Biophysical Journal, Volume *110*

Supplemental Information

A New Method for Determining Structure Ensemble: Application to a RNA Binding Di-Domain Protein

Wei Liu, Jingfeng Zhang, Jing-Song Fan, Giancarlo Tria, Gerhard Grüber, and Daiwen Yang

Supporting Information

Wei Liu[1], Jingfeng Zhang[1, $], Jing-Song Fan[1], Giancarlo Tria[2], Gerhard Grüber[2], and Daiwen Yang[1, *]

[1]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543

[2]Nanyang Technological University, School of Biological Sciences, 60 Nanyang Drive, Singapore 637551

[$]Present address: State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Wuhan Institute of Physics and Mathematics, The Chinese Academy of Sciences, 430071 Wuhan，China

*To whom correspondence should be addressed, email: dbsydw@nus.edu.sg, phone: 65-65161014

Table S1. Difference between calculated and reference spin-label conformers

Labeling site	Average difference for fixed population (Å)[a]	Average difference for unfixed population (Å)[a]
M107	0.8	2.0
H123	0.5	1.3
N148	1.0	1.4
S190	0.3	1.0
N218	0.5	2.4

[a]The average difference between calculated and reference spin-label conformers (Δ) is given by:
$\Delta = \sum_i p_i \times D_i$, where p_i is the population of the calculated i^{th} spin-label conformer, D_i is the distance between the calculated i^{th} conformer and its closest reference (or input) conformer. Note that one spin-label conformer is represented by one point in space.

Table S2. NMR data and structure determination details for PubRRM12

All NOE distance restraints[a]	1967
Intra-residue	478
Sequential (|i–j| = 1)	610
Medium-range (1 <|i–j|<5)	284
Long-range (|i–j|≥5)	595
Hydrogen bonds restraints	96
Dihedral angle restraints(φ, ψ)[b]	299
Energy statistics	
E_{noe}	66.79±0.83
E_{dih}	3.78±0.50
Deviations from idealized covalent geometry [c]	
RMSD of bond lengths (Å)	0.0022±0.0000
RMSD of bond angles (°)	0.313±0.005
RMSD of improper angles (°)	0.302±0.006
Deviations from experimental restraints	
RMSD of distance restraints (Å)	0.0255±0.0005
RMSD of dihedral angle restraints (°)	0.45±0.03
Ramachandran plot analysis (%)[d]	
Residues in allowed region	97.2%
Residues in generally allowed regions	2.5%
Residues in disallowed regions	0.3%
Average RMSD from mean structure (Å)[e]	RRM1 RRM2
Heavy atoms	1.22±0.15 1.28±0.16
Backbone atoms (N, CA, C',O)	0.28±0.06 0.45±0.17

[a] The distance restraints were obtained by classifying NOE cross peaks into three categories: strong (1.8–2.9 Å), medium (1.8–3.5 Å), and weak (1.8–5.0 Å).

[b] Dihedral angles of backbone ϕ and ψ were predicted by TALOS using the chemical shifts of Cα, Cβ, Hα, N, and HN.

[c] Twenty lowest-energy conformers with no NOE violations greater than 0.3 Å and no torsion angle violations greater than 3° were selected from 100 conformers to represent the NMR ensembles.

[d] Calculated with PROCHECK-NMR.

[e] Calculated with MOLMOL over secondary structure regions β1 (75-80), α1 (88-96), β2 (103-108), β3 (116-121), α2 (125-134), and β4 (146-149) for N-terminal RRM1 domain; β5 (162-167), α3(175-182), β6 (190-195), β7 (206-210), α4 (213-223), and β8 (234-237) for C-terminal RRM2 domain.

Table S3. SAXS-data collection and scattering derived parameters of PubRRM12

Data collection parameters	
Instrument (source)	Bruker NanoStar equipped with MetalJet eXcillum
Instrument (detector)	VÅNTEC-2000
Beam geometry	100 μm slit
Wavelength (Å)	1.3414
s range ($Å^{-1}$)	0.016-0.4
Exposure time (min)	20 (5 frames x 4min)
Concentration range (mg ml-1)	0.5-2
Temperature (K)	288.15
Structural parameters	
I(0) (cm^{-1}) [from P(r)]	76.81±0.97
R_g (Å) [from P(r)]	19.35±0.23
I(0) (cm^{-1}) (from Guinier)	75.40±1.55
R_g (Å) (from Guinier)	18.66±0.53
D_{max} (Å)	60±5
Porod volume estimate ($Å^3$)	~23200±3000
Dammif excluded volume ($Å^3$)	~31000±3000
Dry volume calculated from sequence ($Å^3$) ‡	~23227
Molecular mass determination	
Calculated monomeric *MM* (kDa) from sequence*	~19
Molecular mass *MM* (kDa) [from *Porod invariant*]	15±3
Molecular mass *MM* (kDa) [from *excluded volume*]	16±3
Software employed	
Primary data reduction	SAXS
Data processing	PRIMUS
Ab initio analysis	DAMMIN
Validation and averaging	DAMAVER
Computation of model intensities	CRYSOL
Flexibility	EOM 2.0
3D graphics representations	PyMOL

* http://web.expasy.org/compute_pi/
‡ http://www.basic.northwestern.edu/biotools/proteincalc.html

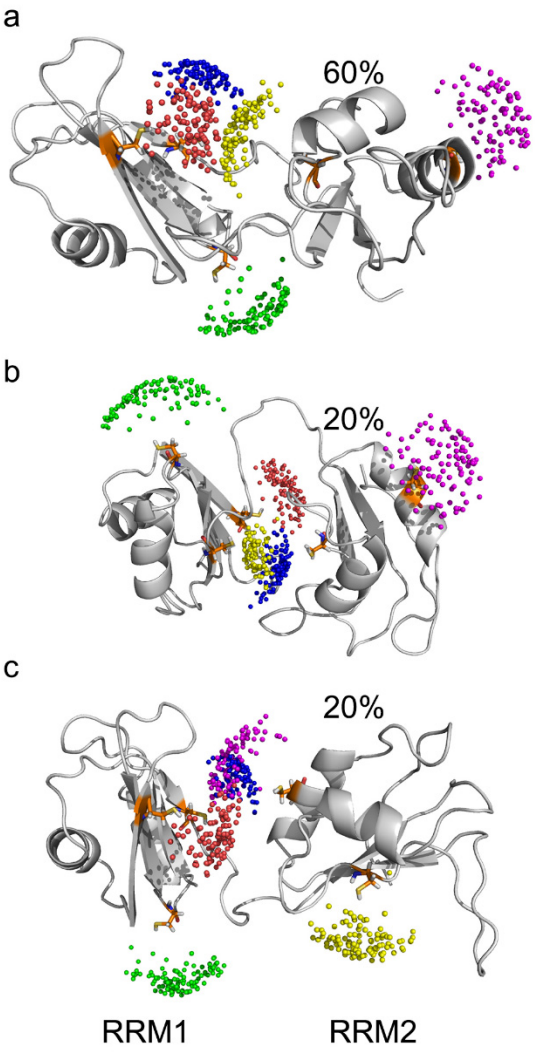

Figure S1. Three predefined structures of PubRRM12 (with populations of 60%, 20%, and 20%), which were used in PRE synthesis. 100 positions of the free electron in each MTSL are shown as small spheres. Red, green, blue, yellow, magenta spheres represent the MTSL at residues M107, H123, N148, S190, and N218, respectively.

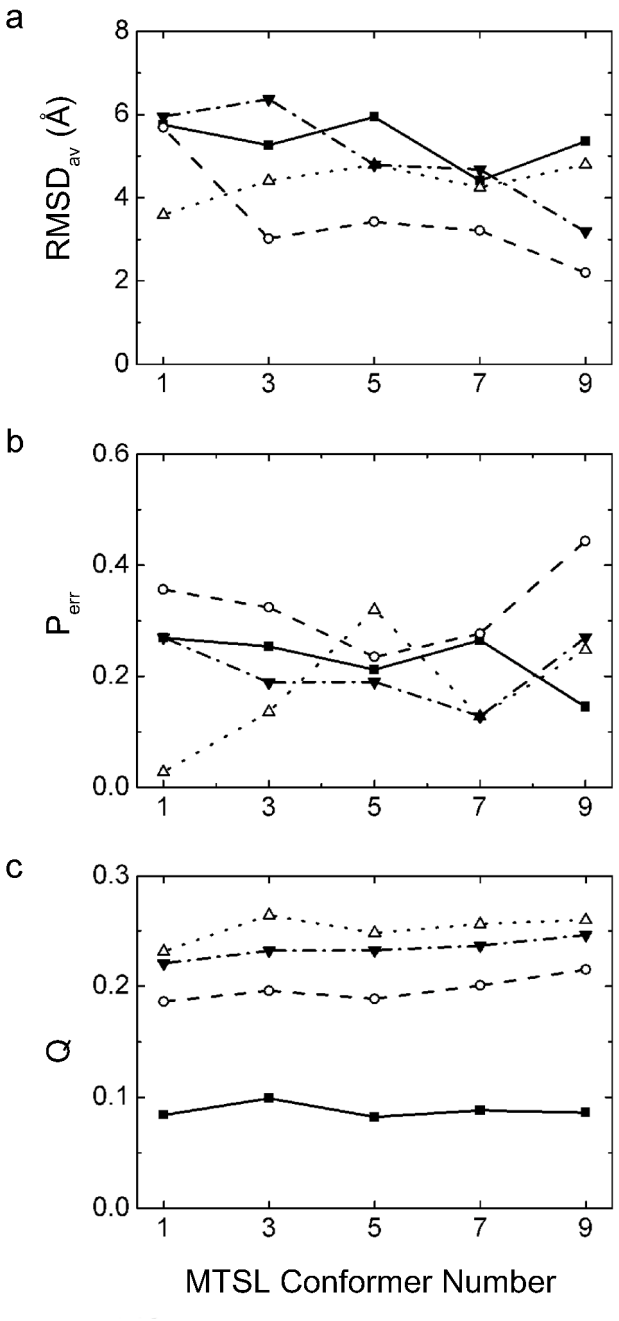

Figure S2. Influence of the number of pseudo MTSL conformers on calculated protein structure ensembles as revealed from four groups of synthetic PRE data (■, without error; ○, error set 1; Δ, error set 2; ▼, error set 3). (a) Structure difference. (b) Population difference: $P_{err} = \sum_i \left| P_i - Pc_i \right|$, where P_i (Pc_i) is the population of the i^{th} reference (calculated) conformer. (c) Q factor values.

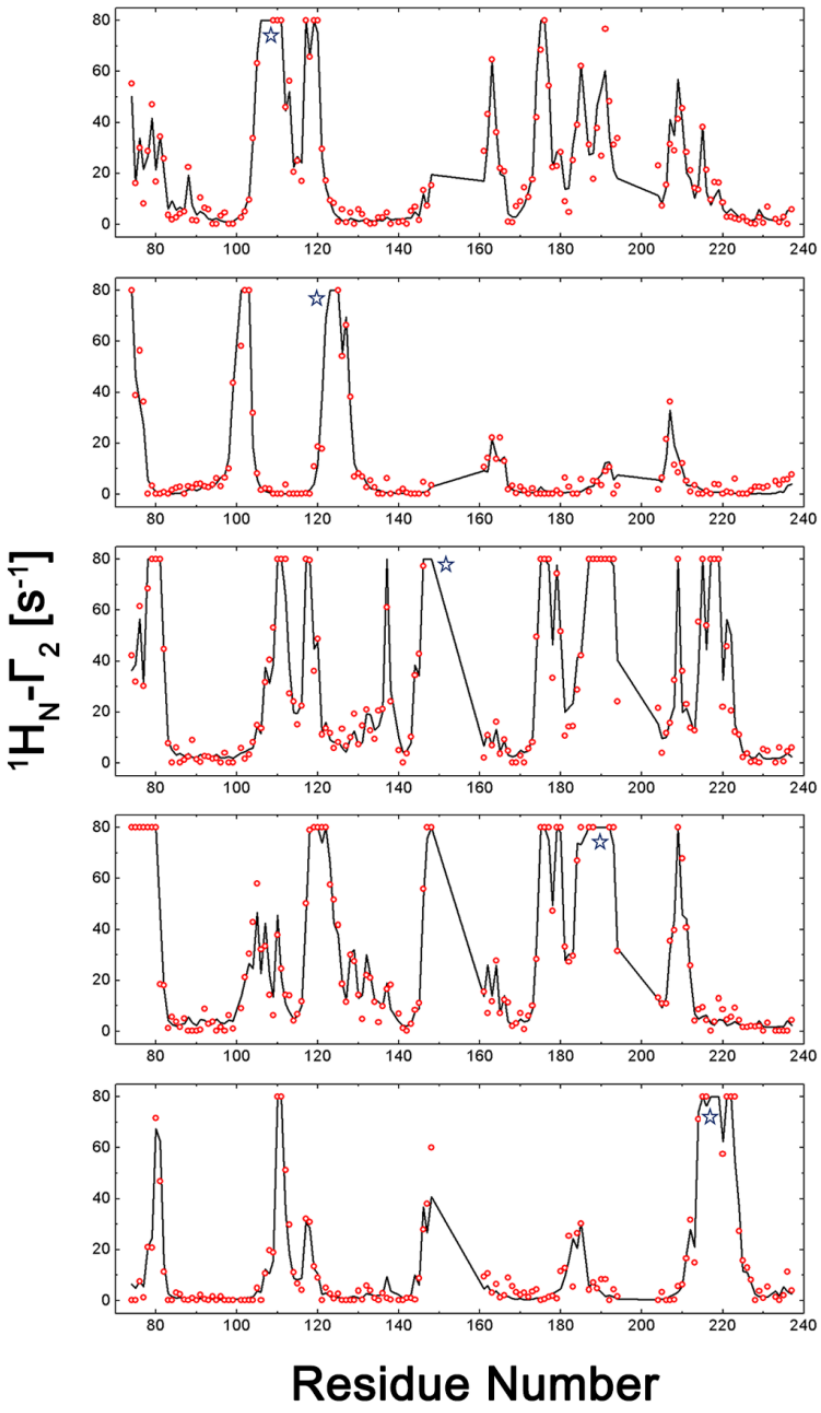

Figure S3. Comparison of synthetic PRE data without errors (○) and calculated PRE data (black line) when 3 pseudo MTSL conformers were used to represent 100 equally populated conformers. Spin labels are located at respective residues M107, H123, N148, S190, and N218, of which locations are indicated by stars. Data from residues in the linker (150-160) and flexible region in RRM2 (195-203) were not used in the calculations.

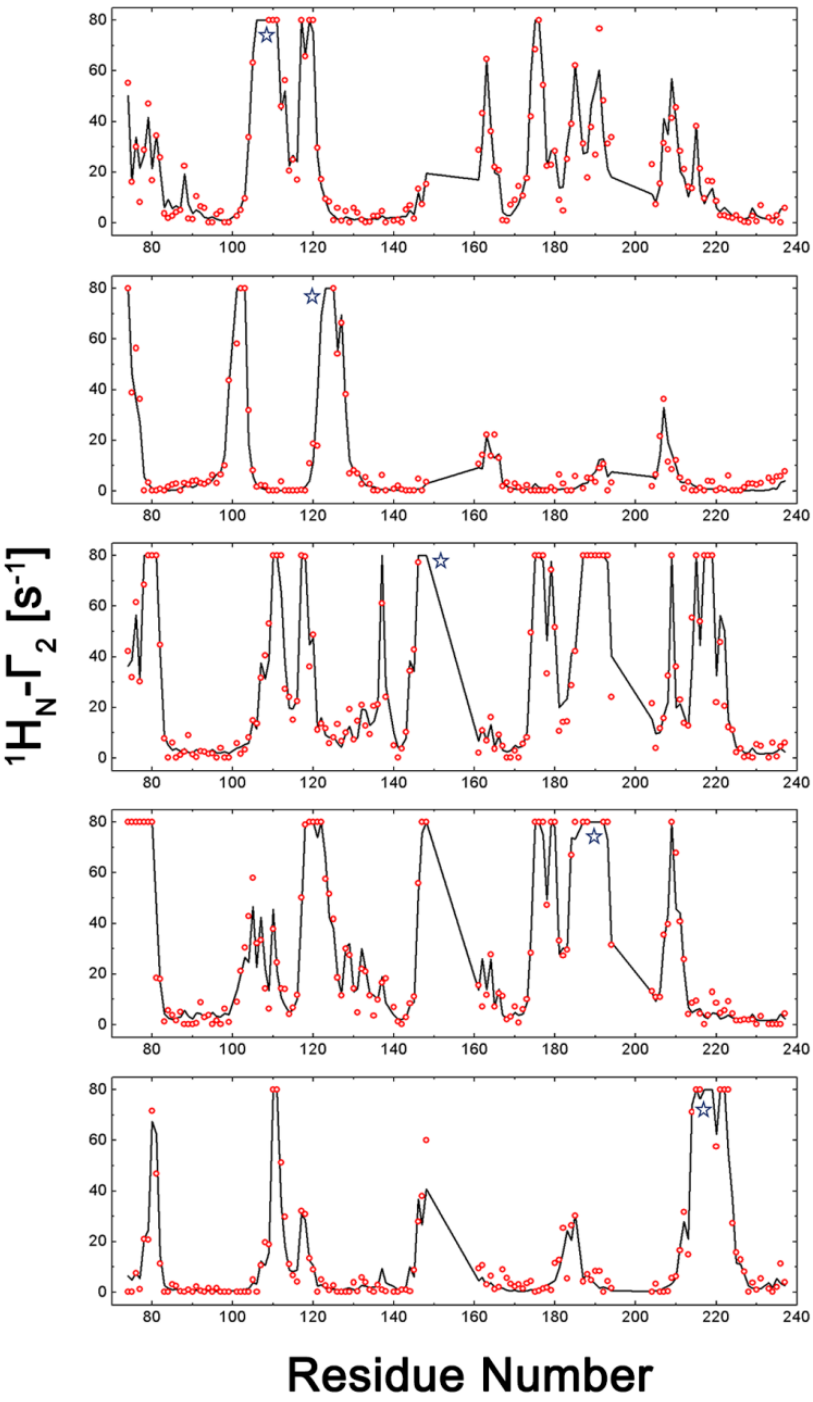

Figure S4. Comparison of synthetic PRE data with error set 1 (○) and calculated PRE data (black line) when 3 pseudo MTSL conformers were used to represent 100 equally populated conformers. Spin labels are located at respective residues M107, H123, N148, S190, and N218, of which locations are indicated by stars. Data from residues in the linker (150-160) and flexible region in RRM2 (195-203) were not used in the calculations.

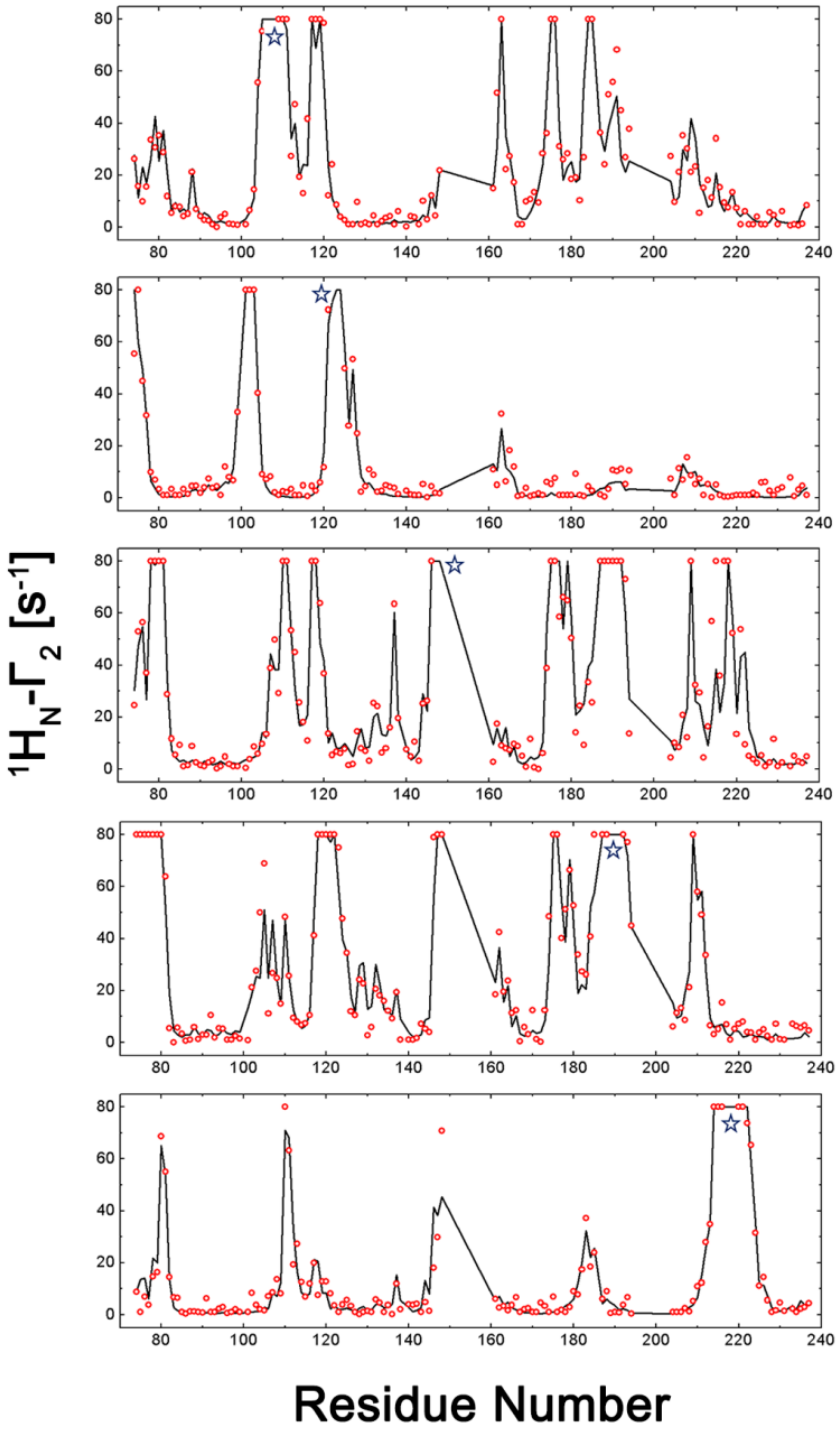

Figure S5. Comparison of synthetic PRE data with error set 2 (○) and calculated PRE data (black line) when 3 pseudo MTSL conformers were used to represent 100 equally populated conformers. Spin labels are located at respective residues M107, H123, N148, S190, and N218, of which locations are indicated by stars. Data from residues in the linker (150-160) and flexible region in RRM2 (195-203) were not used in the calculations.

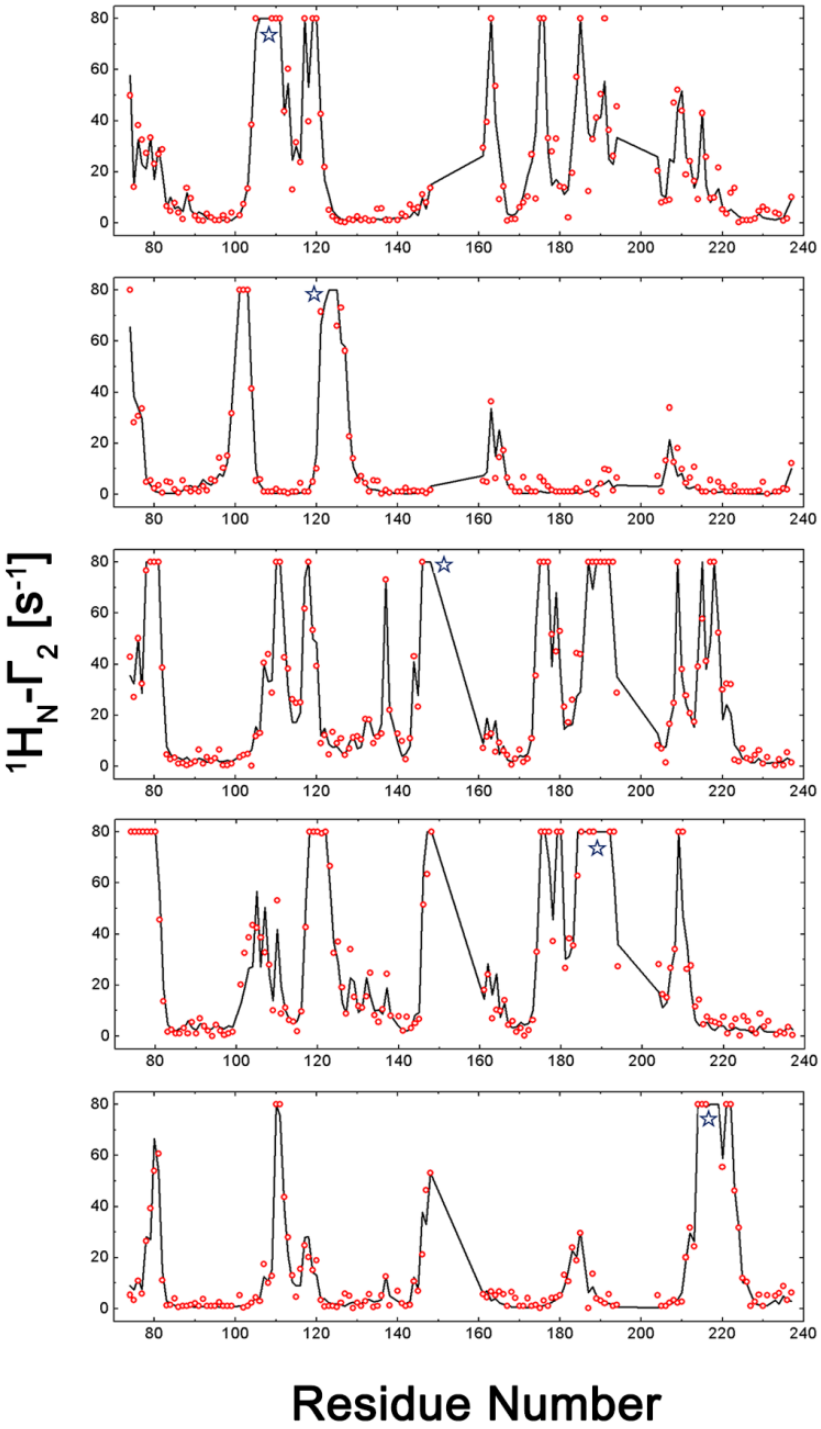

Figure S6. Comparison of synthetic PRE data with error set 3 (○) and calculated PRE data (black line) when 3 pseudo MTSL conformers were used to represent 100 equally populated conformers. Spin labels are located at respective residues M107, H123, N148, S190, and N218, of which locations are indicated by stars. Data from residues in the linker (150-160) and flexible region in RRM2 (195-203) were not used in the calculations.

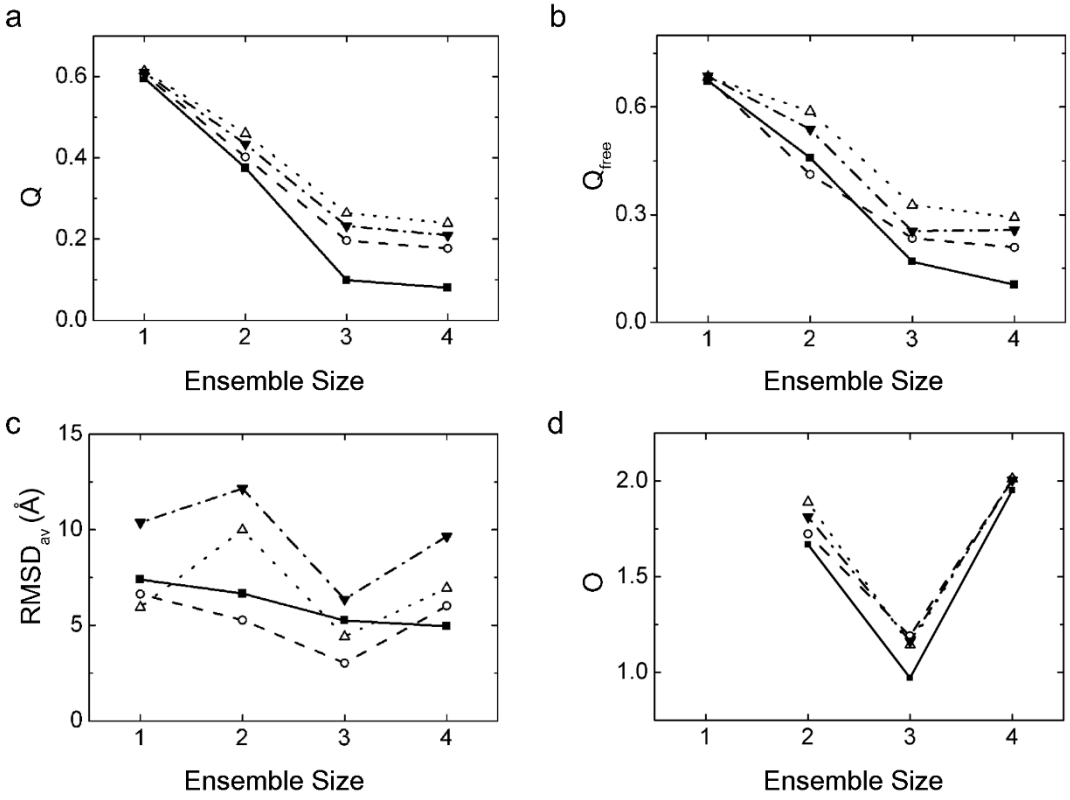

Figure S7. Dependences of Q (a), Q_{free} (b), $RMSD_{av}$ (c), and O (d) factors on the ensemble size when each spin-label with 100 conformers was represented by three pseudo conformers. PRE data without error (■), with errors (○, Δ, ▼).

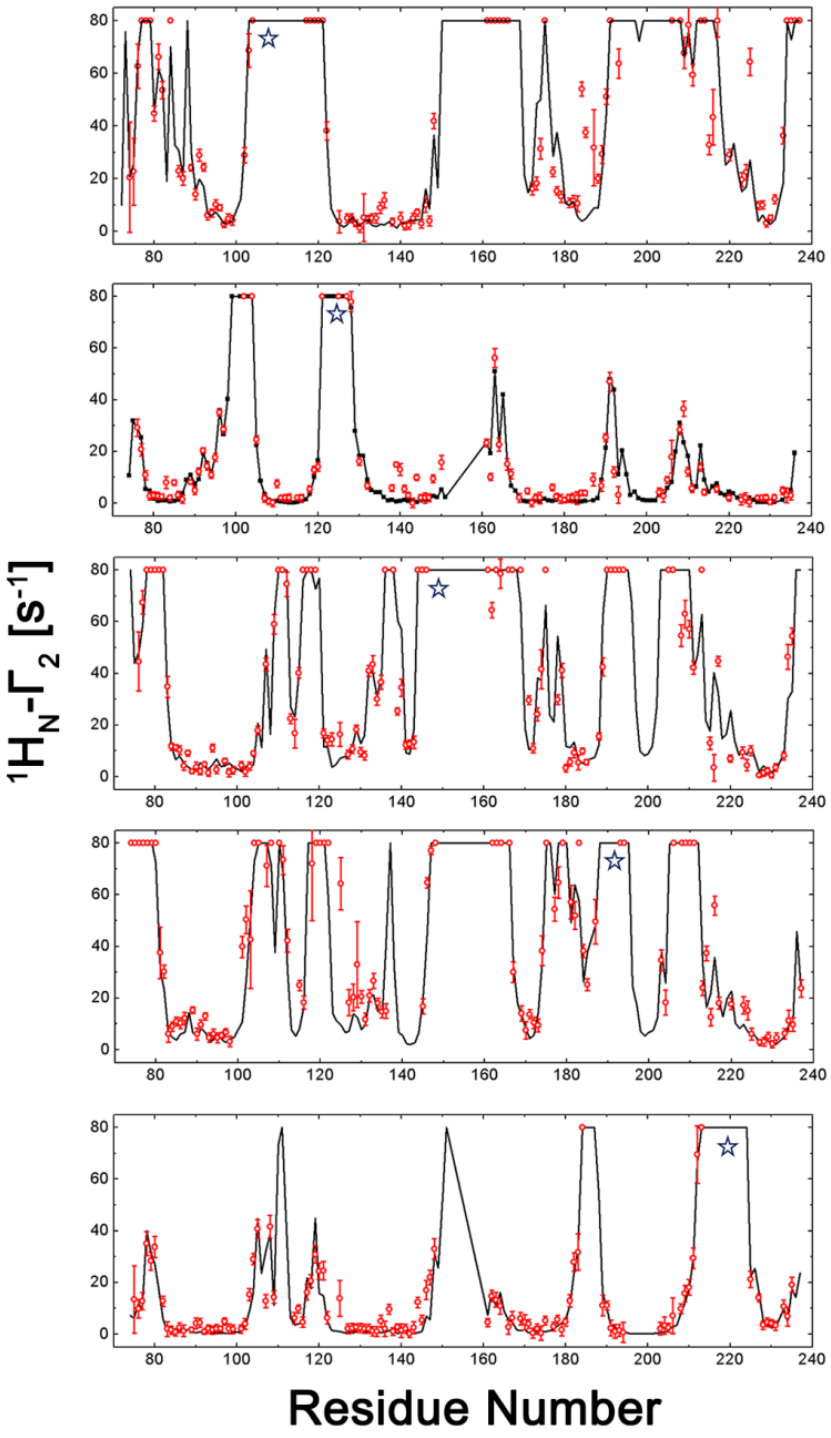

Figure S8. Comparison of experimental PRE data (○) and calculated PRE data (black line). Spin labels are located at respective residues M107, H123, N148, S190, and N218, of which locations are indicated by stars. Data from residues in the linker (150-160) and flexible region in RRM2 (195-203) were not used in the calculations. The measurement errors are indicated by bars.

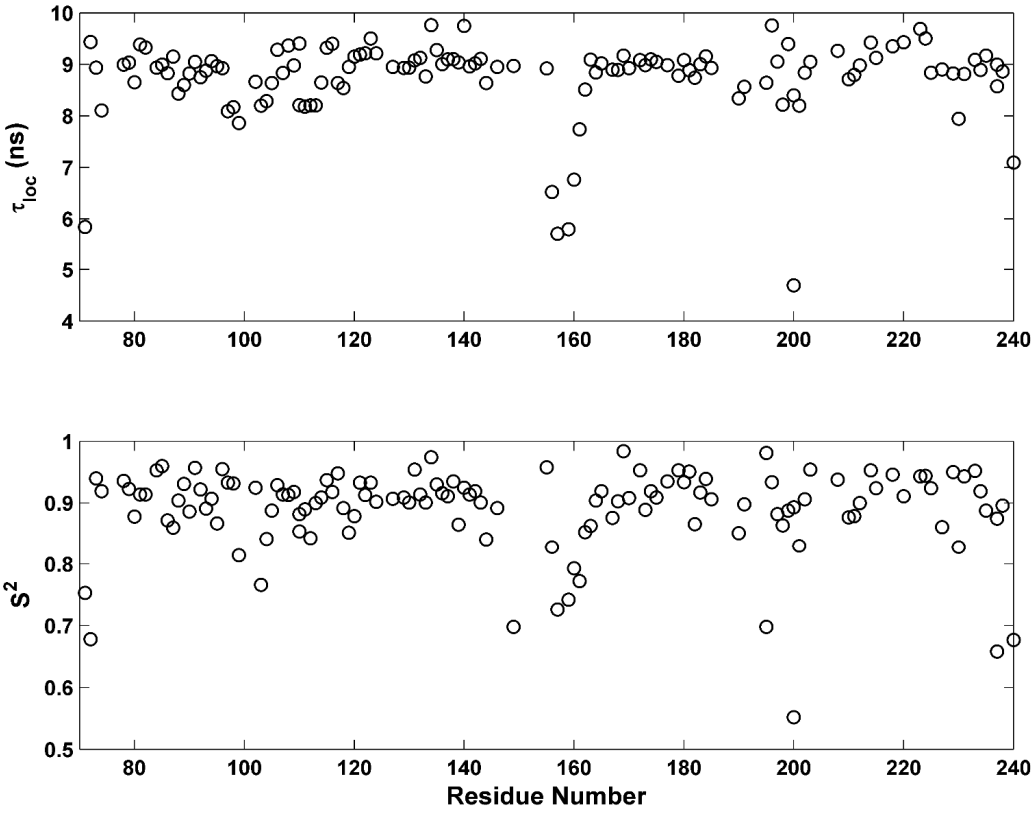

Figure S9. Localized correlation times τ_{loc} (upper panel) and generalized order parameters S^2 (lower panel) for PubRRM12 as measured at 25 °C.

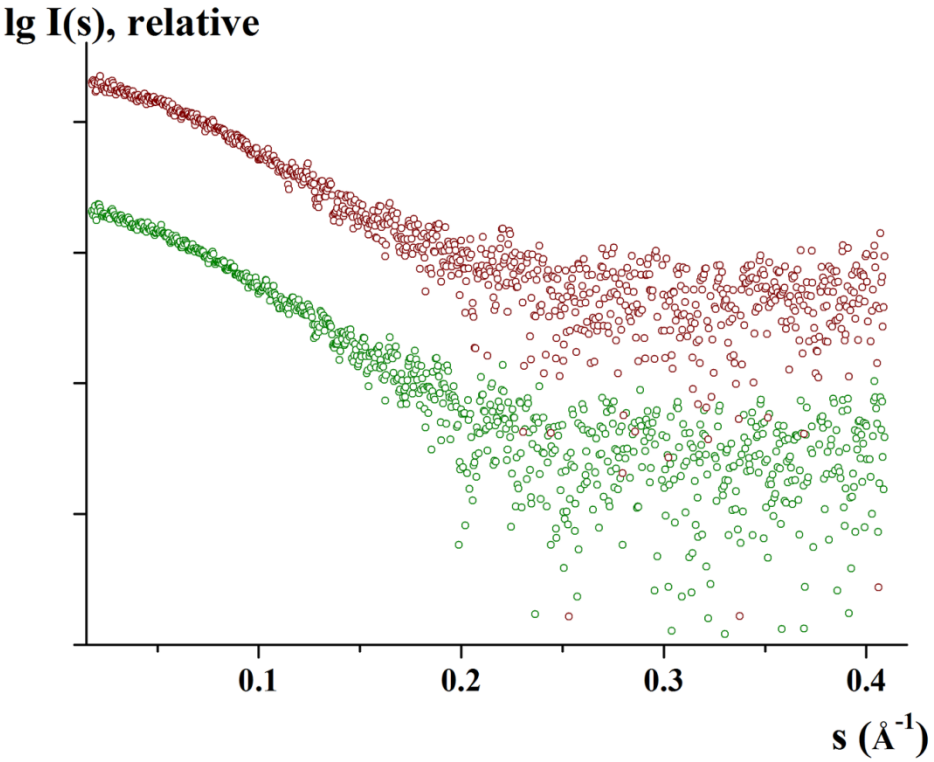

Figure S10. Comparison of SAXS data from wild type PubRRM12 (red dots) and a variant spin-labeled at N148 (green dots).

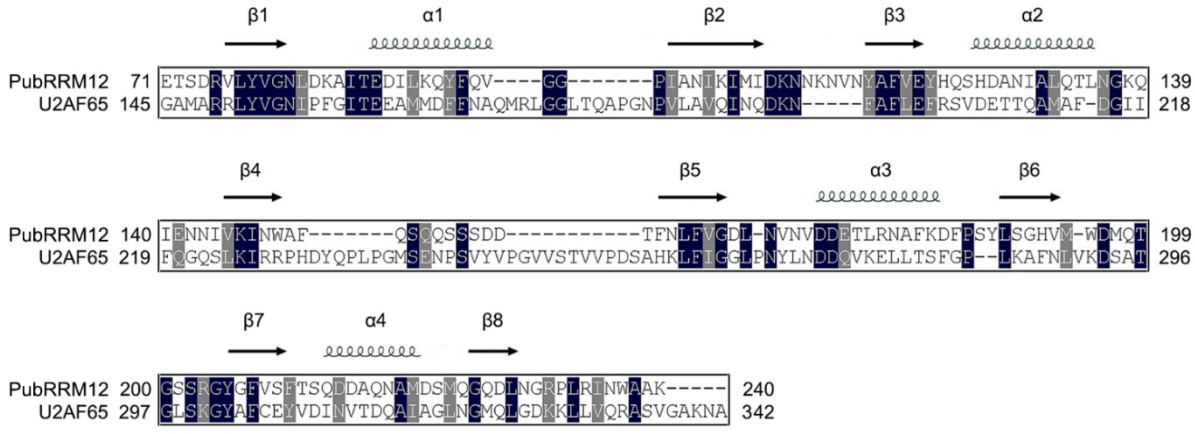

Figure S11. Sequence alignment and secondary structure of PubRRM12 and U2AF65.