

# Application of the random coil index to studying protein flexibility

Mark V. Berjanskii · David S. Wishart

Received: 27 August 2007 / Accepted: 10 October 2007 / Published online: 6 November 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** Protein flexibility lies at the heart of many protein–ligand binding events and enzymatic activities. However, the experimental measurement of protein motions is often difficult, tedious and error-prone. As a result, there is a considerable interest in developing simpler and faster ways of quantifying protein flexibility. Recently, we described a method, called Random Coil Index (RCI), which appears to be able to quantitatively estimate model-free order parameters and flexibility in protein structural ensembles using only backbone chemical shifts. Because of its potential utility, we have undertaken a more detailed investigation of the RCI method in an attempt to ascertain its underlying principles, its general utility, its sensitivity to chemical shift errors, its sensitivity to data completeness, its applicability to other proteins, and its general strengths and weaknesses. Overall, we find that the RCI method is very robust and that it represents a useful addition to traditional methods of studying protein flexibility. We have implemented many of the findings and refinements reported here into a web server that allows facile, automated predictions of model-free order parameters, MD RMSF and NMR RMSD values directly from backbone  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shift assignments. The server is available at <http://wishart.biology.ualberta.ca/rci>.

**Keywords** NMR · Chemical shift · Protein · Flexibility · Order parameters

## Introduction

Over the past 30 years, NMR spectroscopy has emerged as one of the most useful tools to investigate protein flexibility due to its remarkable ability to provide site-specific information about protein motions over a large range of time scales. In particular, residue-specific measurements of hydrogen exchange (spanning periods of seconds to hours), conformational exchange ( $\mu\text{s}$  to  $\text{ms}$ ) and order parameters (ps to ns) can be performed via NMR (Kay 1998). These methods have proven to be very robust and are now routinely used in most NMR studies of protein dynamics (Ishima and Torchia 2000). However, all of these approaches typically require conducting additional non-trivial NMR experiments followed by extensive and complex data analysis (Lacroix et al. 1997; Palmer 2001). This makes many NMR-derived dynamic measurements a relatively tedious and error-prone process. Recently, we described a simple approach for measuring protein flexibility that doesn't require any more information than backbone chemical shift assignments. The method, called the Random Coil Index (RCI), is based on an empirically derived relationship between secondary chemical shifts and protein mobility (Berjanskii and Wishart 2005). Most importantly, the RCI can be quantitatively related to other standard measures of protein motions such as model-free order parameters and per-residue root mean square fluctuations (RMSF) calculated from molecular dynamics simulations and per-residue root mean square deviations (RMSD) calculated from NMR ensembles.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10858-007-9208-0) contains supplementary material, which is available to authorized users.

M. V. Berjanskii · D. S. Wishart  
Department of Computing Science, University of Alberta,  
Edmonton, AB, Canada

D. S. Wishart (✉)  
NRC National Institute for Nanotechnology, University of  
Alberta, 2-21 Athabasca Hall, T6G 2E8 Edmonton, AB, Canada  
e-mail: david.wishart@ualberta.ca

The original RCI paper, which was published as a brief communication, presented the concept as an empirically driven hypothesis with little formal justification or rationale. At the time, relatively few details could be provided about time-scales, limitations, advantages and minimal data requirements for this method. Since then, we have undertaken a far more detailed study of the RCI concept and have developed a more complete understanding of how chemical shifts can (and cannot) be used to monitor protein flexibility. In the work presented here, we (1) explain how the RCI was derived and why it works; (2) demonstrate the efficacy and accuracy of the RCI on a much larger set of proteins; (3) assess the sensitivity of the RCI method to chemical shift referencing errors, random coil reference values and sequential corrections; (4) establish the minimal data requirement for reliable RCI predictions; (5) ascertain the time-scale of RCI-detected motions; (6) identify the theoretical limitations of the method, and (7) show the advantages of the RCI method over existing methods used to study protein dynamics. We believe that a detailed investigation of these issues is critical to establish the utility and legitimacy of the RCI concept within the NMR community. Furthermore, in an effort to encourage greater use of the RCI method in routine protein analysis, we created a freely accessible RCI web server that allows facile, automated predictions of model-free order parameters, MD RMSF and NMR RMSD values directly from backbone  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shift assignments (available at <http://www.wishart.biology.ualberta.ca/rci>).

## Methods

### RCI derivation and principles

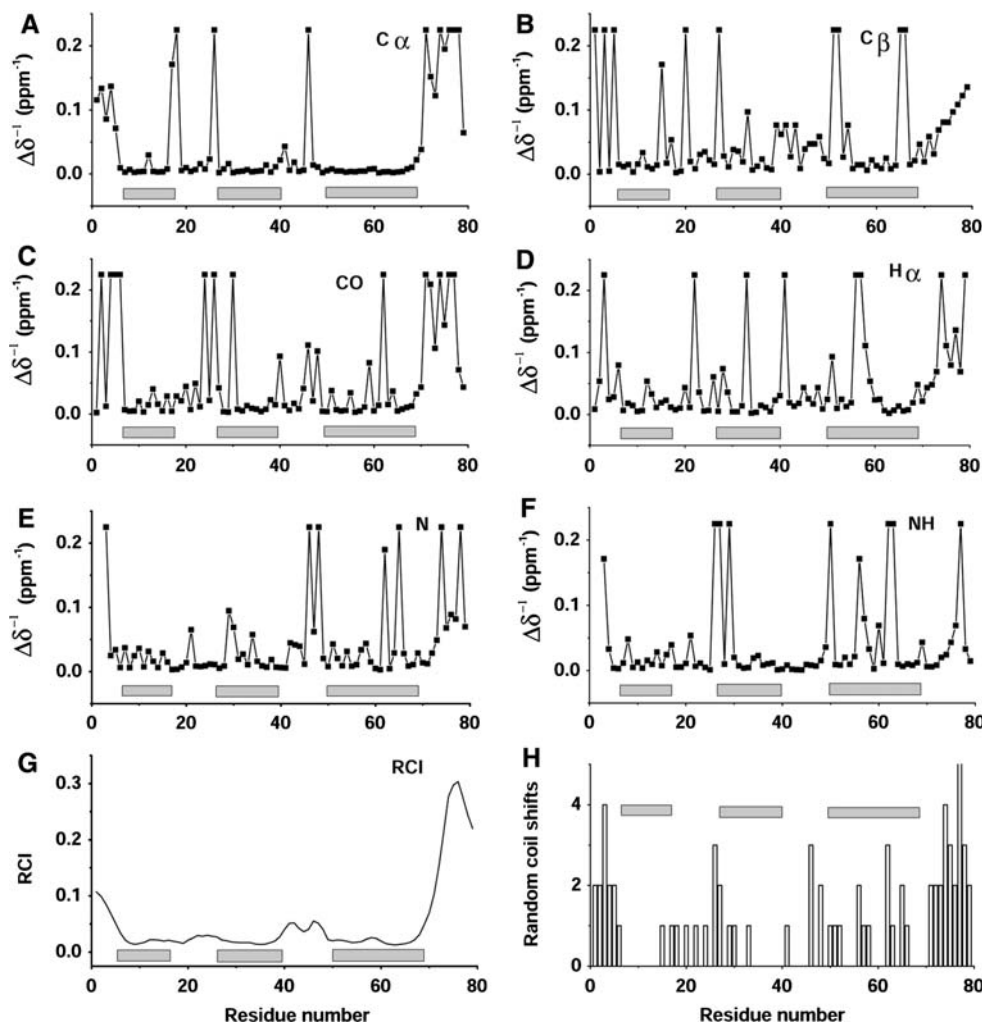
The connection between random coil chemical shifts and protein flexibility has been known for several decades (Grathwoh and Wuthrich 1974; Bundi et al. 1975; Wishart et al. 1991; Wishart and Sykes 1994a, b). Indeed, random coil chemical shifts can be formally defined as shifts that result from a fast conformational exchange among energy-weighted populations of all theoretically possible conformations of an unfolded polypeptide chain in the absence of long-range inter-residue interactions (Bundi and Wuthrich 1979; Vila et al. 2002). As the structure and mobility of a protein segment approaches the random coil state, the chemical shifts of its atoms tend to move toward their corresponding random coil values. In fact, the proximity of amino acid chemical shifts to their random coil values has been used by many research groups to qualitatively estimate the level of protein structural disorder (Lecroisey et al. 1997; Chou et al. 2002; Fiaux et al. 2002). In addition to using chemical shifts to approximately assess

protein flexibility, chemical shifts can also be used to identify regions displaying order or segmental rigidity. Over the past two decades, several approaches have been developed to identify rigid secondary structure elements from secondary chemical shifts and these methods are now routinely applied to the analysis of protein structure (Dalgarno et al. 1983; Pardi et al. 1983; Wishart et al. 1992; Wishart and Sykes 1994a, b; Wang and Jardetzky 2002a, b; Eghbalnia et al. 2005).

The application of secondary chemical shifts to characterize protein flexibility is based on an assumption that the close proximity of chemical shifts to random coil values is a manifestation of increased protein mobility, while significant differences from random coil values is an indication of a relatively rigid structure. Indeed, an intriguing correlation between secondary chemical shifts ( $\text{Hz}$ ) and motional amplitudes (as implied from X-Ray B-factors) was first demonstrated for *E. coli* thioredoxin (Wishart et al. 1991) and, later, for ubiquitin (Wishart and Sykes 1994a, b) more than a decade ago. However, the widespread acceptance of secondary chemical shifts as a tool for predicting protein flexibility was hampered by the fact that the aforementioned assumption for a single type of chemical shift (e.g.,  $\text{C}\alpha$ ) would not always hold. Indeed, chemical shift hypersurfaces (Wishart and Nip 1998; Xu and Case 2002; Wang and Jardetzky 2004) and published distributions of secondary chemical shifts in secondary structure elements (Spera and Bax 1991; Wang and Jardetzky 2002a, b) suggest that the nuclei of certain amino-acid residues may have random coil-like shifts as a consequence of particular combinations of  $\phi$  and  $\psi$  angles regardless of a given residue's dynamic properties. For instance, inspection of the inverse absolute secondary chemical shifts of a small protein, the J domain of polyomavirus T antigen (Berjanskii et al. 2000; Berjanskii et al. 2002) reveals the presence of random coil shifts among certain nuclei even in rigid  $\alpha$ -helical regions (Fig. 1a–f).

Our initial attempts to overcome this problem indicated that, by combining the chemical shifts from multiple nuclei ( $\text{C}\alpha$ ,  $\text{C}\beta$ , CO, N, NH, and  $\text{H}\alpha$ ) into a single parameter, one is able to decrease the level of unwanted “random coil shift noise” (Fig. 1g). The improved performance originates from the different probabilities of random coil chemical shifts from different nuclei being found among amino acid residues in flexible regions versus rigid regions. Typically, residues in rigid helices or rigid  $\beta$ -strands are less likely to have more than one random coil chemical shift among their backbone and  $\text{C}\beta$  shifts than residues in mobile regions (Fig. 1h). After testing a variety of mathematical expressions, it was determined that the simplest expression for combining chemical shifts in a way that

**Fig. 1** Random coil chemical shifts and Random Coil Index of PyJ. Helical regions are shown with gray bars. (a–f) Inverse absolute secondary chemical shifts ( $\Delta\delta^{-1}$ ) of C $\alpha$ , C $\beta$ , CO, H $\alpha$ , N and NH. Carbon, nitrogen and proton absolute secondary chemical shifts were scaled by 18.75, 7.5 and 75.0, respectively. High values of  $\Delta\delta^{-1}$  correspond to chemical shifts that approach their corresponding random coil values. (g) Random Coil Index that was calculated using Eq. 1. (h) Per-residue occurrence of random coil chemical shifts of C $\alpha$ , C $\beta$ , CO, H $\alpha$ , N and NH. If after aforementioned scaling of secondary chemical shifts, absolute  $\Delta\delta^{-1}$  above 0.1 ppm $^{-1}$  Chemical shifts were considered “random-coil”-like if, after aforementioned scaling of secondary chemical shifts, absolute  $\Delta\delta^{-1}$  values were above 0.1 ppm $^{-1}$



correlated well with motional amplitudes was the inverse of the averaged, weighted secondary backbone shifts. We named this parameter the Random Coil Index (or RCI) since it quantitatively tracks the relative degree to which a protein segment matches the random coil state.

The actual calculation of the RCI involves several additional steps including the smoothing of secondary shifts over several adjacent residues, the use of neighboring residue corrections, chemical shift re-referencing, gap filling, chemical shift scaling and numeric adjustments to prevent divide-by-zero problems.  $^{13}\text{C}$ ,  $^{15}\text{N}$  and  $^1\text{H}$  secondary chemical shifts are then scaled to account for the characteristic resonance frequencies of these nuclei and to provide numeric consistency among different parts of the protocol. A detailed description of these steps was published elsewhere (Berjanskii and Wishart 2006). Once these scaling corrections have been done, the RCI is calculated using the following expression:

$$\text{RCI} = [(A|\Delta\delta_{\text{C}\alpha}| + B|\Delta\delta_{\text{CO}}| + C|\Delta\delta_{\text{H}\beta}| + D|\Delta\delta_{\text{N}}| + E|\Delta\delta_{\text{NH}}| + F|\Delta\delta_{\text{H}\alpha}|)/n]^{-1} \tag{1}$$

where  $|\Delta\delta_{\text{X}\alpha}|$ ,  $|\Delta\delta_{\text{XO}}|$ ,  $|\Delta\delta_{\text{X}\beta}|$ ,  $|\Delta\delta_{\text{N}}|$ ,  $|\Delta\delta_{\text{NH}}|$ , and  $|\Delta\delta_{\text{H}\alpha}|$  are the absolute scaled values of the secondary chemical shifts (in ppm) of C $\alpha$ , CO, C $\beta$ , N, NH and H $\alpha$  respectively. A, B, C, D, E, and F are nucleus-specific weighting coefficients (Supplemental Table 1) and  $n$  is the number of chemical shift types (e.g., 6 if all six types of chemical shifts are used to calculate RCI). When all 6 chemical shifts are available, the RCI is calculated as

$$\text{RCI} = [(0.74|\Delta\delta_{\text{C}\alpha}| + 0.72|\Delta\delta_{\text{CO}}| + 0.13|\Delta\delta_{\text{H}\beta}| + 0.38|\Delta\delta_{\text{N}}| + 0.15|\Delta\delta_{\text{NH}}| + 0.91|\Delta\delta_{\text{H}\alpha}|)/6]^{-1} \tag{2}$$

The “end-effect correction” can also be applied at this point. The last step of the protocol involves smoothing the

initial set of RCI values by three-point averaging (Berjanskii and Wishart 2006).

#### Optimization and validation of RCI with a larger data set

Originally, the RCI method was trained on a set of just 14 proteins [Berjanskii 2005, (1529)]. Because of concerns that this data set may have been too small and that it incompletely sampled the full dynamic or structural range found in proteins, we enlarged the original training set of proteins to 28 well-resolved proteins with complete  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  backbone assignment to ascertain the robustness and extensibility of the RCI method (Supplemental Table 2). The new training set consisted of 2988 residues, spanning a range of sizes (46–203 residues) and including all major protein fold classes (all  $\alpha$ , all  $\beta$ , mixed  $\alpha/\beta$ ). This set also contained proteins with widely varying flexibility content (a flexible residue is defined as one with an order parameter less than 0.85) covering between 15% and 65% of the protein length (Supplemental Table 3). To obtain detailed, residue-specific information on the backbone mobility of these proteins, we calculated a set of 4-ns MD trajectories for each protein using GROMACS 3.2.1 (Lindahl et al. 2001) and GROMOS96 43a1 force field (Scott et al. 1999). Details of the MD simulation protocol have been published elsewhere (Berjanskii and Wishart 2005). The full set of MD simulations conducted for the study reported here required more than 8,000 CPU hours in total on a cluster of 840 3.0 GHz Xeon processors (WestGrid supercomputer, Canada). Residue-specific amide nitrogen RMSFs were calculated for each protein trajectory with GROMACS. These RMSF values served as a proxy measure of that protein's backbone mobility. Weighting coefficients were determined by optimizing the correlation between the calculated RCI values and the MD RMSF using a simple grid search.

#### Assessing minimal data requirements and the influence of random chemical shift reference values and nearest neighbor corrections

Chemical shift data from peptides and protein is often “noisy”, with missing assignments or problems from improper chemical shift referencing. There are also a large number of methods for correcting or calibrating chemical shifts and comparing them to different sets of random coil (or reference) shifts. Some of these are known to work well, while others do not. We undertook a number of comparative studies to assess how these chemical shift effects affect the RCI accuracy and to optimize the

performance of the RCI method so that it would be more robust to this chemical shift “noise”.

Note that the RCI method was originally developed to deal only with proteins having a full set of 6 types of backbone chemical shift assignments ( $C\alpha$ , CO,  $C\beta$ , N, NH and  $H\alpha$ ). However, not all proteins are routinely assigned to such an extent. Consequently, we decided to investigate the level to which the RCI method could perform in situations where assignments of some of these nuclei are missing. The effect of excluding one or more types of nuclei was evaluated by calculating the correlation between the RCI (generated without a particular set of shifts) and its corresponding MD RMSF for all 28-test proteins. For each of the 63 different chemical shift combinations, we optimized the RCI weighting coefficients and calculated the mean correlation coefficients using a simple grid search strategy.

To assess the influence of the choice of random coil reference shifts and nearest neighbor corrections on the accuracy of RCI-predicted protein flexibility, we tested eight combinations of four sets of random coil shift values (Wishart et al. 1995; Lukin et al. 1997; Schwarzinger et al. 2000; Wang and Jardetzky 2002a, b) and two sets of  $i \pm 1$  neighboring residue corrections (Schwarzinger et al. 2001; Wang and Jardetzky 2002a, b). In all cases,  $i \pm 2$  neighboring correction values published by Schwarzinger et al. (2001) were used. Note that the reference chemical shifts published by Lukin et al. (1997) lack proton chemical shifts and were supplemented with the  $H\alpha$  and NH chemical shifts generated by Wang and Jardetzky (2002a, b). Weighting coefficients in the RCI equation (for each combination of random coil values and nearest neighbor correction factors) were optimized against the correlation coefficient calculated between the RCI and MD RMSF values using 33 proteins (Supplemental Table 2).

#### The RCI web server

Results from the performance optimizations, reference corrections, missing data compensation, automatic renormalization, and nearest neighbor corrections described in the preceding pages were implemented into both a stand-alone program and a publicly available web server, called the RCI web server. The stand-alone program, which generates results identical to the server, was used to generate the data shown in the following pages. The RCI web server, which is designed for general access and single protein queries, accepts both SHIFTY and BMRB NMR-STAR formatted chemical shift files as input and provides both text files and graphical plots of the RCI values, predicted MD RMSF, NMR RMSD and order parameters as output. Additionally, the simple-to-use interface allows

users to select the set of random coil reference values, the type of nearest-neighbor residue corrections, the method of chemical shift reference correction, the treatment of end-effects and the treatment of assignment gaps. The backend for the RCI web server is written in Python, while the graphical user interface is coded in Python and HTML. The server is available at <http://www.wishart.biology.ualberta.ca/rci>.

## Results and discussion

### RCI performance and minimal data requirements

A leave-one-out strategy was employed to test the performance of the RCI algorithm (and RCI server) in predicting protein flexibility. In particular, when the RCI values were calculated for each protein in the test set, the chemical shifts for that protein were excluded from the test set. This avoids problems of over-fitting and prevents any bias in the results. The average correlation coefficient between the RCI and MD RMSF was 0.81 (identical to that obtained using the whole data set). To ensure that the good correlation was not a result of over-fitting, another 5 proteins, not included in the grid search and spanning a range of sizes from 106 to 286 residues (830 residues in total), were analyzed (Supplemental Table 2, italics). The average correlation between RCI and the MD RMSF values of these proteins was also 0.81. Note that nearly identical results ( $r = 0.82$ ) were obtained with the smaller set of 14 proteins used in the original RCI paper. This suggests that the RCI method is robust and extensible to any other protein.

The minimal data requirements for optimal performance of the RCI program (and server) were assessed for each of the 63 possible chemical shift combinations. The results, which are too extensive to present here, are available in Supplemental Table 1. This table includes weighting coefficients and calculated the mean correlation coefficients for each chemical shift combination. Nevertheless, some interesting trends are worth noting. In particular, the omission of  $C\beta$ , NH or N shifts from the RCI calculation had a minimal effect on the correlation between the per-residue RCI and MD RMSF values. In fact, the correlation coefficient dropped by only 0.04. On the other hand, including only  $C\beta$ , NH and N shifts in the RCI calculations decreased the average correlation coefficient by 0.17. This result is consistent with the smaller weighting coefficients for  $C\beta$ , NH and N shifts seen in the RCI expression (Eq. 2, Supplemental Table 1). It is also consistent with the smaller correlation coefficient between MD RMSF values and single-atom RCI values for these nuclei (Supplemental Table 1).

The low correlation between protein flexibility and the RCI derived from NH or N shifts could be partially explained by the high sensitivity of NH and N shifts to long-range shielding/de-shielding determinants, such as ring currents and local charges; see (Szilagyi 1995) for a review. Furthermore, the sensitivity of NH and N shifts to experimental conditions, such as temperature, pH, ionic strength and certain buffer components (Szilagyi 1995), may also be responsible for contributing to the disagreement between their secondary chemical shifts and calculated protein flexibility.

The poor correlation between  $C\beta$  secondary chemical shifts and backbone dynamics can potentially be explained by the sensitivity of  $C\beta$  shifts to side-chain mobility. For instance, calculations of Glu chemical shifts using density functional theory demonstrated that fluctuations of the  $\chi_3$  angle can result in significant changes to the  $C\beta$  chemical shift (4 ppm) while having only a modest effect (<1 ppm) on  $C\alpha$ , CO, and  $H\alpha$  shifts (Xu and Case 2002). The same study revealed that nitrogen chemical shifts, which also correlate poorly with backbone mobility, can vary as much as 2 ppm due to  $\chi_3$  fluctuations. Given that side chain mobility is often completely uncoupled from backbone motions (Wand 2001), one would expect nuclei that are particularly sensitive to side chain motions to be poorer predictors of backbone motions. Weighting coefficients optimized for each possible combination of assigned nuclei have been integrated into the latest version of the RCI web server. These weighting coefficients were also scaled for each chemical shift combination in order to make the RCI value (Eq. 1) independent from the number of chemical shifts used in the calculation. To achieve this, weighting coefficients for every combination of chemical shifts were uniformly scaled to make the average weighting coefficient equal to the average of weighting coefficients for the six-shift RCI (i.e., when  $C\alpha$ , CO,  $C\beta$ , N, NH and  $H\alpha$  are used). As a result, users will be able to obtain comparable RCI values independent of the number of chemical shift assignments used. However, we recommend using, at least,  $C\alpha$ ,  $H\alpha$ , and CO shifts to obtain accurate predictions from the RCI server.

### Influence of random coil shift values and neighboring residue corrections

It was shown recently that the outcome of secondary structure identification by chemical shifts may depend critically on the choice of random coil reference values (Mielke and Krishnan 2004). We investigated how the choice of random coil reference shifts and nearest neighbor corrections affects the accuracy of RCI-predicted protein flexibility. Several different sets of



statistically derived (Wishart et al. 1991; Wishart and Sykes 1994a, b; Lukin et al. 1997; Wang and Jardetzky 2002a, b; Wang and Jardetzky 2002a, b) and experimentally derived (Richarz and Wuthrich 1978; Bundi and Wuthrich 1979; Glushka et al. 1989; Braun et al. 1994; Thanabal et al. 1994; Merutka et al. 1995; Wishart et al. 1995; Schwarzingler et al. 2000) random coil shifts have been published over the past two decades. The experimentally derived sets differ from each other due to a number of experimental design issues, including nearest neighbor effects, the choice of chemical shift referencing methods, the influence of end effects and experimental conditions (temperature, pH, buffer composition and ionic strength). Statistically derived sets of random coil chemical shifts differ from each other primarily due to the size of their reference databases (currently ranging from 36 to 415 proteins), from which they were derived. To date, only one of the statistical sets (Wang and Jardetzky 2002a, b) includes any form of nearest-neighbor residue correction, while two experimentally derived sets include neighboring residue corrections (Wishart et al. 1995; Schwarzingler et al. 2001).

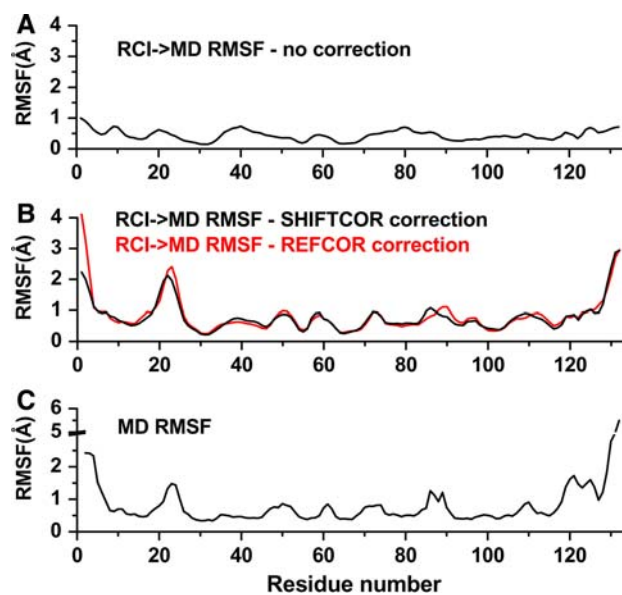
As noted before, eight combinations of four sets of random coil shift values (Wishart et al. 1995; Lukin et al. 1997; Schwarzingler et al. 2000; Wang and Jardetzky 2002a, b) and two sets of  $i \pm 1$  neighboring residue corrections (Schwarzingler et al. 2001; Wang and Jardetzky 2002a, b) were used to assess the performance of the RCI program. In all cases,  $i \pm 2$  neighboring correction values published by Schwarzingler et al. (2001) were used. The results from this analysis revealed that the mean correlation coefficients vary over a very small range (between 0.79 and 0.81) for every combination of reference random coil chemical shift values and  $i \pm 1$  neighboring corrections (Supplemental Table 4). This result demonstrates that RCI-derived flexibility is relatively insensitive to the differences among published reference random chemical shifts and nearest neighbor sequential corrections. The RCI web server utilizes the random coil shifts published by Schwarzingler et al. (2001) as the default random coil values because this is currently the only set of reference shifts for which both  $i \pm 1$  and  $i \pm 2$  neighboring corrections were determined. However, the RCI web server allows a user to choose any of the eight combinations of random coil values and  $i \pm 1$  nearest-neighbor corrections.

#### Effect of different methods to correct mis-referenced chemical shifts

As noted earlier, secondary shifts ( $\Delta\delta$ ) depend not only on the choice of random coil chemical shifts or the inclusion of nearest neighbor corrections, but also on the accuracy of

the reported chemical shifts (i.e., correct chemical shift referencing). Chemical shift referencing, particularly for  $^{13}\text{C}$  and  $^{15}\text{N}$  shifts, continues to be a problem with about 20% of newly deposited shifts being referenced in a non-standard way (Zhang et al. 2003). To ensure uniformity in chemical shift referencing for both the test and training set of proteins used in the RCI calculations, we re-referenced all 33 sets of chemical shifts using the structure-based SHIFTCOR protocol (Zhang et al. 2003). This kind of reference correction was critical in developing and refining the RCI protocol, but it also led to two important questions: (1) What are the consequences of using mis-referenced chemical shifts in an RCI calculation? (2) How can the RCI server correct chemical shift referencing problems in the absence of a 3D structure?

To measure the effect of chemical shift mis-referencing, we selected several proteins from our training and testing sets that were significantly mis-referenced (as determined by SHIFTCOR). One example is the SV40 T Antigen DNA-Binding Domain (BMRB ID: 4127, Fig. 2). In this case, SHIFTCOR detected that the  $^{13}\text{C}$  and  $^{15}\text{N}$  shifts required referencing corrections of between 2.4 and 4.7 ppm (the CO, C $\alpha$ , C $\beta$  and N chemical shifts had to be adjusted by 3.3, 3.1, 2.4, and 4.7 ppm, respectively). Implementing these reference corrections resulted in a significantly improved correlation (from 0.45 to 0.72)



**Fig. 2** Effect of spectrum mis-referencing and reference corrections on Random Coil Index of SV40 T antigen DNA-binding domain. (a) MD RMSF predicted from RCI using NMR assignments of SV40 T antigen DNA-binding domain (BMRB ID: 4127) without reference correction. (b) MD RMSF predicted from RCI of the same NMR assignment set (BMRB ID: 4127) after reference correction with SHIFTCOR (black line) and REFCOR (red line). (c) RMSF of the structural ensemble obtained with 4-ns MD simulation of solution structure of SV40 T antigen DNA-binding domain (PDB ID: 2TBD)

between the per-residue RCI-predicted and observed MD RMSF values. Improvements of correlation coefficients between RCI-predicted and observed MD RMSF were also observed for several other proteins, most noticeably, Staphylococcal nuclease ( $\Delta r = 0.25$ , BMRB ID: 4052) and the N-terminal domain of DNA polymerase  $\beta$  ( $\Delta r = 0.07$ , BMRB ID: 4326). Clearly, mis-referenced chemical shifts can diminish the performance of the RCI method.

While chemical shift referencing errors can be corrected with SHIFTCOR, this reference correction protocol requires that the protein's tertiary structure already be known. Since the RCI method is expected to be used prior the determination of a protein structure, a structure-independent method of reference correction needs to be found. Fortunately one exists. The PSSI re-referencing protocol (Wang and Wishart 2005) uses  $^1\text{H}\alpha$  shifts (which are rarely mis-referenced) to correct and calibrate mis-referenced  $^{13}\text{C}$  and  $^{15}\text{N}$  shifts. We have extended this method to perform reference shift corrections in the absence of  $^1\text{H}\alpha$  shifts (that are often unavailable in the published assignments of large proteins). This re-referencing protocol (called REFCOR) was implemented in the RCI web server. A comparison between the REFCOR re-referencing and SHIFTCOR reference corrections on the RCI of SV40 T-antigen DNA-binding domain is shown on Fig. 2. A more detailed description of the REFCOR protocol has been published elsewhere (Berjanskii and Wishart 2006).

#### RCI versus other methods to investigate protein dynamics

The Random Coil Index was originally developed and refined on data obtained from MD simulations. One may wonder how realistic these simulations are and whether the RCI results would correlate with other measures of protein mobility such as order parameters, B-factors, and NMR RMSD values. To assess these relationships, we analyzed all 33 proteins and determined the correlation of the RCI values with experimentally determined model-free (Lipari and Szabo 1982; Clore et al. 1990) order parameters ( $S_{\text{exp}}^2$ ), theoretically predicted (Zhang and Bruschweiler 2002) order parameters ( $S_{\text{pred}}^2$ ), amide nitrogen RMSD of NMR ensembles and B-factors from crystallographic structures.

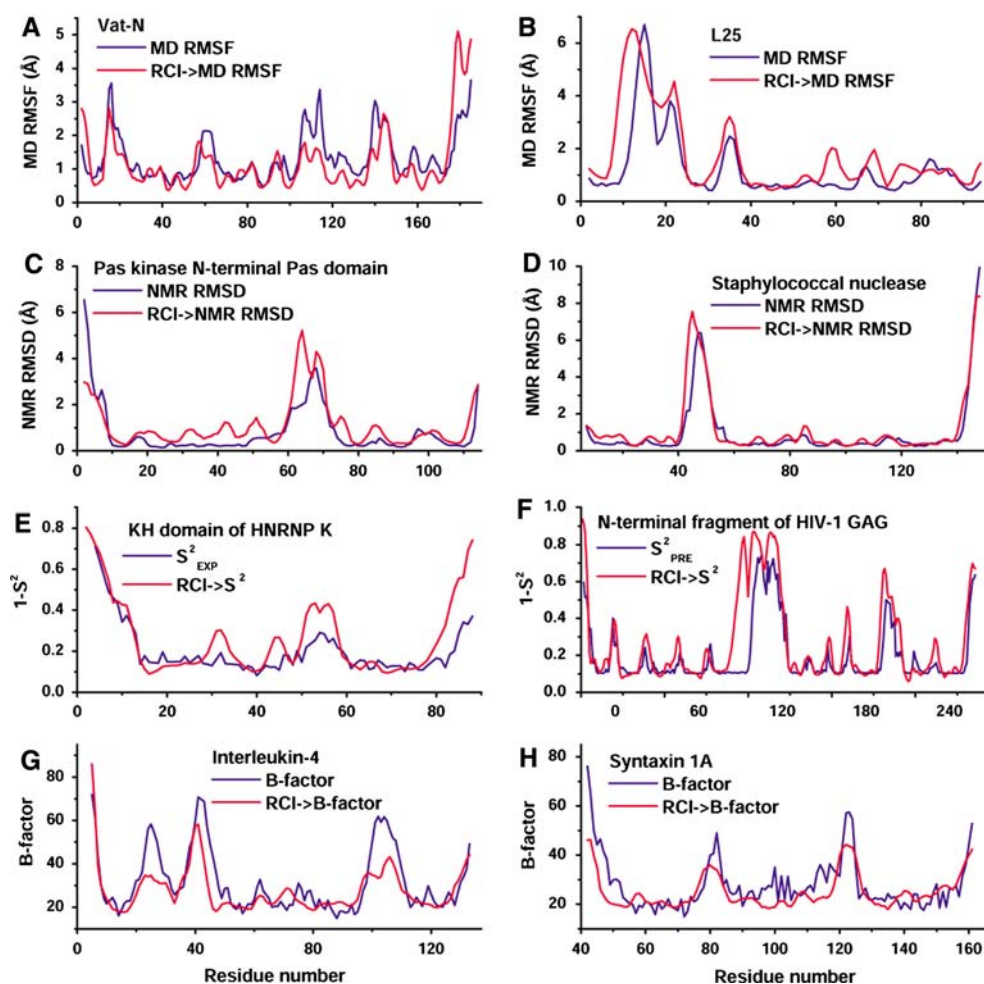
Note that the difference between NMR RMSD (root mean square deviation) and MD RMSF (root mean square fluctuation) is that MD RMSF is a measure of in silico spatial fluctuation calculated over a period of time. MD RMSF depends on the protein model, the length of MD simulations, and the quality of the MD force-field. On the other hand, NMR RMSD is calculated from an NMR-derived structural ensemble and reflects the quality and quantity of NMR restraints, the NMR structure

determination protocol and the manner of assembling the structural ensemble. Typically, per-residue RMSDs of NMR ensembles are not used to characterize protein dynamics, but rather to serve as a measure of ensemble precision or uncertainty of ensemble's average structure. However, the local variability of NMR ensembles also depends on the number of structural restraints (e.g., NOEs, dihedral restraints) and these can be significantly reduced in number when spectral peaks are broadened or diminished in intensity due to conformational exchange. Also, the number of NOEs is typically decreased (and, as a result, the NMR RMSD is increased) in flexible regions due to reduced local atomic density. Therefore, we have decided that the sensitivity of disorder in NMR ensembles to protein dynamics justifies the usage of per-residue NMR RMSD as a reporter of protein flexibility in the current work.

Analysis of the NMR ensembles was done using MOLMOL (Koradi et al. 1996). As seen in Table 1, the RCI values correlate well with all of these parameters except the thermal B-factors. The average correlation coefficients between RCI and NMR RMSD,  $S_{\text{exp}}^2$ ,  $S_{\text{pred}}^2$ , and B-factor values are 0.77, 0.75, 0.71, and 0.60, respectively. Interestingly, we found that the correlation between any two of these conventional measures of flexibility was never more than 0.84 (MD RMSF vs. NMR RMSD) and that in some cases the correlation proved to be remarkably poor ( $S_{\text{pred}}^2$  vs. B-factor with a correlation coefficient of only 0.43—see Table 1). Of all the methods tested, we found that the RCI, MD RMSF and NMR RMSD exhibited the best agreement with the other experimental and theoretical measures of mobility (mean correlation coefficients of 0.72, 0.72, and 0.71, respectively), while the B-factor exhibited the worst (mean correlation coefficients of 0.54). Figure 3 illustrates the typical correlation seen between the Random Coil Index and values obtained for MD RMSF, model-free  $S^2$ , and RMSD from NMR ensembles as well as two examples of good correlations between RCI and B-factors. While analytical relationships connecting RCI with these parameters have yet to be established, it is possible to make rough estimates of MD RMSF, NMR RMSD,  $S^2$ , and

**Table 1** Correlation among different measures of motional amplitudes in proteins

|                       | MD RMSF | NMR RMSD | $1-S_{\text{exp}}^2$ | $1-S_{\text{pred}}^2$ | B-factor |
|-----------------------|---------|----------|----------------------|-----------------------|----------|
| RCI                   | 0.81    | 0.77     | 0.75                 | 0.71                  | 0.60     |
| MD RMSF               |         | 0.84     | 0.63                 | 0.72                  | 0.59     |
| NMR RMSD              |         |          | 0.68                 | 0.71                  | 0.61     |
| $1-S_{\text{exp}}^2$  |         |          |                      | 0.66                  | 0.48     |
| $1-S_{\text{pred}}^2$ |         |          |                      |                       | 0.43     |



**Fig. 3** Correlations of RCI with MD RMSF, NMR RMSD, model-free order parameter ( $S^2$ ) and B-factors. RCI→MD RMSF, RCI→NMR RMSD, RCI→ $S^2$ , RCI→B-factor indicate MD RMSF, NMR RMSD,  $S^2$ , B-factor, respectively, predicted from RCI with Eqs. 3–6.  $S_{\text{EXP}}^2$  and  $S_{\text{PRE}}^2$  are experimentally obtained and structure-derived model-free order parameters, respectively. The following BMRB IDs and PDB IDs were used to obtain RCI, RMSF, RMSD,  $S^2$  and B-factors: (a) 4376 and 1CZ5 for Vat-N. (b) 4395 and 1B75 for

L25. (c) 5354 and 1LL8 for N-terminal PAS domain of PAS. (d) 4052 and 1JOR for Staphylococcal Nuclease. (e) 4405 for KH domain of HNRNP K. Order parameters published by the Tjandra group (Baber et al. 2000) were compared with RCI-derived  $S^2$  (f) 5316 parameters for N-terminal fragment of HIV-1 GAG. Order parameters were predicted from the solution structure with PDB ID 1L6N using the contact model (Zhang and Bruschweiler 2002). (g) 4094 and 1RCB for Interleukin-4. (h) 4198 and 1EZ3 for Syntaxin 1A

B-factors that should be expected for particular RCI values. Using both the protein training and test sets (Supplemental Table 2), we have identified several empirical expressions that allow the conversion of RCI values into the aforementioned parameters with a satisfactory level of agreement. When assignments for all six types of nuclei are available, these scaling relationships are as follows:

$$S^2 = 1 - 0.4 \ln(1 + \text{RCI} * 17.7) \quad (3)$$

$$\text{MD RMSF} = \text{RCI} * 28.3 \text{ \AA} \quad (4)$$

$$\text{NMR RMSD} = \text{RCI} * 16.7 \text{ \AA} \quad (5)$$

$$\text{B - factor} = \text{RCI}^{1/2} * 142.0 \quad (6)$$

The average absolute errors for  $S^2$ , MD RMSF, NMR RMSD, and B-factor predicted from RCI values are 0.05, 0.40 Å, 0.43 Å, and 16.9, respectively.

The imperfect agreement between original and RCI-predicted measures of motional amplitudes ( $S^2$ , MD RMSF, and NMR RMSD) should not be viewed negatively. Indeed, no single method discussed in this paper gives a complete and faultless picture of protein dynamics. For example, NMR ensembles may be affected by numerous factors unrelated to protein dynamics. They may sample conformational space incompletely due to investigator bias and the inclusion of unrecognized spin-diffusion NOEs (yielding upper bounds that are too tightly



restrained). Likewise, NMR ensembles may be corrupted or structurally biased due to the mis-assignment of ambiguous NOEs, the mis-interpretation of spectral artifacts or noise peaks as NOEs. On the other hand, NMR ensembles may also over-sample conformation space due to investigator bias (neglect or inexperience) or a lack of NOEs and other conformational restraints.

The problems with NMR RMSD as a proxy measure of protein mobility are not unique. For instance, experimental order parameters may not properly reflect amplitudes of protein motions due to the poor separation of overall tumbling and internal dynamics (Korzhnev et al. 1997). Likewise, opposing effects arising from motions on different time-scales for both transverse (Palmer et al. 2001) and longitudinal (Fushman et al. 1997) relaxation processes, site-specific variations of  $^{15}\text{N}$  CSA (Damberg et al. 2005) and other artifacts (Case 2002) can lead to incorrect order parameter estimates. Predicted order parameters are not without their problems either. These theoretical measures may miss long-range correlated motions (Zhang and Bruschweiler 2002) and appear to be quite sensitive to sub-angstrom inaccuracies in the 3D model, from which they are derived.

B-factors also have their faults as they can manifest not only the internal dynamic disorder of a protein but also multiple conformations in different unit cells (internal static disorder). B-factors can also be corrupted by refinement errors, the contributions of more than one atom to a particular electron density, intermolecular crystal packing contacts, lattice defects and lattice vibrations (Petsko and Ringe 1984; Carugo and Argos 1999). Likewise, MD simulations are not immune to criticism. MD RMSF can suffer from a number of problems including incomplete conformational sampling (Elofsson and Nilsson 1993) and large uncertainties of motional amplitudes in mobile regions (Horita et al. 2000). Insufficient system equilibration, numerical rounding errors, limitations of energy functions and treatments of long-range non-bonded interactions can all contribute to inaccuracies of MD-derived amplitudes of protein motions.

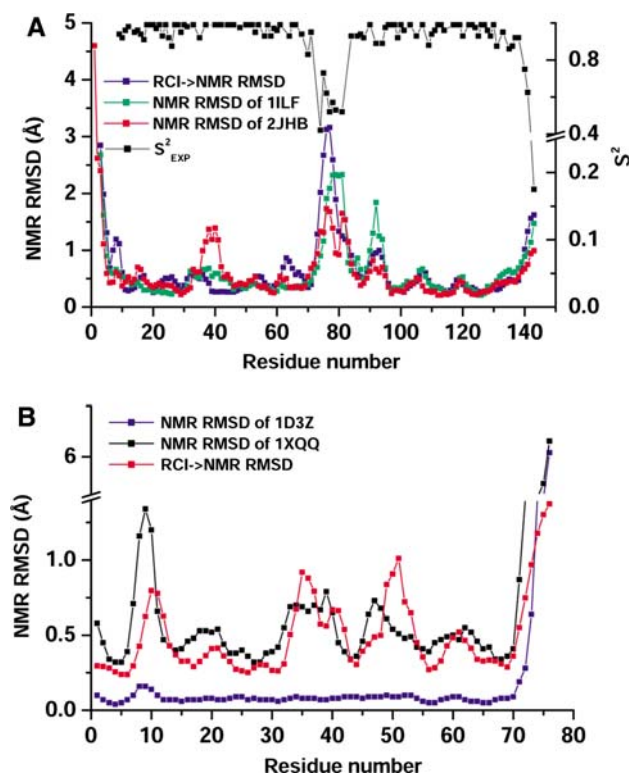
The inherent errors and uncertainties associated with these techniques are not the only contributors to the imperfect level of agreement. Discrepancies also arise from the different time-scales, for which these methods are most sensitive. MD simulations are computationally limited to monitoring ps to ns motions. Experimental model-free order parameters reflect the amplitude of motions on a picosecond to nanosecond time-scale (Lipari and Szabo 1982; Clore et al. 1990). Their precision and accuracy decreases when the time-scale of internal motions approaches that of overall tumbling (Jin et al. 1998; Chen et al. 2004). In contrast to experimental order parameters, B-factors can be affected by uncorrelated (Petsko and

Ringe 1984) motions that may occur over much longer period of time (hours to days) during the acquisition of X-ray data. The expression for predicted order parameter (Zhang and Bruschweiler 2002) has been empirically optimized to convert protein-packing density into a model-free order parameter (ps–ns time scale). However, it has not been ruled out that the predicted order parameter can represent motions on other time scales. In fact, it has been shown that protein packing can also be related to crystallographic B-factors (Halle 2002), which are sensitive to motions over much longer time scales than the time-scale of model-free order parameters. Chemical shifts, as manifested by the Random Coil Index, are expected to be most sensitive to motions on time-scales ranging from 100's of ps to possibly ms. This time scale appears to depend on the coalescence conditions of contributing nuclei and the heterogeneity of sampled chemical shift space (see a discussion about RCI time-scale later in the paper).

### Case studies

Despite their sensitivity to different time-scales, all of the previously mentioned methods, including the RCI, share the ability to identify protein “hot spots” with increased mobility. Given the ease with which RCI values can be calculated, we believe the RCI method could prove to be particularly useful in conducting quick “sanity” checks of the motional amplitudes obtained with other methods (e.g., NMR RMSD,  $S^2$ , and B-factors).

To demonstrate the utility of the RCI in validating NMR ensembles, we will use the RCI method to evaluate flexibility of a protein called Core Binding Factor  $\beta$  (CBF $\beta$ ) and compare it with the structural diversity of two NMR ensembles: 2JHB (Huang et al. 1999) and 1ILF. The latter structure also had its order parameters determined using a model-free analysis of its  $^{15}\text{N}$  relaxation data. As seen in Fig. 4a, a plot of the RMSD for the 2JHB NMR ensemble suggests that the flexibility of CBF $\beta$  region 38–44 is comparable with mobility of the “long loop” (residues 72–84) and significantly higher than the motional amplitudes observed in other regular secondary structures of this protein. In contrast, the RMSD values of the 1ILF ensemble and the model-free analysis of its  $^{15}\text{N}$  relaxation data suggests that the “long loop” (residues 72–84) has a much higher level of disorder than that of the 38–44 region. Furthermore, the structural diversity of the 38–44 region in the 1ILF model does not significantly exceed the conformational sampling seen in its  $\alpha$ -helices and  $\beta$ -sheets. The disagreement between these two NMR ensembles could be attributed to differences in NMR restraints or/and small differences in protein constructs and NMR experimental conditions (Wolf-Watz et al. 2001). However, as seen in

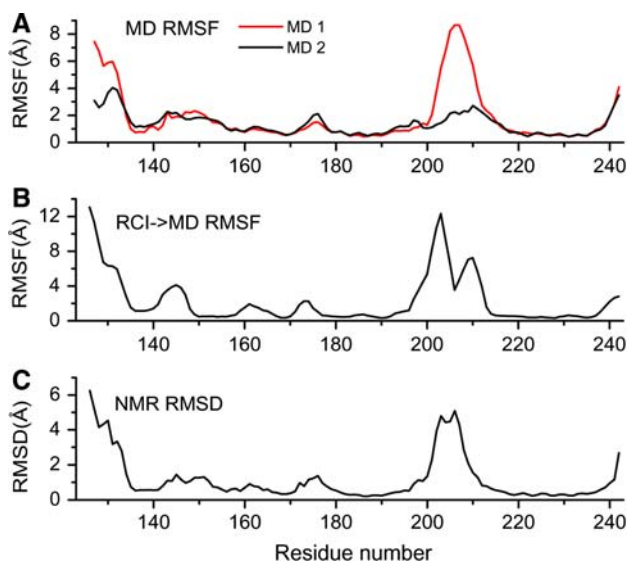


**Fig. 4** Using RCI to identify “dynamically correct” NMR ensembles. (a) Comparison of NMR RMSD predicted from RCI of Core Binding Factor  $\beta$  with RMSD of NMR ensembles and model-free order parameters reveals the under-restraining of 38–44 region in the 2JHB ensemble. NMR assignments with BMRB accession number 4092 were used to calculate RCI. NMR RMSDs were predicted using equation 5. Two NMR ensembles of Core Binding Factor  $\beta$  (PDB IDs: 2JHB and 1ILF) were used to calculate RMSD of NMR ensembles. Model-free order parameters were taken from (Wolf-Watz, Grundstrom et al. 2001). (b) RCI of ubiquitin (NMR assignment with BMRB ID 5387) indicates the over-restraining of RDC-refined NMR ensemble (PDB ID: 1D3Z) and demonstrates good correlation with  $S^2$ -refined NMR ensemble (PDB ID: 1XQQ) of ubiquitin

Fig. 4a, the Random Coil Index for CBF $\beta$  is consistent with the RMSD data of the 1ILF ensemble and the model-free analysis. This fact points to the structure determination protocols as the primary source of the differences between the two CBF $\beta$  models. Indeed, the experiments for NMR assignments (BMRB (4092) used in these RCI calculations were conducted under the same buffer conditions as those of the experiments for 2JHB structure determination. Still, the RCI-predicted NMR RMSD values of CBF $\beta$  correlate better with the RMSD of 1ILF ( $r = 0.84$ ) than with RMSD of 2JHB ( $r = 0.62$ ), and indicate a more flexible “long loop” and a more rigid 38–44 region than those observed in 2JHB. Combined with the results of model-free analysis, the RCI data suggest that the 1ILF model is a more “dynamically correct” representation of CBF $\beta$  solution structure. In particular, the “long loop” was likely over-restrained and the 38–44 region was likely under-restrained during the structure calculations of 2JHB.

In the next example, we will show how the RCI method can be used to determine whether an NMR ensemble has reached the appropriate degree of overall structural diversity. Figure 4b compares the RMSD values for two different NMR ensembles of ubiquitin. One ensemble (PDB ID: 1XQQ) has a total of 128 different conformers and was determined via a novel dynamically optimized structure generation process (Lindorff-Larsen et al. 2005). The other ensemble (PDB ID: 1D3Z) consists of just 10 conformers and was determined with additional residual dipolar coupling (RDC) restraints (Cornilescu et al. 1999). As seen in Fig. 4b, the RCI plot correlates better ( $r = 0.74$ ) with the dynamically optimized NMR ensemble of ubiquitin (PDB ID: 1XQQ) than it does with the RDC-refined ensemble of ubiquitin ( $r = 0.60$ ). This plot clearly shows that the difference between RMSD of the 1D3Z model and RCI-predicted RMSD is much larger than the difference between RCI-predicted RMSD and the RMSD of the dynamically optimized 1XQQ (mean difference of 0.42 Å vs. 0.11 Å for regular secondary structures). As indicated by this example, we believe that the RCI method offers a considerable promise for the detection of insufficient or exaggerated structural variations in NMR ensembles.

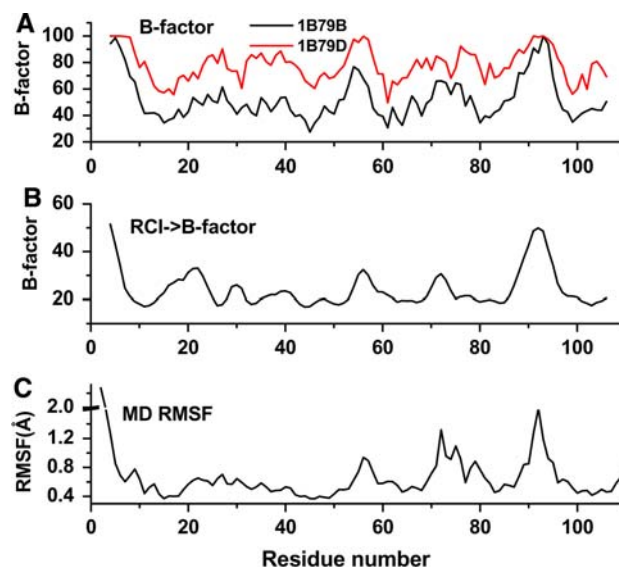
In addition to its use in validating NMR ensembles, the RCI method can also be applied to assessing the quality or completeness of MD simulations. MD simulations can be affected by the insufficient sampling of conformational space, limitations of MD force fields or incorrect usage of MD software (van Gunsteren and Mark 1998) and have to be validated with experimental data, such as NMR order parameters or/and X-ray B-factors. Figure 5 shows an example of RMSF variability observed in two 5-ns MD simulations of the chicken prion protein (PDB ID 1U3M) as designated by MD 1 and MD 2. The starting structures in both simulations were identical and the set-up for the MD simulations was described previously (Berjanskii and Wishart 2005). In both cases, the initial velocities were chosen randomly based on the starting temperature. While RMSF profiles of these two simulations are comparable throughout most of the protein, the amplitudes of motions in the N-terminus and the loop between helices 2 and 3 (residues 196–211) are much smaller in MD 2 versus MD 1. Similar variability in MD motional amplitudes were observed by other groups (Horita et al. 2000) and likely originate from the presence of motions on the time-scale much longer than the duration of MD simulations. Comparison of the MD RMSF with both the per-residue RMSD of the 1U3M NMR ensemble and the MD RMSF predicted from RCI suggest that the MD 2 simulation does not sample the conformational space of the chicken prion completely and the reduced mobility is likely an artifact. It is important to realize that a comparison of MD



**Fig. 5** The RCI method allows identification of MD simulation that incompletely samples conformational space of chicken prion. (a) RMSF values from two MD simulations (MD 1 and MD2) with identical set-up. Model 1U3M was used as a starting structure. All MD ensembles were generated using identical MD protocols (but different randomly-chosen initial velocities) published elsewhere (Berjanskii and Wishart 2005). (b) MD RMSF predicted from RCI (RCI→MD RMSF) using NMR assignments with BMRB ID 6269. (c) RMSD of NMR ensemble of chicken prion solution structure with PDB ID 1U3M

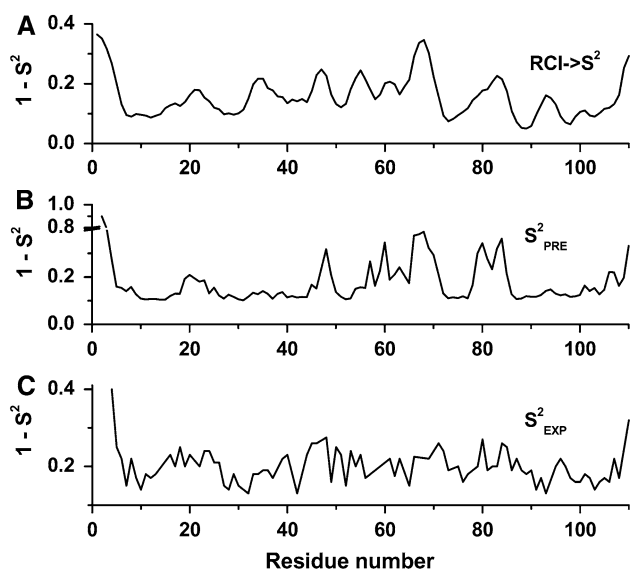
RMSF with the RCI-predicted RMSF only is sufficient to reach this conclusion and that RCI can serve as a stand-alone method of validating structural diversity of MD ensembles.

The RCI method can also be used to identify which chains within multi-chain X-ray structures have dynamically correct B-factors. As it was mentioned earlier, B-factors can be affected by variety of factors unrelated to protein dynamics and may often provide misleading information about protein flexibility. The DnaB helicase is a protein that has had both an X-ray structure (PDB ID: 1B79) and a set of NMR assignments (BMRB ID: 4297). For the X-ray structure the B-factors of the A, B, C and D chains have mean coefficients of correlation with conventional dynamic parameters (MD RMSF, NMR RMSD, and predicted  $S^2$ ) of 0.65, 0.76, 0.70, and 0.58, respectively. Figure 6 compares the experimental B-factors of the DnaB helicase with MD RMSF and the RCI-predicted B-factors for chains B and D. Interestingly, the correlation of RCI-predicted B-factors with experimental B-factors is able to capture this trend almost exactly ( $r = 0.71, 0.77, 0.69, 0.52$  for A, B, C and D chains, respectively) and allows the identification of chains B and D as having the most and the least dynamically correct B-factors, respectively.



**Fig. 6** RCI identifies 1B79B as the X-RAY model of DnaB helicase with the most realistic B-factor. (a) B-factors of chains B and D from the X-RAY model of DnaB helicase with PDB ID 1B79. (b) B-factor predicted from RCI using NMR assignments with BMRB ID 4297 and Eq. 6. (c) RMSF of a structural ensemble obtained from 4-ns MD trajectory of DnaB helicase. The solution structure of the protein (PDB ID 1JWE) served as the starting structure for the MD simulation

The RCI method can also be used to validate model-free order parameters. For this example, we chose to compare RCI-derived (Fig. 7a) and experimental (Fig. 7c)  $^{15}\text{N}$  order parameters of barnase (Sahu et al. 2000). Barnase is 84% identical to binase, a protein, whose  $^{15}\text{N}$  order parameters are known to poorly represent this protein's flexibility (Pang et al. 2002; Wang et al. 2003). Similar to binase, the experimental  $^{15}\text{N}$  order parameters of barnase do not correlate well with other reporters of protein flexibility, such as predicted order parameters ( $r = 0.50$ , Fig. 7, panels b and c), MD RMSF ( $r = 0.03$ ), NMR RMSD ( $r = 0.34$ , RMSD's calculated using 1FW7), and B-factors ( $r = 0.05$ , B-factor of 1A2P). Interestingly, the RCI values do not correlate particularly well with the experimental order parameters in barnase either ( $r = 0.45$ , Eq. 3, BMRB ID: 4964). However, the RCI demonstrates a significantly better agreement with MD RMSF, NMR RMSD, predicted order parameters and B-factors than the experimental order parameters do (0.64 vs. 0.03, 0.69 vs. 0.34, 0.73 vs. 0.50, and 0.58 vs. 0.05, respectively; see Supplemental Table 2) Cumulatively, these data suggest that the standard model-free analysis of  $^{15}\text{N}$  relaxation cannot produce a proper profile of backbone flexibility for barnase. Generally, if the correlation coefficient between the RCI value and another measure of flexibility drops below 0.50 (as seen in the example above) it may



**Fig. 7** RCI allows identification of a set of model-free order parameters that poorly characterize flexibility of barnase. **(a)** Model-free order parameter ( $\text{RCI} \rightarrow S^2$ ) predicted from the NMR assignment set with BMRB ID 4964 using Eq. 3. **(b)** Order parameter ( $S^2_{\text{PRE}}$ ) predicted from barnase solution structure (PDB ID 1FW7) using the contact model (Zhang and Bruschweiler 2002). **(c)** Order parameter ( $S^2_{\text{EXP}}$ ) derived by the model-free analysis of barnase  $^{15}\text{N}$  relaxation rates (Sahu, Bhuyan et al. 2000)

be a good idea to re-examine the procedure for calculating or measuring that parameter.

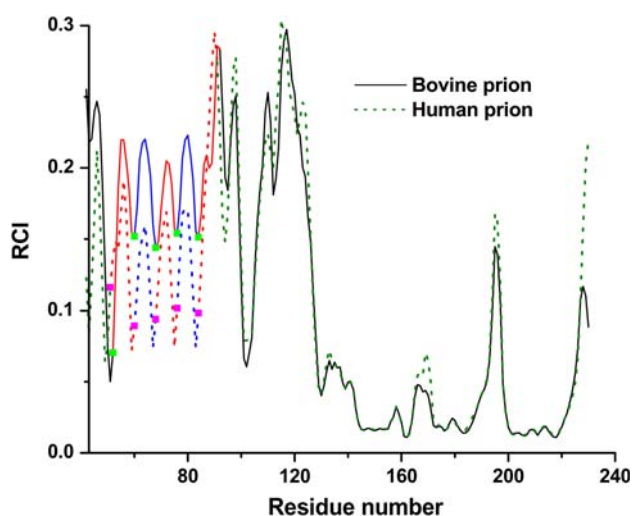
#### RCI: its applicability to large and unfolded proteins

One of the key advantages of the RCI method is that it appears to be capable of predicting protein flexibility regardless of the protein weight, shape or domain composition. In particular, we have found that RCI values correlates very well with per-residue MD RMSF for proteins ranging in size from 56 to 283 amino acids (Supplemental Table 2) and in relative flexibility ranging from 15 to 65% (Supplemental Table 3). The insensitivity of the RCI method to protein molecular weight makes it a particularly useful tool to study the flexibility of large proteins. This is because larger proteins often present a challenge to conventional NMR relaxation measurements (which are required for model-free analysis) due to spectral overlap and weak signal intensity. Figure 3f shows an example of the high level of agreement between model-free order parameters predicted from the RCI method and the MD RMSF of a structural model (1L6N) of the 32.2 kDa HIV-1 Gag protein. This example also demonstrates the rather impressive performance of the RCI method in identifying mobile regions in a protein that consists of several domains connected by a flexible linker.

Multi-domain proteins often prove to be quite problematic for order parameter calculations. In particular, if the time-scales of individual domain motions and the overall tumbling are close, one may experience difficulties in characterizing the frequency and anisotropy of the overall rotation, which are needed for the model-free analysis. This can lead to difficulties in the accurate calculation of order parameters (Korzhnev et al. 1997). For large-amplitude domain motions, the model-free approach may often have to be replaced with a complex model-dependent analysis such as a Triple-Exponential Wobble-In-A-Cone approximation (Chang and Tjandra 2001). In contrast, the RCI approach requires no model fitting and provides an excellent alternative to explore and quantify intra-domain dynamics of flexible multi-domain proteins. As evident from the aforementioned example, RCI values appear to be insensitive to domain reorientation in the absence of frequent domain collisions. This is likely due to the dominant role of very local de-/shielding effects on chemical shifts.

Fully or partially unfolded proteins are another group of biomolecules that are amenable to RCI analysis. The application of model-free analysis to unstructured proteins is difficult because this approach requires knowledge of the three-dimensional structure to estimate rotational anisotropy and its effect on internal dynamics (Palmer 2001). It is commonly accepted that the overall rotation of an unfolded protein cannot be considered as a single motional process with a unique correlation time. As a result, the model-free formalism, in its original form, cannot be used in such cases (Farrow et al. 1997; Dyson and Wright 1998; Penkett et al. 1998). In contrast, RCI values, derived from isotropic chemical shifts, are not affected by overall tumbling and do not require any knowledge of the three-dimensional structure and its rotational anisotropy. Therefore, we believe that the RCI approach could be a powerful new tool in investigating residual structure in intrinsically disordered proteins or partially folded proteins. Figure 8 demonstrates how the RCI method can be used to detect semi-rigid areas around proline residues in the presumably disordered octapeptide repeats of bovine and human prion proteins. However, because the RCI method was optimized primarily for folded proteins, we would advise users to exercise some caution in interpreting RCI results for completely unfolded proteins. The different sensitivities between folded and unfolded proteins to imperfections in the MD force-field(s) justifies the need for further verification and, possibly, a separate optimization of the RCI expression for unfolded proteins. Unfortunately, experimental dynamic measurements that could be used to validate MD simulations of unfolded proteins are too sparse to allow us to properly tune the RCI expression for this protein class.





**Fig. 8** RCI identifies octarepeats with semi-rigid structure near proline residues in unfolded N-terminal domains of bovine and human prions. NMR assignments with BMRB accession codes 4564 and 4402 were used to calculate RCI of bovine and human prions, respectively. Octarepeats are colored with red and blue colors. Position of proline residues in octarepeats of bovine and human prions are shown with green and magenta filled squares, respectively

#### The RCI time scale

A key question that has not yet been addressed in our previous work is: What is the time-scale of motions captured by the RCI method? The answer to this intriguing question is tightly bound to our understanding the effects of protein dynamics on chemical shifts in proteins. The conformational averaging that affects the positions of individual peaks in an NMR spectrum and, therefore, the shifts used to calculate the RCI occur on the time-scale of fast conformational exchange. Changes in intermediate and slow exchange processes affect peak intensities while having no influence on peak positions (chemical shifts). In a simple case of *two-site* conformational exchange, the upper time limit of fast exchange depends on the coalescence conditions (the frequency of exchange, at which two slow-exchange peaks merge into a single peak) for a particular chemical shift difference between the two sites (Levitt 2001). It is reasonable to assume that the upper limit of the RCI time scale should also depend on the

coalescence conditions of *multi-site* conformational exchange that each nucleus in the RCI formula (Eq. 1) experiences. In practical terms, protein NMR resonances are generally too weak to be detected at this theoretical coalescence point. Therefore the actual upper limit of the RCI time-scale will correspond to the frequency of fast-intermediate exchange—a point at which peaks become observable in an NMR spectrum. In principle, the exchange rate (or frequency) corresponding to this transition point can be identified by calculating the amplitude of NMR resonances using McConnell's extension of the Bloch equations (McConnell 1958) and compared with the expected noise level. However, the NMR signal in real experiments also depends on numerous factors, such sample concentration, the efficiency of magnetization transfer in a given NMR experiment, the magnetic field of NMR spectrometer, and the number of scans. The necessity to make assumptions about these factors makes the value of purely theoretical calculations rather limited. Moreover, one should realize that the application of the RCI method and, thus, the RCI time-scale will depend not only on the presence of visible signals in the spectra, but also on the probability of these signals to be assigned to particular nuclei in the protein.

Despite these limitations, we decided to assess the RCI time-scale using the feasibility of obtaining good quality NMR assignments as the criterion. A detailed description of this assessment can be found in the supplemental material. To summarize our results, the upper limit of the RCI time-scale can vary from microseconds to milliseconds for  $^{13}\text{C}$  and  $^{15}\text{N}$  nuclei and from hundreds nanoseconds to hundreds microseconds for protons depending on the parameters associated with exchange rates (i.e., motions), protein molecular weights and experimental conditions (Table 2). These data suggest that the RCI method can be sensitive to relatively slow motions.

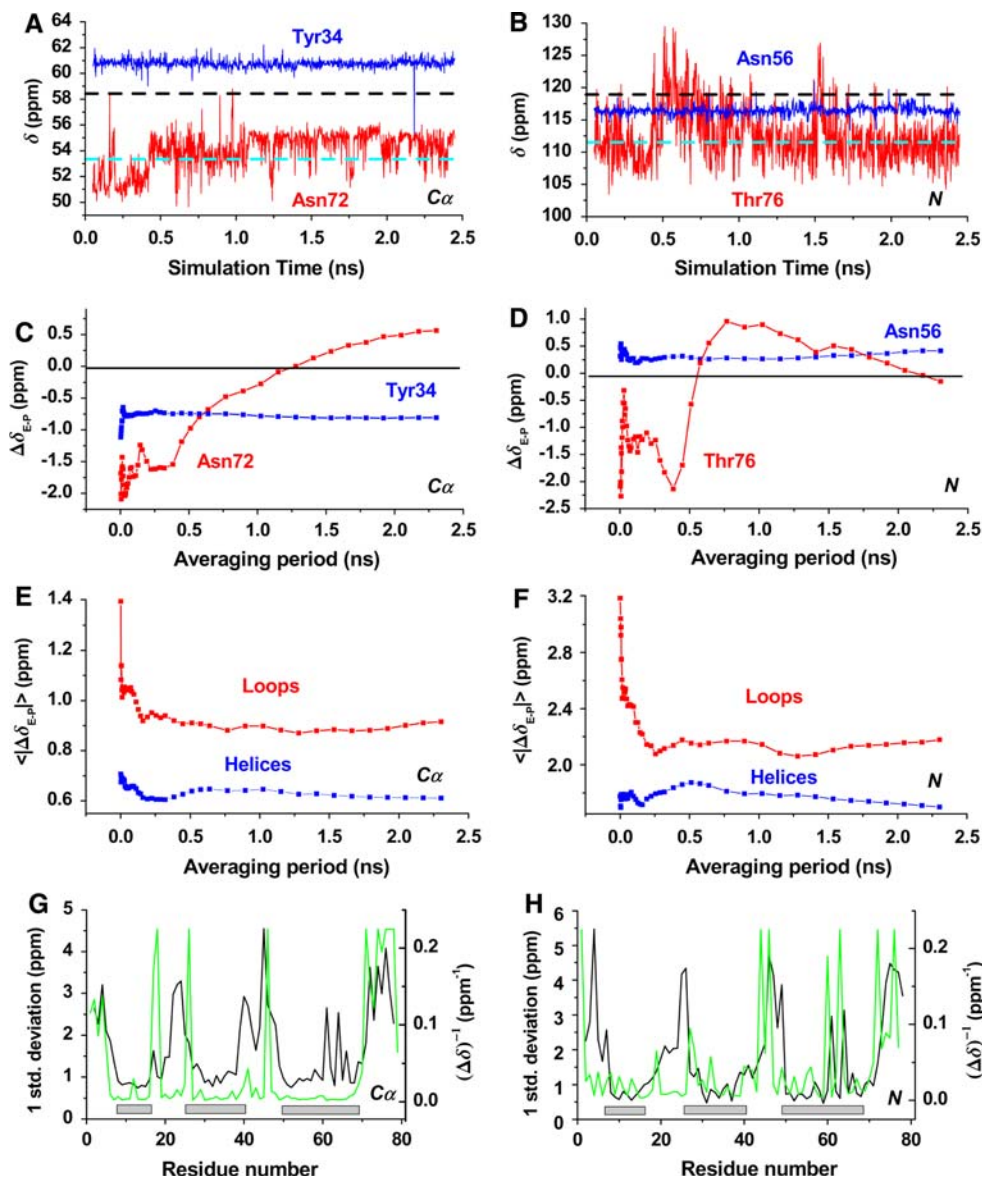
In order to estimate the span of the RCI time-scale, we also attempted to get insights into the lower time limits of the fast conformational exchange that affects chemical shifts (and RCI). This was done by calculating the temporal evolution of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts of different residues during an extended MD simulation (Fig. 9a and b). Specifically, MD simulations on a small protein (PyJ)

**Table 2** The upper limit of the RCI time-scale determined from the life-time ranges of exchanging states

| Types of nuclei             | Life-time of exchanging states (s) |          |                             |          |
|-----------------------------|------------------------------------|----------|-----------------------------|----------|
|                             | 10 Hz frequency difference         |          | 100 Hz frequency difference |          |
|                             | Minimum                            | Maximum  | Minimum                     | Maximum  |
| $C\alpha$ , $C\beta$ , $CO$ | 2.04E-04                           | 1.54E-03 | 2.04E-06                    | 1.54E-05 |
| N                           | 3.44E-04                           | 2.60E-03 | 3.44E-06                    | 2.60E-05 |
| $H_N$                       | 1.47E-05                           | 1.11E-04 | 1.47E-07                    | 1.11E-06 |
| $H\alpha$                   | 2.33E-05                           | 1.76E-04 | 2.33E-07                    | 1.76E-06 |

were done using its solution structure (PDB ID: 1FAF). The MD protocol was described elsewhere (Berjanskii and Wishart 2005). Chemical shifts were predicted for each

MD snapshot using ShiftX (Neal et al. 2003). Analysis of individual MD snapshots separated by 6-ps simulation periods revealed significant (e.g., 1–3 ppm for



**Fig. 9** Predictions of chemical shifts from MD trajectories of PyJ with ShiftX (Neal et al. 2003). (a–b) Changes of hypothetical “evolved” chemical shifts of C $\alpha$  (panel a) and N (panel b) during MD simulations as predicted by ShiftX. Chemical shift trajectories of residues from helical and loop regions are colored blue and red, respectively. Positions of random coil values of helical residues (Tyr34 and Asn56) and coil residues (Asn72 and Thr76) with respect to the fluctuating chemical shifts are shown with black solid and cyan dashed lines, respectively. (c–d) Dependence of the mean absolute prediction error on the simulation period, over which the predictions of chemical shifts (C $\alpha$  on panel c and N on panel d) were averaged. To calculate the mean absolute prediction error ( $\langle|\Delta\delta_{E-P}|$ ), the absolute differences between predicted ( $\delta_P$ ) and experimental ( $\delta_E$ ) chemical shifts are determined and averaged for different lengths of MD simulations (starting from time zero and gradually increasing the

length of analyzed MD trajectory). Averaged chemical shift errors of residues from helical (Tyr34 and Asn56) and loop (Asn72 and Thr76) regions are colored blue and red, respectively. Zero error is shown with solid black line. (e–f) Dependence of the mean absolute error of chemical shift predictions (C $\alpha$  on panel e and N on panel f) on the averaged simulation period for different types of secondary structure. Averaged chemical shift errors for helices and loops are colored blue and red, respectively. (g–h) Co-location of experimental random coil chemical shifts and large amplitudes of predicted chemical shift fluctuations in PyJ primary sequence. Per-residue distributions of one standard deviations of C $\alpha$  (panel g) and N (panel h) chemical shifts predicted from the MD trajectory of PyJ are shown with black lines. Inverse absolute secondary chemical shifts are colored green. Location of helices in PyJ are shown with gray bars

$C\alpha$ (of Thr76) changes to predicted chemical shifts for all residues in flexible (non-helical) regions of PyJ. This result suggests that the lower limit of the RCI time scale is likely to be on femtosecond time-scale because even 6 ps periods of MD simulations are sufficient to produce PyJ conformations with significantly different characteristic chemical shifts (Fig. 9a and b).

In addition to its utility in assessing the lower limit of the RCI time-scale, it is also possible to use MD methods to identify some of the slower motions affecting RCI values and compare their frequency with the results of aforementioned theoretical calculations of the upper limit of the RCI time-scale. This was done by monitoring the chemical shift averaging process for different residues during MD simulations. Figure 9 (Panels c and d) show the dependence of the averaged  $C\alpha$  and N chemical shifts on the length of averaging period for two helical and two coil residues. The averaging process normally starts from an initial (<100 ps) period of large amplitude chemical shift changes, followed by high-frequency chemical shift oscillations that trend towards the averaged chemical shifts after about 1 ns of temporal evolution. When the differences between the experimental and the predicted average chemical shifts from different residues are combined, it is clear that the averaging process makes the predicted chemical shifts somewhat more accurate (Fig. 9e and f). In helices, the amplitude is much smaller (Fig. 9e and f) and the average chemical shift often reaches its plateau value within 300–500 ps (Fig. 9c and d, Tyr34 and Asn56). In contrast, the average chemical shifts in coil regions often continue to experience significant changes (Fig. 9c and d, Asn72 and Thr76) throughout the course of their MD simulations (2.5 ns). This result suggests that the RCI can manifest motions on the time scale of nanoseconds and above, and is consistent with the sensitivity of the RCI methods to slower motions (with sub-state life-times on  $\mu$ s-ms and shorter time-scales) as suggested by earlier theoretical calculations in this paper.

Simulation of chemical shift averaging in PyJ using MD and ShiftX revealed that the motions in loop regions can result in significant fluctuations of chemical shifts (with one standard deviation up to 5 ppm for  $^{13}\text{C}$  and  $^{15}\text{N}$  shifts and up to 10 ppm for  $^1\text{H}$ ; Fig. 9g and h). These fluctuations are comparable with the effects of torsion angle variations on chemical shifts observed in quantum mechanical calculations for  $H\alpha$  (Osapay and Case 1994),  $C\alpha$  (Oldfield 1995; Sun et al. 2002; Xu and Case 2002),  $C\beta$  (Oldfield 1995; Sun, Sanders et al. 2002; Xu and Case 2002), CO (Xu and Case 2002) and N (Le and Oldfield 1996; Xu and Case 2002). Large chemical shift deviations co-localize with random coil chemical shifts in the PyJ structure (Fig. 9g and h, green line) and are comparable with *3D specific* contributions, such as effect of hydrogen bonding

(Dedios and Oldfield 1994) that are rare in loops. It is reasonable to conclude that random coil-like chemical shifts in mobile regions originate from chemical shift averaging and are not just a mere result of the absence of *3D specific* de-/shielding. Hence, the upper and lower limits of the RCI time scale are primarily associated with the time scale of fast conformational exchange as discussed above.

## Conclusions and RCI limitations

The RCI method is not without its faults. As discussed below, the RCI method is somewhat limited in (1) the detection of domain movements; (2) the handling of strongly shielded/deshielded residues or nuclei; and (3) the analysis of magnetically aligned proteins. In general, a dynamic process will change an RCI value only if it changes the averaged chemical shift of a particular nucleus. It is certainly possible that certain motions in a protein may increase the range of sampled chemical shifts without changing the averaged shift. In these cases, the RCI will not be sensitive to such motions.

Obviously, the local environment of a given residue or nucleus plays a critical role in the success and sensitivity of the RCI method. Fast conformational exchange will only alter the chemical shift distribution if it significantly changes nuclear shielding. A complex interplay between opposing shielding/de-shielding effects may result in rigid protein regions ( $\alpha$ -helices or  $\beta$ -sheets) displaying small (loop-like) secondary chemical shifts. A recent review of residue-specific secondary chemical shifts revealed that a significant number of residues have relatively small characteristic secondary shifts for N, NH, and  $C\beta$  nuclei in  $\alpha$ -helices or  $\beta$ -sheets (Wang and Jardetzky 2002a, b). For these residues, the RCI method may be able to predict flexibility only if the chemical shifts are affected by additional long-range shielding. To counter these effects, we recommend the inclusion of  $C\alpha$ , CO, and  $H\alpha$  shifts in RCI calculations to obtain more reliable results.

Certain types of correlated dynamics, such as domain movements or concerted motions of large protein segments, may not result in significant changes in the local structure and environment of moving residues (e.g., residues in the hydrophobic core of a domain). Such motions will not be detected by RCI. However, this can also be an advantage, if one is interested to separate correlated motions and local dynamics (see an example of HIV-1 Gag analysis above).

If both a nucleus and an environmental element, which has a dominant contribution into nucleus chemical shift (e.g., an aromatic ring, a side-chain charge, and certain types of spin-labels), experience the same motional event,

this dynamic process will likely fail to change the chemical shift (and hence the RCI) despite fluctuations of local structure. On the other hand, if a residue in a rigid area is in close proximity to a flexible protein segment with high shielding/de-shielding capabilities (e.g., N- and C-termini, aromatic and charged side-chains), motions of that segment may result in “coil-like” large amplitudes of fast conformational exchange of the rigid residue and elevated RCI values. In such a case, the RCI may incorrectly identify the rigid residue as a mobile one. The use of multiple chemical shifts (which are sensitive to different environmental factors) in combination with data smoothing are employed in the RCI protocol to decrease the likelihood of such occurrences.

Some caution should be exercised when interpreting the RCI values of magnetically aligned proteins. Incomplete averaging of the chemical shift anisotropy component of the Hamiltonian may result in significant chemical shift offsets. For example, it was shown that CO chemical shifts of leucine enkephalin may change by more than 3.0 ppm as the degree of alignment increases (Sanders and Landis 1994). Nitrogen shielding is expected to vary due to changes of peptide plane orientation with respect to the external magnetic field as much as 200 ppm (Case 1998). In such cases, efforts should be made to assess the effect of magnetic alignment on chemical shifts and, if necessary, to predict the values of corresponding isotropic chemical shifts prior to any RCI calculations.

Despite these caveats, we believe that the RCI method represents a very simple and robust addition to traditional methods of studying protein flexibility. While it cannot substitute for the actual measurement of relaxation parameters, it is particularly useful for gaining insights into protein dynamics in the absence of these data and for comparison with other kinds of experimentally or computationally acquired dynamics data.

**Acknowledgements** This work was supported by the Natural Sciences and Engineering Research Council (NSERC), the National Research Council’s National Institute for Nanotechnology (NINT), the Protein Engineering Network of Centres of Excellence (PENEC), Alberta Prion Research Institute, and PrioNet Canada.

## References

- Baber JL, Levens D, Libutti D et al (2000) Chemical shift mapped DNA-binding sites and <sup>15</sup>N relaxation analysis of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K. *Biochemistry* 39:6022–6032
- Berjanskii MV, Wishart DS (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc* 127:14970–14971
- Berjanskii M, Wishart DS (2006) NMR: prediction of protein flexibility. *Nature Protocols* 1:683–688
- Berjanskii MV, Riley MI, Xie A et al (2000) NMR structure of the N-terminal J domain of polyomavirus T antigens: implications for DnaJ-like domains and for T antigen mutations. *J Biol Chem* 275:36094–36103
- Berjanskii M, Riley M, Van Doren SR (2002) Hsc70-interacting HPD loop of the J domain of polyomavirus T antigens fluctuates in ps to ns and micros to ms. *J Mol Biol* 321:503–516
- Braun D, Wider G, Wuthrich K (1994) Sequence-corrected N-15 random coil chemical-shifts. *J Am Chem Soc* 116:8466–8469
- Bundi A, Wuthrich K (1979) H-1-NMR parameters of the common amino-acid residues measured in aqueous-solutions of the linear tetrapeptides H-GLY-GLY-X-L-ALA-OH. *Biopolymers* 18:285–297
- Bundi A, Grathwohl C, Hochmann J et al (1975) Proton NMR of protected tetrapeptides TFA-GLY-GLY-L-X-L-ALA-OCH<sub>3</sub>, where X stands for one of 20 common amino-acids. *J Magn Reson* 18:191–198
- Carugo O, Argos P (1999) Reliability of atomic displacement parameters in protein crystal structures. *Acta Crystallogr D Biol Crystallogr* 55(Pt 2):473–478
- Case DA (1998) The use of chemical shifts and their anisotropies in biomolecular structure determination. *Curr Opin Struct Biol* 8:624–630
- Case DA (2002) Molecular dynamics and NMR spin relaxation in proteins. *Acc Chem Res* 35:325–331
- Chang SL, Tjandra N (2001) Analysis of NMR relaxation data of biomolecules with slow domain motions using wobble-in-a-cone approximation. *J Am Chem Soc* 123:11484–11485
- Chen J, Brooks CL 3rd, Wright PE (2004) Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation. *J Biomol NMR* 29:243–257
- Chou YT, Swain JF, Gierasch LM (2002) Functionally significant mobile regions of Escherichia coli SecA ATPase identified by NMR. *J Biol Chem* 277:50985–50990
- Clore GM, Szabo A, Bax A et al (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J Am Chem Soc* 112:4989–4991
- Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289–302
- Dalgarno DC, Levine BA, Williams RJ (1983) Structural information from NMR secondary chemical shifts of peptide alpha C-H protons in proteins. *Biosci Rep* 3:443–452
- Damberg P, Jarvet J, Graslund A (2005) Limited variations in <sup>15</sup>N CSA magnitudes and orientations in ubiquitin are revealed by joint analysis of longitudinal and transverse NMR relaxation. *J Am Chem Soc* 127:1995–2005
- Dedios AC, Oldfield E (1994) Chemical-shifts of carbonyl carbons in peptides and proteins. *J Am Chem Soc* 116:11485–11488
- Dyson HJ, Wright PE (1998) Equilibrium NMR studies of unfolded and partially folded proteins. *Nat Struct Biol* 5(Suppl):499–503
- Eghbalian HR, Wang L, Bahrami A et al (2005) Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements. *J Biomol NMR* 32:71–81
- Elofsson A, Nilsson L (1993) How consistent are molecular-dynamics simulations - comparing structure and dynamics in reduced and oxidized Escherichia-Coli thioredoxin. *J Mol Biol* 233:766–780
- Farrow NA, Zhang O, Forman-Kay JD et al (1997) Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry* 36:2390–2402
- Fiaux J, Bertelsen EB, Horwich AL et al (2002) NMR analysis of a 900K GroEL GroES complex. *Nature* 418:207–211



- Fushman D, Cahill S, Cowburn D (1997) The main-chain dynamics of the dynamin pleckstrin homology (Ph) domain in solution – analysis of N-15 relaxation with monomer/dimer equilibration. *J Mol Biol* 266:173–194
- Glushka J, Lee M, Coffin S et al (1989) N-15 chemical-shifts of backbone amides in bovine pancreatic trypsin-Inhibitor and apamin. *J Am Chem Soc* 111:7716–7722
- Grathwoh C, Wuthrich K (1974) C-13 NMR of protected tetrapeptides TFA-GLY-GLY-L-X-L-ALA-OCH<sub>3</sub>, where X stands for 20 common amino-acids. *J Magn Reson* 13:217–225
- Halle B (2002) Flexibility and packing in proteins. *Proc Natl Acad Sci USA* 99:1274–1279
- Horita DA, Zhang W, Smithgall TE et al (2000) Dynamics of the Hck-SH3 domain: comparison of experiment with multiple molecular dynamics simulations. *Protein Sci* 9:95–103
- Huang X, Peng JW, Speck NA et al (1999) Solution structure of core binding factor beta and map of the CBF alpha binding site. *Nat Struct Biol* 6:624–627
- Ishima R, Torchia DA (2000) Protein dynamics from NMR. *Nat Struct Biol* 7:740–743
- Jin DQ, Andrec M, Montelione GT et al (1998) Propagation of experimental uncertainties using the Lipari-Szabo model-free analysis of protein dynamics. *J Biomol NMR* 12:471–492
- Kay LE (1998) Protein dynamics from NMR. *Nat Struct Biol* 5(Suppl):513–517
- Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graphics* 14:51–55
- Korzhev DM, Orekhov VY, Arseniev AS (1997) Model-free approach beyond the borders of its applicability. *J Magn Reson* 127:184–191
- Lacroix E, Bruix M, Lopez-Hernandez E et al (1997) Amide hydrogen exchange and internal dynamics in the chemotactic protein CheY from *Escherichia coli*. *J Mol Biol* 271:472–487
- Le HB, Oldfield E (1996) Ab initio studies of amide-N-15 chemical shifts in dipeptides: Applications to protein NMR spectroscopy. *J Phys Chem* 100:16423–16428
- Lecroisey A, Martineau P, Hofnung M et al (1997) NMR studies on the flexibility of the poliovirus C3 linear epitope inserted into different sites of the maltose-binding protein. *J Biol Chem* 272:362–368
- Levitt MH (2001) Spin dynamics: basics of nuclear magnetic resonance. Wiley, Chichester
- Lindahl E, Hess B, van der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 7:306–317
- Lindorff-Larsen K, Best RB, Depristo MA et al (2005) Simultaneous determination of protein structure and dynamics. *Nature* 433:128–132
- Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* 104:4546–4559
- Lukin JA, Gove AP, Talukdar SN et al (1997) Automated probabilistic method for assigning backbone resonances of (C-13,N-15)-labeled proteins. *J Biomol NMR* 9:151–166
- McConnell HM (1958) Reaction Rates by Nuclear Magnetic Resonance. *J Chem Phys* 28:430–431
- Merutka G, Dyson HJ, Wright PE (1995) Random Coil H-1 Chemical-Shifts Obtained as a Function of Temperature and Trifluoroethanol Concentration for the Peptide Series Gxg<sub>2</sub>g. *J Biomol NMR* 5:14–24
- Mielke SP, Krishnan VV (2004) An evaluation of chemical shift index-based secondary structure determination in proteins: influence of random coil chemical shifts. *J Biomol NMR* 30:143–153
- Neal S, Nip AM, Zhang H et al (2003) Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. *J Biomol NMR* 26:215–240
- Oldfield E (1995) Chemical shifts and three-dimensional protein structures. *J Biomol NMR* 5:217–225
- Osapay K, Case DA (1994) Analysis of proton chemical shifts in regular secondary structure of proteins. *J Biomol NMR* 4: 215–230
- Palmer AG 3rd (2001) NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu Rev Biophys Biomol Struct* 30:129–155
- Palmer AG 3rd, Kroenke CD, Loria JP (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* 339:204–238
- Pang Y, Buck M, Zuiderweg ER (2002) Backbone dynamics of the ribonuclease binase active site area using multinuclear ((15)N and (13)CO) NMR relaxation and computational molecular dynamics. *Biochemistry* 41:2655–2666
- Pardi A, Wagner G, Wuthrich K (1983) Protein conformation and proton nuclear-magnetic-resonance chemical shifts. *Eur J Biochem* 137:445–454
- Penkett CJ, Redfield C, Jones JA et al (1998) Structural and dynamical characterization of a biologically active unfolded fibronectin-binding protein from *Staphylococcus aureus*. *Biochemistry* 37:17054–17067
- Petsko GA, Ringe D (1984) Fluctuations in protein structure from X-ray diffraction. *Annu Rev Biophys Bioeng* 13:331–371
- Richarz R, Wuthrich K (1978) C-13 Nmr Chemical-Shifts of Common Amino-Acid Residues Measured in Aqueous-Solutions of Linear Tetrapeptides H-Gly-Gly-X-L-Ala-Oh. *Biopolymers* 17:2133–2141
- Sahu SC, Bhuyan AK, Udgaonkar JB et al (2000) Backbone dynamics of free barnase and its complex with barstar determined by 15N NMR relaxation study. *J Biomol NMR* 18:107–118
- Sanders CR, Landis GC (1994) Facile acquisition and assignment of oriented sample NMR-spectra for bilayer surface-associated proteins. *J Am Chem Soc* 116:6470–6471
- Schwarzinger S, Kroon GJ, Foss TR et al (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* 123:2970–2978
- Schwarzinger S, Kroon GJ, Foss TR et al (2000) Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J Biomol NMR* 18:43–48
- Scott WRP, Hunenberger PH, Tironi IG et al (1999) The GROMOS biomolecular simulation program package. *J Phys Chem A* 103:3596–3607
- Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and C-alpha and C-beta C-13 nuclear magnetic resonance chemical shifts. *J Am Chem Soc* 113:5490–5492
- Sun HH, Sanders LK, Oldfield E (2002) Carbon-13 NMR shielding in the twenty common amino acids: Comparisons with experimental results in proteins. *J Am Chem Soc* 124:5486–5495
- Szilagyi L (1995) Chemical-shifts in proteins come of age. *Prog Nucl Magn Reson Spectrosc* 27:325–443
- Thanabal V, Omecinsky DO, Reilly MD et al (1994) The 13C chemical shifts of amino acids in aqueous solution containing organic solvents: application to the secondary structure characterization of peptides in aqueous trifluoroethanol solution. *J Biomol NMR* 4:47–59
- van Gunsteren WF, Mark AE (1998) Validation of molecular dynamics simulation. *J Chem Phys* 108:6109–6116
- Vila JA, Ripoll DR, Baldoni HA et al (2002) Unblocked statistical-coil tetrapeptides and pentapeptides in aqueous solution: a theoretical study. *J Biomol NMR* 24:245–262

- Wand AJ (2001) Dynamic activation of protein function: a view emerging from NMR spectroscopy. *Nat Struct Biol* 8:926–931
- Wang Y, Jardetzky O (2002a) Investigation of the neighboring residue effects on protein chemical shifts. *J Am Chem Soc* 124:14075–14084
- Wang Y, Jardetzky O (2002b) Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* 11:852–861
- Wang Y, Jardetzky O (2004) Predicting  $^{15}\text{N}$  chemical shifts in proteins using the preceding residue-specific individual shielding surfaces from PHI, PSI i-1, and CHI-1 torsion angles. *J Biomol NMR* 28:327–340
- Wang Y, Wishart DS (2005) A simple method to adjust inconsistently referenced  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shift assignments of proteins. *J Biomol NMR* 31:143–148
- Wang T, Cai S, Zuiderweg ER (2003) Temperature dependence of anisotropic protein backbone dynamics. *J Am Chem Soc* 125: 8639–8643
- Wishart DS, Nip AM (1998) Protein chemical shift analysis: a practical guide. *Biochem Cell Biol* 76:153–163
- Wishart DS, Sykes BD (1994) The  $^{13}\text{C}$  chemical-shift index: a simple method for the identification of protein secondary structure using  $^{13}\text{C}$  chemical-shift data. *J Biomol NMR* 4:171–180
- Wishart DS, Sykes BD (1994) Chemical shifts as a tool for structure determination. *Methods Enzymol* 239:363–392
- Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* 222:311–333
- Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* 31:1647–1651
- Wishart DS, Bigam CG, Holm A et al (1995)  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR* 5:67–81
- Wolf-Watz M, Grundstrom T, Hard T (2001) Structure and backbone dynamics of Apo-CBFbeta in solution. *Biochemistry* 40: 11423–11432
- Xu XP, Case DA (2002) Probing multiple effects on  $^{15}\text{N}$ ,  $^{13}\text{C}$  alpha,  $^{13}\text{C}$  beta, and  $^{13}\text{C}'$  chemical shifts in peptides using density functional theory. *Biopolymers* 65:408–423
- Zhang F, Bruschweiler R (2002) Contact model for the prediction of NMR N-H order parameters in globular proteins. *J Am Chem Soc* 124:12654–12655
- Zhang H, Neal S, Wishart DS (2003) RefDB: a database of uniformly referenced protein chemical shifts. *J Biomol NMR* 25:173–195