Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Predicting the Rotational Tumbling of Dynamic Multidomain Proteins and Supramolecular Complexes**

Nasrollah Rezaei-Ghaleh, Frederik Klama, Francesca Munari, and Markus Zweckstetter**

anie_201305094_sm_miscellaneous_information.pdf

Supporting Figures

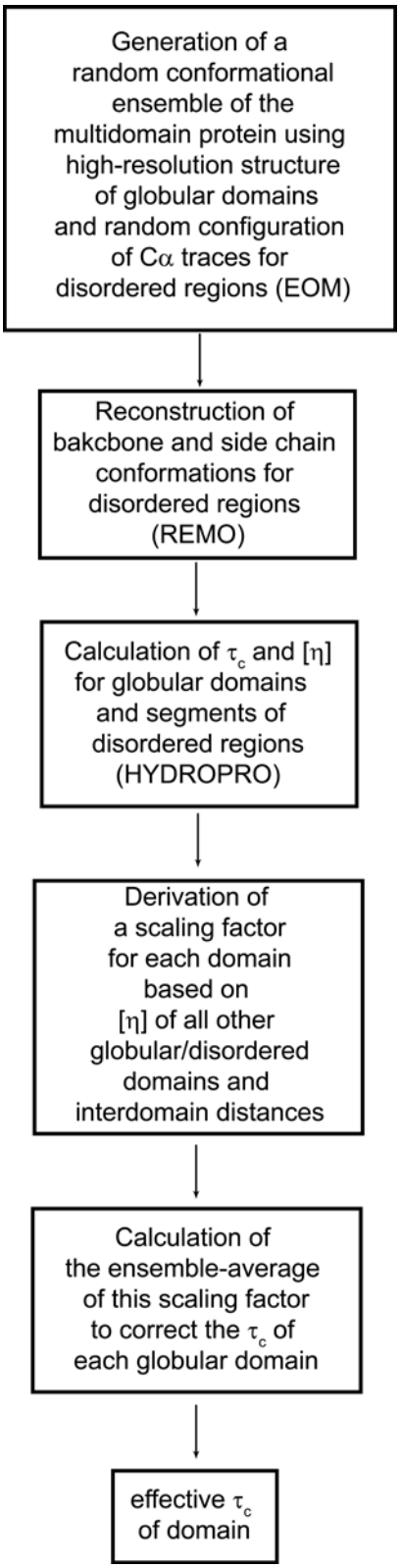

Figure S1. A diagrammatic representation of consecutive steps of HYCUD calculations.

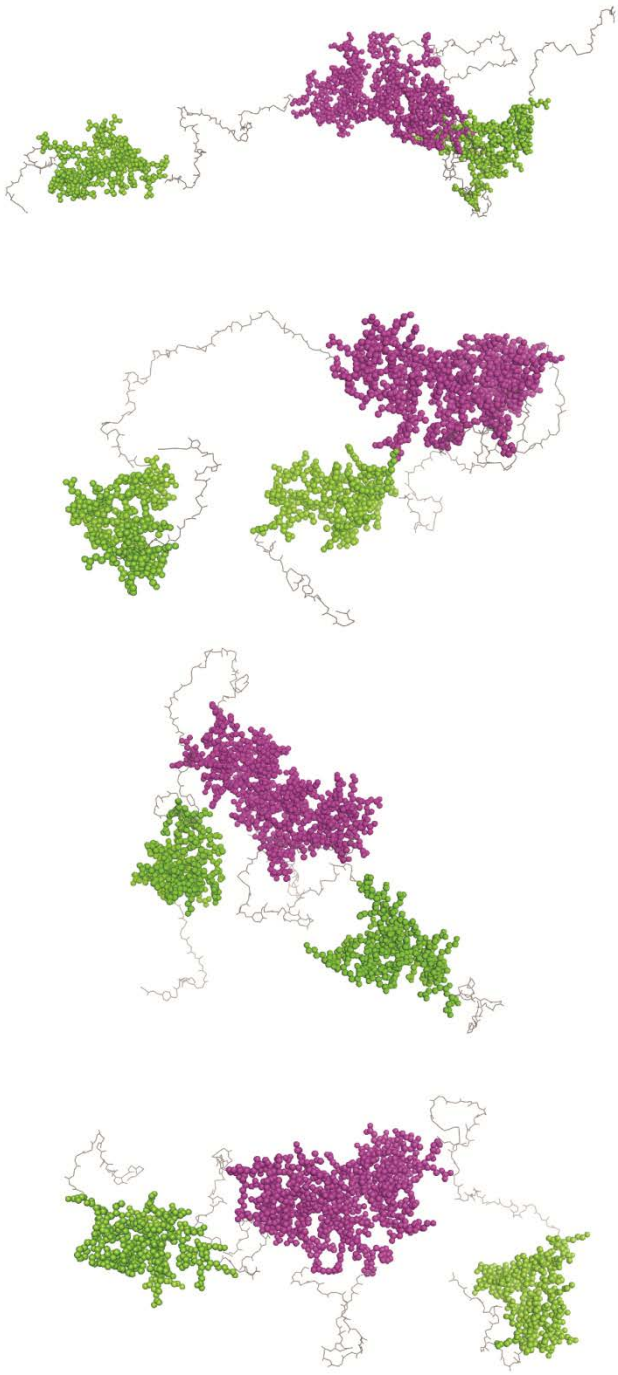

Figure S2. Four representative models of the full-length HP1 ensemble. Globular domains are marked in green (CD) and magenta (CSD), while disordered tails and linkers are shown in gray.

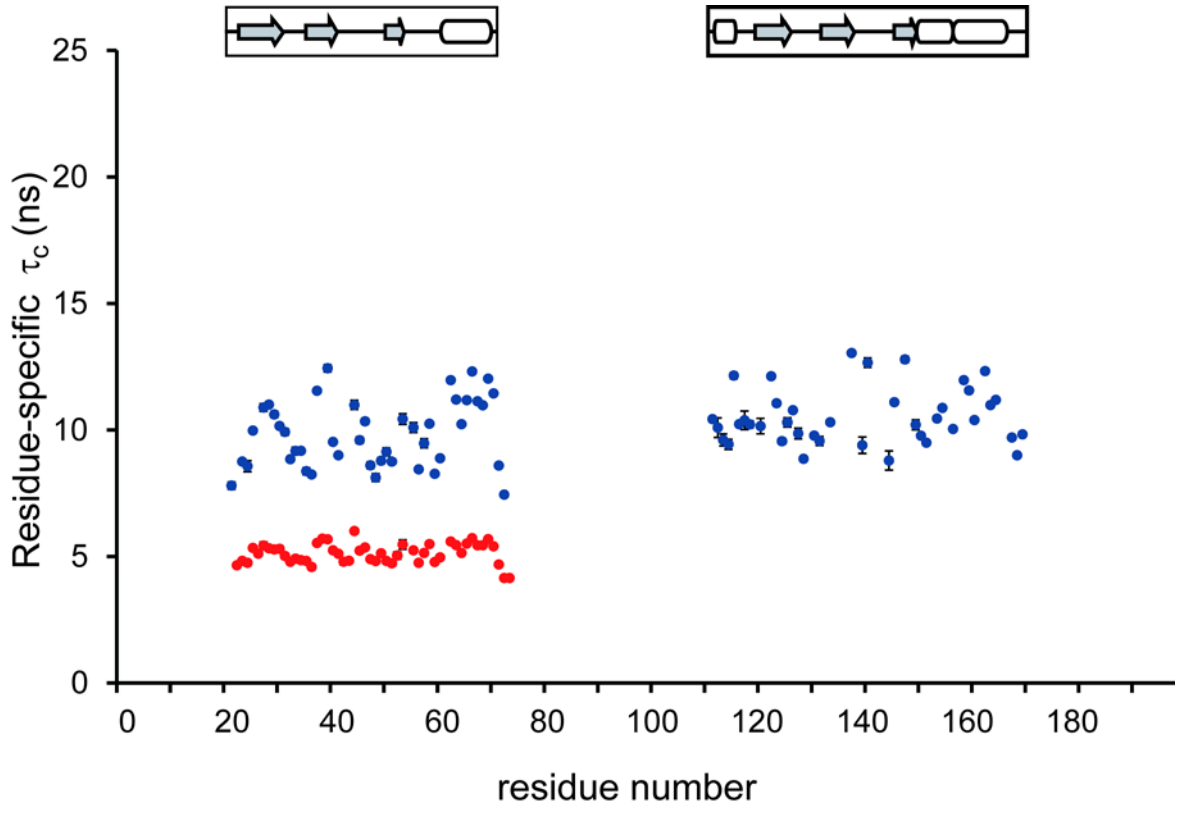

Figure S3. Residue-specific rotational correlation time (τ_c) of CD (19-79) (red) and monomeric HP1 (blue), as determined from experimental ^{15}N R_2/R_1 NMR spin relaxation ratios. The domain organization of HP1 is schematically shown at the top. Error bars are often smaller than the symbol size.

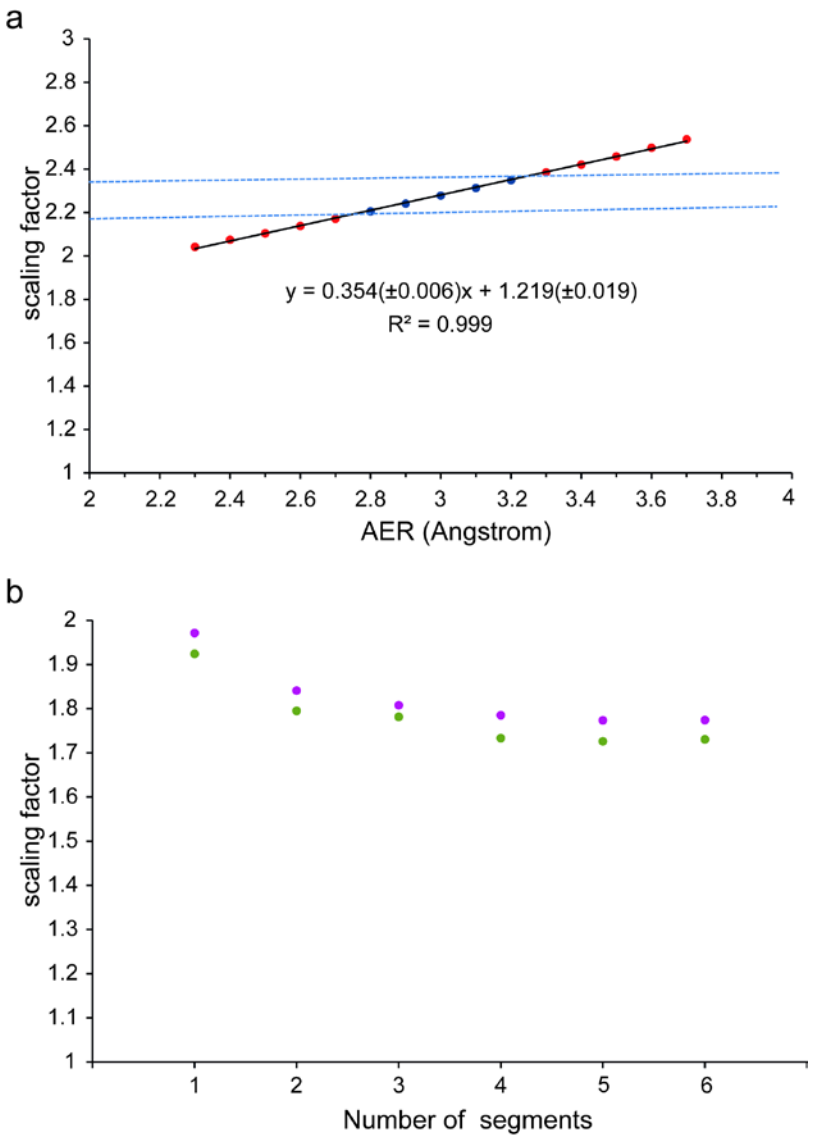

Figure S4. Influence of adjustable parameters on HYCUD calculations. a) HYCUD-derived scaling factor for the effective rotational correlation time (τ_c) of chromodomain (green) and choromshadow domain (magenta) in the context of monomeric HP1, as a function of AER of disordered segments used in hydrodynamic calculations. At a relatively wide range of 2.75-3.25 Å (dashed lines), the scaling factor changes only within the experimental range of error. b) HYCUD-derived scaling factor for the effective rotational correlation time (τ_c) of chromodomain (green) and choromshadow domain (magenta) in the context of monomeric HP1, as a function of the number of fragments in the 37-residue long hinge region. Only contribution of the hinge region to the scaling factor is shown. The scaling factor changes only slightly when the size of fragments increases from ~ 6 residues (6 fragments) to 18 residues (2 fragments).

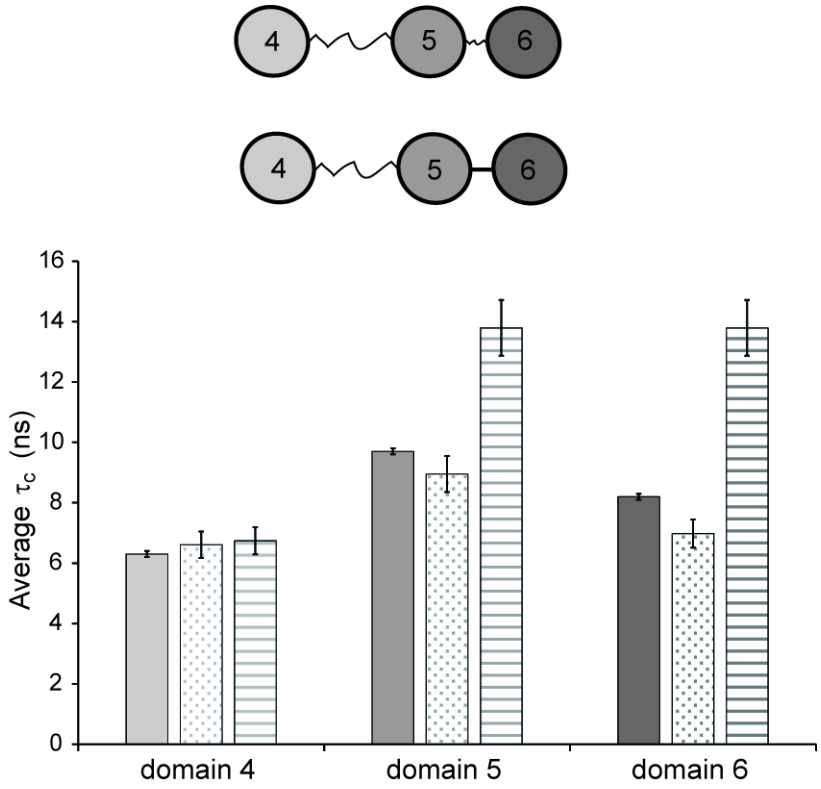

Figure S5. Analysis of rotational correlation times in the three-domain construct of Wilson disease protein. Experimental τ_c are shown as solid column. Values calculated by HYCUD for completely flexible and rigid states of the interdomain linker are indicated by dotted and lined columns, respectively. A cartoon representation of the protein is shown on top.

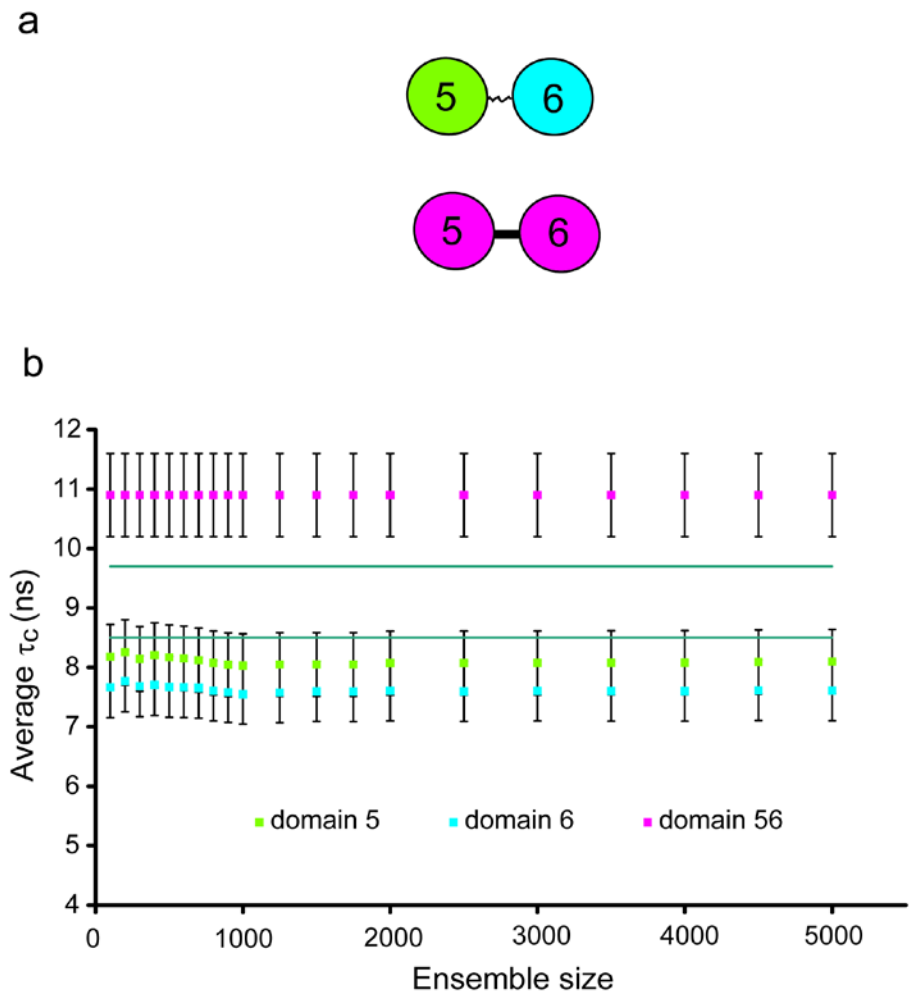

Figure S6. Analysis of rotational correlation times in the two-domain construct of Wilson disease protein. a) Cartoon representation with the interdomain linker in a completely flexible (wavy lines) or rigid (straight line) state. b) Effective rotational correlation times (τ_c) of domains 5 and 6 as a function of the number of models included in the HYCUD calculation. The experimental value of the two domains, shown as lines (average ± standard deviation), lies between the two extremes of a fully flexible and rigid state.

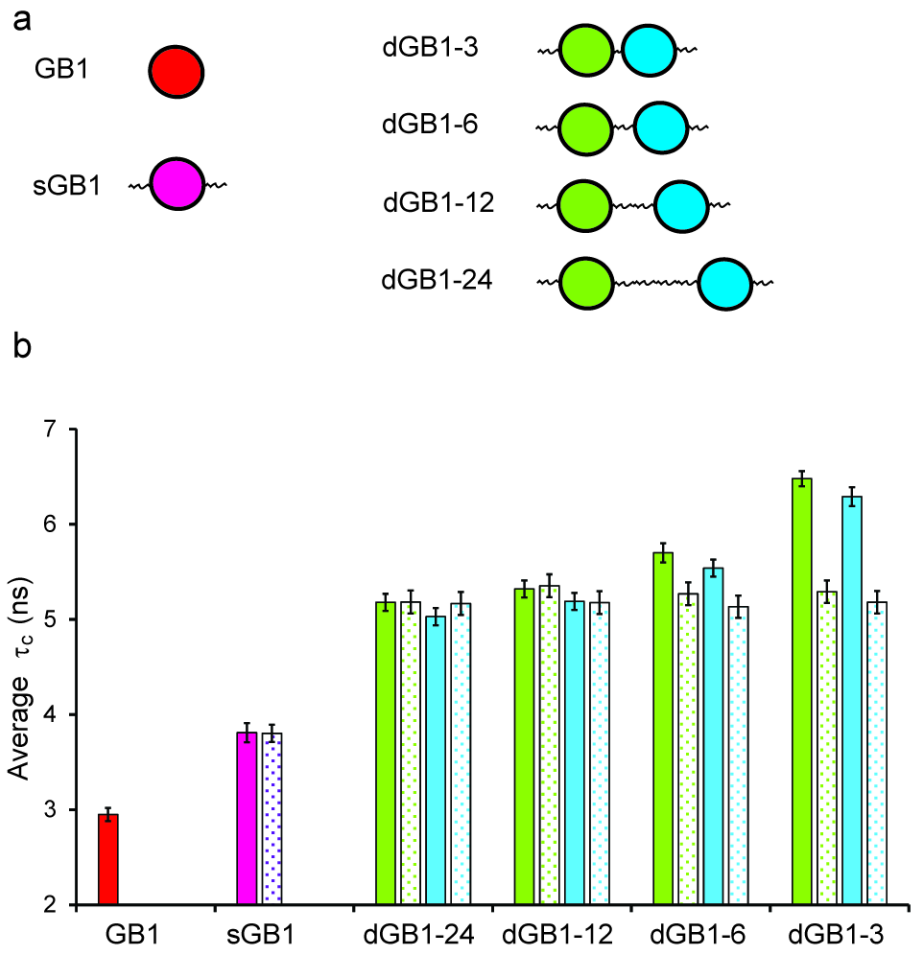

Figure S7. Influence of linker length on hydrodynamic coupling. a) Cartoon representation of GB1 protein constructs: tail-less GB1, single GB1 construct (sGB1) with short N- and C-terminal tails and double GB1 constructs (dGB1) with increasing linker length between the N-terminal and C-terminal domains (NTD, CTD). b) Experimental and HYCUD-calculated values of the effective rotational correlation time (τ_c) of the GB1 domain in sGB1 and dGB1 (separately for NTD and CTD) are shown as solid and dotted columns, respectively.

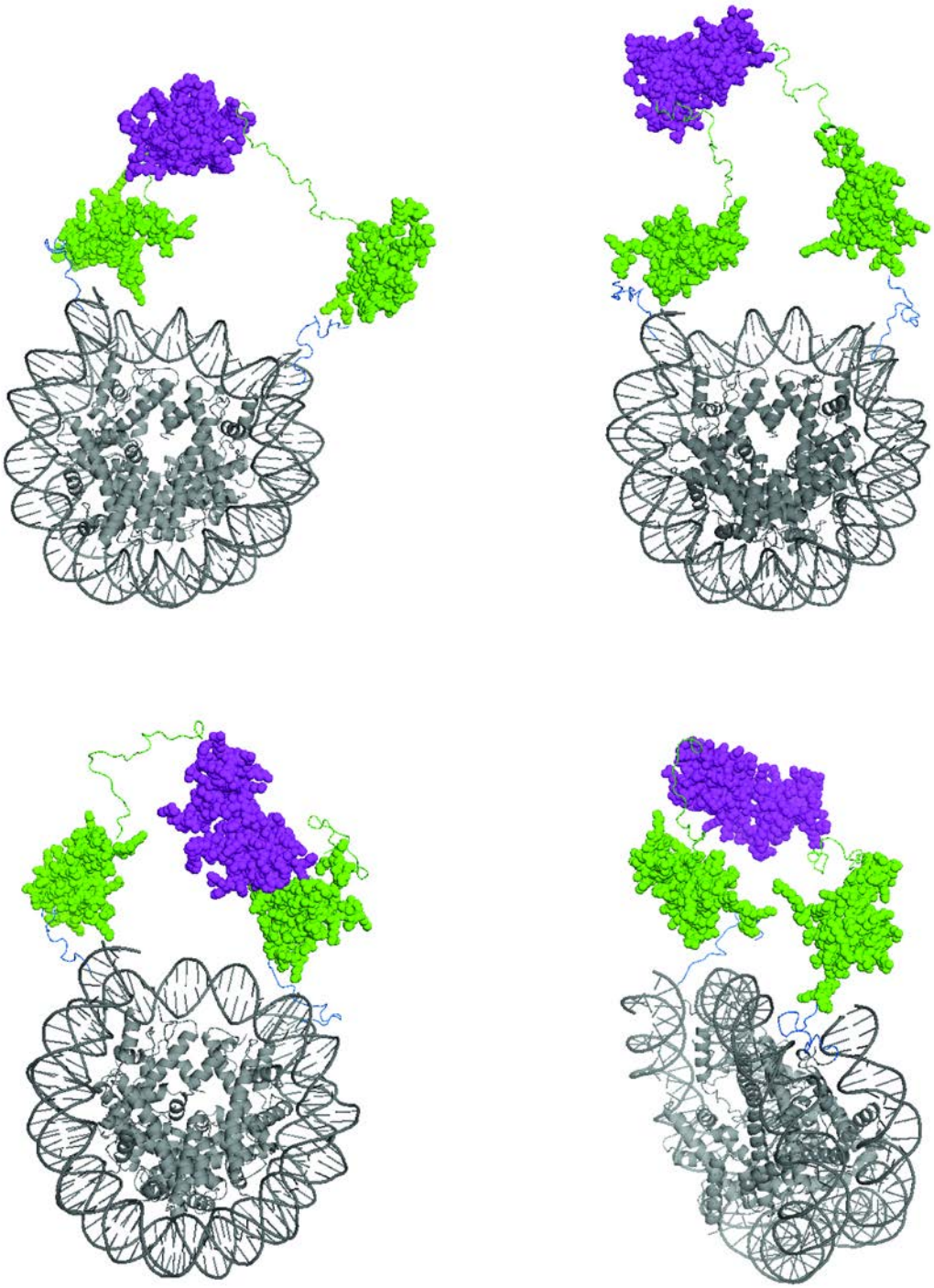

Figure S8. Four representative models of the HP1-nucleosome complex ensemble. Globular domains of HP1 are marked in green (CD) and magenta (CSD), nucleosome (histones+DNA) in gray and H3 tail protruding out of nucleosome core in blue.

Supporting text

Influence of the length of interdomain linkers on reorientational coupling

An intriguing example of domain diffusion in a modular protein are the changes that were observed in the rotational diffusion tensor of the protein GB1, when the domain was connected to an identical second domain [1]. A detailed NMR study showed that the rotational correlation time of the two globular domains is most sensitive to the length of the interdomain linker, while diffusion anisotropy and especially the orientation of the rotational diffusion tensors were much less affected [1]. The fact that it is primarily the rotational correlation time that is changed in flexible modular systems, provides support for the HYCUD approach that approximates the influence of the additional domains by a scaling factor for τ_c (equation 2).

The two 56-residue GB1 domains had been connected through Gly-Ser-Gly linkers comprising 3, 6, 12 and 24 residues (Figure S7a) [1]. The Gly-Ser-Gly repeats were used to minimize residual structure in the linker. NMR relaxation measurements showed that tumbling of the folded GB1 domains is slowed down in the context of the flexible two-domain arrangement (dGB1) [1]. Even in case of a 24-residue linker, domain diffusion differed substantially from the value of its isolated domain (τ_c = 3.0±0.1 ns at 25 °C [1]). In addition, a construct in which a seven- and six-residue tail was attached to the N- and C-terminus of a single GB1 domain (sGB1), respectively, showed a τ_c increase to 3.8±0.1 ns. In the two-domain constructs, the τ_c of the N-terminal domain (NTD) decreased from 6.5±0.1 ns to 5.7±0.1, 5.3±0.1 and 5.2±0.1 ns, when the two domains were separated by 3, 6, 12 and 24 residues. For the C-terminal domain (CTD), the corresponding values were slightly smaller: 6.3±0.10, 5.5±0.1, 5.2±0.1 and 5.0±0.1 ns [1].

We used the high-resolution crystal structure of the isolated GB domain refined with residual dipolar couplings (PDB code: 1P7E [2]) to generate conformer ensembles for the tail-extended GB1 (sGB1) and the different dGB1 constructs. Using the AER value giving the best agreement between the HYDROPRO-calculated τ_c of the isolated GB1 domain (i.e. without additional tails) and its experimental τ_c, the value for sGB1 was predicted at 3.8±0.1 ns, in perfect agreement with experiment (Figure S7b). For the NTD in dGB1, HYCUD gave 5.3±0.1, 5.3±0.1, 5.4±0.1 and 5.2±0.1 ns for linker lengths of 3, 6, 12 and 24 residues. The corresponding values for CTD were 5.2±0.1, 5.1±0.1, 5.2±0.1 and 5.2±0.1 ns (Figure S7b). The agreement with experimental values is good for dGB1 of linker lengths 12 and 24. In contrast, when the linker length is reduced to 6 and 3 residues HYCUD does not predict an increase in τ_c. Considering the persistence length of disordered polypeptide chains (~7 residues) [3], it is likely that the diffusion of the two domains is partially correlated for very short linkers.

Supporting Methods

Production of recombinant HP1 proteins

HP1 variants were based on the sequence of the β isoform of human heterochromatin protein 1 (hHP1β). Monomeric HP1 (residues 2-185 with an I161A mutation; hHP1βI161A) and chromodomain (residues 19-79; CD (19-79)) were cloned into pET16b expression vectors (Novagen) together with an N-terminal His-affinity tag and TEV cleavage site [4]. Protein expression was done in BL21DE3 cultures overnight at 24 °C in minimal medium containing $^{15}NH_4Cl$ as unique source of nitrogen to obtain protonated ^{15}N-labeled samples. Protein purification was carried out with affinity chromatography (Ni-NTA, Qiagen), His-tag removal by TEV cleavage and size-exclusion chromatography (Superdex75, GE Healthcare) following standard procedures. Samples were dialyzed against sodium phosphate buffer (20 mM, pH 6.5) containing 50 mM NaCl, 2 mM DTT and 0.02% NaN_3.

NMR experiments

^{15}N spin relaxation experiments were conducted on a 600 MHz Bruker spectrometer at 298 K using samples with 0.5 mM protein concentration. ^{15}N longitudinal relaxation rates (R_1) were measured with relaxation delays of 8, 40, 80, 120, 220, 320, 500, 800 and 1200 ms for hHP1βI161A, and 8, 60, 150, 300, 500, 800 and 1000 ms for CD (19-79). ^{15}N longitudinal relaxation rates in the rotating frame ($R_{1\rho}$) were measured with a ^{15}N spin-lock field strength of 2 kHz, using relaxation delays of 20, 40, 60, 80, 100, 140, 180 and 240 ms for hHP1βI161A, and 20, 40, 60, 80, 100, 120, 160, 200 and 240 ms for CD(19-79). ^{15}N transverse relaxation rates (R_2) were calculated from R_1 and $R_{1\rho}$ according to:

$$R_2 = \frac{(R_{1\rho} - R_1 cos^2\theta)}{sin^2\theta},$$

where $\theta = tan^{-1}(\frac{\nu_1}{\Omega})$. ν_1 is the ^{15}N spin-lock field strength (in Hz) and Ω the resonance offset from the spin-lock carrier (Hz) [5]. Residue-specific rotational correlation times ($\tau_{c,loc}$) were calculated from the observed R_2/R_1 ratios [6]. The overall rotational correlation time (τ_c) was determined as the average of $\tau_{c,loc}$ after exclusion of residues with peak overlap or insufficient signal-to-noise ratio.

Ensemble generation and HYCUD calculations

For each system, an ensemble of 5000 conformers was generated using the EOM program included in ATSAS 2.5 package. [7] The following PDB structures were used as rigid bodies: 1AP0 [8] for CD and chains A,B of 2FMM [9] for CSD of HP1, 2ROP [10] for domains 3-4 and 2EW9 [11] for domains 5-6 of Wilson disease protein, 1P7E [2] for GB domain, 1GUW [12] for CD bound to H3 tail, and 3LZ1 [13] for the nucleosome core particle. When the protein structure was solved by NMR, the average structure over the NMR ensemble was used. The boundaries of the folded domains were determined based on the available NMR data.[1, 14] Experimentally determined τ_c values of isolated globular domains were used to determine the AER required for hydrodynamic calculations. Uncertainty ranges of experimental τ_c values were transformed into an uncertainty range for AER, which was later propagated to determine the range of error for HYCUD calculations.

A key step in HYCUD calculation is to convert the interdomain distance r_{ij} to the effective concentration of domain *j*, i.e. C_j, experienced by domain *i*. The effective concentration C_j was defined as the concentration for which the expected value to find the domain *j* at the distance r_{ij} from domain *i* is 1. The *Cj* was then estimated in equivalence to a simple cubic lattice arrangement with the edge length of $\sqrt[3]{6}\,r_{ij}$ in which we can find 6 number of domain *j* in the distance $\sqrt[3]{6}\,r_{ij}$, hence 1 domain *j* at the distance r_{ij}.

The HYCUD method was implemented using Python 2.7. HYCUD calculations were performed using Linux machines with 8 or 12 cores. Typical HYCUD calculation for an ensemble of 500 protein models of ~ 400 residues takes approximately 2-3 hours. Below, more details are provided for each studied example.

HP1 proteins

To generate the ensemble of CD (19-79), we used the high-resolution average NMR structure of 1AP0.pdb [8]. Residues 19-72 defined the rigid globular domain and residues 73-79 were allowed to sample random conformations according to their conformational propensities in the disordered state.

Monomeric HP1 was modeled using as two rigid domains the above-mentioned structure for CD (19-72) and the crystal structure of CSD (110-171) monomer, as defined in chain A of 2FMM.pdb [9]. Residues 1-18, 73-109 and 172-185 were added and treated as disordered parts during ensemble generation. For hydrodynamic calculations, the N-terminal tail 1-18 was split into two 9-residue fragments, the hinge region (residues 73-109) into three fragments of 12, 13 and 12 residues and the C-terminal tail was considered as a single fragment of 14 residues. To evaluate the effect of segment size onto the HYCUD predictions, we changed the fragmentation pattern of residues 73-109 as follows: one fragment of 37 residues; two fragments of 18 and 19 residues; three fragments of 12, 13 and 12 residues; four fragments of 9, 9, 10 and 9 residues; five fragments of 7, 8, 7, 8 and 7 residues; and six fragments of 6, 6, 6, 7, 6 and 6 residues (see Figure S4b).

Dimeric HP1 was modeled considering the CD and dimeric CSD as rigid bodies. The average NMR structure of CD (19-72) in 1AP0.pdb was taken for the two CD domains of dimeric HP1 and the CSD (110-171) dimer was defined according to chains A,B of 2FMM.pdb. In order to generate the ensemble, we attached the two HP1β chains through their C-termini and reverted the aminoacid sequence of the second chain. Then, we fixed the position of CSD and let the rest of the chains sample a large structural space through conformational freedom of the disordered parts. For hydrodynamic calculation, we considered the two CDs as separate diffusional entities and the CSD dimer as a single object. The disordered parts were treated as in monomeric HP1 (see above).

In the ensemble generation of dimeric HP1 described above, the C-terminal tails of the two proteins were attached to each other. This restricts the freedom of the C-termini to move away from the globular dimeric CSD and leads to a slight overestimation of CSD rotational correlation times (τ_c). To eliminate this bias, a different ensemble of dimeric CSD was generated as follows: the structure of the globular CSD dimer was defined as described before, but the sequence of the first CSD molecule was reverted so that its C-terminal tail was in the beginning of the sequence and the C-terminal tail of the second CSD molecule was at the end of the sequence. In this way, the two C-terminal tails were free to sample conformational space separate from each other. We made the hydrodynamic calculations for this ensemble and obtained τ_c of globular CSD dimer in the ensemble after correction for the effects of C-terminal tails. This value was fixed as the un-corrected τ_c (τ_{c0}, see equation 2 in the main text) of the CSD dimer in HYCUD calculations for dimeric HP1.

Wilson disease protein

The two-domain (WLN5-6) protein construct was modeled using the average high-resolution NMR structure 2EW9.pdb [11] with domains 5 (72 residues) and 6 (69 residues) connected through an 8-residue linker. The three-domain construct (WLN4-6) ensemble was built using the average NMR structure of domain 4 in 2ROP.pdb (72 residues) [10], connected through a 56-residue linker to domains 5 and 6 as defined above. For hydrodynamic calculations, the linker between domains 4 and 5 was divided into 4 segments of 15, 14, 14 and 15 residues. The 8-residue linker between domains 5 and 6 was considered as a single domain.

Two-domain GB1 variants

The high-resolution crystal structure of GB protein refined with residual dipolar couplings, as in 1P7E.pdb [2], was used to define the structure of globular GB1 domain (56 residues). The ensemble of sGB1 was built adding short flexible tails of 7 and 6 residues upstream and downstream of the domain. The ensembles of two-domain GB1 constructs were generated with two GB1 domains, a flexible N-terminal tail of 7-residues before the first GB1 domain, a flexible inter-domain linker of various lengths (3, 6, 12 and 24 residues), and a flexible C-terminal tail of 6 residues, as described in [1]. For hydrodynamic calculations, the 24-residue linker was split to two fragments of 12 residues. In all other cases, the linker was treated as a single fragment.

The HP1-nucleosome complex

The dimeric HP1β in complex with H3K9me3-nucleosome (tri-methylated at lysine 9 of H3 tail) was modeled using the following high-resolution structures, 1GUW.pdb [12], chains A,B of 2FMM.pdb [9] and 3LZ1.pdb [13], for two CDs bound to H3 tail, the dimeric CSD and the histone octamer bound to DNA, respectively. We generated the ensemble of the HP1-nucleosome complex for two different scenarios, (i) the symmetric recognition in which the two CDs of one HP1 dimer bind to the two H3K9me3 H3-tails of a single mono-nucleosome, and (ii) the bridging mode in which one HP1 molecule links two mono-nucleosomes.

For symmetric binding, we first removed the DNA part from 3LZ1.pdb and split the histone octamer into two tetramers. To generate the ensemble, the protein sequence was defined in the following order: the first histone tetramer ended with H3, then H3 tail residues 18-38, then H3 tail residues 1-17 bound to CD, then CD, the hinge region between CD and CSD, and the dimeric CSD. The hinge region, the second CD, the H3 tail bound to CD, the rest of H3 tail and the second histone tetramer starting with H3 followed. The two histone tetramers, two CDs in complex with H3 tail and the dimeric CSD were treated as rigid bodies, while the other parts as disordered. During ensemble generation, the coordinates of the two histone tetramers were kept fixed and the second tetramer was considered as rigidly bound to DNA.

For binding of one HP1 molecule to two mono-nucleosomes, we defined the protein sequence as follows: the first nucleosome octamer ended with H3, then H3 tail residues 18-38, then H3 tail residues 1-17 bound to CD, then CD, the hinge region, dimeric CSD, the second hinge region and CD, the H3 tail residues 1-17 bound to CD, H3 tail residues 18-38 and finally the second nucleosome octamer starting with H3.

In hydrodynamic calculations, the disordered parts were treated as follows: each of the two hinge regions was split into two segments of 14 and 15 residues and the two flexible H3 tails were divided into fragments of 11 and 12 residues. The hydrodynamic parameters of nucleosome were calculated for the whole particle as in 3LZ1.pdb [13], i.e. histone octamer and DNA. The hydrodynamic effect of flexible C-terminal tails of CSDs on the rotational correlation time of CSD dimer was included as for the dimeric unbound HP1 (see above).

References

[1] J. D. Walsh, K. Meier, R. Ishima, A. M. Gronenborn, *Biophys J* **2010**, *99*, 2636-2646.

[2] T. S. Ulmer, B. E. Ramirez, F. Delaglio, A. Bax, *J Am Chem Soc* **2003**, *125*, 9179-9191.

[3] H. Schwalbe, K. M. Fiebig, M. Buck, J. A. Jones, S. B. Grimshaw, A. Spencer, S. J. Glaser, L. J. Smith, C. M. Dobson, *Biochemistry* **1997**, *36*, 8977-8991.

[4] F. Munari, S. Soeroes, H. M. Zenn, A. Schomburg, N. Kost, S. Schroder, R. Klingberg, N. Rezaei-Ghaleh, A. Stutzer, K. A. Gelato, P. J. Walla, S. Becker, D. Schwarzer, B. Zimmermann, W. Fischle, M. Zweckstetter, *J Biol Chem* **2012**, *287*, 33756-33765.

[5] D. M. Korzhnev, N. R. Skrynnikov, O. Millet, D. A. Torchia, L. E. Kay, *J Am Chem Soc* **2002**, *124*, 10743-10753.

[6] L. E. Kay, D. A. Torchia, A. Bax, *Biochemistry* **1989**, *28*, 8972-8979.

[7] P. Bernado, E. Mylonas, M. V. Petoukhov, M. Blackledge, D. I. Svergun, *J Am Chem Soc* **2007**, *129*, 5656-5664.

[8] L. J. Ball, N. V. Murzina, R. W. Broadhurst, A. R. Raine, S. J. Archer, F. J. Stott, A. G. Murzin, P. B. Singh, P. J. Domaille, E. D. Laue, *EMBO J* **1997**, *16*, 2473-2481.

[9] Y. Huang, M. P. Myers, R. M. Xu, *Structure* **2006**, *14*, 703-712.

[10] L. Banci, I. Bertini, F. Cantini, A. C. Rosenzweig, L. A. Yatsunyk, *Biochemistry* **2008**, *47*, 7423-7429.

[11] D. Achila, L. Banci, I. Bertini, J. Bunce, S. Ciofi-Baffoni, D. L. Huffman, *Proc Natl Acad Sci U S A* **2006**, *103*, 5729-5734.

[12] P. R. Nielsen, D. Nietlispach, H. R. Mott, J. Callaghan, A. Bannister, T. Kouzarides, A. G. Murzin, N. V. Murzina, E. D. Laue, *Nature* **2002**, *416*, 103-107.

[13] D. Vasudevan, E. Y. Chua, C. A. Davey, *J Mol Biol* **2010**, *403*, 1-10.

[14] aF. Munari, N. Rezaei-Ghaleh, S. Xiang, W. Fischle, M. Zweckstetter, *PLoS One* **2013**, *8*, e60887; bN. Fatemi, D. M. Korzhnev, A. Velyvis, B. Sarkar, J. D. Forman-Kay, *Biochemistry* **2010**, *49*, 8468-8477.