

Supplementary material for:

# **Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics.**

Charlotte S. Sørensen,<sup>1,2</sup> and Magnus Kjaergaard<sup>1-4\*</sup>

<sup>1</sup> Department of Molecular Biology and Genetics, Aarhus University

<sup>2</sup> The Danish Research Institute for Translational Neuroscience (DANDRITE)

<sup>3</sup> Aarhus Institute of Advanced Studies (AIAS)

<sup>4</sup> Center for Proteins in Memory - PROMEMO, Danish National Research Foundation

\* Corresponding author: [magnus@mbg.au.dk](mailto:magnus@mbg.au.dk)

## Protein sequences of constructs used:

### Color code for features:

SA-Strep tag

mClover3

mRuby3

Unique restriction sites

### Interaction partners:

MBD2

p66 $\alpha$

### Fusion proteins:

```
MGSSHHHHHH SSGLVPRGSH MVSKGEELFT GVPILVELD GDVNGHKFSV RGEGEDATN
GKLTLLKFICT TGKLPVPWPT LVTTFGYGVA CFSRYPDHMK QHDFFKSAMP EGYVQERTIS
FKDDGTYKTR AEVKFEEDTL VNRIELKGID FKEDGNILGH KLEYNFNSHY VYITADKQKN
CIKANFKIRH NVEDGSVQLA DHYQQNTPIG DGPVLLPDNH YLSHQSKLSK DPNEKRDHNV
LLEFVTAAL E SGGEDPMVST GQSQSQSQSQ SVTDEDIRKQ EERAQQVRKK LEEALMADAS
(Variable linker)
GTPEERERMI KQLKEELRLE EAKLVLLKLL RQSTQSQSQ SQSQSMVSKG EELIKENMRM
KVVMEGSVNG HQFKCTGEGE GRPYEGVQTM RIKVIEGGPL PFAFDILATS FMYGSRTFIK
YPADIPDFEK QSFPEGFTWE RVTRYEDGGV VTVTQDTSLE DGELVYNVKV RGVNFPSNGP
VMQKKTGWE PNTEMMYPAD GGLRGYTDIA LKVDGGGHLH CNFVTYRSK KTVGNIKMPG
VHAVDHRLER IEESDNETYV VQREVAVAKY SNLGGGMDL YKQSQSQSWS HPQFEK
```

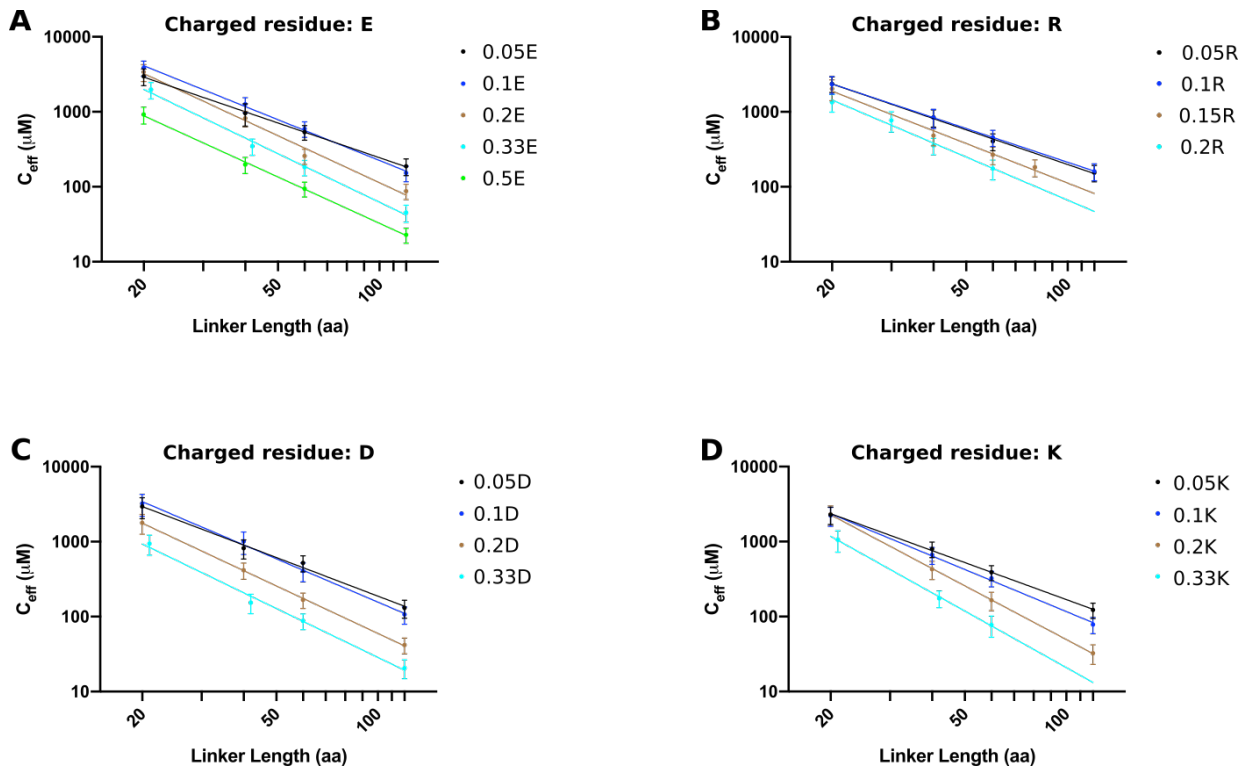
### MBD2 WT peptide:

```
MGSSHHHHHH SSGLVPRGSH MQSQSQSQSQ S VTDIEDIRKQ EERVQQVRKK LEEALMADAS
GSGSGSGSGS Y
```

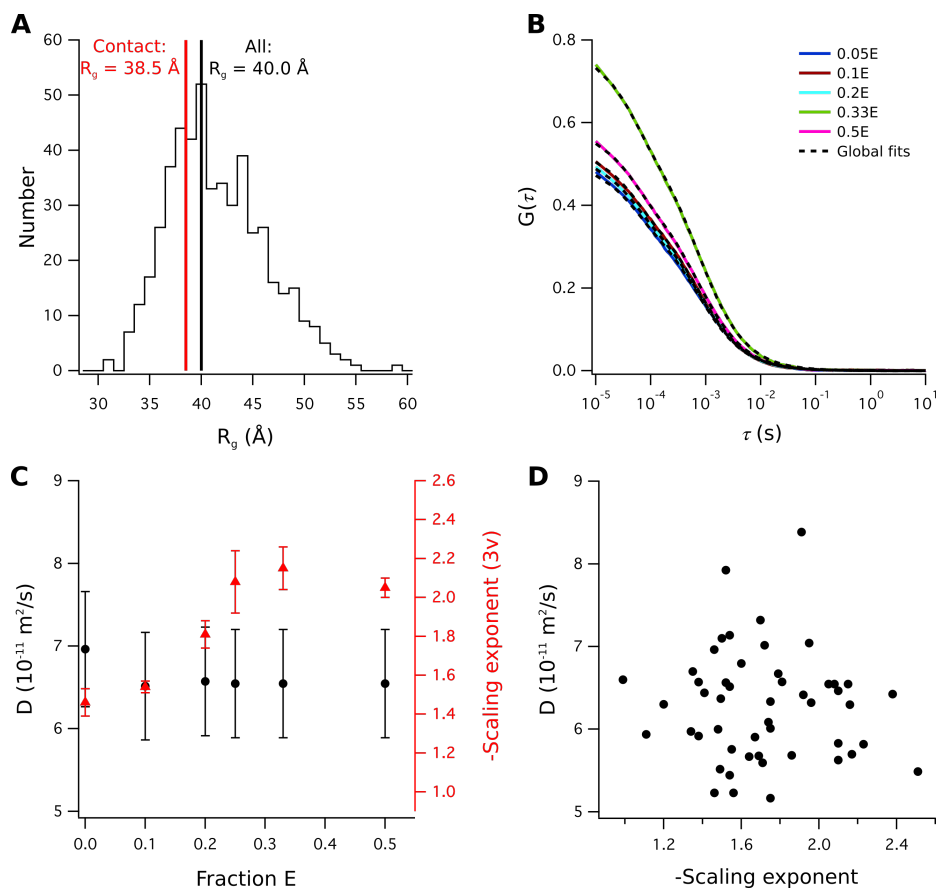
**Supplementary methods:**

**Ensemble simulations.** We generated an 500 conformer ensemble of a the biosensor containing a 40-residue linker using the Ensemble Optimization Method 2.<sup>1,2</sup> The fluorescent protein domains and the MBD2:p66 $\alpha$  were treated as rigid bodies, and the linkers were modelled as a string of beads representing C $\alpha$ -atoms with a compactness typical for intrinsically disordered proteins. Input structures for the fluorescent proteins were generated using homology modelling by SWISS-MODEL<sup>3</sup> and the MBD2-p66 $\alpha$  was based on the NMR structure of the complex.<sup>4</sup> To estimate the effects of an interaction between the central linker and the fluorescent domains, we selected the subset of conformers where any atoms of the linker approach the fluorescent proteins. Contact conformers were selected as any conformer where a pseudo-atom from the linker approached closer than an arbitrarily defined threshold of 8 Å. This likely shows the effect of such an interaction, although it may underestimate the effect of a full absorption.

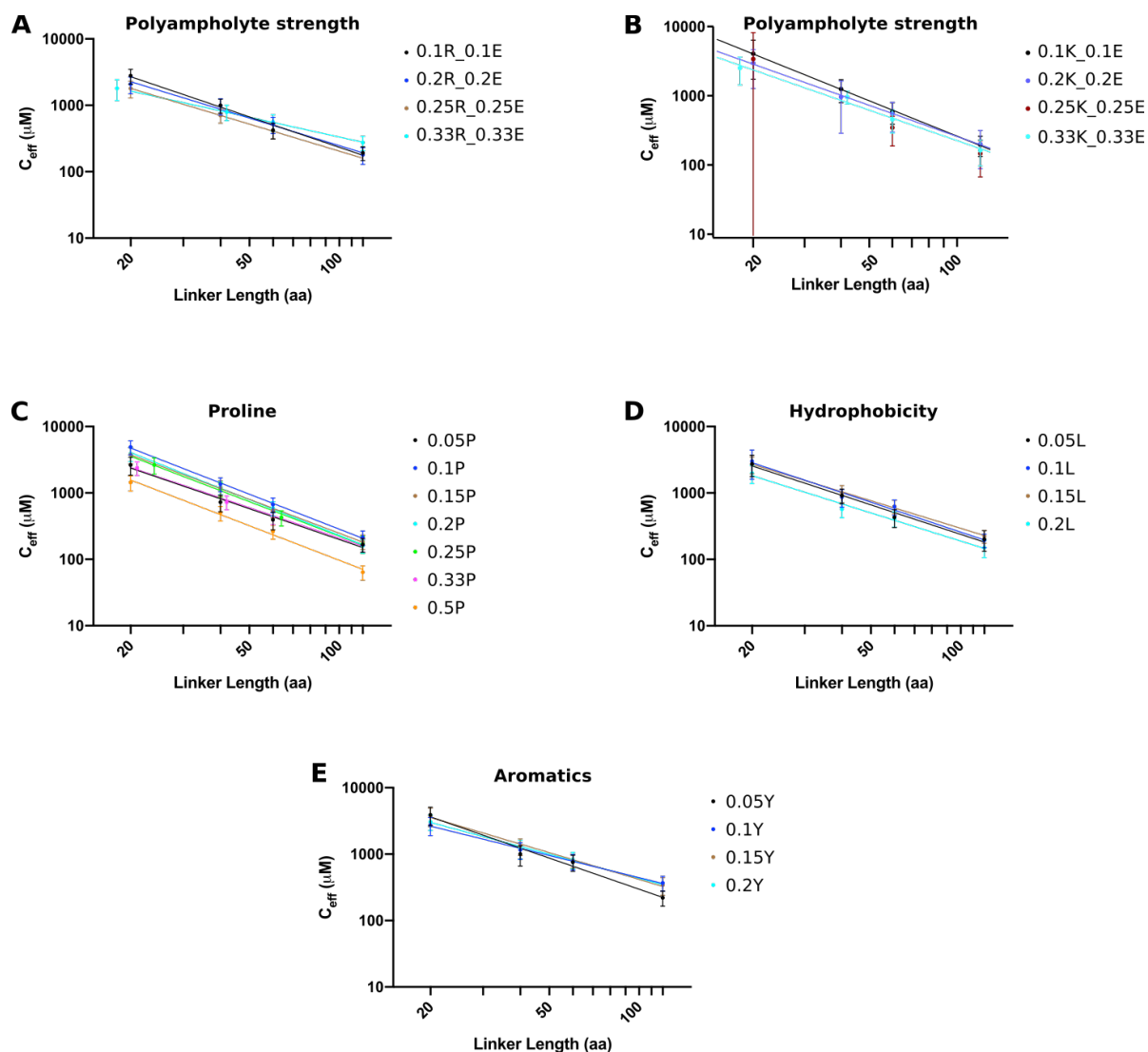
## Supplementary figures:



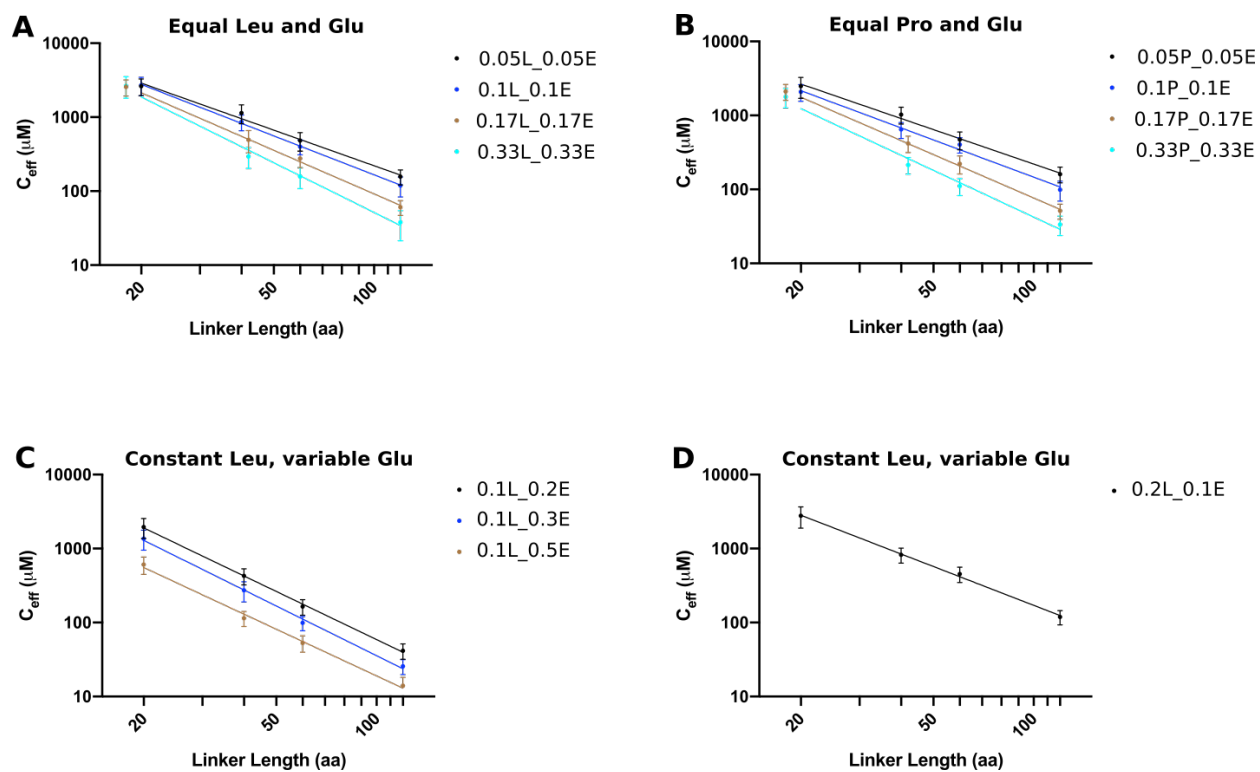
**Fig. S1:** Power law fits to experimentally determined effective concentrations for linkers containing different types of charged residues



**Fig. S2:** Ensemble dimension estimated by simulation and a diffusion measurements. (A) Distribution of the radius of gyration in an random ensemble of biosensors with realistic geometries. Conformers with a distance  $< 8\text{\AA}$  between the central linker and either of the fluorescent proteins were selected to estimate the effect of a hypothetical interaction between the linker and the fluorescent proteins. Linker interactions with the FP results in a slight contraction of the ensemble. (B) Auto-correlation curves for biosensors containing glutamate residues (C) Comparison of the measured diffusion coefficients and scaling exponents for effective concentrations. The rise in scaling exponent is not accompanied by a change in the diffusion coefficient. (D) Correlation between scaling exponents and diffusion times shows no correlation between the two parameters (Pearson coefficient = -0.06).



**Fig. S3:** Power law fits to experimentally determined effective concentrations for linkers with variations in polyampholyte strength, chain flexibility, hydrophobicity and aromatic residue content. The quality of the 0.25K\_0.25E data series did not allow determination of a scaling coefficient.



**Fig. S4:** Power law fits to experimentally determined effective concentrations for linkers with combinations of charged residues and leucine or proline.

<b>Table S1: Full sequence of all linkers used:</b>		
<b>Linker name</b>	<b>Linker sequence</b>	<b>Estimated helicity (%)</b>
<b>GS</b>		
GS <sub>20</sub>	GS <sub>20</sub>	-
GS <sub>30</sub>	GS <sub>30</sub>	-
GS <sub>40</sub>	GS <sub>40</sub>	-
GS <sub>60</sub>	GS <sub>60</sub>	0.01
GS <sub>120</sub>	GS <sub>120</sub>	-
<b>E</b>		
0.05E <sub>20</sub>	0.05E <sub>20</sub>	-
0.05E <sub>40</sub>	0.05E <sub>40</sub>	-
0.05E <sub>60</sub>	0.05E <sub>60</sub>	0.02
0.05E <sub>120</sub>	0.05E <sub>120</sub>	-
0.1E <sub>20</sub>	0.1E <sub>20</sub>	-
0.1E <sub>40</sub>	0.1E <sub>40</sub>	-
0.1E <sub>60</sub>	0.1E <sub>60</sub>	0.04
0.1E <sub>120</sub>	0.1E <sub>120</sub>	-
0.2E <sub>20</sub>	0.2E <sub>20</sub>	-
0.2E <sub>40</sub>	0.2E <sub>40</sub>	-
0.2E <sub>60</sub>	0.2E <sub>60</sub>	0,06
0.2E <sub>120</sub>	0.2E <sub>120</sub>	-
0.33E <sub>21</sub>	0.33E <sub>21</sub>	-
0.33E <sub>42</sub>	0.33E <sub>42</sub>	-
0.33E <sub>60</sub>	0.33E <sub>60</sub>	0,06
0.33E <sub>120</sub>	0.33E <sub>120</sub>	-
0.5E <sub>20</sub>	0.5E <sub>20</sub>	-
0.5E <sub>40</sub>	0.5E <sub>40</sub>	-
0.5E <sub>60</sub>	0.5E <sub>60</sub>	0.12
0.5E <sub>120</sub>	0.5E <sub>120</sub>	-
<b>R</b>		
0.05R <sub>20</sub>	0.05R <sub>20</sub>	-
0.05R <sub>40</sub>	0.05R <sub>40</sub>	-
0.05R <sub>60</sub>	0.05R <sub>60</sub>	0.03
0.05R <sub>120</sub>	0.05R <sub>120</sub>	-
0.1R <sub>20</sub>	0.1R <sub>20</sub>	-
0.1R <sub>40</sub>	0.1R <sub>40</sub>	-
0.1R <sub>60</sub>	0.1R <sub>60</sub>	0.04













**Table S2: Power law fitting parameters and diffusion coefficients.**

Linker series	Scaling coefficient (S.E.)	Pre-factor [mM] (relative S.E.)	D [ $10^{-11}$ m <sup>2</sup> /s]* (40 residue linker)
<b>GS</b>	-1.46 (0.07)	330 (31%)	6.96
<b>E</b>			
0.05E	-1.54 (0.03)	291 (12%)	6.51
0.1E	-1.81 (0.07)	922 (30%)	6.57
0.2E	-2.08 (0.16)	1652 (84%)	6.55
0.33E	-2.15 (0.11)	1253 (54%)	6.55
0.5E	-2.05 (0.05)	414 (21%)	6.55
<b>R</b>			
0.05R	-1.54 (0.05)	237 (20%)	7.14
0.1R	-1.49 (0.004)	207 (2%)	5.52
0.15R	-1.75 (0.13)	360 (64%)	5.17
0.2R	-1.91 (0.16)	443 (77%)	8.39
<b>L</b>			
0.05L	-1.48 (0.11)	253 (55%)	6.00
0.1L	-1.50 (0.11)	253 (55%)	7.10
0.15L	-1.38 (0.04)	169 (16%)	6.57
0.2L	-1.41 (0.12)	124 (59%)	6.44
<b>P</b>			
0.05P	-1.54 (0.12)	239 (58%)	5.44
0.1P	-1.75 (0.04)	893 (17%)	6.33
0.15P	-1.69 (0.04)	598 (18%)	5.68
0.2P	-1.79 (0.09)	853 (18%)	6.67
0.25P	-1.72 (0.09)	615 (42%)	7.02
0.33P	-1.52 (0.06)	230 (29%)	7.92
0.5P	-1.73 (0.10)	278 (48%)	n.d.
<b>PE</b>			
0.05PE	-1.55 (0.08)	274 (37%)	5.76
0.1PE	-1.67 (0.11)	324 (51%)	5.90
0.167PE	-1.95 (0.05)	605 (19%)	7.04
0.33PE	-2.10 (0.16)	664 (87%)	5.83
<b>LE</b>			
0.05LE	-1.60 (0.12)	345 (58%)	6.80
0.1LE	-1.75 (0.03)	521 (12%)	6.01
0.167LE	-1.96 (0.06)	750 (28%)	6.32
0.33LE	-2.23 (0.13)	1507 (65%)	n.d.
<b>L+E</b>			
0.1L_0.2E	-2.16 (0.05)	1230 (23%)	6.30
0.1L_0.3E	-2.23 (0.08)	1033 (38%)	5.82

0.1L_0.5E	-2.10 (0.10)	296 (47%)	5.63
<b>E+L</b>			
0.1E_0.2L	-1.74 (0.06)	524 (24%)	6.08
<b>RE</b>			
0.1RE	-1.52 (0.12)	257 (57%)	6.56
0.2RE	-1.38 (0.09)	141 (43%)	5.92
0.25RE	-1.35 (0.002)	103 (0.5%)	6.70
0.33RE	-0.99 (0.012)	32 (5%)	6.60
<b>KE</b>			
0.1KE	1.69 (0.069)	1109 (36%)	7.32
0.2KE	1.49 (0.083)	478 (46%)	6.37
0.25KE	Did not allow reliable fitting		6.70
0.33KE	-1.54 (0.11)	265 (71%)	5.23
<b>D</b>			
0.05D	-1.71 (0.10)	494 (47%)	5.59
0.1D	-1.92 (0.07)	1072 (32%)	6.42
0.2D	-2.10 (0.03)	964 (13%)	6.46
0.33D	-2.17 (0.13)	614 (69%)	5.70
<b>K</b>			
0.05K	-1.64 (0.03)	321 (14%)	5.67
0.1K	-1.86 (0.06)	624 (27%)	5.68
0.2K	-2.38 (0.02)	2838 (7%)	6.42
0.33K	-2.51 (0.07)	2138 (31%)	5.49
<b>Y</b>			
0.05Y	-1.56 (0.14)	385 (75%)	5.23
0.1Y	-1.11 (0.04)	73 (15%)	5.94
0.15Y	-1.34 (0.06)	195 (24%)	5.97
0.2Y	-1.2 (0.09)	111 (39%)	6.30

\* Repeats of a single construct indicates a standard error of ~10%, which is likely to be representative for the whole series.  
n.d. = not determined.

### Supplementary references:

- (1) Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* *129*, 5656–5664.
- (2) Tria, G., Mertens, H. D. T., Kachala, M., and Svergun, D. I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* *2*, 207–217.
- (3) Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., Beer, T. A. P. De, Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. (2018) SWISS-MODEL : homology modelling of protein structures and complexes *46*, 296–303.
- (4) Gnanapragasam, M. N., Scarsdale, J. N., Amaya, M. L., Webb, H. D., Desai, M. a, Walavalkar, N. M., Wang, S. Z., Zu Zhu, S., Ginder, G. D., and Williams, D. C. (2011) p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 7487–92.
- (5) Munoz, V., and Serrano, L. (1994) Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol.* *1*, 399–409.
- (6) Muñoz, V., and Serrano, L. (1995) Elucidating the Folding Problem of Helical Peptides using Empirical Parameters. II†. Helix Macrodipole Effects and Rational Modification of the Helical Content of Natural Peptides. *J. Mol. Biol.* *245*, 275–296.