LETTER

Native, sequential protein folding via anchored N and C protein termini

Saverio Alberti[a,1]

In a recent article in PNAS, Zhuravlev et al. (1) determine unfolding trajectories of the Src tyrosine kinase SH3 (Src homology 3) domain [Protein Data Base (PDB) ID code 1SRL]. Using laser optical tweezing, constant force was applied to the SH3 N and C termini (residues 9 and 59), and *f*-dependent unfolding rates were computed. Notably, a switch between distinct unfolding pathways was detected. The question arises as to what extent such molecular mechanisms apply to living cells/physiological settings.

The approach by Zhuravlev et al. (1) relies on proteins unfolding upon application of force to N and C protein termini. A notable parallel appears to be the folding/unfolding cycles, as guided by chaperones, such as Hsp70 or GroES/GroEL (2). Chaperone interactions indeed greatly lower the activation energy required for protein unfolding, through a sequential sliding of progressively unfolded/folding polypeptides in folding apparatuses.

The issue of whether N and C termini anchoring for folding/unfolding can occur in living cells was thus challenged. Such an occurrence rests on a basic tenet, that is, that N and C termini for anchored folding/unfolding should be structurally available for binding (i.e., they should be exposed to the solvent on the protein surface). This model was tested on a sample of randomly chosen protein crystals (Table 1). Rather remarkably, all examined proteins were found to possess surface-accessible, exposed N and C termini. Distinct protein domains can behave as independent folding units (3). Consistent, surface-exposed N and C termini were identified in SH2, SH3, PH, PTB, EGF, and DNA-binding domains [including the PDB ID code 1SRL SH3 domain in (1)].

However, cotranslational protein folding (4, 5) occurs in a strictly sequential manner, from the N- to C-terminal ends of newly synthesized polypeptides. Investigations on the speed limit of protein folding have identified examples where folding occurs in microseconds to nanoseconds (6). Main conformational changes of a protein backbone can be complete after only 20 ps (7). The protein synthesis apparatus can add one amino acid every 50 ms. Thus, protein synthesis is several orders of magnitude slower than the folding process. Translation-coupled folding is thus rate-limiting, leading to a quasiequilibrium, restricted sampling of the conformational space during translation, which plays a key role in folding (5). Correspondingly, protein folding was shown to proceed through a compact conformation in the peptide tunnel, to then reach a native-like structure after emergence from the ribosome (5). Such processivity is largely preserved also in unfolding/folding chaperone-assisted processes (2, 8), suggesting this feature to be fundamental for folding in living cells.

Such N to C terminus folding processivity is entirely missing in typical ensemble folding or unfolding experiments (1, 7, 9). Nucleation mechanisms can be shared by the ensemble vs. sequential folding processes (3, 10). However, all subsequent folding steps remain missing. Hence, additional technology quantum leaps are called for, to extend the validity of the mechanisms investigated by Zhuravlev et al. (1) to physiological settings. The potential is there for fundamental insights into protein folding/unfolding processes.

[a]Unit of Cancer Pathology, Center of Sciences of Aging and Translational Medicine, University 'G. d'Annunzio,' Chieti Scalo, Chieti 66100, Italy

Author contributions: S.A. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

[1]Email: s.alberti@unich.it.

Table 1. Protein surface-exposure of N and C termini

Protein species	N terminus	C terminus
SH2, SH3 domains		
Src SH3 1SRL	Surface	Surface
SEM-5 C-terminal SH3 1SEM	Surface	Surface
SH2 1cwd	Surface	Surface
PH, PTB domains		
PLCδ PH 1mai	Surface	Surface
Pleckstrin PH 1pls	Surface	Surface
Dynamin PH 1dyn	Surface	Surface
Spectrin PH 1btn	Surface	Surface
Spectrin PH 1dro	Surface	Surface
PTB 1shc	Surface	Surface
EGF domains		
IGF-BP5 1boe	Surface	Surface
E-selectin 1esl	Surface	Surface
Gromos 1apo	Surface	Surface
Signaling complexes		
Gαβγ 1got (all subunits)	Surface	Surface
Gβ 2trc	Surface	Surface
Cytoskeleton		
β-actin 2oan	Surface	Surface
Profilin 1hlu	Surface	Surface
Gelsolin 1d0n	Surface	Surface
Severin 1svq	Surface	Surface
Spectrin repeat 2spc	Surface	Surface
Villin 1vil	Surface	Surface
Enzymes		
Acetylcholinesterase 2ace	Surface	Surface
Cathepsin D 1lya	Surface	Surface
Ferredoxin 1awd	Surface	Surface
Fructose-1,6-bisphosphatase 1fpi	Surface	Surface
Glucose oxidase 1gog	Surface	Surface
Glutaredoxin (phage T4 thioredoxin) 1aba	Surface	Surface
Inositol polyphosphate 1-phosphatase 1inp	Surface	Surface
PI-specific PLC 1gym	Surface	Surface
PLCδ1 1djx	Surface	Surface
Insulin receptor catalytic domain 1irk	Surface	Surface
PKA 1cmk	Surface	Surface
PKCβ1 1rlw	Surface	Surface
Metallothionein-2 2mhu (EGF domain)	Surface	Surface
Transcription factors		
1a1	Surface	Surface
1a5t	Surface	Surface
1a6b	Surface	Surface
1aaf	Surface	Surface
1aay	Surface	Surface
1ard	Surface	Surface
1are	Surface	Surface
1arf	Surface	Surface
1bbo	Surface	Surface
1bhi	Surface	Surface
1bj6	Surface	Surface
1dsq	Surface	Surface
1dsv	Surface	Surface
1dvp	Surface	Surface
1fre	Surface	Surface
1gnf	Surface	Surface
1hcp	Surface	Surface
1hra	Surface	Surface
1hvn	Surface	Surface
1hvo	Surface	Surface
1ile	Surface	Surface
1mey	Surface	Surface
1mfs	Surface	Surface

Table 1. Cont.

Protein species	N terminus	C terminus
1nc8	Surface	Surface
1ncs	Surface	Surface
1paa	Surface	Surface
1pyi	Surface	Surface
1qf8	Surface	Surface
1rgd	Surface	Surface
1rmd	Surface	Surface
1sp1	Surface	Surface
1sp2	Surface	Surface
1tf3	Surface	Surface
1tf6	Surface	Surface
1ubd	Surface	Surface
1zaa	Surface	Surface
1zfd	Surface	Surface
1zin	Surface	Surface
1znf	Surface	Surface
1znm	Surface	Surface
2adr	Surface	Surface
2gli	Surface	Surface
2znf	Surface	Surface
3znf	Surface	Surface
4znf	Surface	Surface
5znf	Surface	Surface
7znf	Surface	Surface
DNA binding proteins		
GCN4 1A02	Surface	Surface
TAF12 1QB3	Surface	Surface
RFC2 1IQP (all subunits)	Surface	Surface
Cytoplasmic proteins		
Haemoglobin 1nih	Surface	Surface
MCP-1 1don	Surface	Surface
Myoglobin 1do1	Surface	Surface
NEF 1efn	Surface	Surface
Streptavidin 1rst	Surface	Surface
GFP 1ema	Surface	Surface

1 Zhuravlev PI, Hinczewski M, Chakrabarti S, Marqusee S, Thirumalai D (2016) Force-dependent switch in protein unfolding pathways and transition-state movements. *Proc Natl Acad Sci USA* 113(6):E715–E724.
2 Willmund F, et al. (2013) The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis. *Cell* 152(1-2):196–209.
3 Imoto T (2001) Effective protein folding in simple random search. *Biopolymers* 58(1):46–49.
4 Kim SJ, et al. (2015) Protein folding. Translational tuning optimizes nascent protein folding in cells. *Science* 348(6233):444–448.
5 Holtkamp W, et al. (2015) Cotranslational protein folding on the ribosome monitored in real time. *Science* 350(6264):1104–1107.
6 Mayor U, et al. (2003) The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* 421(6925):863–867.
7 Chang IJ, Lee JC, Winkler JR, Gray HB (2003) The protein-folding speed limit: Intrachain diffusion times set by electron-transfer rates in denatured Ru(NH3)5(His-33)-Zn-cytochrome c. *Proc Natl Acad Sci USA* 100(7):3838–3840.
8 Thulasiraman V, Yang CF, Frydman J (1999) In vivo newly translated polypeptides are sequestered in a protected folding environment. *EMBO J* 18(1):85–95.
9 Camacho CJ, Thirumalai D (1995) Theoretical predictions of folding pathways by using the proximity rule, with applications to bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA* 92(5):1277–1281.
10 Finkelstein AV (2002) Cunning simplicity of a hierarchical folding. *J Biomol Struct Dyn* 20(3):311–313.