

## ***HYCUD*: A computational tool for prediction of effective rotational correlation time in flexible proteins**

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

### **SUPPLEMENTARY INTRODUCTION**

Rotational motion of flexible multidomain proteins represents a complex problem as it is contributed by several motion types occurring on different timescales. Indeed, it has frequently been observed that rotational motion of protein domains is effectively slowed down in the context of flexible multidomain proteins, despite that the interdomain linker is completely disordered, its length is bigger than the persistence length of disordered polypeptide chains, and there is not any persisting domain-domain or domain-linker contact (Fatemi, et al., 2010; Munari, et al., 2013; Munari, et al., 2012; Rezaei-Ghaleh, et al., 2013; Walsh, et al., 2010). The rigorous treatment of this complex problem requires Brownian dynamic simulation with the full inclusion of hydrodynamic interaction as treated in (Amoros, et al., 2013). Nevertheless, it has been shown that the orientational time correlation function  $C(t)$  in such systems follows a sum of a limited number of exponentials (Halle, 2009; Wong, et al., 2009), and when the amplitude of interdomain motion is large, the  $C(t)$  effectively behaves as a single-exponential decaying function but with a slower rate than that of the isolated domain (Chen and Tjandra, 2008). In addition, it has been suggested that rotational deceleration in multidomain proteins is mediated by the hydrodynamic continuum and has its origin in the hydrodynamic drag forces (Bae, et al., 2009). Moreover, it has been observed that the difference between rotational diffusion of protein domains in isolation and within multidomain proteins is largely determined by a scaling factor, i.e. the differences are mainly observed in the rotational correlation times but not the anisotropy or orientation of rotational diffusion tensors (Walsh et al., 2010). The *HYCUD* method therefore intends to quantify this scaling factor in terms of a local “effective viscosity” contributed by the nearby domains.

## SUPPLEMENTARY METHODS

### *Ensemble generation by EOM*

Ensembles of 5000 random models of monomeric and dimeric L7/L12 and 500 random models of protein X from Sendai virus were generated using the EOM (Ensemble Optimization Method) program (Bernado, et al., 2007). High-resolution structure of the rigid domains (NTD: residues 3-30, and CTD: residues 53-120) of L7/L12 protein were taken from 1RQU.PDB (Bocharov, et al., 2004), and the boundaries between rigid and disordered domains were defined on the basis of NMR spin relaxation data (Bocharov, et al., 2004). In case of dimeric L7/L12, the protein sequence was defined as a single chain as if the two protein molecules were connected through their N-termini. For protein X, the structure of the rigid domain (residues 43-95) was obtained as the average structure of the NMR ensemble reported in 1R4G.pdb (Blanchard, et al., 2004).

### *Generation of PDT file*

The PDB files generated in the previous step were merged into a single PDT file (my\_ensemble.pdt). Thereafter, an in-house script (fixResidueNumbering.py) was used to ensure that residue numbering in each protein model within the PDT file started at 1 and incremented consecutively.

### *Hydrodynamic calculation for isolated domains*

The program HYDROPRO (v. 10) (Ortega, et al., 2011) was used to calculate the hydrodynamic properties of isolated protein domains at the specified temperatures, 30 °C for L7/L12 as in (Bocharov, et al., 2004) and 25 °C for protein X as in (Houben, et al., 2007), and corresponding solvent viscosities. A key parameter for the HYDROPRO calculations is the atomic effective radius (AER). A value of 2.9 Å recommended for folded proteins was used (Ortega, et al., 2011), although its choice could be in principle refined. Special care was taken with respect to the parameters SIGMAX and SIGMIN in the *hydropro.dat* parameter file to ensure that the number of mini-beads used to build a shell model of the interested domain ranges between ~ 300 and 2000. The calculated rotational correlation time ( $\tau_{c,0}$ ) and intrinsic viscosity  $[\eta_0]$  of protein domains were used in a later step of *HYCUD* calculations.

### *HYCUD calculation*

Starting from the my\_ensemble.pdt file, the *HYCUD* calculations passed steps 1-5 as described in the main text. Further details are provided here: At step 1, a third-party program REMO (Li and Zhang, 2009) was called to reconstruct the backbone and side-chain coordinates of protein molecules from their CA traces. This was required because the PDB models generated by EOM contain only CA coordinates of the flexible tails/linkers. At step 2, the interdomain linker (residues 31-52) of L7/L12 protein was split into two 11-residue fragments. For protein x, the N-terminal unstructured 42-residue domain was split into three 14-residue fragments. At step 3, the  $\tau_{c,0}$  and  $[\eta_0]$  of the rigid domains were fixed to the values calculated above. In case of L7/L12 protein, the effect of residues 1-2 was switched off by fixing the  $[\eta_0]$  of the corresponding fragment at 0. Step 4 involves a key *HYCUD* calculation where the interdomain distance  $r_{ij}$  (i.e. the distance between the center of mass of fragments  $i$  and  $j$ ) is converted 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  $C_j$  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 numbers of domain  $j$  within a distance of  $\sqrt[3]{6} r_{ij}$ , hence 1 domain  $j$  at the distance  $r_{ij}$ . The effective “relative viscosity” experienced by fragment  $i$  was then obtained as a first-order approximation through:

$$\frac{\eta_i}{\eta_0} = 1 + \sum_{j \neq i} [\eta]_j c_j$$

where  $\eta_0$  is the reference solvent viscosity. The corrected rotational correlation time of fragment  $i$ ,  $\tau_{c,i}$ , was calculated according to:

$$\frac{\tau_{c,i}}{\tau_{c,0i}} = \frac{\eta_i}{\eta_0}$$

At step 5, the outliers in predicted  $\tau_c$  distributions were detected and eliminated as follows: the distribution of  $\tau_c$  was trimmed from the rightmost side (largest rotational correlation time) until the relative change in the standard deviation of distribution upon removal of the last point decreased to <1%. In other words, models were sorted on the basis of their fragment-specific  $\tau_c$ . The standard deviation of  $\tau_c$  was calculated with and without the inclusion of the model with the

longest  $\tau_c$ . If the relative change in standard deviation (i.e.  $(SD_{\text{with}} - SD_{\text{without}}) / SD_{\text{with}}$ ) exceeded the specified cut-off value, the related model was excluded. The same procedure was then iterated until the relative change in standard deviation fell below the cut-off value. To evaluate the precision of *HYCUD* predictions, the averaging of  $\tau_c$  was performed over non-overlapping sub-ensembles each containing 500 (for L7/L12) or 100 (for PX) models, and the standard deviation among predicted  $\tau_c$  was considered as the uncertainty of *HYCUD*-predicted  $\tau_c$  (reported in supplementary Table S1).

**Table S1.** Comparison of *HYCUD*-predicted and experimental rotational correlation times ( $\tau_c$ ) for the N- (NTD) and C-terminal domains (CTD) of L7/L12 protein obtained at 30 °C and the C-terminal domain of protein X (PX-CTD) obtained at 25 °C.

	Predicted $\langle\tau_c\rangle^*$	Predicted $\langle\tau_c\rangle^*$	Experimental $\langle\tau_c\rangle^{**}$
Isolated	AER 2.9	AER 2.8	
NTD(3-30) monomer	2.2 ns	2.1 ns	-
NTD(3-30) dimer	3.4 ns	3.4 ns	-
CTD(53-120)	3.5 ns	3.5 ns	-
PX-CTD(43-95)	4.5 ns	4.4 ns	-
L7/L12 monomer			
NTD	5.2±0.2 ns	5.1±0.2 ns	-
CTD	5.5±0.0 ns	5.5±0.0 ns	-
L7/L12 monomer			
NTD	10.6±0.2 ns	10.4±0.2 ns	9.5±0.3 ns
CTD	6.8±0.1 ns	6.7±0.1 ns	7.3±0.8 ns
Protein X			
PX-CTD	7.4±0.1 ns	7.2±0.1 ns	7.3 ns

\*. The uncertainty of *HYCUD* predictions was estimated as described in Supp. Methods.

\*\* Ref.: Bocharov, et al., 2004. Houben, et al., 2007.

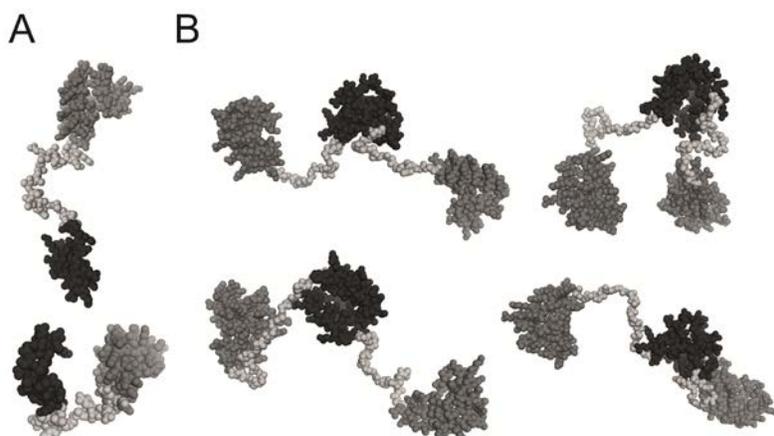
**Table S2.** Comparison of *HYCUD*-predicted and experimental rotational correlation times ( $\tau_c$ ) in a number of flexible multidomain proteins\* used for validation of *HYCUD* in (Rezaei-Ghaleh, et al., 2013).

	Predicted < $\tau_c$ >**	Experimental < $\tau_c$ >	Temperature (°C)
<b>HP1 monomer</b>			
Chromo domain (19-72)	9.4±0.7 ns	9.8±1.3 ns	25
Chromoshadow domain (110-171)	10.7±0.7 ns	10.5±1.1 ns	25
<b>HP1 dimer</b>			
Chromo domain (19-72)	10.9±0.7 ns	10.2±1.2 ns	25
Dimerized chromoshadow domain (110-171)	28.6±2.0 ns	24.6±5.0 ns	25
<b>GB1</b>			
sGB1	3.8±0.1 ns	3.8±0.1 ns	25
dGB1-12: N-terminal domain	5.4±0.1 ns	5.3±0.1 ns	25
dGB1-12: C-terminal domain	5.2±0.1 ns	5.2±0.1 ns	25
dGB1-24: N-terminal domain	5.2±0.1 ns	5.2±0.1 ns	25
dGB1-24: C-terminal domain	5.2±0.1 ns	5.0±0.1 ns	25
<b>ATP7B</b>			
Domain 4 (72-residue)	6.6±0.4 ns	6.3±0.1 ns	35

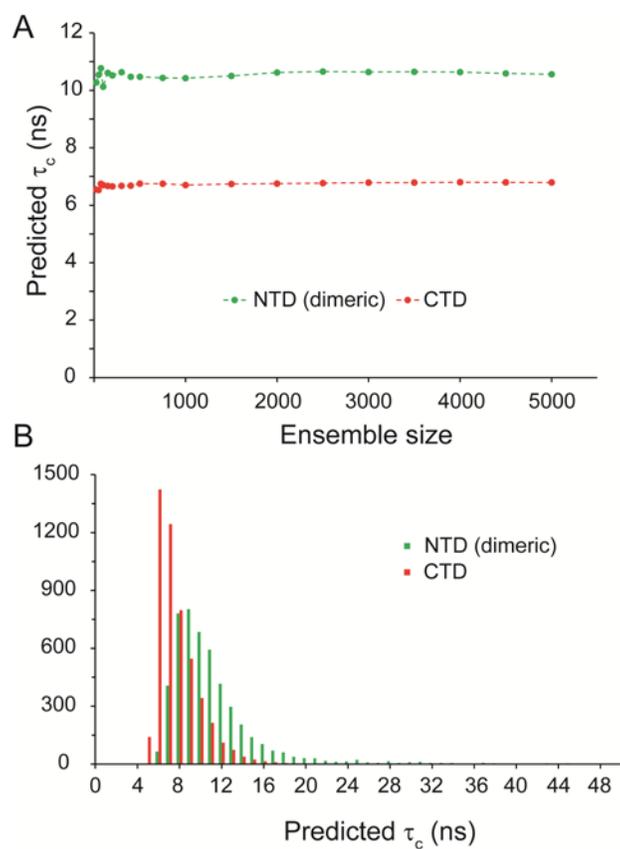
\*. Heterochromatin protein 1 (HP1), single (sGB1) and double-GB1 (dGB1) protein constructs with interdomain linkers of 12 or 24 residues and the three-domain (domains 4-6) construct of Wilson's disease protein (ATP7B).

\*\* An AER of 2.9 Å was used.

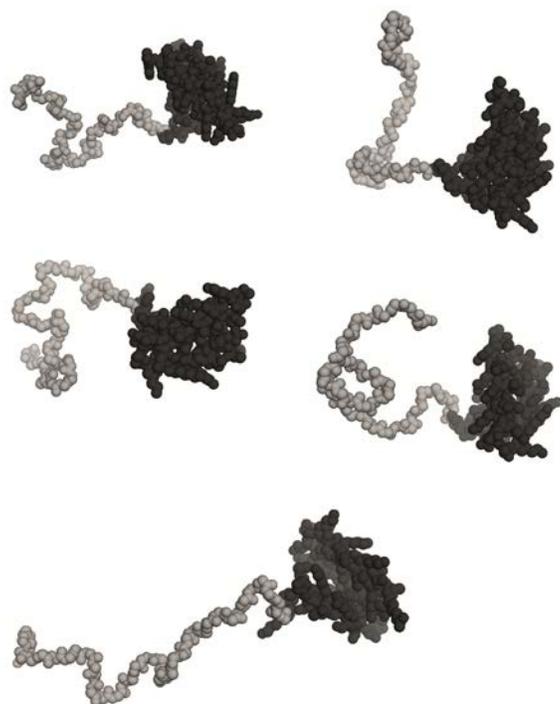
## SUPPLEMENTARY FIGURES



**Fig. S1.** Representative models of the ribosomal L7/L12 protein ensemble in its monomeric (A) and dimeric (B) states. The globular N- (black) and C-terminal (dark gray) domains are shown in black and dark grey, respectively, while, the unfolded interdomain linkers are in light grey. In (B), the dimeric N-terminal domain (residues 3-30) and the two C-terminal domains (residues 53-120) (PDB code: 1RQU) were treated as rigid bodies, while the two interdomain linkers (residues 31-52) as well as residues 1-2 were allowed to sample the conformational space according to their amino acid conformational propensities.



**Fig. S2.** (A) *HYCUD*-predicted rotational correlation time ( $\tau_c$ ) of the N- (NTD) and C-terminal domains (CTD) of L7/L12 protein. (B) Histogram of predicted  $\tau_c$  for the NTD and CTD in an ensemble of 5000 structures of the L7/L12 protein. The  $\tau_c$  distribution within the ensemble of dimeric L7/L12 is right-skewed with a minimum value of about 5 ns and reaching up to  $\sim 37$  ns.



**Fig. S3.** Representative models of the protein X of Sendai virus. This protein is the RNA-binding domain of phosphoprotein P, which is a component of RNA polymerase. The unstructured N-terminal tail (residues 1-42) and the globular C-terminal domain (residues 43-95) are shown respectively in light gray and black.

## References

- Amoros, D., Ortega, A. and de la Torre, J.G. (2013) Prediction of Hydrodynamic and Other Solution Properties of Partially Disordered Proteins with a Simple, Coarse-Grained Model, *J. Chem. Theory Comput.*, **9**, 1678-1685.
- Bae, S.H., Dyson, H.J. and Wright, P.E. (2009) Prediction of the rotational tumbling time for proteins with disordered segments, *J. Am. Chem. Soc.*, **131**, 6814-6821.
- Bernado, P., *et al.* (2007) Structural characterization of flexible proteins using small-angle X-ray scattering, *J. Am. Chem. Soc.*, **129**, 5656-5664.
- Blanchard, L., *et al.* (2004) Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution, *Virology*, **319**, 201-211.
- Bocharov, E.V., *et al.* (2004) From structure and dynamics of protein L7/L12 to molecular switching in ribosome, *J. Biol. Chem.*, **279**, 17697-17706.
- Chen, K. and Tjandra, N. (2008) Extended model free approach to analyze correlation functions of multidomain proteins in the presence of motional coupling, *J. Am. Chem. Soc.*, **130**, 12745-12751.
- Fatemi, N., *et al.* (2010) NMR characterization of copper-binding domains 4-6 of ATP7B, *Biochemistry*, **49**, 8468-8477.
- Halle, B. (2009) The physical basis of model-free analysis of NMR relaxation data from proteins and complex fluids, *J. Chem. Phys.*, **131**, 224507.
- Houben, K., *et al.* (2007) Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase cofactor P, *Biophys. J.*, **93**, 2830-2844.
- Li, Y. and Zhang, Y. (2009) REMO: A new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks, *Proteins*, **76**, 665-676.
- Munari, F., *et al.* (2013) Structural Plasticity in Human Heterochromatin Protein 1 $\beta$ , *PLoS One*, **8**, e60887.
- Munari, F., *et al.* (2012) Methylation of lysine 9 in histone H3 directs alternative modes of highly dynamic interaction of heterochromatin protein hHP1beta with the nucleosome, *J. Biol. Chem.*, **287**, 33756-33765.
- Ortega, A., Amoros, D. and Garcia de la Torre, J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models, *Biophys. J.*, **101**, 892-898.
- Rezaei-Ghaleh, N., *et al.* (2013) Predicting the rotational tumbling of dynamic multidomain proteins and supramolecular complexes, *Angew. Chem. Int. Ed. Engl.*, **52**, 11410-11414.
- Walsh, J.D., *et al.* (2010) NMR studies on domain diffusion and alignment in modular GB1 repeats, *Biophys. J.*, **99**, 2636-2646.
- Wong, V., Case, D.A. and Szabo, A. (2009) Influence of the coupling of interdomain and overall motions on NMR relaxation, *Proc. Natl. Acad. Sci. USA*, **106**, 11016-11021.