Supplementary Materials

Protein folding *in vitro* and in the cell: from a solitary journey to a team effort

Miranda F. Mecha,[a,†] Rachel B. Hutchinson,[a,†] Jung Ho Lee,[a,1] Silvia Cavagnero[a,*]

[a] Department of Chemistry, University of Wisconsin-Madison, Madison, WI 53706

[1] Present address: Department of Chemistry, Seoul National University, Seoul 08826, Korea

[†]These authors contributed equally to this work

*Corresponding author: Silvia Cavagnero
Email Address: cavagnero@chem.wisc.edu

SUPPLEMENTARY TABLES:

Table S1. Protein length (number of amino acids) and folding rate constant (k_f) values of two-state folders plotted in Figure 3A.

Name	PDB	Length	ln k_f	Reference
C-terminal β-hairpin of protein GB1	1PGB	16	12.0	[1, 2]
Trp-cage protein	1L2Y	20	12.5	[1, 3]
Alanine-based peptide	NA	21	15.5	[1, 4]
BBA5 mini-protein	1T8J	23	11.8	[1, 5]
Pin WW domain	1PIN	34	9.5	[1, 6]
Villin headpeace subdomain	1VII	36	9.4	[1, 7]
Formin-binding protein	1E0L	37	10.6	[1, 8]
Prototype WW domain	1E0M	38	8.9	[1, 8]
Yes kinase-associated protein	1JMQ	40	8.4	[1, 8]
Peripheral subunit-binding domain	2PDD	41	9.8	[1, 9]
E3-binding domain of BBL	2WXC	47	11.2	[1]
GA module of albumin binding domain	1PRB	47	13.8	[1, 10]
POB	1W4J	51	12.3	[1]
TRF1 Myb domain	1BA5	53	5.9	[1, 11]
c-Myb-transforming protein	1GV2	55	8.7	[1, 12]
N-terminal domain from ribosomal protein L9	1DIV	56	6.6	[1, 13]
B-domain of staphylococcal proteinA	1BDD	58	11.7	[1, 14]
RAP1 Myb domain	1FEX	59	8.2	[1, 11]

B1 domain of streptococcal protein G	3GB1	62	6.3	[1]
Src SH3 domain	1RLQ	62	4.4	[1, 13]
α-spectrin SH3 domain	1SHG	62	1.1	[1, 13]
B1 domain of streptococcal protein G	3GB1	62	6.3	[13]
Sso7d protein, Y34W	1BNZ	64	7.0	[1]
Sso7d protein, Y34W	1C8C	64	6.95	[15]
Chymotrypsin inhibitor 2	2CI2	65	5.8	[1, 13]
Cold shock protein B	1C9O	66	7.2	[1, 16]
Cold shock protein B	1G6P	66	6.3	[1, 16]
Cold shock protein B	1CSP	67	6.5	[1, 16]
LysM domain	1E0G	66	7.0	[1]
Actin binding protein ABP1 SH3 domain	1JO8	68	2.5	[1, 13]
Photosystem I accessory protein E	1PSF	69	3.2	[1]
Cold shock protein A	1MJC	70	5.3	[1, 17]
Immunoglobulin light chain-binding domain of protein L	1HZ5	72	4.1	[1]
α3D	2A3D	73	12.2	[1, 18]
Sho1 SH3 domain	2VKN	76	2.1	[1, 13]
Ubiquitin	1UBQ	76	7.3	[1, 13]
Fyn SH3 domain	1AVZ	78	4.9	[1, 13]
Immunoglobulin light chain-binding domain of protein L	2PTL	79	4.1	[13]

Ras-binding domain of C-raf-1	1RFA	80	8.4	[1, 13]
l-repressor	1LMB	80	10.4	[13]
Activation domain of procarboxypeptidase A2	1O6X	81	6.8	[1, 13]
l-repressor	1LMB	81	10.4	[1]
Histidine-containing phosphocarrier protein	1POH	85	2.7	[1, 19]
SH3-like domain of virulence protein internalin B	1M9S	85	4.0	[1, 13]
Acyl-coenzyme A binding protein	1NTI	86	7.0	[1, 13]
CAfn2 (B. circulans)	1K85	88	1.4	[1]
N-domain of spore coat protein S	1PRS	88	3.0	[1]
Hypothetical protein encoded by the Yjbj gene from E.coli	1JYG	89	9.1	[1, 13]
Ninth fibronectin type III module of fibronectin	1FNF	90	-0.9	[1, 20]
SH3 domain of the p85α subunit of phosphatidylinositol 3'-kinase	1PNJ	90	-1.0	[1, 21]
Third fibronectin type III repeat of tenascin	1TEN	90	1.1	[22]
C-domain of spore coat protein S	1PRS	91	-2.0	[1]
C-terminal domain from ribosomal protein L9	1DIV	92	3.3	[1, 13]

Third fibronectin type III repeat of tenascin	1TEN	92	1.1	[1]
18th module of muscle protein twitchin	1WIT	93	0.4	[1]
Colicin E9 immunity protein	1IMQ	93	7.3	[1, 13]
hbLBD	1K8M	93	-0.7	[1]
18th module of muscle protein twitchin	1WIT	93	0.4	[23]
Colicin E7 immunity protein	1AYI	94	7.2	[1, 13]
Ribosomal protein L23	1N88	96	2.0	[1, 13]
Common-type acylphosphatase	2VH7	98	0.8	[1]
Common-type acylphosphatase	2ACY	98	0.8	[24]
Muscle acylphosphatase	1APS	99	-1.6	[1, 13]
Death domain	1E41	100	6.9	[1]
Ribosomal protein S6	1RIS	101	6.1	[1, 13]
Ubiquitin related modifyer 1	2PKO	101	2.6	[1, 13]
Spliceosomal protein U1A	1AUD	102	4.6	[1, 13]
apocytochrome b5	1HKO	104	3.0	[1]
Cytochrome b562	256B	106	12.3	[1, 25]
FK506 binding protein	1FKF	109	1.6	[1, 13]
Src SH2 domain	1IS0	110	8.7	[1, 13]
P13	1QTU	117	-0.4	[1]
15th domain of brain α-spectrin	1U5P	118	11.0	[1, 26]
16th domain of brain α-spectrin	1CUN	118	4.8	[1, 26]
17th domain of brain α-spectrin	1CUN	118	3.4	[1, 26]

Hypothetical protein Tm1083	1J5U	124	6.9	[1, 13]
Chemotaxis protein CheW	1K0S	151	7.4	[1, 13]
Cyclophilin A	1LOP	164	6.6	[1, 27]
Apoflavodoxin (Anabaena sp.)	1RCF	169	0.8	[1]
Lyme desease variable surface antigen	1L8W	338	2.0	[1, 13]

Table S2. Protein chain length and folding rate constant (k_f) values of multi-state folders of Figure 3B.

Name	PDB	Length	ln k_f	Reference
Engrailed homeodomain	1ENH	61	10.5	[1, 11]
Phage 434 cro protein	2CRO	71	3.7	[1, 28]
FF domain	1UZC	71	7.7	[1]
ACBP (Yeast)	1ST7	86	8.5	[1]
Immunoglobulin domain of cardiac titin	1TIT	89	3.6	[1, 29]
Barstar	1BRS	89	3.4	[1, 30]
N-terminal domain of HypF	1GXT	91	4.4	[1, 31]
Third PDZ domain from PSD-95	1TP3	93	3.0	[1]
Tenth fibronectin type III domain of fibronectin	1FNF	94	5.5	[1, 32]
PDZ2 domain from PTP-BL	1GM1	94	1.0	[1]
FRB	1AUE	95	6.0	[1]
C-terminal domain of the cell-surface receptor protein CD2	1HNG	98	1.8	[1, 33]
Barnase	1BNI	110	2.6	[1, 34]
Cell-cycle regulatory protein p13suc1	SUC1	113	4.2	[1, 35]
Hisactophilin	1HCD	118	4.6	[1]
Myotrophin	2MYO	118	4.8	[1]
Villin 14T, N-terminal domain of villin	2VIK	126	5	[1, 36]
Ileal lipid binding protein	1EAL	127	1.3	[1, 37]

Chemotactic protein	3CHY	129	1	[1, 38]
Intestinal fatty acid binding protein	1IFC	131	3.4	[1, 39]
Cellular retinol binding protein II	1OPA	134	1.4	[1, 39]
Cellular retinoic acid binding protein I	1CBI	138	-3.2	[1, 39]
RNase-H (C. tepidum)	3H08	146	1.9	[1]
Tumour suppressor protein p16	2A5E	148	3.5	[1, 40]
Apoflavodoxin (D. desulfuricans)	3F6R	148	3.5	[1]
Staphylococcal nuclease	1SNQ	149	2.3	[1]
Apomyoglobin	1A6N	151	1.1	[1, 41]
C-terminally truncated fragment of GroEL apical domain	1AON	155	-1.5	[1, 42]
Ribonuclease HI	2RN2	155	0.1	[1, 43]
Dihydrofolate reductase	1RA9	159	-3.2	[1, 44]
T4 lysozyme	2LZM	164	4.1	[1, 43]
p19INK4d	1BI8	166	2.2	[1]
N-terminal domain of phosphoglycerate kinase	1PHP	175	2.3	[1, 43]
C-terminal domain of phosphoglycerate kinase	1PHP	221	-3.5	[1, 45]
Sigps	1IGS	222	-7.8	[1]
Carbonic anhydrase	5A25	260	-4.2	[1]
Tryptophan synthase α subunit	1QOP	268	-2.5	[1, 46]
Tryptophan synthase β2 subunit	1QOP	396	-6.9	[1, 47]

SUPPLEMENTARY REFERENCES:

[1] M. Corrales, P. Cuscó, D.R. Usmanova, H.-C. Chen, N.S. Bogatyreva, G.J. Filion, D.N. Ivankov, Machine learning: How much does it tell about protein folding rates?, PLoS ONE 10 (2015) e0143166.

[2] V. Muñoz, P.A. Thompson, J. Hofrichter, W.A. Eaton, Folding dynamics and mechanism of β-hairpin formation, Nature 390 (1997) 196-199.

[3] L.L. Qiu, S.A. Pabit, A.E. Roitberg, S.J. Hagen, Smaller and faster: The 20-residue Trp-cage protein folds in 4 mu s, J. Am. Chem. Soc. 124 (2002) 12952-12953.

[4] P.A. Thompson, W.A. Eaton, J. Hofrichter, Laser Temperature Jump Study of the Helix⇌Coil Kinetics of an Alanine Peptide Interpreted with a ‘Kinetic Zipper' Model, Biochemistry 36 (1997) 9200-9210.

[5] C.D. Snow, N. Nguyen, V.S. Pande, M. Gruebele, Absolute comparison of simulated and experimental protein-folding dynamics, Nature 420 (2002) 102-106.

[6] M. Jäger, H. Nguyen, J.C. Crane, J.W. Kelly, M. Gruebele, The folding mechanism of a β-sheet: the WW domain, J. Mol. Biol. 311 (2001) 373-393.

[7] M.H. Wang, Y.F. Tang, S.S. Sato, L. Vugmeyster, C.J. McKnight, D.P. Raleigh, Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale, J. Am. Chem. Soc. 125 (2003) 6032-6033.

[8] N. Ferguson, C.M. Johnson, M. Macias, H. Oschkinat, A. Fersht, Ultrafast folding of WW domains without structured aromatic clusters in the denatured state, Proc. Natl. Acad. Sci. USA 98 (2001) 13002-13007.

[9] S. Spector, D.P. Raleigh, Submillisecond folding of the peripheral subunit-binding domain, J. Mol. Biol. 293 (1999) 763-768.

[10] T. Wang, Y.J. Zhu, F. Gai, Folding of a three-helix bundle at the folding speed limit, J. Phys. Chem. B 108 (2004) 3694-3697.

[11] S. Gianni, N.R. Guydosh, F. Khan, T.D. Caldas, U. Mayor, G.W.N. White, M.L. DeMarco, V. Daggett, A.R. Fersht, Unifying features in protein-folding mechanisms, Proc. Natl. Acad. Sci. USA 100 (2003) 13286-13291.

[12] S. Gianni, Y. Ivarsson, P. Jemth, M. Brunori, C. Travaglini-Allocatelli, Identification and characterization of protein folding intermediates, Biophys. Chem. 128 (2007) 105-113.

[13] K.L. Maxwell, D. Wildes, A. Zarrine-Afsar, M.A. de los Rios, A.G. Brown, C.T. Friel, L. Hedberg, J.C. Horng, D. Bona, E.J. Miller, A. Vallee-Belisle, E.R.G. Main, F. Bemporad, L.L. Qiu, K. Teilum, N.D. Vu, A.M. Edwards, I. Ruczinski, F.M. Poulsen, B.B. Kragelund, S.W. Michnick, F. Chiti, Y.W. Bai, S.J. Hagen, L. Serrano, M. Oliveberg, D.P. Raleigh, P. Wittung-Stafshede, S.E. Radford, S.E. Jackson, T.R. Sosnick, S. Marqusee, A.R. Davidson, K.W. Plaxco, Protein folding: Defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins, Protein Sci. 14 (2005) 602-616.

[14] J.K. Myers, T.G. Oas, Preorganized secondary structure as an important determinant of fast protein folding, Nat. Struct. Biol. 8 (2001) 552-558.

[15] R. Guerois, L. Serrano, The SH3-fold Family: Experimental Evidence and Prediction of Variations in the Folding Pathways, J. Mol. Biol. 304 (2000) 967-982.

[16] D. Perl, C. Welker, T. Schindler, K. Schroder, M.A. Marahiel, R. Jaenicke, F.X. Schmid, Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins, Nat. Struct. Biol. 5 (1998) 229-235.

[17] K.L. Reid, H.M. Rodriguez, B.J. Hillier, L.M. Gregoret, Stability and folding properties of a model beta-sheet protein, Escherichia coli CspA, Protein Sci. 7 (1998) 470-479.
[18] Y. Zhu, D.O.V. Alonso, K. Maki, C.Y. Huang, S.J. Lahr, V. Daggett, H. Roder, W.F. DeGrado, F. Gai, Ultrafast folding of alpha(3): A de novo designed three-helix bundle protein, Proc. Natl. Acad. Sci. USA 100 (2003) 15486-15491.
[19] N.A.J. Van Nuland, W. Meijberg, J. Warner, V. Forge, R.M. Scheek, G.T. Robillard, C.M. Dobson, Slow cooperative folding of a small globular protein HPr, Biochemistry 37 (1998) 622-637.
[20] K.W. Plaxco, C. Spitzfaden, I.D. Campbell, C.M. Dobson, A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules, J. Mol. Biol. 270 (1997) 763-770.
[21] J.I. Guijarro, C.J. Morton, K.W. Plaxco, I.D. Campbell, C.M. Dobson, Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy11Edited by P. E. Wright, J. Mol. Biol. 276 (1998) 657-667.
[22] J. Clarke, S.J. Hamill, C.M. Johnson, Folding and stability of a fibronectin type III domain of human tenascin, J. Mol. Biol. 270 (1997) 771-778.
[23] J. Clarke, E. Cota, S.B. Fowler, S.J. Hamill, Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway, Structure 7 (1999) 1145-1153.
[24] N. Taddei, F. Chiti, P. Paoli, T. Fiaschi, M. Bucciantini, M. Stefani, C.M. Dobson, G. Ramponi, Thermodynamics and kinetics of folding of common-type acylphosphatase: Comparison to the highly homologous muscle isoenzyme, Biochemistry 38 (1999) 2135-2142.
[25] P. Wittung-Stafshede, J.C. Lee, J.R. Winkler, H.B. Gray, Cytochrome folding triggered by electron transfer: Approaching the speed limit for formation of a four-helix-bundle protein, Proc. Natl. Acad. Sci. USA 96 (1999) 6587-6590.
[26] K.A. Scott, S. Batey, K.A. Hooton, J. Clarke, The folding of spectrin domains I: wild-type domains have the same stability but very different kinetic properties, J. Mol. Biol. 344 (2004) 195-205.
[27] T. Ikura, T. Hayano, N. Takahashi, K. Kuwajima, Fast folding of Escherichia coli cyclophilin A: a hypothesis of a unique hydrophobic core with a phenylalanine cluster11Edited by C. R. Matthews, J. Mol. Biol. 297 (2000) 791-802.
[28] D.V. Laurents, S. Corrales, M. Elías-Arnanz, P. Sevilla, M. Rico, S. Padmanabhan, Folding Kinetics of Phage 434 Cro Protein, Biochemistry 39 (2000) 13963-13973.
[29] S.B. Fowler, J. Clarke, Mapping the folding pathway of an immunoglobulin domain: Structural detail from phi value analysis and movement of the transition state, Structure 9 (2001) 355-366.
[30] G. Schreiber, A.R. Fersht, The refolding of cis-peptidylprolyl and trans-peptidylprolyl isomers of barstar, Biochemistry 32 (1993) 11195-11203.
[31] G. Calloni, N. Taddei, K.W. Plaxco, G. Ramponi, M. Stefani, F. Chiti, Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding, J. Mol. Biol. 330 (2003) 577-591.
[32] E. Cota, J. Clarke, Folding of beta-sandwich proteins: Three-state transition of a fibronectin type III module, Protein Sci. 9 (2000) 112-120.

[33] M.J. Parker, C.E. Dempsey, M. Lorch, A.R. Clarke, Acquisition of native beta-strand topology during the rapid collapse phase of protein folding, Biochemistry 36 (1997) 13396-13405.
[34] A. Matouschek, J.T. Kellis, L. Serrano, M. Bycroft, A.R. Fersht, Transient folding intermediates characterized by protein engineering, Nature 346 (1990) 440-445.
[35] J.W.H. Schymkowitz, F. Rousseau, L.R. Irvine, L.S. Itzhaki, The folding pathway of the cell-cycle regulatory protein p13suc1: clues for the mechanism of domain swapping, Structure 8 (2000) 89-100.
[36] S.E. Choe, P.T. Matsudaira, J. Osterhout, G. Wagner, E.I. Shakhnovich, Folding kinetics of villin 14T, a protein domain with a central beta-sheet and two hydrophobic cores, Biochemistry 37 (1998) 14508-14518.
[37] P.M. Dalessio, I.J. Ropson, beta-sheet proteins with nearly identical structures have different folding intermediates, Biochemistry 39 (2000) 860-871.
[38] V. Munoz, E.M. Lopez, M. Jager, L. Serrano, Kinetic characterization of the chemotactic protein from escherichia-coli, chey - kinetic-analysis of the inverse hydrophobic effect, Biochemistry 33 (1994) 5858-5866.
[39] L.L. Burns, P.M. Dalessio, I.J. Ropson, Folding mechanism of three structurally similar beta-sheet proteins, Proteins: Struct. Funct. Genet. 33 (1998) 107-118.
[40] K.S. Tang, B.J. Guralnick, W.K. Wang, A.R. Fersht, L.S. Itzhaki, Stability and folding of the tumour suppressor protein p1611, J. Mol. Biol. 285 (1999) 1869-1886.
[41] S. Cavagnero, H.J. Dyson, P.E. Wright, Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin11Edited by F. Cohen, J. Mol. Biol. 285 (1999) 269-282.
[42] R. Golbik, R. Zahn, S.E. Harding, A.R. Fersht, Thermodynamic stability and folding of GroEL minichaperones, J. Mol. Biol. 276 (1998) 505-515.
[43] M.J. Parker, S. Marqusee, The cooperativity of burst phase reactions explored, J. Mol. Biol. 293 (1999) 1195-1210.
[44] P.A. Jennings, P.E. Wright, Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin, Science 262 (1993) 892.
[45] M.J. Parker, R.B. Sessions, I.G. Badcoe, A.R. Clarke, The development of tertiary interactions during the folding of a large protein, Fold Des. 1 (1996) 145-156.
[46] K. Ogasahara, K. Yutani, Unfolding-refolding kinetics of the tryptophan synthase alpha-subunit by CD and fluorescence measurements, J. Mol. Biol. 236 (1994) 1227-1240.
[47] M.E. Goldberg, G.V. Semisotnov, B. Friguet, K. Kuwajima, O.B. Ptitsyn, S. Sugai, An early immunoreactive folding intermediate of the tryptophan synthase β2 subunit is a ‘molten globule’, FEBS Lett. 263 (1990) 51-56.