Referee 1.

We thank the referee for the careful reading of the manuscript.

1.   Could you please comment on the number of field strengths required to successfully implement bootstrap aggregation. Assuming I get the math right, using 3 field strengths, the total number of possible samples is $7^3 = 343$, which seems to be at the low end of the number of samples normally employed, but perhaps this would suffice?

> **Response**: The referee's math is correct. Reducing the number of field strengths from four to three has three potential consequences for data analysis:
>
> a. Depending on the three fields chosen, the total range of frequencies can be affected. For example, three fields in the range 600 MHz – 1.2 GHz are likely to be more powerful than three fields in the range 600 MHz to 800 MHz.
> b. Reducing the number of fields reduces the number of data points and thus the number of statistical degrees of freedom. The number of degrees of freedom influences data analysis, whether conventional approaches or the bootstrap aggregation method are used.
> c. Most pertinent to the referee's point, three fields provide only 343 bootstrap samples, compared to the 6859 samples obtained for four fields – a 20-fold increase provided by one additional field.
>
> To address the referee's question, the analysis for the residues in the basic region (the disordered sites) was repeated using only the 600 MHz, 800 MHz and 900 MHz data, dropping the 700 MHz data, but keeping the same total range of frequencies. The reduced size of the data set resulted in simpler models being selected for individual residues in the basic (disordered region), because of the reduced number of spectral density values available for fitting. Essentially, the data were not sufficient to determine both τ_f and τ_s, so most residues were fit with only τ_s (models 3 and 4). However, even with the reduced number of bootstrap samples, bootstrap aggregation was effective in smoothing the effects of model selection error arising from choices between these two models.

2.   How do you determine the correlation time for overall rotational diffusion, tau_m (in the case of a globular protein)? The suggested protocol fits tau_m individually for each residue. Do you forego the concept of fitting a global tau_m as part of the MF fits? And subsequently fit a rotational diffusion model (isotropic or anisotropic) to these individual values (while taking into account the orientation of the HN bond vectors in the molecular frame in the case of anistropic models)?

Response: As noted by the referee, the approach adopted in the manuscript fits individual values of τ_m for residues in the ordered domain of the protein. This approach was adopted to correspond to the strategy used in the earlier paper by Gill and coworkers in which the relaxation rates were originally reported (Gill, et al., Phys Chem Chem Phys, 18, 5839–49, 2016). In a subsequent step, the global diffusion tensor could be determined from the values of τ_m as described in many

publications (for example, Lee et al., J. Biomol. NMR, 9, 287–298, 1997). This approach has the advantage of decoupling the determination of internal motional parameters from overall rotational diffusion (as discussed by d'Auvergene and Gooley (Mol. Biosyst., 3, 483–94, 2007), but increases the number of fitted parameters (in the simplest case, multiple local values of τ_m, one for each residue, must be fit rather than one global value). The bootstrap aggregation approach can be applied equally well in a data analysis strategy that optimizes a global rotational diffusion time or tensor. The two approaches for determining the overall rotational diffusion model differ how 'structural noise' arising from the reference N-H bond vector orientations (for example, in an x-ray crystal structure) differ from the time/ensemble average in solution. The present manuscript is not intended to assess the relative merits of either approach.

Minor points:

line 19, Suggestion: spell out Akaike and Bayesian Information Criterion when introducing AIC and BIC.

Response: Done

l. 145-148, the mean J(0) is presumably only used in the conventional MF protocol with MC error analysis(?) This should be stated here to avoid confusion.

Response: The mean J(0) was used in all analyses. However, the mean was the mean of the bootstrap samples for J(0) in the bootstrap analyses. We have clarified this point in the text.

The bootstrap aggregation protocol is well described on p. 6, but I still feel that it might be beneficial to include flow-chart type figure outlining the construction of the bootstrap sample datasets.

Response: We have prepared a flowchart for the revised manuscript.

The tables are not easily interpreted without referring back to the text. Please add footnotes to define p_ij and Y_ij in words (Table 1). Please add text to indicate that "Smooth" refers to the percentage of selected models in Tables 2-4.

Response: We have clarified the tables as requested by all referees.

Typos: "paramaters" (l. 86); "interogating conformional" (l. 250)

Response: Thank you for the careful reading.