays rapidly to a plateau value S^2 with a characteristic internal correlation time. As MD trajectories are of finite length, estimation of $C(t)$ is not precise for $t_i \rightarrow t_N$, since $N - i \rightarrow 0$, only few snapshots contribute to the average and ergodicity cannot be assumed any longer due to the sampling problem.

For $C(t)$ to converge, we require it to stay within a certain range of a plateau value S^2 , without large fluctuations, for an extended range of t . We define an error function $\varepsilon(t) = |C(t) - S^2|$ and aim at estimating the longest interval $[t_i; t_j]$ such that the mean $\tilde{\varepsilon}_{ij} = \langle \varepsilon \rangle_{ij}$ in this time interval and the corresponding standard deviation σ_{ij} do not exceed 0.05 respectively, as well as $\tilde{\varepsilon}_{i'j'} \leq 0.05$ and $\sigma_{i'j'} \leq 0.05$ for all sub-intervals $[i'; j'] \in [i; j]^2$ with $i, j = 1 \dots N$ such that $i' < j'$. If $2(i - j) > N$, i.e. if C is close to S^2 for a consecutive time of at least half of its domain definition, we consider C to have converged.

Values of $\tilde{\varepsilon}$ and σ can efficiently be computed using dynamic programming and on-the-fly computation of means and standard deviations in a single pass. However, since we need to compute $0.5 N(N + 1)$ values of $\tilde{\varepsilon}$ and in our case N is in the range of 2.5×10^4 , with $\geq 10^4$ internal correlation functions to be examined, we decided to divide $[1; N]$ into 10^2 non-overlapping stretches of equal size, compute the averages of ε on these intervals, and use the 10^2 averages instead of the 2.5×10^4 original values for further analysis. This has the additional benefit of smoothening the data, without changing the characteristic course of a particular correlation function.

For human ubiquitin, 56.6 % of the 5,449 correlation functions C_{kl} converge, while 92.4 % of 489 individual protons considered have at least one correlation function which converges; for human FYN tyrosine kinase SH3 domain, 40.5 % of the 2,992 correlation functions converge with 89.6 % of 345 individual protons having at least one converging correlation function; for fibronectin type III domain, 51.6 % correlation functions converge out of 5295 and 93.5 % of 539 protons have at least one converging correlation function; for murine adipocyte lipid binding protein, 50.1 % of 7556 correlation functions

converge and 93.1 % of 796 individual protons have at least one converging correlation function.

Determination of generic order parameters

Distance-dependent order parameters for all pairs of non-exchanging protons with mutual distance less than 5 Å were extracted from the trajectories of four globular proteins, as described above. For the order parameter of proton pair (k, l) S_{kl}^2 , a matrix \mathbf{A} with $A_{kl} = S_{kl}^2$ was constructed. By applying a conjugant gradient algorithm in MATLAB[®] (2007a, The MathWorks, Natick, MA), a vector \mathbf{x} was determined minimizing the l^2 -norm $\|\mathbf{A} - \mathbf{x}^T \mathbf{x}\|$. This vector holds the S -factors S_k for all nuclei k , approximating $S_{kl}^2 \approx S_k \cdot S_l$ for all pairs of nuclei (k, l) , belonging to either the same residue (“intra-residue” dataset) or different residues (“inter-residue” dataset). Protons were classified as belonging to one of 9 different chemical groups, depending on the carbon they are attached to: C- α , CH₃, CH₂- β , CH₂- γ , CH₂- δ , CH₂- ε , CH₂-proline, CH₁, or aromatic. For each group, the S -factors of the protons were averaged over all pairs in the group and over four globular proteins to yield the generic S -factor for this group. To restore generic order parameters, protons were assigned generic S -factors according to their connectivity, and two S -factors were multiplied to retrieve the generic order parameter.

INPHARMA calculations

As described previously (Orts et al. 2009; Orts et al. 2008), INPHARMA NOEs between the two exchanging ligands for Protein Kinase A (PKA) are computed employing the full relaxation matrix approach (Kalk and Berendsen 1976; Keepers and James 1984; London 1999; Nilges et al. 1991) to account for all possible pathways of spin diffusion, thus allowing for rigorous, quantitative treatment of the NOE transfer. The differential equation

$$\frac{d\mathbf{M}(t)}{dt} = -(\mathbf{R} + \mathbf{K}) \cdot \mathbf{M}(t) \quad (11)$$

is solved for a given NOESY mixing time τ_M , yielding $\mathbf{M}(t)$, the magnetization matrix at time t , as

$$\mathbf{M}(\tau_M) = \exp(-(\mathbf{R} + \mathbf{K})\tau_M) \cdot \mathbf{M}(0) \quad (12)$$

The kinetic matrix \mathbf{K} represents the chemical exchange according to the kinetic model $TL_A + L_B \leftrightarrow L_A + TL_B$ with T being the target protein and L_A and L_B being the respective ligands. The relaxation matrix \mathbf{R} contains the auto-relaxation rates $R_{kk} = \rho_k$ and cross-relaxation rates $R_{kl} = \sigma_{kl}$ for all nuclei (k, l) . The spectral density function used has the form of the first term of Eq. (4), as described in the Theory section.

A more thorough theoretical treatment of INPHARMA can be found elsewhere (Orts et al. 2009).

Equation (11) was solved using a matrix exponential routine of the SciPY library in Python 2.4.3 for mixing times 150, 300, 450, 600 and 750 ms, if not stated differently. Adjustable parameters were set to $\omega = 800$ MHz, proton resonance frequency; $\tau_c = 17$ ns, correlation time of the protein; $\tau_L = 100$ ps, correlation time of the free ligands; $k_{AB} = 3,000$ s⁻¹ and $k_{BA} = 1,000$ s⁻¹ exchange rates of the respective ligands according to the kinetic model; $L_A = 450$ μ M and $L_B = 150$ μ M, respective ligand concentrations, and 25 μ M (for the NOESY experiments with 450 and 750 ms mixing time) or 30 μ M (else) protein concentration, to recapitulate the experimental setup. INPHARMA NOEs were normalized to the intensities of the diagonal peaks of L_A in a NOESY spectrum at 150 ms mixing time. Normalized INPHARMA NOEs computed for mixing times 300–750 ms were compared to normalized experimental intensities obtained at the same mixing times, and a simple linear regression was performed to yield the Pearson correlation coefficient and the slope a of the regression line $y = ax$.

Linear regression

Pearson correlation coefficients between samples x and y and slopes of the best fit line $y = ax$ were calculated as $R = \text{cov}(x,y)/\sigma_x\sigma_y$ and $a = \sum_i x_i y_i / \sum_i x_i^2$, respectively, with σ the sample standard deviation.

Acknowledgments This research was supported by the EMBL and by the DFG grant CA 294/6-1 to TC. The authors thank Frank Thommen for excellent technical assistance and Bernd Simon and John P. Overington for valuable discussions on the subject. BS is a scholar of Robinson College, University of Cambridge, and gratefully acknowledges funding by the EMBL International PhD Program.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28(1):235–242
- Best RB, Rutherford TJ, Freund SM, Clarke J (2004) Hydrophobic core fluidity of homologous protein domains: relation of side-chain dynamics to core composition and packing. *Biochemistry* 43(5):1145–1155
- Blommers MJJ, Stark W, Jones CE, Head D, Owen CE, Jahnke W (1999) Transferred cross-correlated relaxation complements transferred NOE: structure of an IL-4R-derived peptide bound to STAT-6. *J Am Chem Soc* 121:1949–1953

- Brueschweiler R, Roux B, Blackledge M, Griesinger C, Karplus M, Ernst RR (1992) Influence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution. *J Am Chem Soc* 114(7):2289–2302
- Carlomagno T (2005) Ligand-target interactions: what can we learn from NMR? *Annu Rev Biophys Biomol Struct* 34:245–266
- Carlomagno T, Felli IC, Czech M, Fischer R, Sprinzl M, Griesinger C (1999) Transferred cross-correlated relaxation: application to the determination of sugar pucker in an aminoacylated tRNA-mimetic weakly bound to EF-Tu. *J Am Chem Soc* 121(9):1945–1948. doi:10.1021/ja9835887
- Case DA (2002) Molecular dynamics and NMR spin relaxation in proteins. *Acc Chem Res* 35(6):325–331
- Constantine KL, Friedrichs MS, Wittekind M, Jamil H, Chu CH, Parker RA, Goldfarb V, Mueller L, Farmer BT II (1998) Backbone and side chain dynamics of uncomplexed human adipocyte and muscle fatty acid-binding proteins. *Biochemistry* 37(22):7965–7980
- Ernst RR, Bodenhausen G, Wokaun A (1987) Principles of NMR in one and two dimensions. Clarendon Press, Oxford
- Hamill S, Meekhof A, Clarke J (1998). The effect of boundary selection on the stability and folding of the third fibronectin type III domain from human tenascin. *Biochemistry* 37(22):8071–8079
- Hub JS, de Groot BL (2009) Detection of functional modes in protein dynamics. *PLoS Comput Biol* 5(8):e1000480
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14(1):33–38, 27–38
- Jayalakshmi V, Krishna NR (2002) Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand–receptor complexes. *J Magn Reson* 155(1):106–118
- Kalk A, Berendsen HJC (1976) Proton magnetic-relaxation and spin diffusion in proteins. *J Magn Reson* 24(3):343–366. doi:10.1016/0022-2364(76)90115-3
- Karplus M, Kuriyan J (2005) Molecular dynamics and protein function. *Proc Natl Acad Sci U S A* 102(19):6679–6685
- Kay LE, Muhandiram DR, Wolf G, Shoelson SE, Forman-Kay JD (1998) Correlation between binding and dynamics at SH2 domain interfaces. *Nat Struct Biol* 5(2):156–163
- Keepers JW, James TL (1984) A theoretical-study of distances determination from NMR—two-dimensional nuclear overhauser effect spectra. *J Magn Reson* 57(3):404–426. doi:10.1016/0022-2364(84)90257-9
- Kohn JE, Afonine PV, Ruscio JZ, Adams PD, Head-Gordon T (2010) Evidence of functional protein dynamics from X-ray crystallographic ensembles. *PLoS Comput Biol* 6(8):1–5
- Leahy DJ, Hendrickson WA, Aukhil I, Erickson HP (1992) Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* 258(5084):987–991
- Lee AL, Flynn PF, Wand AJ (1999) *J Am Chem Soc* 121:2891–2902
- Lipari G, Szabo A (1982) *J Am Chem Soc* 104:4546
- London RE (1999) Theoretical analysis of the inter-ligand overhauser effect: a new approach for mapping structural relationships of macromolecular ligands. *J Magn Reson* 141(2):301–311
- MacKerell AD et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616
- Markwick PR, Malliavin T, Nilges M (2008) Structural biology by NMR: structure, dynamics, and interactions. *PLoS Comput Biol* 4(9):e1000168
- Mayer M, Meyer B (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J Am Chem Soc* 123(25):6108–6117

- McCoy MA, Wyss DF (2002) Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. *J Am Chem Soc* 124(39):11758–11763
- Ming D, Brueschweiler R (2004) Prediction of methyl-side chain dynamics in proteins. *J Biomol NMR* 29(3):363–368
- Mittermaier A, Kay LE (2006) New tools provide new insights in NMR studies of protein dynamics. *Science* 312(5771):224–228
- Mittermaier A, Davidson AR, Kay LE (2003) Correlation between ²H NMR side-chain order parameters and sequence conservation in globular proteins. *J Am Chem Soc* 125(30):9004–9005
- Ni F (1994) Recent developments in transferred NOE methods. *Prog NMR Spectrosc* 26:517–606
- Nilges M, Habazettl J, Brunger AT, Holak TA (1991) Relaxation matrix refinement of the solution structure of squash trypsin inhibitor. *J Mol Biol* 219(3):499–510
- Noble ME, Musacchio A, Saraste M, Courtneidge SA, Wierenga RK (1993) Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin. *EMBO J* 12(7):2617–2624
- Orts J, Tuma J, Reese M, Grimm SK, Monecke P, Bartoschek S, Schiffer A, Wendt KU, Griesinger C, Carlomagno T (2008) Crystallography-independent determination of ligand binding modes. *Angew Chem Int Ed Engl* 47(40):7736–7740
- Orts J, Griesinger C, Carlomagno T (2009) The INPHARMA technique for pharmacophore mapping: a theoretical guide to the method. *J Magn Reson* 200(1):64–73
- Orts J, Bartoschek S, Griesinger C, Monecke P, Carlomagno T (2012) An NMR-based scoring function improves the accuracy of binding pose predictions by docking by two orders of magnitude. *J Biomol NMR* 52(1):23–30
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802
- Reese M, Sanchez-Pedregal V, Kubicek K, Meiler J, Blommers M, Griesinger C, Carlomagno T (2007) Structural basis of the activity of the microtubule-stabilizing agent epothilone A studied by NMR spectroscopy in solution. *Angew Chem Int Ed Engl* 46(11):864–1868
- Sanchez-Pedregal VM, Reese M, Meiler J, Blommers MJ, Griesinger C, Carlomagno T (2005) The INPHARMA method: protein-mediated interligand NOEs for pharmacophore mapping. *Angew Chem Int Ed Engl* 44(27):4172–4175
- Schneider TR, Brunger AT, Nilges M (1999) Influence of internal dynamics on accuracy of protein NMR structures: derivation of realistic model distance data from a long molecular dynamics trajectory. *J Mol Biol* 285(2):727–740
- Showalter SA, Brueschweiler R (2007) Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: an application to the AMBER99SB force field. *J Chem Theory Comput* 3(3):961–975
- Showalter SA, Johnson E, Rance M, Bruschweiler R (2007) Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. *J Am Chem Soc* 129(46):14146–14147
- Trbovic N, Kim B, Friesner RA, Palmer AG III (2008) Structural analysis of protein dynamics by MD simulations and NMR spin-relaxation. *Proteins* 71(2):684–694
- Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol* 194(3):531–544
- Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 31(4):671–690
- Vogeli B, Segawa TF, Leitz D, Sobol A, Choutko A, Trzesniak D, van Gunsteren W, Riek R (2009) Exact distances and internal dynamics of perdeuterated ubiquitin from NOE buildups. *J Am Chem Soc* 131(47):17215–17225
- Wallach D (1967) *J Chem Phys* 47:5258
- Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* 285(4):1735–1747
- Xu Z, Bernlohr DA, Banaszak LJ (1993) The adipocyte lipid-binding protein at 1.6-Å resolution. Crystal structures of the apoprotein and with bound saturated and unsaturated fatty acids. *J Biol Chem* 268(11):7874–7884
- Zheng J, Post CB (1993) Protein indirect relaxation effects in exchange-transferred NOESY by a rate-matrix analysis. *J Magn Reson* 101:262–270