PNAS
www.pnas.org

Supplementary Information for

Extreme stability in de novo designed repeat arrays is determined by unusually stable short-range interactions

Kathryn Geiger-Schuller[1,2]*, Kevin Sforza[1]*, Max Yuhas[1,3], Fabio Parmeggiani[4,5], David Baker[4], & Doug Barrick[1]§

[1]Department of Biophysics and Program in Molecular Biophysics, Johns Hopkins University, 3400 N. Charles St, Baltimore MD 21218.
[2]Current address: Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142
[3]Current address: Applied Mathematics, Yale University, New Haven CN 06520
[4]Institute for Protein Design, Department of Biochemistry and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195
[5]Current address: School of Chemistry and School of Biochemistry, University of Bristol. Life Sciences Building, 24 Tyndall Avenue, Bristol BS8 TQ1

*These authors contributed equally to this work.
§Corresponding Author: barrick@jhu.edu, (410) 516-0409

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1800283115

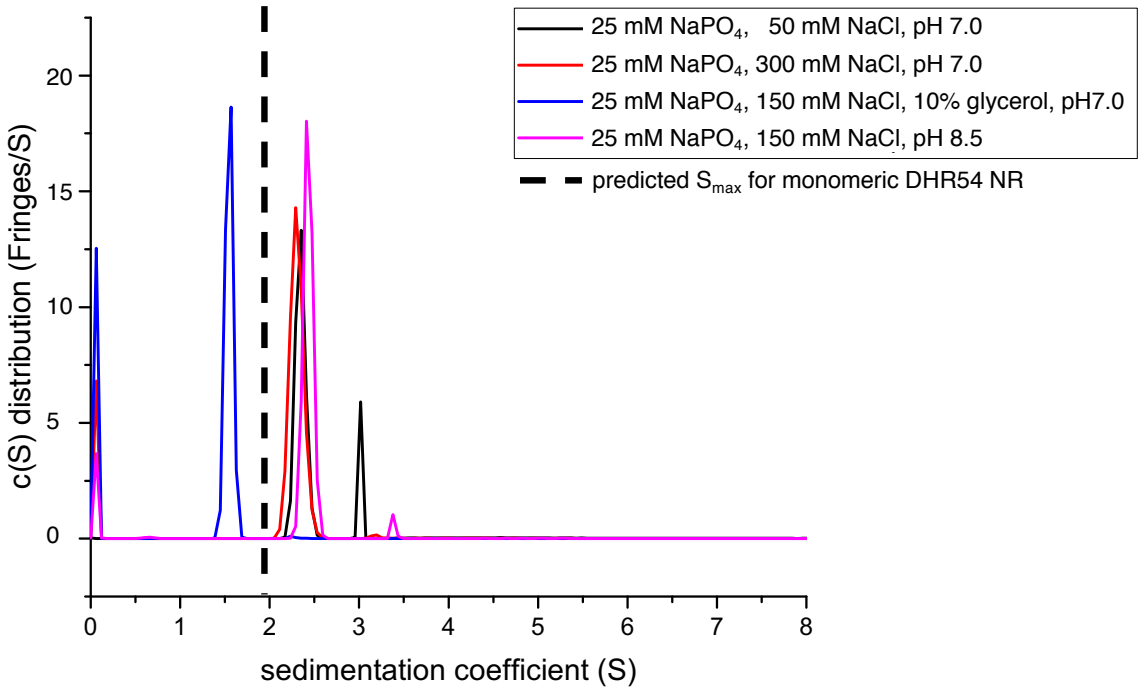

Fig. S1. Sedimentation Velocity c(S) plot for DHR54 NR in the absence and presence of glycerol. Data were processed and fitted in Sedfit[1] as previously described[2]. The predicted S_{max} (dotted vertical line) was calculated for DHR54 NR using Sednterp[3]. In the presence of 10 % glycerol, the c(S) distributions are consistent with monomers.

Table S1. Sequence features of designed helical repeat proteins used in this study.

Construct	Sequence (NR_2C)	n_{rep} (n_{tot})	NCPR	Z	m_{Ising} (m_{2s})	m_{pred}
DHR10.2	SSE**KE**EL**RE**RL**K**K**E**V**R**ENA**K**R**K**G**D**D**TE**EA**RE**AA**R**EAF**ER**V**R**EAA**E**RAGI**D** SSEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGIT SSEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGIT **SSETLKRAIEEIRKRVEEAQREGN**DISEAAR**Q**AAE**EFRKK**AE**ELK**RR**G**DG**	50 (200)	0.47	-0.02	4.92	4.75
DHR54	TTEDERRELEKVARKAIEAAREGNTDEVREQLQRALEIARESG TTEAVKLALEVVARVAIEAARRGNTDAVREALEVALEIARESG TTEAVKLALEVVARVAIEAARRGNTDAVREALEVALEIARESG TEEAVRLALEVVKRVSDEAKKQGNEDAVKEAEEVRKKIEEESG	43 (172)	0.41	-0.06	4.19	4.19
DHR71	DPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEG DPELVLEAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEG DPELVLEAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEG DPELVEEAAKVAEEVRKLAKKQGDEEVYEKARETAREVKEELKRVREEKG	50 (200)	0.475	0.045	4.75	4.75
DHR79	SSDEEEARELIERAKEAAERAQEAAERTGDPRVRELARELKRLAQEAAEEVKRDPS SSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRNPS SSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRNPS SEEVNEALKKIVKAIQEAVESLREAEESGDPEKREKARERVREAVERAEEVQRDPS	56 (224)	0.44	-0.10	5.24	5.24

n_{rep}: number of residues per repeat. n_{tot}: number of residues in NR_2C construct. NCPR: the net fractional charge per residue, $(n_K+n_R+n_D+n_E)/n_{tot}$. Z: the total fractional charge, $(n_K+n_R-n_D-n_E)/n_{tot}$. m_{Ising}: experimental m-values from the fitted Ising parameters for NR_2C (Table 2; $3xm_{Gdn,\,i} + m_{Gdn,\,C}$ for DHR71, $4xm_{Gdn,\,i}$ for the other three constructs. m_{2s}: experimental m-values from fitting a two-state model to the guanidine HCl-induced unfolding of NR_2C constructs (Figure 2; units of kcal mol^{-1} M^{-1}). m_{calc}: m-values estimated from the empirical correlation between m-values for guanidine HCl-induced folding and chain-length (Myers Pace Scholtz[4]; units of kcal mol^{-1} M^{-1}).

Table S2. Structural features of designed helical repeat proteins used in this study.

Construct	Contacts within repeats[a]	Contacts between repeats[a]	SASA per naked repeat[b]	SASA between repeats[b]	Twist[c] (radians)	Rise[c] (Å)	Radius[c] (Å)	ΔG_i	$\Delta G_{i,i+1}$
DHR10.2	56.75	143.7	4,340	2,341	0.03	9.22	138.2	-2.51	-4.80
DHR54	46.0	92.3	3,654	1,874	-0.26	8.76	15.1	-2.04	-6.76
DHR71	66.5	108.2	4.090	1,904	0.23	5.12	49.02	-1.41	-9.93
DHR79	50.5	120.0	4,641	2,311	0.42	6.86	22.51	-3.48	-4.83

[a]Contacts were counted between non-hydrogen pairs closer than 4.2 Å, at a sequence separation of five or more residues from pdb files 5cwg, 5cwl, 5cwn, and 5cwp. For regions where multiple conformations were modeled, the major conformer was used in the contact calculation. Contacts within repeats were averaged over the four (NR_2C) repeats. Contacts between adjacent repeats were averaged over the three (NR, RR, RC) interfaces.

[b]Solvent accessible surface areas (SASA) were calculated with the get_area command of pymol with the dot_solvent option set to one, using a solvent radius of 1.4 Å. To resolve SASA into naked-repeat (i.e., SASA for a single repeat fragment with no neighboring repeats) from SASA buried between repeats, SASA values were determined for one, two, and three repeat fragments, and for the full four-repeat constructs. For these ten values, SASA was plotted versus repeat number and fitted with a line. The intercept of the best-fit line represents the average SASA buried per interface (with contributions from both repeats forming the interface). The slope of the best-fit line represents the SASA per naked repeat minus the SASA per interface, and is rearranged to determine the former. This procedure gives very similar values to those obtained by summing the SASA of adjacent naked repeats and subtracting the SASA of the corresponding two-repeat fragment. SASA per internal repeat was averaged over the two internal (R) repeats, where each repeat was represented as a pymol "selection" within the four-repeat parent construct (rather than a separate "object").

[c]The geometric parameters twist, rise, and radius describe the superhelical geometry of each DHR, and are from Brunette *et al.*[15].

References

1. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78,** 1606–1619 (2000).
2. Marold, J. D., Kavran, J. M., Bowman, G. D. & Barrick, D. A Naturally Occurring Repeat Protein with High Internal Sequence Identity Defines a New Class of TPR-like Proteins. *Struct. Lond. Engl. 1993* (2015). doi:10.1016/j.str.2015.07.022
3. John Philo's Software Home Page. Available at: http://www.jphilo.mailway.com/. (Accessed: 6th October 2017)
4. Myers, J. K., Pace, C. N. & Scholtz, J. M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci. Publ. Protein Soc.* **4,** 2138–2148 (1995).