

MERA: a webserver for evaluating backbone torsion angle distributions in dynamic and disordered proteins from NMR data

Alexey B. Mantsyzov¹ · Yang Shen² · Jung Ho Lee² · Gerhard Hummer³ · Ad Bax²

Received: 6 June 2015 / Accepted: 20 July 2015 / Published online: 29 July 2015
© Springer Science+Business Media Dordrecht (outside the USA) 2015

Abstract MERA (Maximum Entropy Ramachandran map Analysis from NMR data) is a new webserver that generates residue-by-residue Ramachandran map distributions for disordered proteins or disordered regions in proteins on the basis of experimental NMR parameters. As input data, the program currently utilizes up to 12 different parameters. These include three different types of short-range NOEs, three types of backbone chemical shifts (^{15}N , $^{13}\text{C}^\alpha$, and $^{13}\text{C}'$), six types of J couplings ($^3\text{J}_{\text{HNH}\alpha}$, $^3\text{J}_{\text{C}'\text{C}'}$, $^3\text{J}_{\text{C}'\text{H}\alpha}$, $^1\text{J}_{\text{H}\alpha\text{C}\alpha}$, $^2\text{J}_{\text{C}\alpha\text{N}}$ and $^1\text{J}_{\text{C}\alpha\text{N}}$), as well as the ^{15}N -relaxation derived $J(0)$ spectral density. The Ramachandran map distributions are reported in terms of populations of their $15^\circ \times 15^\circ$ voxels, and an adjustable maximum entropy weight factor is available to ensure that the obtained distributions will not deviate more from a newly derived coil library distribution than required to account for the experimental data. MERA output includes the agreement between each input parameter and its distribution-derived value. As an application, we demonstrate performance of the program for several residues in the intrinsically disordered protein α -synuclein, as well as for several static and dynamic residues in the folded protein GB3.

Keywords Coil library · IDP · Karplus curve · Random coil · Short-range NOE · α -Synuclein

Introduction

It is estimated that more than 30 % of the mammalian genome codes for intrinsically disordered proteins (IDPs) or proteins with large intrinsically disordered regions (IDRs) (Dyson and Wright 2005; Sickmeier et al. 2007; Uversky and Dunker 2010; van der Lee et al. 2014). Inherently, such proteins are poorly suited for detailed structural analysis by X-ray crystallography, but the intrinsic conformational propensities of the disordered backbone are amenable to solution state NMR studies (Dyson and Wright 2004; Shi et al. 2006; Mittag and Forman-Kay 2007; Mittag et al. 2010; Ball et al. 2011; Rezaei-Ghaleh et al. 2012). Such NMR studies build on extensive prior work that analyzed backbone torsion angle propensities in synthetic peptides, and their analysis in terms of ^1H - ^1H NOE data, chemical shifts, as well as homo- and heteronuclear J couplings (Dyson and Wright 1991; Smith et al. 1996; Long and Tycko 1998; Baldwin and Rose 1999; Graf et al. 2007; Hagarman et al. 2010). With long-range NOEs, corresponding to close contacts between residues that are not proximate in the amino acid sequence, usually being unobservable in such proteins, the non-local distance restraint information is often limited to paramagnetic relaxation effects (Bertoncini et al. 2007) and/or small angle X-ray scattering data (Bernado et al. 2007), both of which can be challenging to interpret in quantitative structural terms. Nevertheless, innovative approaches have been introduced to provide a pictorial representation of the structure of IDPs or IDRs in terms of ensemble representations (Varadi et al. 2014). Among

Alexey B. Mantsyzov and Yang Shen have contributed equally.

✉ Ad Bax
bax@nih.gov

¹ Faculty of Fundamental Medicine, M.V. Lomonosov Moscow State University, Moscow, Russian Federation 119991

² Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

³ Max Planck Institute of Biophysics, Max-von-Laue-Straße 3, 60438 Frankfurt am Main, Germany

these methods, the ENSEMBLE program (Krzeminski et al. 2013) and the ASTEROIDS program (Salmon et al. 2010), which select suitable subsets of models from much larger ensembles, have become quite widely used. However, with few if any of the experimental parameters reporting on the correlation between the backbone torsion angles sampled at any given point in time by distant residues i and j , the a priori size of the conformational ensemble grows exponentially with the number N of residues. Even under the tight restriction of allowing only three dominant conformers per residue, a candidate ensemble would contain 3^N conformers overall in a crude representation of the space sampled by the protein. Clearly, the number of experimental parameters is vanishingly small compared to the number of possible conformers whose populations need to be defined by the data, making this a hugely under determined problem.

Instead, we here simply focus on deriving the distributions of the backbone ϕ/ψ angles of individual residues on the basis of the strictly local NMR parameters, i.e., backbone chemical shifts, J couplings, and intra-residue and sequential NOE data. For this purpose, we divide the $360^\circ \times 360^\circ$ Ramachandran space into 24×24 voxels of $15^\circ \times 15^\circ$ each. As we have shown previously (Mantsyzov et al. 2014), 10 independent NMR parameters are readily available for this purpose. Relative to our earlier study (Mantsyzov et al. 2014), two additional restraints are now also available. First, ${}^3J_{C'C}(i-1, i)$, which can be measured at very high precision in IDPs and IDRs, and which follows a very tight Karplus relation (Li et al. 2015) has been added as a possible restraint. The Karplus curve for this coupling is 60° out of phase relative to ${}^3J_{HNH\alpha}$, and ${}^3J_{C'C}$ therefore carries independent restraining information (Lee et al. 2015). The second, newly added and very useful backbone torsion angle restraint is ${}^3J_{C'H\alpha}$ (Wang and Bax 1996). Although its Karplus relation is 180° out of phase relative to ${}^3J_{HNH\alpha}$, its actual Karplus coefficients are very different and lead to a large ($> \sim 6.5$ Hz) value for conformations with $\phi \approx +60^\circ$ values in the α_L region, whereas for the most populated region of negative ϕ values, this coupling is restricted to a narrow range of ~ 1 – 3 Hz (Wang and Bax 1996). This coupling is therefore particularly important for defining the α_L population of Ramachandran space.

Deriving Ramachandran map distributions from NMR data

The up to 12 parameters per residue mentioned above clearly remain insufficient for defining the populations of 576 voxels. However, two additional approximations can be made that permit us to solve this problem. First, we use

the common sense approximation that the distributions sampled in solution must bear some similarity to that observed in the non-secondary structure regions of crystallized proteins, i.e. the coil database. Also, depending on residue type, in this coil database only ca 100–120 of the $15^\circ \times 15^\circ$ voxels have non-negligible populations (about double that for Gly residues, which are not considered in our study). By setting the weights of the remaining ca 460 voxels to zero, we can strongly reduce the number of adjustable parameters or weights, w_k ($\sum_k w_k = 1$), when searching for optimal agreement between the calculated weight-averaged NMR constraints, $I_{calc}(q)$, for any given residue in the sequence and the corresponding experimental observables $I_{exp}(q)$. Here, $I_{exp}(q)$ is any of the q ($q \leq 12$) types of different NMR parameters, and $I_{calc}(q)$ is given by

$$I_{calc}(q) = \sum_{k=1}^{N_c} w_k I_k(q) \quad (1)$$

where the summation extends over all N_c voxels that have non-vanishing populations in the coil database and $I_k(q)$ is the value calculated for parameter q at the center of voxel, k .

Development of a random coil structural database

In our earlier work (Mantsyzov et al. 2014), we used the coil library of Fitzkee et al. (2005), which excluded secondary-structured regions based on a “mesostate” evaluation. This coil database served as a reference when calculating the NMR-derived ϕ/ψ distribution, and for deriving the minimal change relative to this database distribution needed to satisfy the NMR restraints. Here, we constructed an updated analogous coil library which includes fragments that are at least three residues in length and are not subject to intramolecular backbone–backbone H-bonding (as defined by an H-bond energy cut-off energy < -0.7 kcal/mol). As input to this coil library, we used all X-ray structures solved at a resolution ≤ 2.0 Å, with an R factor lower than 23 %, and a maximum pairwise sequence identity between coil database proteins of 50 %. A comparison of the newly generated coil library with the original Fitzkee coil library shows a decrease in the population of backbone torsion angles that are commonly found in H-bonded tight β turns and small increases in the β and polyproline-II (PPII) regions of Ramachandran space (Fig. 1). Excluding the terminal residues of the selected fragments, the new coil database contains torsion angles for 195,859 residues. The shift in our coil library distribution relative to the widely used Fitzkee library (Fig. 1) is not surprising, considering that the Fitzkee library was based primarily on excluding α -helical and β -sheet residues,

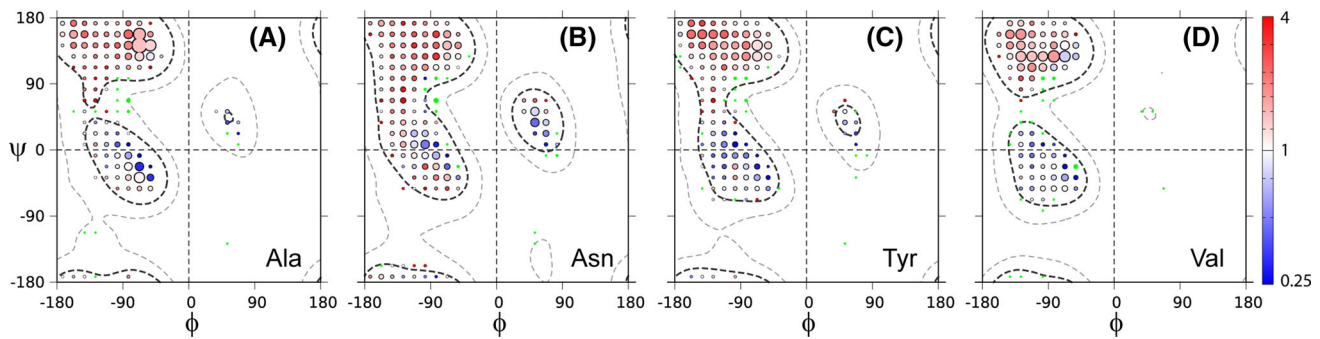


Fig. 1 Backbone torsion angle distributions in the newly generated coil library, illustrated for **A** Ala, **B** Asn, **C** Tyr, and **D** Val. The surface area of each circle is proportional to the population of its $15^\circ \times 15^\circ$ voxel, and its color reflects the ratio relative to that the population seen in the Fitzkee coil library (Fitzkee et al. 2005). Voxels that fall below the 0.1 % population threshold in the newly generated coil library, but observed in the Fitzkee coil library are shown with a size that corresponds to their population in the Fitzkee coil library (green). For

whereas our library additionally eliminates small structured elements such as H-bonded turns. As will be shown below, a smaller deviation from the new coil library generally is needed to obtain a comparable fit to the experimental NMR parameters when evaluating a highly disordered protein such as α -synuclein, suggesting that H-bonded tight turns in highly disordered proteins have a lower population than observed in the Fitzkee coil library.

Generation of the combined experimental restraint function

For each voxel k in the coil library, the J coupling values applicable for the center of the voxel are calculated using standard Karplus equations. Considering that the empirically parameterized Karplus curves already include the effects of small backbone angle fluctuations (Bruschweiler and Case 1994), we use Karplus curves for ${}^3J_{\text{HNH}\alpha}$, ${}^3J_{\text{C}'\text{C}'}$ and ${}^3J_{\text{C}'\text{H}\alpha}$ couplings where the effect of such motions has been factored out (Lee et al. 2015). A similar adjustment can be derived for equations that describe the Karplus-like relations between backbone torsion angles and ${}^1J_{\text{H}\alpha\text{C}\alpha}$, ${}^2J_{\text{C}\alpha\text{N}}$ and ${}^1J_{\text{C}\alpha\text{N}}$ couplings (Vuister et al. 1993; Ding and Gronenborn 2004; Wirmer and Schwalbe 2002; Mantsyzov et al. 2014). However, considering that the rmsd between observed couplings in folded proteins and their corresponding best-fitted Karplus curves is much larger than the effect of any dynamics correction to these Karplus-like equations, no such corrections were made.

Calculating the expected NOE at the center of each voxel is not entirely straightforward as (ϕ/ψ) -dependent small deviations from ideal bond angles are often observed in high resolution X-ray structures (MacArthur and Thornton 1996). As a result, the intra-residue $\text{H}_i^{\text{N}}-\text{H}_i^{\alpha}$

each amino acid type, the favorably and generously allowed ϕ/ψ conformational regions are defined as those with a residue density $d(\phi, \psi)$ above thresholds of 1 and 0.1 %, respectively, in the newly generated coil library. Their boundaries are marked by *dark* and *light dashed lines*, respectively. The normalized residue density, $d(\phi, \psi)$, is derived by convolution of each of the ϕ_k/ψ_k coil database entries with a Gaussian function, $\exp(-((\phi - \phi_k)^2 + (\psi - \psi_k)^2)/450)$, analogous to previous work (Shen and Bax 2013)

distance does not only depend on ϕ_i but also on adjacent torsion angles, and similarly the $\text{H}_i^{\alpha}-\text{H}_{i+1}^{\text{N}}$ distance does not solely depend on ψ_i . In practice, we find that hydrogen atoms built on to the X-ray structure backbone by the program MOLMOL (Koradi et al. 1996) yield somewhat better agreement when evaluating experimental residual dipolar couplings, which are very sensitive to the correct placement of H atoms, than other programs that can be used for this purpose. The representative $\text{H}_i^{\text{N}}-\text{H}_i^{\alpha}$, $\text{H}_i^{\alpha}-\text{H}_{i+1}^{\text{N}}$, and $\text{H}_i^{\text{N}}-\text{H}_{i+1}^{\text{N}}$ distances for voxel k of residue i are then calculated by averaging the corresponding distances for all coil database residues in that voxel, using MOLMOL-added H atoms.

In order to convert the r_{HH} values into NOE cross-relaxation rates, we require $J(0)$ spectral density values derived from reduced spectral density mapping analysis of ${}^{15}\text{N}$ relaxation data (Farrow et al. 1995), and this approach was used in our earlier work (Mantsyzov et al. 2014). Alternatively, as we previously have shown that the intraresidue $\text{H}^{\text{N}}-\text{H}^{\alpha}$ NOE correlates tightly with ${}^{15}\text{N}$ -derived $J(0)$ values (Mantsyzov et al. 2014), this intraresidue NOE can also be used as an internal reference for $r_{\text{H}\alpha, \text{HN}}(i-1, i)$ and $r_{\text{HNHN}}(i, i+1)$. This new option, not used during our previous analysis of α -synuclein, is aimed at reducing the amount of input data required for MERA. For this purpose, we use the intraresidue $\langle r_{\text{HNH}\alpha}^{-6} \rangle^{-1/6} = 2.91 \pm 0.03 \text{ \AA}$ as an internal reference, where the 2.91- \AA and its standard deviation are derived from the Ramachandran distributions obtained by MERA for α -synuclein when using $J(0)$ spectral densities to calculate the ${}^1\text{H}-{}^1\text{H}$ NOEs. Our $\langle r_{\text{HNH}\alpha}^{-6} \rangle^{-1/6} = 2.91\text{-}\text{\AA}$ assumption could potentially lead to erroneous $r_{\text{HNH}\alpha}(i, i-1)$ and $r_{\text{HNHN}}(i, i+1)$ distance restraints if residue i had a

substantial positive ϕ angle population (where $r_{\text{HNH}\alpha}(i, i) \approx 2.3 \text{ \AA}$), and the user is alerted to this possibility when MERA encounters unusually low ratios for the sequential to intraresidue NOE intensities, $d_{\text{H}\alpha\text{HN}}(i-1, i)/d_{\text{HNH}\alpha}(i, i) \leq \sim 1.5$. Substantial positive ϕ angle populations (>20 %) often can also be recognized from below average (< 140 Hz) $^1J_{\text{C}\alpha\text{H}\alpha}$ values, and above average (>3 Hz) $^3J_{\text{C}'\text{H}\alpha}$ values.

The ^{15}N , $^{13}\text{C}^\alpha$ and $^{13}\text{C}'$ chemical shifts for each voxel k in the coil library are calculated from the sum of the residue-specific random coil values, corrected for through-bond nearest neighbor effects (Wang and Jardetzky 2002), and a ϕ/ψ dependence taken from the empirically parameterized program SPARTA (Shen and Bax 2007).

Determining the Ramachandran map distribution sampled by any given residue then is equivalent to assigning weights, w_k ($\sum_k w_k = 1$), to the voxels such that the sum over the calculated NMR-parameters optimally matches the experimental data, i.e., minimizing the normalized χ^2 function:

$$\chi^2 = \frac{1}{N_q} \sum_{q=1}^{N_q} (I_{\text{calc}}(q) - I_{\text{exp}}(q))^2 / \sigma^2(q) \quad (2)$$

Here, we assume for simplicity that the N_q different NMR parameters q are statistically independent. The value for the uncertainty in any given parameter, $\sigma(q)$, determines the effective weight of each parameter q and the choice of their values will be briefly discussed below. For deriving $I_{\text{calc}}(q)$ from Eq. (1), the value $I_k(q)$ is the value calculated for the center position of voxel k . $I_{\text{exp}}(q)$ is the experimentally observed value for parameter q , i.e., the J coupling, chemical shift, or ^1H - ^1H cross-relaxation rate.

Analysis of the molecular dynamics trajectory of the 40-residue N-terminal segment of α -synuclein previously showed that the $J(0)$ spectral density, dominating the ^1H - ^1H NOE buildup, is quite anisotropic (Mantsyzov et al. 2014). Similarly, the $d_{\text{HNH}\alpha}(i, i)$ NOE data in a hexapeptide segment of α -synuclein exhibit the zero-NOE condition ($\omega_{\text{H}}\tau_c \sim 1.1$) at a higher field than the sequential $d_{\text{H}\alpha\text{HN}}(i-1, i)$ NOE, indicative of diffusion anisotropy (Ying et al. 2014). To a first approximation, the anisotropy of $J(0)$ applicable for the sequential $d_{\text{HNH}\alpha}(i, i-1)$ NOE in an IDP can be accounted for by writing (Mantsyzov et al. 2014):

$$J(0) = J(0)_{\text{exp}} \left\{ 1 - 2ke^{-\left[\frac{180}{W}\right]^2} + ke^{-\left[\frac{\psi-120}{W}\right]^2} + ke^{-\left[\frac{\psi+240}{W}\right]^2} \right\} \quad (3)$$

Here, and in the following, all angles are in degrees. $J(0)_{\text{exp}}$ is the ^{15}N -derived $J(0)$ value, $1+k$ reflects the effective spectral density ratio for a $\text{H}_{i-1}^\alpha\text{-H}_i^\text{N}$ vector parallel to the chain direction ($\psi \approx 120^\circ$) over that for a $\text{H}_{i-1}^\alpha\text{-}$

H_i^N vector that is at the magic angle relative the $\text{C}_{i-1}^\alpha\text{-C}_i^\alpha$ direction, as approximately applies for α -helical residues. The parameter W defines the steepness at which the effective $J(0)$ varies with ψ , and was empirically adjusted to 90° . The spectral density for $d_{\text{HNH}\alpha}(i, i)$ and $d_{\text{HNHN}}(i, i+1)$ NOEs, both corresponding to interproton vectors at large angles relative to the $\text{C}_{i-1}^\alpha\text{-C}_i^\alpha$ chain direction, do not require any scaling relative to $J(0)_{\text{exp}}$ because the ^{15}N - ^1H dipolar interaction, which dominates ^{15}N relaxation, is also oriented at a large angle relative to the chain direction. The anisotropy of the chain dynamics is taken into account by MERA when calculating the $d_{\text{HNH}\alpha}(i, i-1)$ NOE contribution for each voxel. The exponential pre-factor in Eq. (3) was empirically adjusted to $k=1$, which appears suitable for IDPs, but possibly could be smaller for IDRs and other less dynamic systems, and be reduced to zero for a well-ordered isotropically tumbling protein. The MERA input file therefore provides the option to adjust this parameter relative to its default setting of 1.0.

Values of the uncertainty parameters, $\sigma(q)$

The values of $\sigma(q)$ in Eq. (2) effectively determine the weight of restraint q in determining the ϕ/ψ distribution. Their optimal values depend both on the error in the experimental measurement of parameter q , and on how accurately $I_k(q)$ can be predicted for voxel k (cf Eq. 1). Below we discuss our choices of the default $\sigma(q)$ parameters, but note that these are user-adjustable in the MERA input file.

For the cross relaxation rates, σ is determined by how precisely the NOE buildup rate can be measured, and by the accuracy of Eq. (3) in accounting for diffusion anisotropy as well as the validity of neglecting the high frequency spectral density terms, $J(2\omega_{\text{H}})$. The latter were previously shown to be very small in α -synuclein (Mantsyzov et al. 2014), but could be significant in smaller proteins, at lower magnetic fields, or at higher temperatures than were used for that protein. Empirically, in order to obtain a comparable contribution to χ^2 (Eq. 2) as from the other terms, the error in the cross relaxation rate was set to 15 % of the measured rate plus 15 % of the rate predicted for the intraresidue $d_{\text{HNH}\alpha}(i, i)$ NOE. As a result, the typically strong sequential $d_{\text{H}\alpha\text{HN}}(i-1, i)$ NOE cross relaxation rate is assigned a total fractional error of ca 20 %, $d_{\text{HNH}\alpha}(i, i)$ is assigned an error of ca 30 %, and the fractional error for the typically weak $d_{\text{HNHN}}(i, i+1)$ NOE will be larger since it will be dominated by the 15 % fraction of the intraresidue $d_{\text{HNH}\alpha}(i, i)$ NOE.

For chemical shifts, barring wrong assignments or erroneous spectral reference calibration, $\sigma(q)$ is completely dominated by the accuracy at which chemical shifts can be

predicted. For a known set of ϕ/ψ angles in folded proteins, the prediction errors typically are ca 1 ppm for ^{13}C and 2–2.5 ppm for ^{15}N (Han et al. 2011). These relatively large uncertainties reflect small deviations from idealized geometry, variations in H-bond lengths and geometry, sidechain torsion angles, ring current effects, etc., which can increase or decrease the chemical shift relative to the value predicted on the basis of only ϕ and ψ (Shen and Bax 2010). In an IDP, such as α -synuclein, measured chemical shift values cluster very tightly around their random coil values (Maltsev et al. 2012). For example, excluding residues preceding Pro and the C-terminal residue, its 17 Ala $^{13}\text{C}^\alpha$ have a standard deviation of only 0.2 ppm relative to their averaged value (52.6 ppm). Comparably small standard deviations for other residues as well as the $^{13}\text{C}'$ nuclei indicate that the effects of the parameters other than ϕ and ψ have a much smaller net impact on chemical shift values than they do in folded proteins, and a $\sigma(q)$ value much smaller than in folded proteins can be used (Table 1).

For the J coupling parameters, an appropriate $\sigma(q)$ choice must reflect both the uncertainty in their measurement and how tightly they correlate with the backbone torsion angles in proteins of known structure. In folded proteins, the scatter between best-fitted and measured $^1\text{J}_{\text{C}\alpha\text{H}\alpha}$ values equals ~ 1.6 Hz (Vuister et al. 1993), and large rmsd's (~ 0.6 Hz) are also obtained for $^3\text{J}_{\text{H}\text{N}\text{H}\alpha}$ when ignoring the out-of-plane position of the amide proton, which in favorable cases can be measured by RDCs (Maltsev et al. 2014). Similarly, rmsd values between observed and predicted $^1\text{J}_{\text{N}\text{C}\alpha}$ or $^2\text{J}_{\text{N}\text{C}\alpha}$ in folded proteins (~ 0.6 Hz) are dominated by prediction errors, resulting from factors other than the backbone torsion angles. For each of these parameters, much tighter clustering around random coil values is again observed in α -synuclein (standard deviations of ca 0.5 Hz for $^1\text{J}_{\text{C}\alpha\text{H}\alpha}$, 0.2 Hz for $^1\text{J}_{\text{N}\text{C}\alpha}$ or $^2\text{J}_{\text{N}\text{C}\alpha}$) indicating that the impact of factors other

than backbone torsion angles again is much smaller in an IDP than in folded proteins.

The above considerations suggest that the values of $\sigma(q)$ depend on the type of protein studied, requiring very low values for highly disordered proteins such as α -synuclein, but considerably higher values in partially ordered systems that may include significant populations of stable, H-bonded conformers. In our work, we approximately scale the value of $\sigma(q)$ for a given residue by the value of its chemical shift order parameter, RCI-S² (Berjanskii and Wishart 2005) as calculated by the program TALOS-N (Shen and Bax 2013), which ranges from a typical value of ~ 0.3 in IDPs to ≥ 0.7 in short loops in folded proteins. Typical $\sigma(q)$ values used in our work are listed in Table 1, but users may adjust these parameters to accurately reflect the degree of flexibility applicable for the residues studied and the errors in the experimental data. As these $\sigma(q)$ values are global parameters, applicable for all residues in the input file, a user may choose to generate separate input files for residues in highly flexible regions of a protein and in (partially) ordered segments. We note, however, that the final Ramachandran distribution reached by MERA is relatively insensitive to $\sigma(q)$ values and mostly impacts the value of χ^2 . Values of $\sigma(q)$ impact the relative importance of the entropy term, discussed below, used to regularize the minimization of Eq. (2).

Inclusion of a maximum entropy term in the restraint function

With only a dozen or less experimental parameters available to define the populations of over 100 voxels in Ramachandran space, this problem is inherently underdetermined. However, a further reduction can be achieved by using the common sense approximation that the distributions sampled in solution must bear some similarity to

Table 1 Values of $\sigma(q)$ recommended for MERA use

Parameter	IDP (RCI-S ² ≈ 0.3)	(Partially) ordered (RCI-S ² ≥ 0.7)
$\delta(^{13}\text{C}^\alpha)/\text{ppm}$	0.25	1.0
$\delta(^{13}\text{C}')/\text{ppm}$	0.25	1.0
$\delta(^{15}\text{N})/\text{ppm}$	0.8	2.5
$^3\text{J}_{\text{H}\text{N}\text{H}\alpha}/\text{Hz}$	0.20	0.6
$^3\text{J}_{\text{C}'\text{C}'}/\text{Hz}$	0.08	0.12
$^3\text{J}_{\text{C}'\text{H}\alpha}/\text{Hz}$	0.1	0.25
$^1\text{J}_{\text{C}\alpha\text{H}\alpha}/\text{Hz}$	0.4	1.6
$^1\text{J}_{\text{N}\text{C}\alpha}/\text{Hz}$	0.2	0.6
$^2\text{J}_{\text{N}\text{C}\alpha}/\text{Hz}$	0.2	0.6
NOE/%	15 + 15 ^a	15 + 15 ^a

^a The error in the cross relaxation rate is calculated as % of the rate provided plus % of the rate expected for a 2.91-Å $d_{\text{H}\text{N}\text{H}\alpha}(i, i)$ interaction

those observed in the non-secondary structure regions of crystallized proteins, i.e. to the coil database. Therefore, we previously introduced an additional, maximum entropy term (Rozycki et al. 2011) that essentially aims to minimize the deviation from coil library voxel populations:

$$S = - \sum_{k=1}^{N_c} w_k \ln \frac{w_k}{w_k^{(0)}} \quad (4)$$

where $w_k^{(0)}$ corresponds to the fractional population of voxel k in the coil library (Mantsyzov et al. 2014). The effective energy function that is minimized by MERA then is given by:

$$G = \chi^2 - \theta S, \quad (5)$$

where the parameter θ controls the importance of the entropy term relative to that of the experimental data. It is worth noting that the term $-S$ of Eq. (4) is equivalent to the Kullback–Leibler information divergence between the MERA-derived Ramachandran map distribution and that of the coil database (Kullback and Leibler 1951).

If the weight, θ , of the entropy term is set to zero, the minimization procedure may not be able to converge on a unique set of w_k values, even though reaching low χ^2 values. As demonstrated previously (Mantsyzov et al. 2014), it then becomes useful to plot the obtained χ^2 as a function of θ , where θ usually is changed in steps of a factor 2. Plotting χ^2 as a function of θ , or as a function of S , then typically will show an “elbow” upturn in χ^2 and the θ value just prior to this upturn (typically in the 0.4–2 range) is chosen for all further data analysis, i.e., as the set of values that most closely resembles the coil database without significantly sacrificing the agreement with the experimental restraints.

Minimization of the free energy function

Minimization of the “free energy” function, Eq. (5), is accomplished by simulated annealing, followed by Powell minimization. The algorithm searches for the set of voxel populations, w_k , that corresponds to the minimum value of G . Minimization starts from a set of w_k with uniform values. Optionally, for faster convergence the program can be started from the coil database populations, or to test robustness an option to start from random initial populations exists. The latter option invariably results in essentially indistinguishable differences relative to starting from a uniform distribution, but requires more iterations during the simulated annealing stages of the minimization. In each round of simulated annealing, perturbations Δx_k are added to each w_k in the set. Perturbation Δx_k is randomly generated within defined boundaries and scaled using the temperature parameter, T . The temperature is gradually

decreased in successive steps of the minimization, thus decreasing the perturbation, and the algorithm narrows the sampling of points around the minimum found at the previous step. The method applies the Metropolis criterion for accepting a new state generated: the state is accepted if the new value of the function G is smaller than that from the previous step ($\Delta G < 0$); if $\Delta G \geq 0$, the probability of acceptance of the perturbed state depends on a Boltzmann weight, $e^{(\Delta G/kT)}$, where k is the Boltzmann constant, but in practice adjusted to a constant value compatible with the typical values encountered for ΔG . The simulated annealing method yields efficient sampling of the energy function landscape and gradually pushes the system towards its global minimum.

Our implementation of the simulated annealing protocol includes 3 separate cooling phases. Phase 1 uses a very fast annealing schedule and 100 steps of minimization, with the perturbation $\Delta x_{k,i}$ for the k -th voxel at the i -th step of minimization and the evolution of the temperature parameter T defined as (Ingber 1989):

$$\Delta x_{k,i} = \text{sgn}(u - 1/2) T_i \left[(1 + 1/T_i)^{2|u-1|} - 1 \right] \quad (6)$$

where $T_i = T_o e^{(-i/2.71)}$, i is the number of the temperature step, T_o is the initial temperature, and u is a random value generated from the uniform distribution in the range $[0, 1]$. Populations $w_{k,i}$ for all voxels are updated simultaneously using the relation $w_{k,i} = |w_{k,i-1} + \Delta x_{k,i}|$. New populations are normalized according to $w_{k,i,norm} = w_{k,i} / \sum_k |w_{k,i}|$. Phases 2 and 3 consist of 500 and 2000 steps each and both use the slow cooling schedule (Szu and Hartley 1987) given by:

$$\Delta x_{k,i} = l \tan(v_{k,i}) / [\sum_k |\tan(v_{k,i})| (1 + i)] \quad (7)$$

where $v_{k,i}$ is a random value generated from a uniform distribution in the range $[-\pi/2, +\pi/2]$ for each voxel and l is the learning rate ($l = 1$ for phase 2; $l = 0.5$ for phase 3). Populations are updated and normalized as in phase 1. The two-fold decrease of the learning rate in phase 3 yields better sampling of points close to the local minimum found at the previous iteration, while phase 2 offers wider sampling. In order to sample the energy landscape efficiently, calculations of perturbations Δx_i are repeated 20 times at each temperature step in all phases.

Although the simulated annealing approach is robust for finding an approximate global minimum of the empirical energy function, it is quite inefficient for reaching the actual minimum of this function. Therefore, after the third phase of simulated annealing has been completed, MERA carries out a Powell minimization to converge to the lowest value in the vicinity of the solution found by the three-stage simulated annealing procedure.

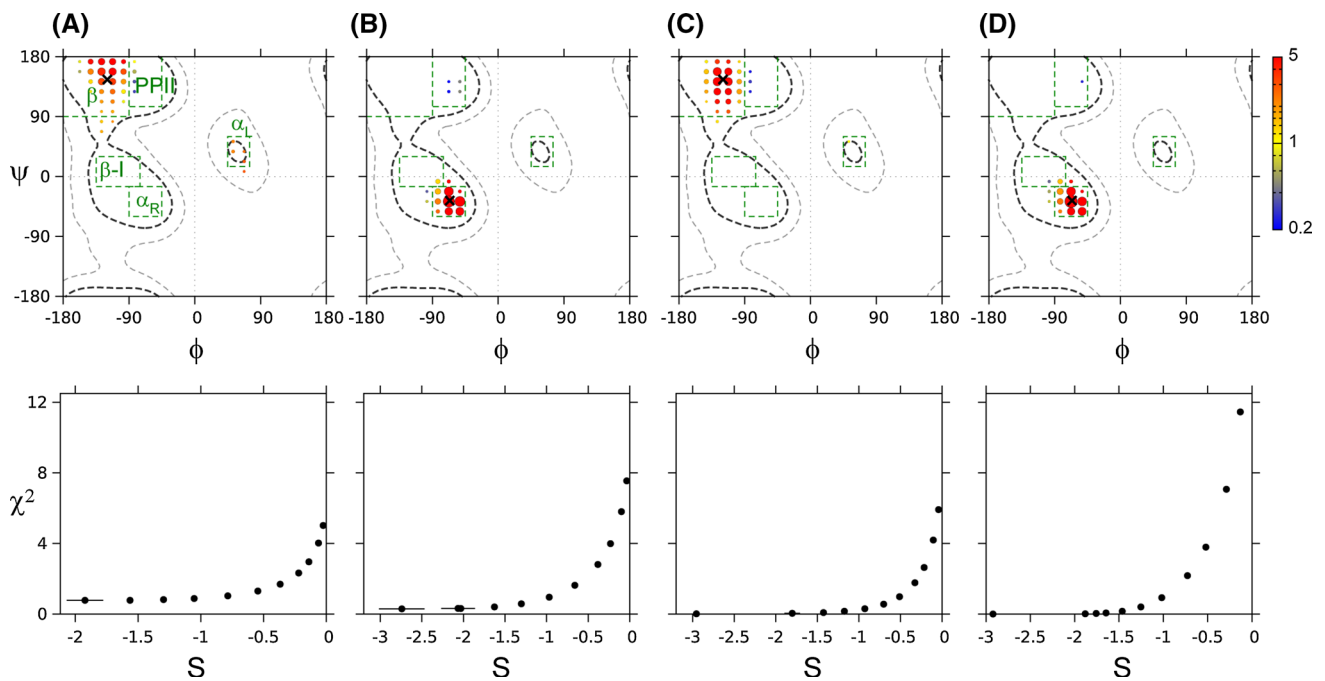


Fig. 2 Examples of MERA-derived ϕ/ψ distributions for well-ordered residues K4 (**A**, **C**) and K31 (**B**, **D**), using experimental input data (**A**, **B**), and ideal simulated input data (**C**, **D**), with the corresponding reference backbone torsion angles (PDB entry 2OED) (Ulmer et al. 2003) marked “x”. The surface area of each circle is proportional to its voxel population, and the color of each circle reflects its fractional deviation from that seen in the coil database. The bottom panels show plots of χ^2 as a function of S , obtained for values of θ of (from left to right) 0, 0.1, 0.2, 0.4, 0.8, 1.6, 3, 6, 10, 20, and 40. The horizontal error bars in the lower panels reflect the lack of convergence of the simulated annealing protocol, seen for very low θ

The program has been written in C++ and can be downloaded from <http://spin.niddk.nih.gov/bax/software/MERA>. The program also runs as a webserver, <http://spin.niddk.nih.gov/bax/nmrserver/mera>, where the user is asked to upload a file with experimentally determined input data, including the estimated uncertainties, and the program emails the results back to the user. The program will run with any number of input parameters, but will return ill-defined results if too few restraints are available, as reflected in large standard deviations on the Kullback–Leibler divergence and near-zero χ^2 values.

Results and discussion

Although MERA is designed for the analysis of the backbone torsion angle distribution of IDPs or of highly flexible regions in a protein, it can also be applied to folded proteins, and potentially is particularly useful for viewing the conformational space sampled by flexible regions in such proteins. Below, we will demonstrate the program for such

values. The value $\theta = 0.4$ is used for deriving the populations in the upper panels. For all analyses, the diffusion anisotropy parameter, k , was set to zero and $\sigma(q)$ values recommended for folded proteins in Table 1 were used. χ^2 values when using the database distributions ($S = 0$) are 7.2 (K4) and 10.8 (K31). Green boxes mark secondary structure regions: β , PPII, α_L , type I β -turn (β -I) and α_R (see green labels in A). Experimental input data for K4 and K31 include all six types of J couplings ($^3J_{\text{HNH}\alpha}$, $^3J_{\text{C}'\text{C}'}$, $^3J_{\text{C}'\text{H}\alpha}$, $^1J_{\text{C}\alpha\text{H}\alpha}$, $^2J_{\text{NC}\alpha}$ and $^1J_{\text{NC}\alpha}$), three types of chemical shifts (^{15}N , $^{13}\text{C}\alpha$, and $^{13}\text{C}'$), and three types of short-range ^1H - ^1H NOEs

applications, as well as for several residues in the previously extensively studied IDP α -synuclein.

Application to ordered and dynamic residues in protein GB3

Nearly complete input data are available for the small, globular protein GB3. This protein contains a Greek key motif β -sheet, with one long α -helix (A23-N37) separating strands β_2 and β_3 (Derrick and Wigley 1994). The amide groups of several of GB3's residues, incl. L12 and G41, exhibit strongly elevated backbone dynamics as judged from ^{15}N NMR relaxation (Hall and Fushman 2003) and RDC analyses (Yao et al. 2008). Here, we show the results of MERA analysis for L12 and D40, as well as for two well-structured residues, K4, located in the first β -strand, and K31 in the middle of GB3's α -helix.

As expected, for both the highly structured residues, K4 and K31, a good fit to the experimental input parameters corresponds to quite high values ($> \sim 1$) of the Kullback–Leibler divergence, S . When increasing the weight of the

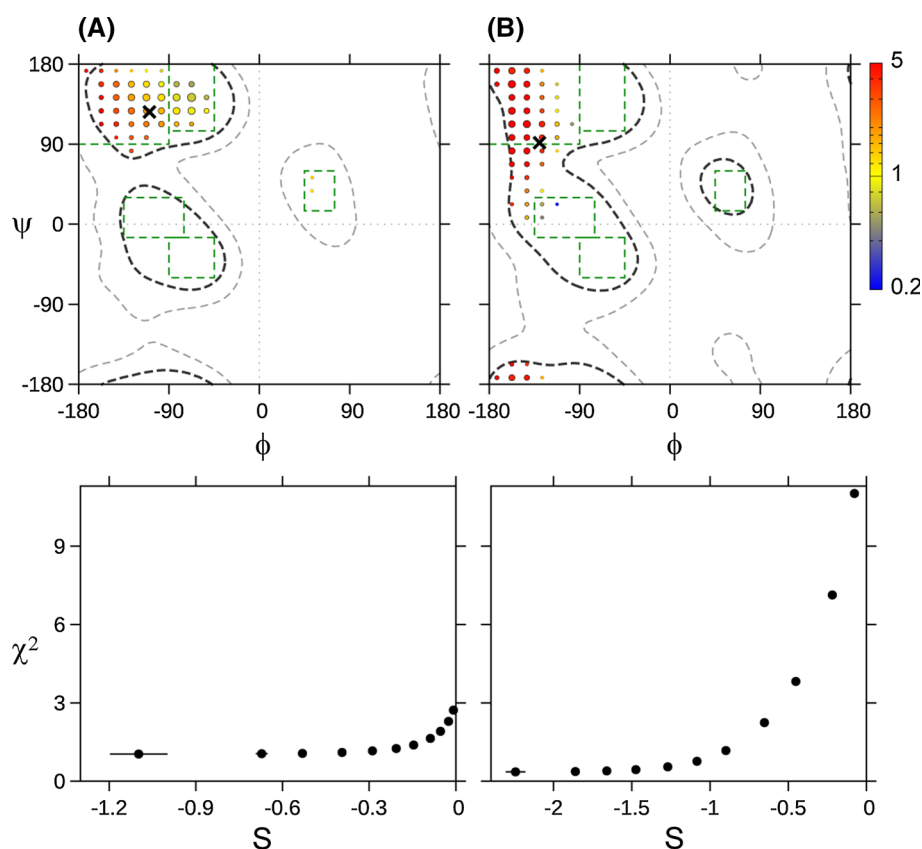


Fig. 3 Examples of MERA-derived ϕ/ψ distributions for dynamically disordered residues L12 (A) and D40 (B) in GB3. The corresponding averaged NMR-derived backbone torsion angles (PDB entry 2OED) (Ulmer et al. 2003) are marked “ \times ”. The bottom panels show plots of χ^2 as a function of S , obtained for values of θ of (from left to right) 0, 0.1, 0.2, 0.4, 0.8, 1.6, 3, 6, 10, and 20. The θ value of 0.4 is used for deriving the populations shown in the upper panels. For all analyses, the diffusion anisotropy parameter, k , was set

to zero. χ^2 values when using the database distributions ($S = 0$) are 3.4 (L12) and 17.7 (D40). The horizontal error bars in the lower panels reflect the lack of convergence of the simulated annealing protocol for very low θ values. Experimental input data for L12 and D40 include all six types of J couplings ($^3J_{\text{HNH}\alpha}$, $^3J_{\text{C}'\text{C}'}$, $^3J_{\text{C}'\text{H}\alpha}$, $^1J_{\text{C}\alpha\text{H}\alpha}$, $^2J_{\text{NC}\alpha}$ and $^1J_{\text{NC}\alpha}$), three types of chemical shifts (^{15}N , $^{13}\text{C}^\alpha$, and $^{13}\text{C}'$), and three types of short-range ^1H - ^1H NOEs

database, θ , the χ^2 value rapidly rises (Fig. 2A, B), confirming that the torsion angle backbone distributions strongly deviate from those seen in the coil library. Indeed narrow clusters of backbone torsion angles are observed for both of these residues, centered around the angles of the PDB X-ray (Derrick and Wigley 1994) and RDC-refined NMR (Ulmer et al. 2003) structures (Fig. 2A, B). It is important to realize that MERA searches for the widest distribution (maximum entropy solution) that is compatible with the experimental data.

Because the individual chemical shift and J coupling parameters, to first order, vary linearly with ϕ and ψ , values predicted for a narrow cluster of ϕ/ψ angles will be very close to those calculated for the center of that cluster. So, even for idealized input parameters, simulated for a single static conformation, the analysis will return a (relatively narrow) distribution of backbone torsion angles (Fig. 2C, D).

For L12 and D40, MERA finds distributions of ϕ/ψ angles compatible with the experimental input data that are much wider than found above for the highly structured residues (Fig. 3). Although the amide group of D40 is relatively well ordered as judged from ^{15}N relaxation (Hall and Fushman 2003) and RDC analyses (Yao et al. 2008), the G41 ^{15}N - ^1H was found to be highly dynamic. Results obtained here for D40 are consistent with this result, as the distribution shown in Fig. 3B indicates it is primarily its ψ angle that varies, thereby impacting the amide of G41.

Application to α -synuclein

Our prior analysis of backbone chemical shifts (Maltsev et al. 2012) showed that deviations from random coil values for this protein were even smaller than for the proteins used to generate the random coil chemical shift database (Kjaergaard et al. 2011). In other words, this protein is

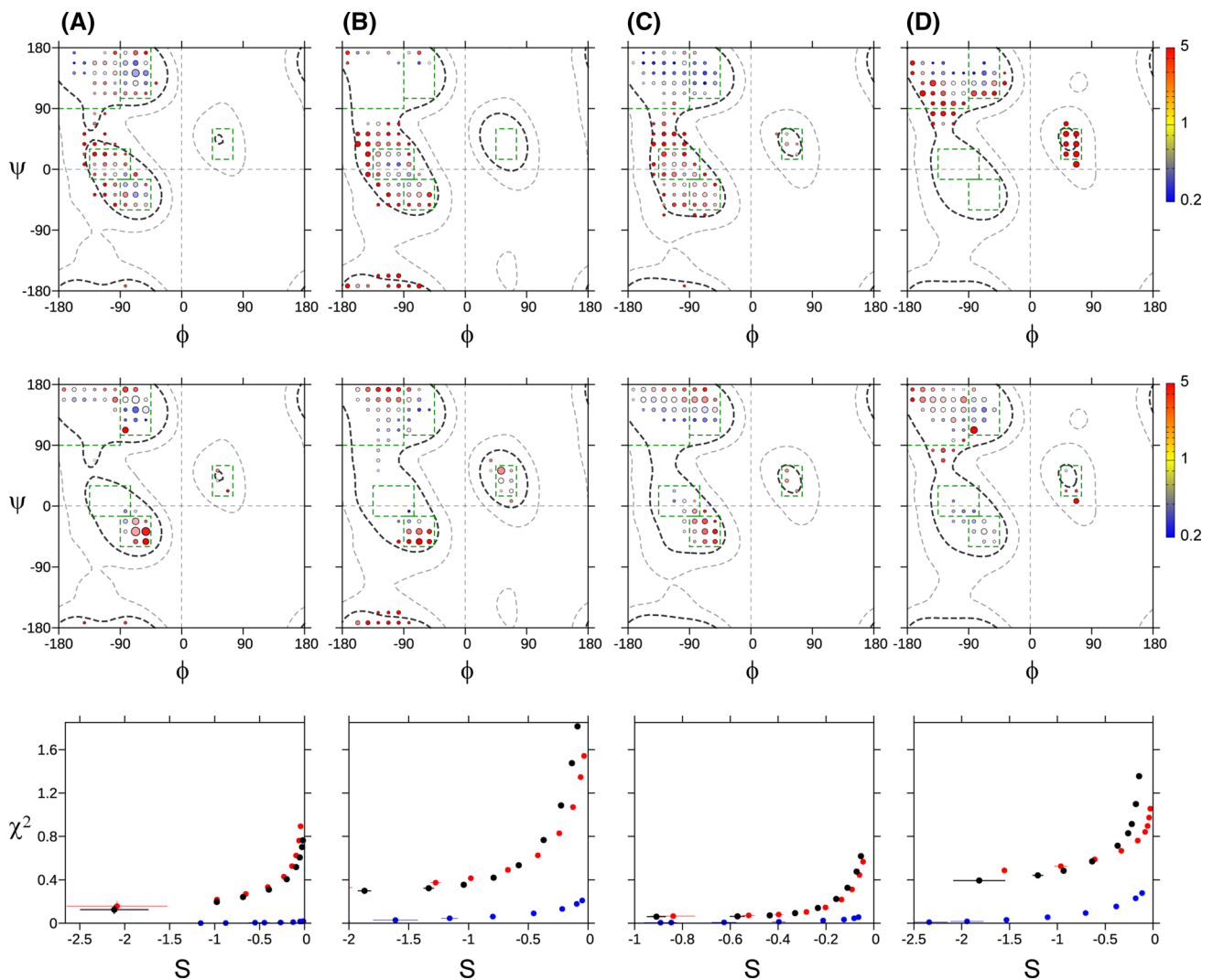


Fig. 4 Examples of ϕ/ψ distributions derived for several α -synuclein residues: **A** A19; **B** N65, **C** Y133; **D** Q134. For each residue, the populations are shown when using chemical shift values (^{15}N , $^{13}\text{C}^\alpha$, and $^{13}\text{C}'$) only (*upper panels*) and when using all available 12 parameters ($^3J_{\text{HNH}\alpha}$, $^3J_{\text{C}'\text{C}'}$, $^3J_{\text{C}'\text{H}\alpha}$, $^1J_{\text{C}\alpha\text{H}\alpha}$, $^2J_{\text{NC}\alpha}$, $^1J_{\text{NC}\alpha}$, $\delta^{15}\text{N}$, $\delta^{13}\text{C}^\alpha$, $\delta^{13}\text{C}'$, $d_{\text{HNH}\alpha}(i, i)$, $d_{\text{H}\alpha\text{HN}}(i-1, i)$ and $d_{\text{HNHN}}(i, i+1)$) for each residue (*middle panels*). $\theta = 0.4$ was used for the distributions shown, with all $\sigma(q)$ values set to their default (Table 1, IDP) values and using a diffusion anisotropy parameter $k = 1.0$. The plots of χ^2 versus S (*lower panels*) are displayed for θ values of (*left to right*) 0, 0.1, 0.2,

0.4, 0.8, 1.6, 3, 6, and 10, showing that on average slightly lower χ^2 values and higher S values are obtained when using the new coil database (*red symbols*) than when using the Fitzkee coil library (*black*) when fitting all 12 experimental parameters for each residue. *Blue dots* in the *lower panel* correspond to the S and χ^2 values obtained when using only chemical shift data as restraint inputs. The *horizontal error bars* in the *lower panels* reflect the lack of convergence of the simulated annealing protocol for very low θ values

representative of an IDP with very high backbone disorder and virtually no propensity to adopt stable secondary structure in the absence of binding partners.

We illustrate the results obtained by MERA for four representative residues: A19, located at the end of a stretch of three Ala residues; N65, a residue type that has a known higher propensity for adopting α_L backbone torsion angles, and a sequential pair of residues in the highly acidic C-terminal tail of the protein, Y133 and Q134.

When only using chemical shift data as input to MERA, the resulting χ^2 values are very low (blue dots in bottom

panels of Fig. 4) but the Ramachandran distributions show relatively large deviations from the coil distribution (top panels), as exemplified by Q134 which shows a substantial α_L -population. The absence of a significant increase in χ^2 when θ is increased strongly indicates that these chemical-shift-only calculations are overfitted, and that the coil database populations are actually in good agreement with the chemical shifts.

Remarkably, when all 12 experimental input parameters are used for each residue, the normalized χ^2 values remain low, but more “reasonable” deviations from the coil

distributions are observed (center panels in Fig. 4). For example, the α_L -population for Q134 virtually disappears, and A19 converges on the α_R and PPII regions, as expected for an Ala preceded by two other Ala residues. For all four residues, the backbone angles are much closer to the coil database distribution than for the above GB3 examples. This result is also reflected in Kullback–Leibler divergences (S) in the 0.1–0.5 range, versus 0.5–1.5 for the GB3 residues.

For most residues in α -synuclein, the S values are lower when using the newly developed coil database (red dots in the lower panels of Fig. 4) than when using the original Fitzkee coil library (black dots). Residue A19 is a minor exception to this rule, and shows a structural preference that matches slightly better to the Fitzkee library. For the vast majority of residues, however, the new coil library results in lower Kullback–Leibler divergences by amounts that fall somewhere between what is seen for residues Y133 and Q134.

Concluding remarks

The MERA program provides a convenient avenue for visualizing the distribution of backbone torsion angles compatible with NMR data. It is applicable to fully disordered proteins, as well as to dynamic regions in ordered proteins. It is important to realize, however, that the program will find the broadest possible distribution of backbone torsion angles that is consistent with the experimental data. Typically, therefore, when fewer experimental input parameters are provided this will result in wider distributions of backbone torsion angles. Care should therefore be exercised not to overinterpret the distributions obtained by MERA as evidence for actual dynamics. On the other hand, when clear indications for dynamics are available from other sources, such as ^{15}N or ^{13}C relaxation experiments, MERA analysis can actually provide some insights into the type of motions that take place. For example, residue L12 in GB3 samples a broad distribution of ϕ/ψ angles, all located in the β /PPII region of Ramachandran space (Fig. 3A). By contrast, motion of loop residue D40 appears to involve primarily its ψ angle, with ϕ remaining close to ca -140° .

The experimental input parameters used by MERA also depend on factors other than ϕ/ψ , which are neglected in our analysis. For example, valence angle distortions and H-bonding impact both J couplings and chemical shifts. These additional factors tend to be significant in well ordered regions of a protein but their effects, which can be positive or negative, will average and therefore be smaller in disordered regions. This typically causes attainable χ^2 values in dynamically disordered regions of a protein to be much lower than for highly ordered residues. Therefore,

considerably lower χ^2 values are commonly obtained for IDPs (Fig. 4). Future work is needed to develop a better quantitative understanding of the effect of H-bonding and valence angles on J couplings and chemical shifts. Overall, the present procedure and its future refinements hold the potential to greatly enhance the structural and dynamic detail that can be extracted from the experimental data.

Acknowledgments This work was supported by the Russian Science Foundation (Grant 14-14-00598) and by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases and the Intramural Antiviral Target Program of the Office of the Director, NIH, and by the Max Planck Society. JHL is the recipient of a KVSTA fellowship.

References

- Baldwin RL, Rose GD (1999) Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem Sci* 24:26–33
- Ball KA, Phillips AH, Nerenberg PS, Fawzi NL, Wemmer DE, Head-Gordon T (2011) Homogeneous and heterogeneous tertiary structure ensembles of amyloid-beta peptides. *Biochemistry* 50:7612–7628
- Berjanskii MV, Wishart DS (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc* 127:14970–14971
- Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129:5656–5664
- Bertoncini CW, Rasia RM, Lamberto GR, Binolfi A, Zweckstetter M, Griesinger C, Fernandez CO (2007) Structural characterization of the intrinsically unfolded protein beta-synuclein, a natural negative regulator of alpha-synuclein aggregation. *J Mol Biol* 372:708–722
- Bruschweiler R, Case DA (1994) Adding harmonic motion to the Karplus relation for spin–spin coupling. *J Am Chem Soc* 116:11199–11200
- Derrick JP, Wigley DB (1994) The 3rd Igg-binding domain from streptococcal protein-G—an analysis by X-ray crystallography of the structure alone and in a complex with fab. *J Mol Biol* 243:906–918
- Ding KY, Gronenborn AM (2004) Protein backbone H-1(N)-C-13(alpha) and N-15-C-13(alpha) residual dipolar and J couplings: new constraints for NMR structure determination. *J Am Chem Soc* 126:6232–6233
- Dyson HJ, Wright PE (1991) Defining solution conformations of small linear peptides. *Annu Rev Biophys Biophys Chem* 20:519–538
- Dyson HJ, Wright PE (2004) Unfolded proteins and protein folding studied by NMR. *Chem Rev* 104:3607–3622
- Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208
- Farrow NA, Zhang OW, Szabo A, Torchia DA, Kay LE (1995) Spectral density-function mapping using N-15 relaxation data exclusively. *J Biomol NMR* 6:153–162
- Fitzkee NC, Fleming PJ, Rose GD (2005) The protein coil library: a structural database of nonhelix, nonstrand fragments derived from the PDB. *Proteins* 58:852–854
- Graf J, Nguyen PH, Stock G, Schwalbe H (2007) Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. *J Am Chem Soc* 129:1179–1189

- Hagarman A, Measey TJ, Mathieu D, Schwalbe H, Schweitzer-Stenner R (2010) Intrinsic propensities of amino acid residues in G×G peptides inferred from amide I' band profiles and NMR scalar coupling constants. *J Am Chem Soc* 132:540–551
- Hall JB, Fushman D (2003) Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. *J Biomol NMR* 27:261–275
- Han B, Liu YF, Gininger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR* 50:43–57
- Ingber L (1989) Very fast simulated re-annealing. *Math Comput Model* 12:967–973
- Kjaergaard M, Brander S, Poulsen FM (2011) Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J Biomol NMR* 49:139–149
- Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 14:51–55
- Krzeminski M, Marsh JA, Neale C, Choy W-Y, Forman-Kay JD (2013) Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* 29:398–399
- Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22:79–86
- Lee JH, Li F, Grishaev A, Bax A (2015) Quantitative residue-specific protein backbone torsion angle dynamics from concerted measurement of 3 J couplings. *J Am Chem Soc* 137:1432–1435
- Li F, Lee JH, Grishaev A, Ying J, Bax A (2015) High accuracy of Karplus equations for relating three-bond J couplings to protein backbone torsion angles. *ChemPhysChem* 16:572–578
- Long HW, Tycko R (1998) Biopolymer conformational distributions from solid-state NMR: alpha-helix and 3(10)-helix contents of a helical peptide. *J Am Chem Soc* 120:7039–7048
- MacArthur MW, Thornton JM (1996) Deviations from planarity of the peptide bond in peptides and proteins. *J Mol Biol* 264:1180–1195
- Maltsev AS, Ying JF, Bax A (2012) Impact of N-terminal acetylation of α -synuclein on its random coil and lipid binding properties. *Biochemistry* 51:5004–5013
- Maltsev AS, Grishaev A, Roche J, Zasloff M, Bax A (2014) Improved cross validation of a static ubiquitin structure derived from high precision residual dipolar couplings measured in a drug-based liquid crystalline phase. *J Am Chem Soc* 136:3752–3755
- Mantsyzov AB, Maltsev AS, Ying J, Shen Y, Hummer G, Bax A (2014) A maximum entropy approach to the study of residue-specific backbone angle distributions in alpha-synuclein, an intrinsically disordered protein. *Protein Sci* 23:1275–1290
- Mittag T, Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17:3–14
- Mittag T, Kay LE, Forman-Kay JD (2010) Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* 23:105–116
- Rezaei-Ghaleh N, Blackledge M, Zweckstetter M (2012) Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery. *ChemBioChem* 13:930–950
- Rozycki B, Kim YC, Hummer G (2011) SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* 19:109–116
- Salmon L, Nodet G, Ozenne V, Yin GW, Jensen MR, Zweckstetter M, Blackledge M (2010) NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc* 132:8407–8418
- Shen Y, Bax A (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J Biomol NMR* 38:289–302
- Shen Y, Bax A (2010) SPARTA plus: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR* 48:13–22
- Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56:227–241
- Shi ZS, Chen K, Liu ZG, Kallenbach NR (2006) Conformation of the backbone in unfolded proteins. *Chem Rev* 106:1877–1897
- Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK (2007) DisProt: the database of disordered proteins. *Nucleic Acids Res* 35:D786–D793
- Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM (1996) Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. *J Mol Biol* 255:494–506
- Szu H, Hartley R (1987) Fast simulated annealing. *Phys Lett A* 122:157–162
- Ulmer TS, Ramirez BE, Delaglio F, Bax A (2003) Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. *J Am Chem Soc* 125:9179–9191
- Uversky VN, Dunker AK (2010) Understanding protein non-folding. *BBA-Proteins. Proteomics* 1804:1231–1264
- van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM (2014) Classification of intrinsically disordered regions and proteins. *Chem Rev* 114:6589–6631
- Varadi M, Kosol S, Lebrun P, Valentini E, Blackledge M, Dunker AK, Felli IC, Forman-Kay JD, Kriwacki RW, Pierattelli R, Sussman J, Svergun DI, Uversky VN, Vendruscolo M, Wishart D, Wright PE, Tompa P (2014) pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Res* 42:D326–D335
- Vuister GW, Delaglio F, Bax A (1993) The use of $^1J_{C\alpha H\alpha}$ coupling constants as a probe for protein backbone conformation. *J Biomol NMR* 3:67–80
- Wang AC, Bax A (1996) Determination of the backbone dihedral angles phi in human ubiquitin from reparametrized empirical Karplus equations. *J Am Chem Soc* 118:2483–2494
- Wang YJ, Jardetzky O (2002) Investigation of the neighboring residue effects on protein chemical shifts. *J Am Chem Soc* 124:14075–14084
- Wirmer J, Schwalbe H (2002) Angular dependence of 1 J(NCa) and 2 J(NCa) coupling constants measured in J-modulated HSQCs. *J Biomol NMR* 23:47–55
- Yao L, Vogeli B, Torchia DA, Bax A (2008) Simultaneous NMR study of protein structure and dynamics using conservative mutagenesis. *J Phys Chem B* 112:6045–6056
- Ying J, Roche J, Bax A (2014) Homonuclear decoupling for enhancing resolution and sensitivity in NOE and RDC measurements of peptides and proteins. *J Magn Reson* 241:97–102