SUPPORTING INFORMATION

A Switch Between Two Intrinsically Disordered Conformational Ensembles Modulates the Active Site of a Basic-Helix-Loop-Helix Transcription Factor

Giuseppe Sicoli,[a]+ Thomas Kress,[b]+ Hervé Vezin,[a] Karin Ledolter,[c] Dennis Kurzbach[d]*

[a] Laboratoire Avancé de Spectroscopie pour les Interactions, la Réactivité et l'Environnement (LASIRE), UMR CNRS 8516, Université de Lille, Avenue Paul Langevin – C4, F-59655 Villeneuve d'Ascq Cedex, France.

[b] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

[c] University Vienna, Department for Structural and Computational Biology, Max F. Perutz Laboratories, Campus Vienna BioCenter 5, 1030 Vienna, Austria.

[d] University of Vienna, Faculty of Chemistry, Institute of Biological Chemistry, Währinger Str. 38, 1090 Vienna, Austria.

* Email: dennis.kurzbach@univie.ac.at

\+ these authors contributed equally

Experimental Section

Protein Expression and Purification. MAX was subcloned into a Pet3d expression vector and transformed into E. coli BL21 pLysS cells. Cells were grown at 37 °C in M9 (for ^{15}N labeling 1 g/L ^{15}N ammonium chloride was added) and induced at an optical density corresponding to *A*(600 nm) = 0.5 with 0.5 mM IPTG prior to incubation at 30 °C overnight. Cell pellets were homogenized in 20 mM PBS, 100 mM NaCl and 1 mM EDTA. For protein purification, fractional $(NH_4)_2SO_4$ precipitation (50% and 80% saturation) was carried out and anion exchange chromatography was applied.

Cysteine mutants and MTSL (S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate) labeled proteins were produced according to methods published earlier.[1] Excess spin label was removed by dialysis into the buffer used for the NMR experiments (25 mM MES, 25 mM NaCl, 100 mM ArgHCl, pH 5.5). The labeling efficiency was always >95% as determined via DTNB assays. PRE referencing was achieved by reduction of the MTSL label through incubation for 1 h with tenfold excess of ascorbic acid at 35 °C.

For DEER EPR, samples were vitrified, *i.e.*, flash frozen at their glass transition temperature by plunging the samples into liquid nitrogen after the addition of 15% glycerol to avoid crystallization.

The final total protein concentration was 0.4 mM for NMR and 0.1 mM for EPR experiments.

EPR experiments. Continuous Wave (CW) X-Band measurements have been recorded using an X-band Bruker E500 instrument (9.4 GHz, TE_{012} resonator). All CW experiments were recorded at room temperature; a 1 mm capillary sealed at one side has been inserted in a 4 mm tube. Data were then recoded with a modulation frequency of 100 KHz and modulation amplitude of 1.0 G. Pulsed EPR experiments at Q-band were performed on a Bruker ELEXYS E-580 spectrometer with a SuperX-FT microwave bridge and a Bruker ER EN4118X-MD4 dielectric resonator. Cryogenic temperatures (50 K) were achieved by an Oxford flow cryostat. The field-swept EPR spectra were recorded by electron spin echo (ESE) detection; electron-spin-echo (ESE)-detected EPR experiments were carried out with the pulse sequence: $\pi/2–\tau–\pi–\tau–$echo. The microwave pulse lengths used were $t_{\pi/2}$ = 8 ns and t_π = 16 ns at a τ value of 204 ns. A two-step phase-cycle was applied to remove all unwanted echoes. PELDOR / DEER measurements were performed at 50 K on a Bruker ELEXSYS E580 Q-AWG (arbitrary waveform generator) pulse Q-band spectrometer equipped with a 50 W amplifier. A 4-pulse sequence with Gaussian, non-selective observer and pump pulses of 8 or 16 ns length with 55 MHz frequency separation was used. An eight-step phase cycling was performed together with $0–\pi$ phase cycling to remove unwanted effects of running echoes from the DEER trace. The evaluation of the DEER data was performed using DeerAnalysis2018.[2] The background of the primary DEER traces was corrected using exponential functions with homogeneous dimensions. A model-free Tikhonov regularization analysis was employed to extract distance distributions from the background corrected form factors.[3-5]

The home-written MatLAB routine (script) for fitting of the room temperature CW EPR spectra can be found in the Supporting Information.

NMR experiments. ^1H-^{15}N transverse relaxation optimized spectroscopy (TROSY) for PRE measurements was recorded at 20 °C using a Bruker HDIII wide-bore 800 MHz spectrometer. Spectra were recorded in the States-TPPI/PFG mode for quadrature detection with carrier frequencies for $^1H^N$ and ^{15}N of 4.73 and 120.0 ppm, respectively. The samples contained 0.4 mM MAX, 25 mM MES, and 25 mM NaCl (pH 5.5) in a 90% H_2O/10% D_2O mixture.

All NMR spectra were processed and analyzed using NMRPipe and SPARKY.[6, 7] A squared and 60° phase-shifted sine bell window function was applied in all dimensions for apodization. Time domain data were zero-filled to twice the data set size, prior to Fourier transformation. ^1H-^{15}N cross peak assignments were obtained from the biological magnetic resonance data base (BMRB) entry 5956 and the work by Sauvé et al.[8, 9]

Full NMR spectra are shown in the Supporting Information Fig. S11a).

Molecular Dynamics. We performed molecular dynamics (MD) simulations using GROMACS 2019.1.[10] The solution structure of a DNA-free MAX_2 homodimer (PDB ID: 1R05) was used in our simulations to build our initial atomistic model. The complex was confined with 37199 water molecules in a dodecahedral box so that the edges of the box were always at least 1 nm away from the complex. The structure was electro-neutralized with Cl^- ions, and Na^+Cl^- ions were added to achieve a salt concentration of 25 mM as used in the experiments. NVT

equilibration was performed at P = 1 bar and T = 310 K in a constrained box for 80 ps, with a step of 2 fs using the Verlet cutoff scheme set to 1.2 nm. The modified Berendsen thermostat temperature scheme coupled the protein and non-protein thermostats. Subsequently, NPT equilibration was performed under similar conditions. Nosé-Hoover thermostat coupling was used, which allows wide fluctuations and produces more natural dynamics than the Berendsen coupling. The extended simple point charge water model (SPC/E) and the AMBER03 protein force field[11] were used. The MD simulations continued NPT equilibration under unconstrained conditions. The MD trajectories were sampled every 2 ns, for a total simulation time up to 200 ns.

The extracted PRE signal suppression ratios V for an entire trajectory is shown in the Supporting Information Fig. S11b.

To simulate the distance distribution with the spin-labeled R5C mutant, the non-native protein was modelled using the YASARA software package. The AMBER 03 force field was employed with periodic boundary conditions.[12] Non-bonded interactions were cut off at 10.5 Å. Long-range Coulombic interactions were treated by a smoothed particle-mesh Ewald method.[13] Non-native amino acids were built using YASARA and semi-quantum-mechanically parameterised (YAPAC-AM1). The modelled structures were subject to energy minimisation in vacuum, subsequently randomly placed in the simulation box and solvated by water at pH 7.4, charge neutralised by addition of 1 % NaCl, and again minimised (steepest descent minimisation followed by simulated annealing). The chosen time increment was 2 fs. An MD trajectory of 50.0 ns length was accumulated. Intermolecular forces were recalculated at every second simulation sub-step. Temperature rescaling was employed with a set-temperature of 37°C. The box dimensions (dodecahedral of 92 Å side length) were controlled so as to yield a solvent pressure of 1 bar. Snapshots of the simulations were taken every 10 000 fs.

Supplementary MD analysis

To further corroborate that the two sub-ensembles depicted in Fig. 1 of the main text are indeed well-distinct, we performed an additional analysis of the MD data using the center of mass of MAX_2 as reference point. Figure S1. shows a scatter plot of distances r(5-5) between the C^{α} positions of residues R5 in both subunits of MAX_2 (yellow dots in the insert in Fig. S1a) vs. distances r(5-M) between the C^{α} position of residue R5 of one chain and the center of mass of MAX_2 (red dot in the insert in Fig. S1a). These distances were extracted from the trajectory underlying Fig. 1 of the main text.

The contour levels in Fig. S1 indicate clusters of data obtained through fitting a Gaussian mixture distribution to the data (implemented in the MATLAB 'fitgmdist' function). Fig. S1 a and b, respectively, represent data for the two different subunits of MAX_2. i.e. two the different r(5-M) vectors. It can clearly bee seen that the hinged sub-ensemble (annotated as cluster 1) is well distinct from the data points representing the extended ensemble (annotated as cluster 2). This indicates that only a small number of intermediate states are sampled during the transition between hinged and extended conformations and that the two states can be considered as distinct sub-ensembles within the conformational space of MAX_2.

Additionally, the reference to the distance to the center of mass, i.e. to r(5-MC) shows that short distances r(5-5) indeed correlate with proximity of residues R5 and the center of the protein. In other words, it can be excluded that the observed short r(5-5) distances stem from transient encounters of the two NTDs in their extended states.

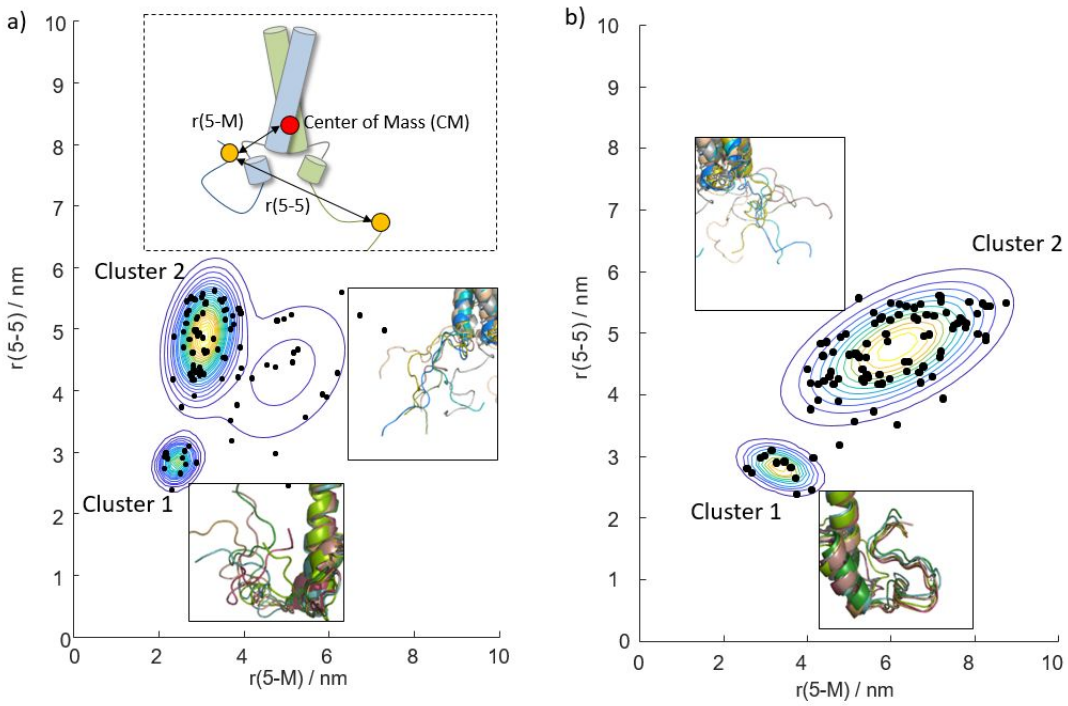

Figure S1. Cluster analysis of the MD data shown in Fig. 1 in the main text for the two sub-chains (shown in a) and b), respectively). The hinged and extended sub-ensemble constitute well-distinct clusters. The distance between C^{α} positions of residues R5 is denoted r(5-5). The distances between the C^{α} position of residue R5 and the center of mass of MAX_2 are denoted as r(5-M).

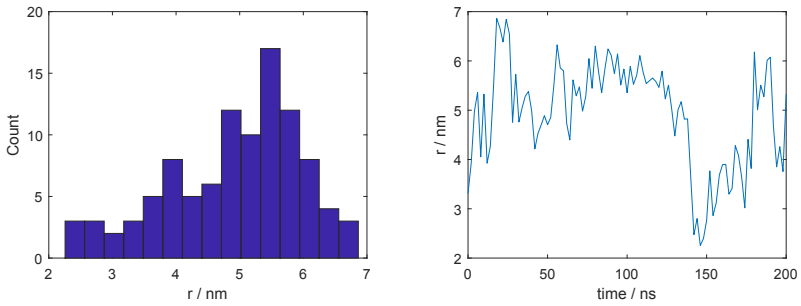

Figure S2. Additional replica run of the MD trajectory in Fig. 1 of the main text and distance histograms. In both cases, the switch between hinged and extended states could be observed.

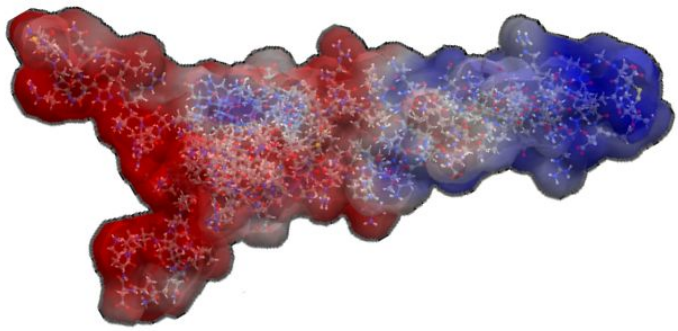

Figure S3. Charge distribution of MAX_2 as obtained from the APBS web-server (http://server.poissonboltzmann.org/). Blue indicates a negative charge and red a positive charge. The dipolar structure of MAX_2, which leads to attraction of the NTD towards the C-terminal region of the transcription factor, can clearly be seen.

Supplementary Information (EPR)

To support the CW EPR data of the main text further assays involving three different mutants of MAX_2 were performed; in addition to R5C-R5C, R5C-G35C and R5C-R55C mutants have also been investigated. Experimental spectra and best fits for CW data of all mutants are shown in Fig. S1-S3. The three mutants exhibit similar trends, i.e. good fits could be obtained assuming a superposition of a faster and a slower correlation time τ_c of the spin labels. The fast species clearly dominates according to the fit results and as summarized in Table S4. Importantly, all three samples yielded a similar contribution of the slow component of ca. 15 %.

Upon DNA binding the CW spectral features are significantly affected (Table S1 and Fig. S4): the conformational flexibility is reduced and the ratios between the two main components of the spectrum shift towards the slow component (Table S4). Indeed, the presence of the DNA ligand inverts the ratios of the fast and slow-motion components for all three mutants.

CW spectra have been fitted by home-written MatLAB routines supported by the EasySpin package as detailed in the main text. To cross-check the choice of two correlation times, additionally, a fit of the CW EPR spectra mutant R5C-R55C has been carried out assuming three components (Supporting information Fig. S5). The improvement achieved by adding a third conformer to the fitting procedure was only marginal as indicated by the final RMSD values (0.0071 vs. 0.0080). Thus, we chose a minimized number of species (*i.e.* two components) for the fitting procedure to remove unnecessary degrees of uncertainty.

In order to also validate the results obtained for the R5C-R5C mutant by DEER EPR, the additional two mutants were also used for a PELDOR/DEER analysis. As for the R5C-R5C mutant, the DEER/PELDOR form factor (dipolar evolution function) exhibits a first maximum of the cosinusoidal form factor around 1.5 µs in the free state of the dimeric structure (Supporting information Fig. S7 for the R5C-G35C); upon binding to EBOX DNA, these maxima diminish, due to the presence of another component with a second maximum at longer times (Supporting information Fig. S8 for the R5C-G35C). This effect is most pronounced for the mutant R5C-R55C (Supporting information Fig. S9); as main effect of the DNA binding, the population of the extended conformation increases. Indeed, the inter-spin distances distributions (and corresponding integrals) centered at 4.8 nm increase. Interestingly, for all mutant we found that in the absence of DNA the distribution features a positive skew and a negative skew after addition of DNA to the sample.

Note that MAX_2 exists in a dynamic equilibrium[1] and the subunits exchange between different dimers. Hence each sample contains all possible permutations of double mutants (*e.g.* sample 'R5C-G35C' consists to 25% of each of the four pairings R5C-R5C, G35C-R5C, R5C-G35C and G35C-G35C). The resulting distance distributions are therefore complex. However clearly, and most importantly, the distances spanned by the distance distributions do not change upon DNA binding in any of the three tested mutants corroborating the conclusions presented in the main text.

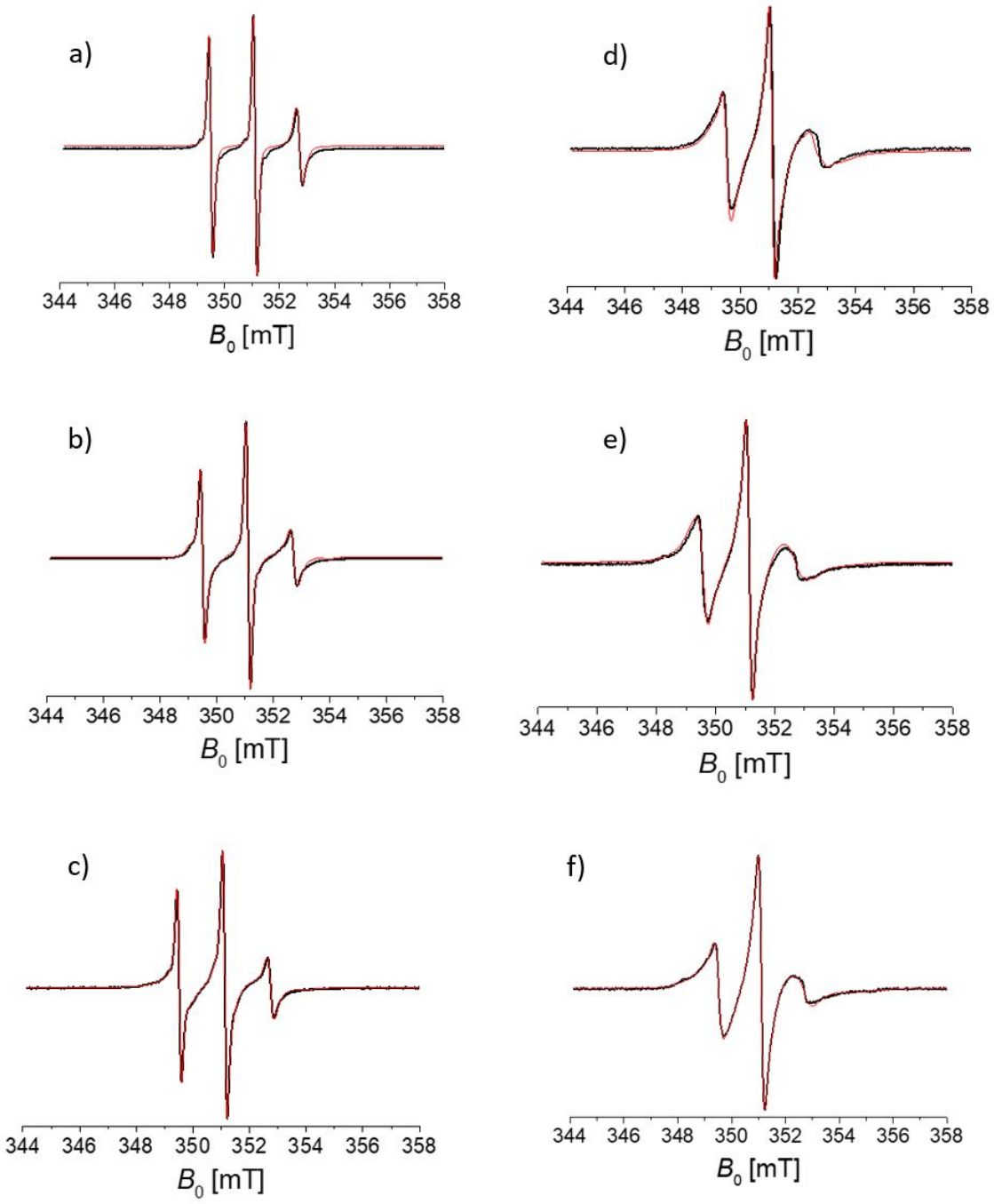

Figure S4. CW EPR spectra (black) and spectral simulations (red) of the different spin-labeled MAX_2 mutants**.** a) Least squares fitting for the R5C dimer (without EBOX DNA). b) Least squares fitting for the R5C-G35C mixture (without EBOX DNA). c) Least squares fitting for the R5C-R55C mixture (without EBOX DNA). d) Least squares fitting for the R5C dimer (with EBOX DNA) e) Least squares fitting for the R5C-G35C dimer (with EBOX DNA) f) Least squares fitting for the R5C-R55C dimer (with EBOX DNA)

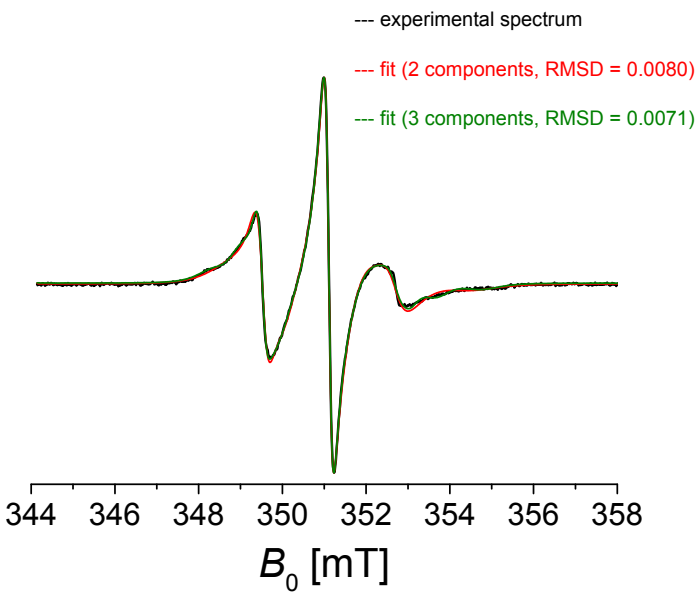

Figure S5. Comparison of least squares fitting (three species *versus* two species) for the R5C-R55C mixture (with EBOX DNA)

Table S1. Rotational correlation time (τ_c) for MTSL spectra of doubly labeled MAX_2, in the absence and presence of $EBOX_{(DNA)}$; relative percentage of the slow and fast components are indicated in bracket.

mutant	**– EBOX$_{(DNA)}$**		**+ EBOX$_{(DNA)}$**	
	τ_{corr}/ns (% slow)	τ_{corr}/ns (% fast)	τ_{corr}/ns (% slow)	τ_{corr}/ns (% fast)
R5C	1.18 (14)	0.14 (86)	1.18 (54)	0.14 (46)
R5C-G35C	4.21 (15)	0.55 (85)	4.21 (65)	0.55 (35)
R5C-R55C	5.6 (17)	0.8 (83)	5.6 (68)	0.8 (32)

Supplementary DEER data

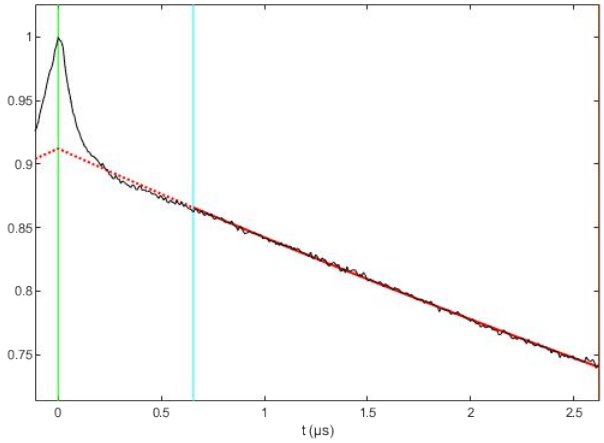

Figure S6. Background fit (red line), zero time (green line) and first data point for the fit of the background of the DEER signal for sample R5C-R5C.

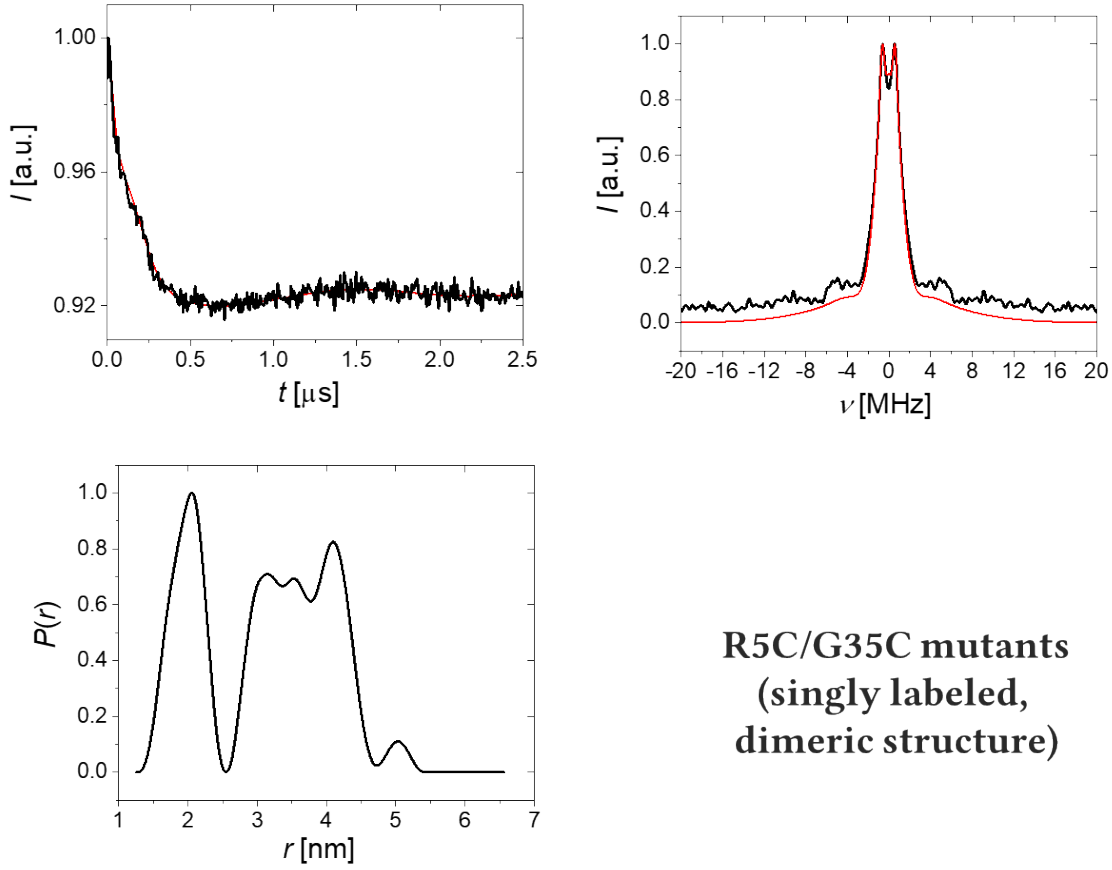

Figure S7. Dipolar evolution function (background corrected) for the PELDOR/DEER signal (*top left*); Pake pattern and corresponding fit (*top right*); distances distribution for the R5C-G35C (without EBOX DNA sequence), upon Tikhonov regularization applied to the Pake pattern (*bottom left*).

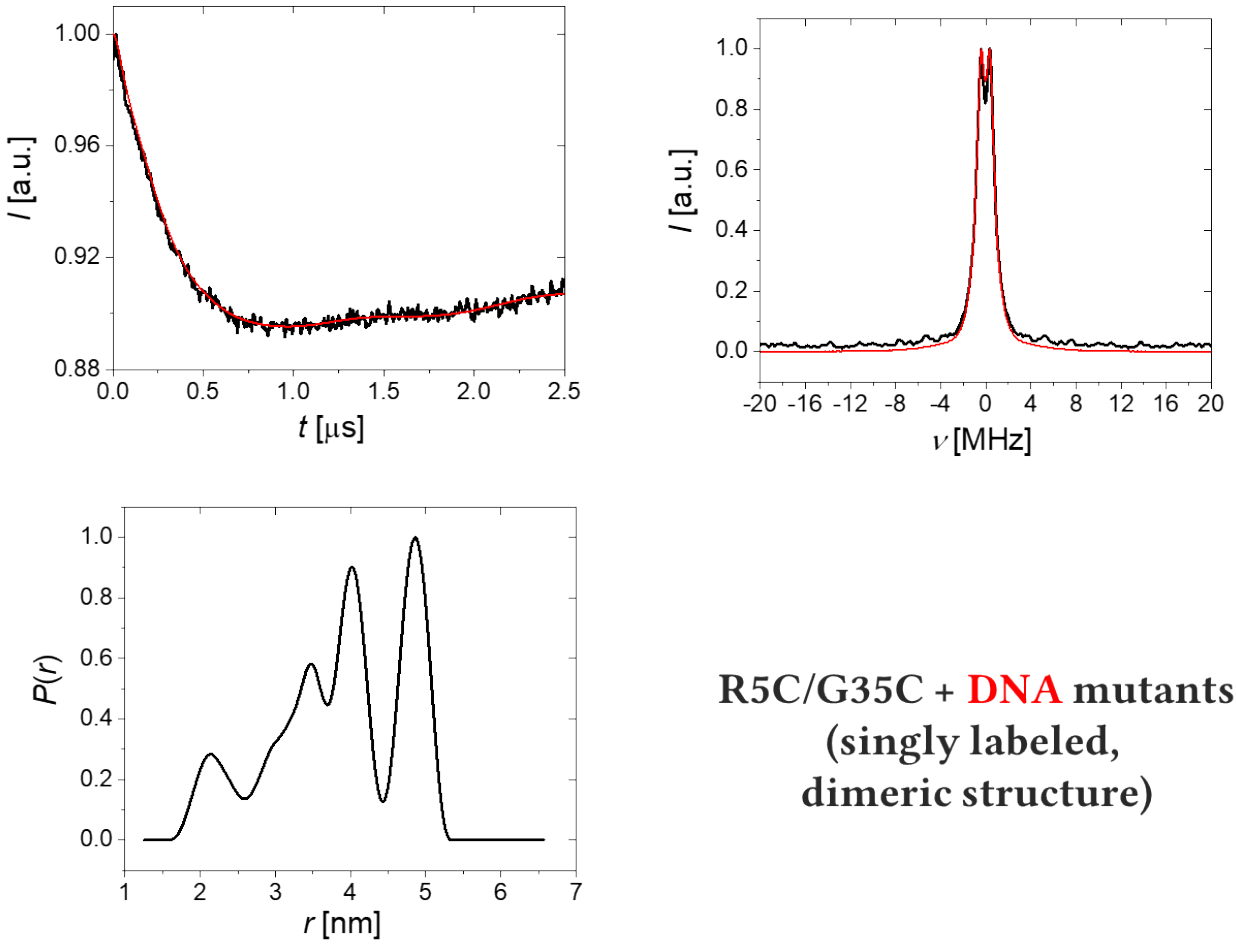

Figure S8. Dipolar evolution function (background corrected) for the PELDOR/DEER signal (*top left*); Pake pattern and corresponding fit (*top right*); distances distribution for the R5C-G35C (with EBOX DNA sequence), upon Tikhonov regularization applied to the Pake pattern (*bottom left*).

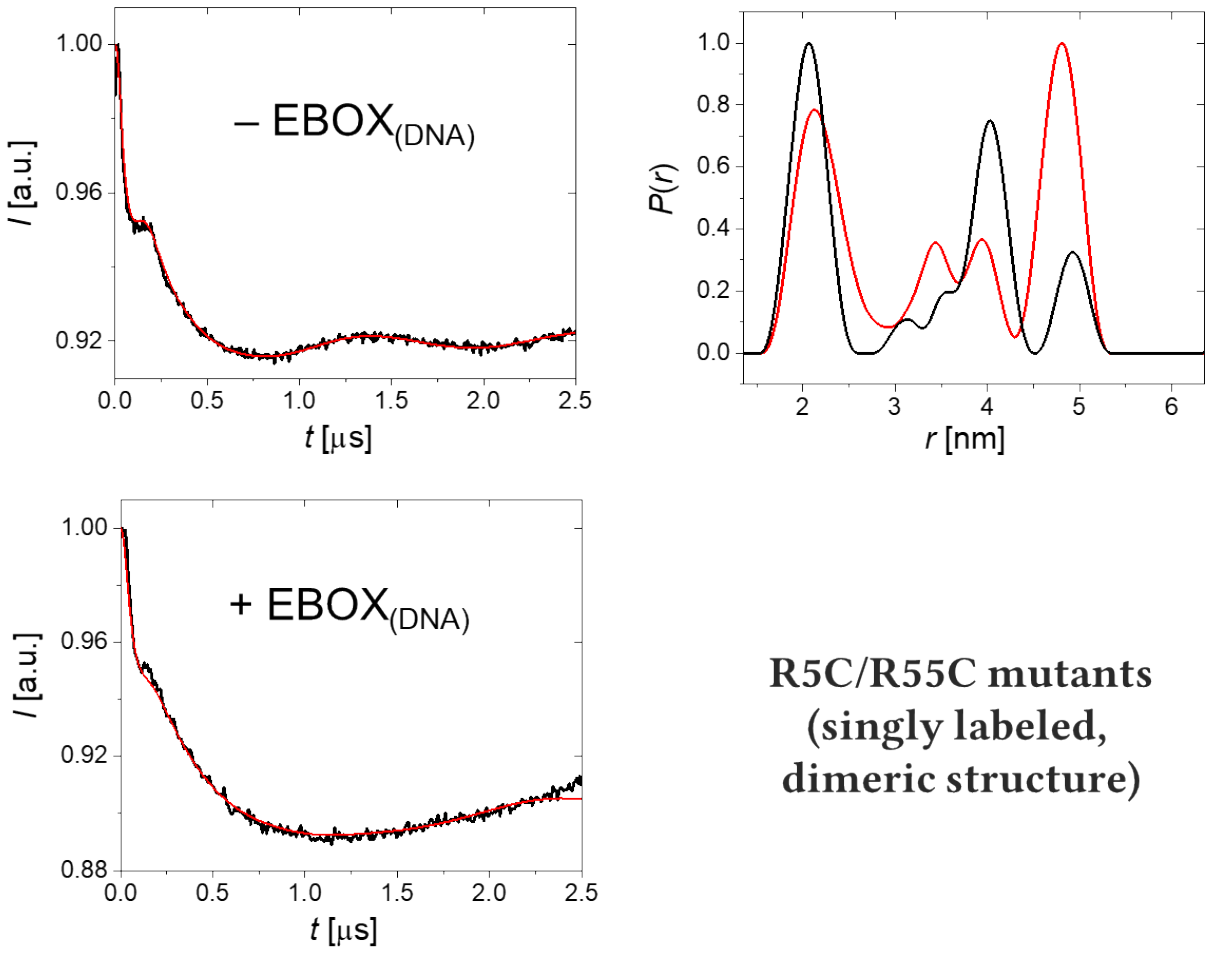

Figure S9. Dipolar evolution function (background corrected) for the PELDOR/DEER signal the R5C-R55C (without EBOX DNA sequence) (*top left*) and for the PELDOR/DEER signal the R5C-R55C (with EBOX DNA sequence) (*bottom left*). On the *top right* of the figure a comparison of distances distribution for those two samples is shown.

Script S1. Mat-Lab script for least squares fitting of CW nitroxide spectra.

```
[Bexp,spec] = eprload(EPR spectrum.DTA);

%------------------------------------------------------
% System
%------------------------------------------------------

Exp.mwFreq=9.858;
Exp.CenterSweep = [351.14 14];
Opt.Verbosity = 1;
%Opt.Method = 'hybrid';

%%%%%%%% Espèce très rapide %%%%%%%%%%%%%

Sys1.S=1/2;
Sys1.g=[g1, g2, g3];
Sys1.lw= [0.09 0.03];% mT
Sys1.Nucs = '14N';
%Amhz=mt2mhz(Amtg); % A en MHz
Sys1.A=[A1 A2 A3];
Sys1.logtcorr=log10(1.3e-8);

Sys1.weight=0.57;



%%%%%%%% Espèce lente %%%%%%%%%%%%%

Sys2.S=1/2;
Sys2.g=[ g1, g2, g3];
Sys2.lw=[0.05 0.01];% mT
Sys2.Nucs = '14N';
%Amtg=[0.460 0.9 4.0];
%Amhz=mt2mhz(Amtg); % A en MHz
Sys2.A= [A1 A2 A3];
Sys2.logtcorr=log10(3.5e-9);
%Sys2.logtcorr=log10(10e-9);
Sys2.weight=0.43;

%%%%%%%% Espèce intermédiaire %%%%%%%%%%%%%
%Sys3.S=1/2;
%Sys3.g=[2.011 2.0051 2.0061];
%Sys3.lw=0.12;% mT
```

```
%Sys3.Nucs = '14N';
%Amtg=[0.460 0.9 4.0];
%Amhz=mt2mhz(Amtg); % A en MHz
%Sys3.A= [15.83 32.95 73.16];
%Sys3.logtcorr=log10(1.01e-8);
%Sys3.weight=0.59;


%%%%%%% paramètres variables %%%%%%%%%%%%
Vary1.weight=0.30;
Vary2.weight=0.7;
%Vary3.weight=0.44;

Vary1.logtcorr=(-6.0);
Vary2.logtcorr=(-9.0);
%Vary3.logtcorr=0.6;

Vary1.lw= [0.05 0.04];
Vary2.lw = [0.10 0.10];
%Vary3.lw=0.2;

Vary1.g=[0.003 0.003 0.003];
Vary2.g=[0.003 0.005 0.005];

Vary1.A=[10 10 10];
Vary2.A=[60 25 15];
%Vary3.A=[6 6 6];

%%%%%%%% fit %%%%%%%%%%
%FitOpt.Method = 'genetic fcn'; % simplex algorithm, integrals of spectra
%FitOpt.Scaling = 'lsq0';
%FitOpt.PopulationSize = 50;
esfit('chili',spec,{Sys1,Sys2},{Vary1,Vary2},Exp);

%esfit('pepper',y,Sys0,Vary,Exp,Opt,FitOpt);

%spc1 = chili(Sys0,Exp,Opt);
%data = [spec(:) spc1(:)];
%save TestAcw2.txt data -ascii;
```

Supplementary NMR Data

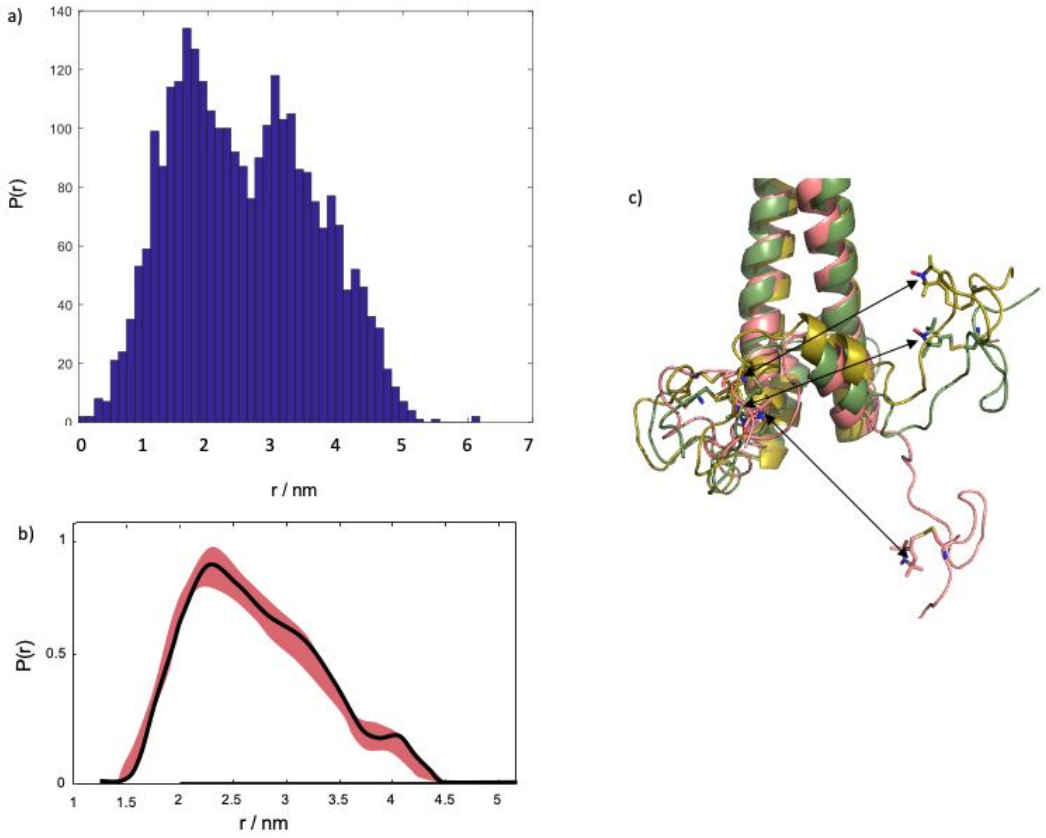

Figure S10. a) Electron-electron distances found in a 50 ns long MD simulation of MTSL spin-labeled R5C MAX_2. The distances correspond to the separation between the nitrogen atoms in the two nitroxide moieties. As can be clearly seen, two distinct distance maxima around 2 and 3 nm were obtained for the hinged and the extended ensembles, respectively. However, the two distributions overlap due to the conformational freedom of the MTSL attached to positions R5C. b) The experimentally obtained distribution for comparison. The regularization approach used to extract the distances from the dipolar evolution function can contribute additional broadening of the distance distributions such that the two maxima cannot be as clearly observed as in panel (a). It should be noted that distances below ca. 1.5 nm cannot be obtained experimentally due to the relatively long pulse length of 8 and 16 ns in the DEER experiments. The experimental distance distribution therefor cannot reflect the shorter distances found in the simulation. c) Representative conformations for hinged (left IDR) and extended (right IDR) conformations in the simulations with MTSL attached to R5C. The black arrows indicate the SL-SL distances.

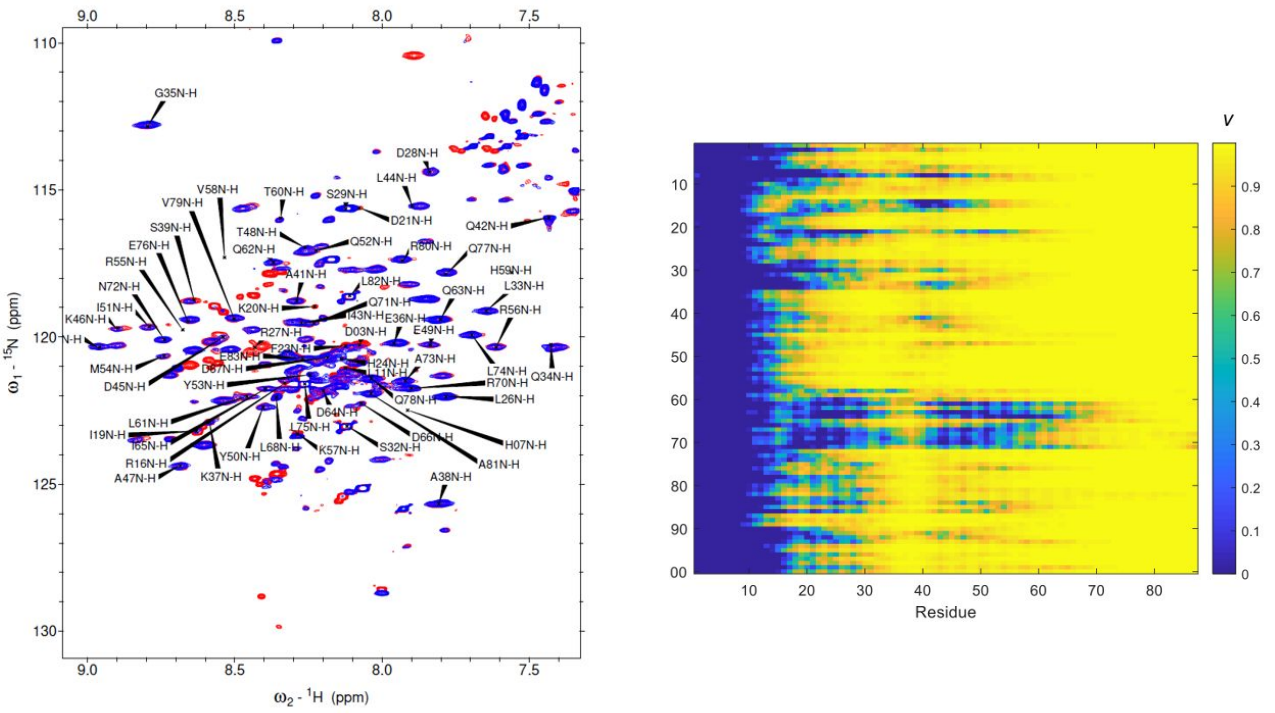

Figure S11. a) TROSY ^{1}H-^{15}N spectra of MAX$_2$ spin-labeled at position R5C. Signal assignments are indicated. ^{1}H-^{15}N cross peak assignments were obtained from the biological magnetic resonance data base (BMRB) entry 5956 and the work by Sauvé et al.[8, 9] Blue spectra were recorded with an active SL and red spectra with a chemically (using ascorbate) deactivated SL. b) Calculated, residue-resolved PRE ratios V for the entire trajectory used for Fig. 1b of the main text and to match the experimental PRE results in Fig. 2a and b of the main text.

References

1. G. Kizilsavas, K. Ledolter and D. Kurzbach, *Biochemistry-Us*, 2017, **56**, 5365-5372.
2. G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banjam, C. R. Timmel, D. Hilger and H. Jung, *Appl. Magn. Reson*, 2006, **30**, 473-498.
3. T. H. Edwards and S. Stoll, *J Magn Reson*, 2018, **288**, 58-68.
4. Y. W. Chiang, P. P. Borbat and J. H. Freed, *J Magn Reson*, 2005, **172**, 279-295.
5. D. Calvetti, S. Morigi, L. Reichel and F. Sgallari, *J Comput Appl Math*, 2000, **123**, 423-446.
6. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax, *J Biomol Nmr*, 1995, **6**, 277-293.
7. T. D. Goddard and D. G. Kneller, *Journal*.
8. S. Sauve, L. Tremblay and P. Lavigne, *J Mol Biol*, 2004, **342**, 813-832.
9. S. Sauve, J. F. Naud and P. Lavigne, *J Mol Biol*, 2007, **365**, 1163-1175.
10. M. J. Abraham, T. Murtola, R. Schulzb, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19-25.
11. Y. Duan, C. Wu, S. Chowdhury, M. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J Comput Chem*, 2003, **24**, 1999-2012.
12. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2013, **3**, 198-210.
13. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.