CrossMark

Polyproteins in structural biology

Thibaut Crépin[1], Christopher Swale[1], Alexandre Monod[1],
Frederic Garzoni[1,2], Maxime Chaillet[1,2] and Imre Berger[1,2,3]

Polyproteins are chains of covalently conjoined smaller proteins that occur in nature as versatile means to organize the proteome of viruses including HIV. During maturation, viral polyproteins are typically cleaved into the constituent proteins with different biological functions by highly specific proteases, and structural analyses at defined stages of this maturation process can provide clues for antiviral intervention strategies. Recombinant polyproteins that use similar mechanisms are emerging as powerful tools for producing hitherto inaccessible protein targets such as the influenza polymerase, for high-resolution structure determination by X-ray crystallography. Conversely, covalent linking of individual protein subunits into single polypeptide chains are exploited to overcome sample preparation bottlenecks. Moreover, synthetic polyproteins provide a promising tool to dissect dynamic folding of polypeptide chains into three-dimensional architectures in single-molecule structure analysis by atomic force microscopy (AFM). The recent use of natural and synthetic polyproteins in structural biology and major achievements are highlighted in this contribution.

Addresses
[1] Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, UMI 3265, 71 Avenue des Martyrs, 38042 Grenoble Cedex 9, France
[2] The European Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, BP181, 38042 Grenoble Cedex 9, France
[3] The School of Biochemistry, University of Bristol, Bristol BS8 1TD, United Kingdom

Corresponding authors: Crépin, Thibaut (crepin@embl.fr) and Berger, Imre (iberger@embl.fr)

Introduction

Polyproteins composed of covalently linked individual proteins with different biological functions are prevalent in nature. For instance SARS coronavirus, the agent that causes severe acute respiratory syndrome, realizes its entire proteome from two large polyproteins, each encoded by a long single open reading frame (ORF) [1]. The expressed SARS polyproteins are then processed into the individual functional protein subunits by the action of highly specific proteases also encoded by the ORFs [1,2]. A further example is human immune deficiency virus (HIV) which causes acquired immune deficiency syndrome (AIDS). The RNA genome of this retrovirus is organized in three major genes *gag*, *pol* and *env*, which encode for polyproteins and undergo proteolytic processing at defined stages during maturation [3]. Recombinant polyprotein approaches mimicking viral polyproteins have recently emerged as a powerful means to express high-value protein complexes for structure determination [4,5•]. Recently, the long elusive influenza polymerase has been produced successfully from a self-processing synthetic polyprotein, enabling high-resolution structure determination [6••,7]. Polyprotein fusions which are not processed by protease, but remain covalently conjoined by engineered linkers have been instrumental to obtain important insight into numerous essential physiological processes including multidrug efflux, co-translational protein targeting or enzymatic processing of chromatin, among others [8••,9,10,11•,12••]. Moreover, single-chain engineering approaches linking protein domains into novel, artificial polyproteins have resulted in new classes of high-affinity binder molecules as potential protein therapeutics [13] and accelerated elucidation of mechanisms governing protein folding by single-molecule techniques [14]. Thus, polyprotein technologies have recently gained prominence as particularly useful tools for unlocking previously often inaccessible protein samples to detailed structural and mechanistic studies, as illustrated in the following. In this contribution, the use of polyproteins in structural biology is discussed, by highlighting recently determined structures of naturally occurring polyproteins on one hand and by reviewing recent structural studies where recombinant polyprotein constructions were utilized on the other.

Natural polyproteins

Polyproteins are used in nature by many viruses to structure their proteome. As a consequence, viral polyproteins are an intense focus of research efforts for numerous reasons [15–17]. For instance, inhibition by small molecules of proteases that process polyproteins during viral maturation can provide a powerful handle to combat viral disease [2,18–20]. In HIV, the physical infrastructure of the virus is provided by the *gag* gene which gives rise to the precursor Gag (group specific antigen) polyprotein [21]. Viral protease processes Gag during maturation into several proteins and spacer peptides dictating immature

and mature viral capsid structure [22]. Maturation is a two-stage process. Precursor Gag polyprotein first forms a hexameric lattice at the plasma membrane of an infected cell. This induces budding and release of immature viral particles. Proteolytic processing of Gag then rearranges the viral structure to the mature form [23]. Inhibition of Gag-processing protease enabled preparation of immature retroviral capsids suitable for structure analysis by electron cryo-tomography and subtomogram averaging methods (Figure 1) [24–26,27••]. Comparison of the structures of HIV Gag protein in reconstituted tubular arrays on one hand and in intact virus particles on the other, provided unprecedented insights into the conformational plasticity of this precursor polyprotein [26,27••]. The studies also revealed that retroviral capsid proteins can adopt different quaternary arrangements during virus assembly, notwithstanding conserved tertiary structures [27••].

A different, less frequently encountered type of natural, non-viral polyproteins are so-called tandemly repetitive polyproteins (TRPs). TRPs are also produced as large precursor proteins and then processed by proteases into several copies of proteins with similar function. TRPs are made of consecutively arranged repeats of amino-acid stretches. Examples of known TRPs include polyprotein lipid binding proteins of nematodes [28] and filaggrins, which are keratinocyte produced TRPs crucial to health and appearance of skin [29,30]. The solution structure of a mature, post-translationally processed repeat unit of a TRP, ABA-1A from the nematode polyprotein allergen of *Ascaris*, was determined, representing the first structure of this class of proteins [31]. ABA-1A adopts a novel fold comprising two juxtaposed four-helical bundles that share a long central alpha-helix (Figure 1). Nematode polyprotein allergens have no known counterpart in humans. The *Ascaris* ABA-1A structure therefore may serve as a starting point for the development of new drugs and therapeutic intervention strategies against disease states caused by these intestinal parasites.

An inverse 'polyprotein' concept of covalently linking functional protein units into long modular polypeptide chains characterizes mega-enzymes that functionally arrange multiple domains into ordered assembly lines for the production of a wide variety of bioactive molecules. Modular polyketide synthases (PKS) and their metazoan homologs, fatty acid synthases (FAS) belong to this class

Figure 1

Current Opinion in Structural Biology
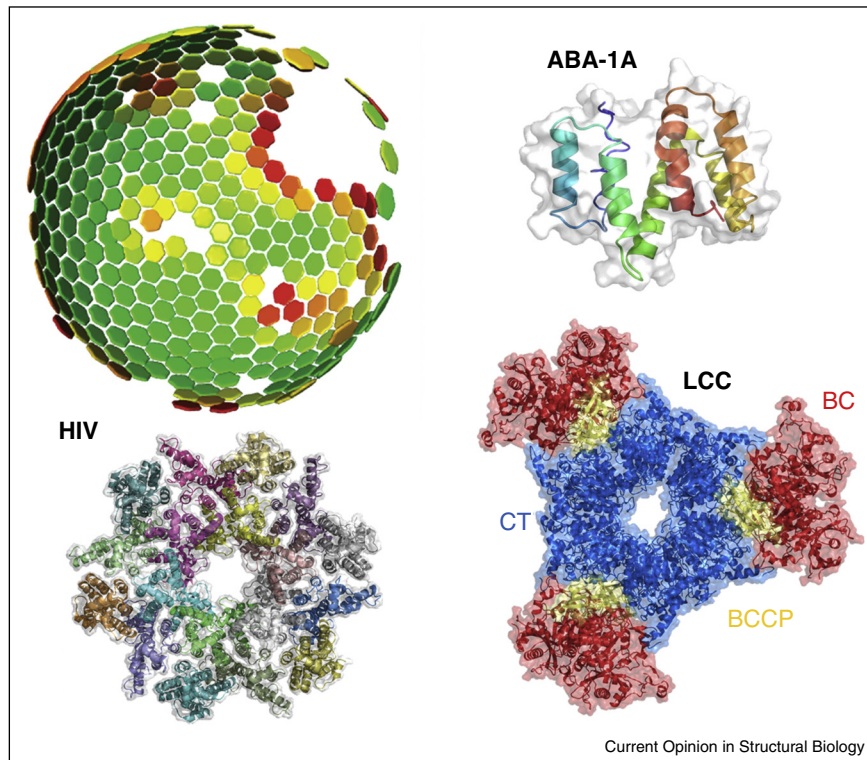
Natural polyprotein structures. Polyproteins, prevalent in biology, are illustrated here by the structures of the human immune deficiency virus (HIV) immature capsid determined by electron cryo-tomography revealing molecular details of hexameric Gag [24,26,27••], the structure of a repeated unit of the ABA-1 nematode polyprotein allergen derived from nuclear magnetic resonance (NMR) spectroscopy [31] and the crystal structure of a single-chain, multi-domain long-chain acyl-CoA carboxylase, LCC [34]. CT stands for carboxyltransferase, BCCP for biotin carboxyl carrier protein, and BC for biotin carboxylase components.

of catalysts [32,33]. Recently, the structure and function of a modular multi-domain long-chain acyl-CoA carboxylase from *Mycobacterium avium* was elucidated [34]. The crystal structure revealed extensive swapping of functional domains in the holo-enzyme which is a homo-hexamer (Figure 1). Thus, four intertwined protomers are involved in completing one catalytic reaction cycle [34].

Synthetic polyproteins

Influenza polymerase produced from a recombinant polyprotein

The highly successful strategy of viruses to utilize polyproteins to their advantage has found its equivalent in recombinant technology. A synthetic polyprotein,

expressed recombinantly in baculovirus-infected insect cells, has enabled the structure determination of influenza polymerase. Despite its common appearance, a detailed understanding of the molecular mechanisms of the virus that causes influenza has remained elusive. More than 40 years ago, the influenza polymerase was discovered, a key protein complex that replicates the genetic material of the virus [35,36]. Atomic resolution information on the structure and function of this protein machine is essential, as it may open up important avenues for drug discovery. However, the influenza polymerase remained inaccessible for decades — to produce this valuable protein complex for detailed analysis, proved to be a seemingly insurmountable technical challenge. This has now changed

Figure 2
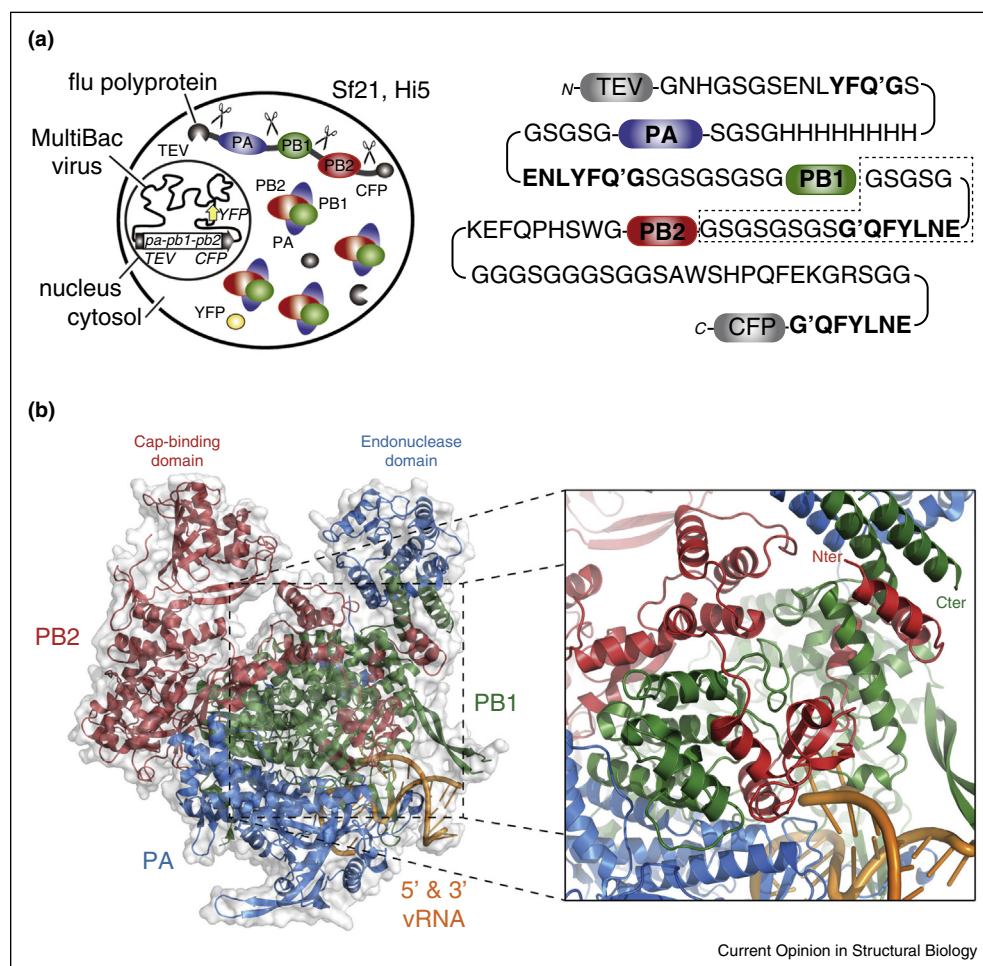
Influenza polymerase. **(a)** A self-processing recombinant polyprotein was used to express influenza polymerase complex in high quality and quantity in insect cells by using the MultiBac system [37] as illustrated (left). The polymerase was expressed from a single open reading frame encoding for tobacco etch virus NIa protease (TEV), the polymerase subunits PA, PB1 and PB2 and a fluorescent protein (CFP). A second, yellow fluorescent protein (YFP) was inserted elsewhere into the MultiBac baculoviral genome as an expression performance marker. The resulting polyprotein is processed into the individual subunits by highly specific proteolysis mediated by TEV. The polyprotein construct is shown schematically (right) illustrating tag placement and design of spacers in between the subunits (adapted from [6••]). The C-terminal part of PB1 and the N-terminal part of PB2 co-fold into a helical bundle necessitating careful linker design (boxed). **(b)** Crystal structure of influenza polymerase bound to cognate viral RNA (vRNA) is shown in a ribbon representation. PA is colored in blue, PB1 in green and PB2 in red. RNA substrate is colored in orange. Cap-binding domain and endonuclease domain are indicated. The structure motif formed by PB1 C-terminal domain and PB2 N-terminal domain is shown in a magnification (right). PB2 N-terminus and PB1 C-terminus are marked.

dramatically with structures of influenza polymerase complex determined by X-ray crystallography [6[··],7]. These break-through studies provide unprecedented insight into the inner workings of this viral protein machine. This revolution in understanding influenza was brought about by applying a polyprotein strategy to produce influenza polymerase recombinantly, in the quality and quantity required for high-resolution structural and functional analysis (Figure 2). The strategy applied recapitulates the mechanism adopted by SARS coronavirus. A single ORF was constructed encoding for a highly specific protease, NIa, from tobacco etch virus (TEV), fused in frame with consecutively arranged PA, PB1 and PB2 subunits of the polymerase. At the far end of the polyprotein, a reporter protein was inserted to track protein production 'in real-time' during expression, by recording fluorescence [5[·]]. All protein units within the polyprotein were spaced apart by customized linkers containing the specific site for cleavage by the TEV protease, purification tag sequences and spacer residues (Figure 2). The polyproteins encoding for influenza polymerases were successfully expressed with the Multi-Bac baculovirus insect cell expression system developed for producing multiprotein complexes [37].

Sample preparation bottlenecks resolved by synthetic polyproteins that remain uncleaved

The polyprotein encoding for influenza polymerase was proteolytically cleaved into the constituent protein sub-units to yield the sample that crystallized. Conversely, engineering of individual proteins into covalently linked polypeptide chains can also accelerate structure determination considerably. These 'polyproteins' remain conjoined as single-chains during sample preparation and structure determination. Particularly prominent examples for such single-chain engineering include the insertion of T4 lysozyme into the primary sequence of G-protein coupled receptors (GPCRs) to facilitate crystallogenesis [38]. In a variation of this approach, the catalytic domain of *Pyrococcus abysii* glycogen synthase was recently used to stabilize an intracellular loop of the human OX_2 orexin receptor to determine its crystal structure bound to an insomnia drug [39].

Elaborate single-chain engineering into polyproteins was applied to determine the architecture of the bacterial multidrug efflux pump AcrABZ-TolC [8[··]]. In this study, the efflux pump was assembled by preparing two single-chain polypeptide fusions, AcrB-AcrA-AcrB and

Figure 3
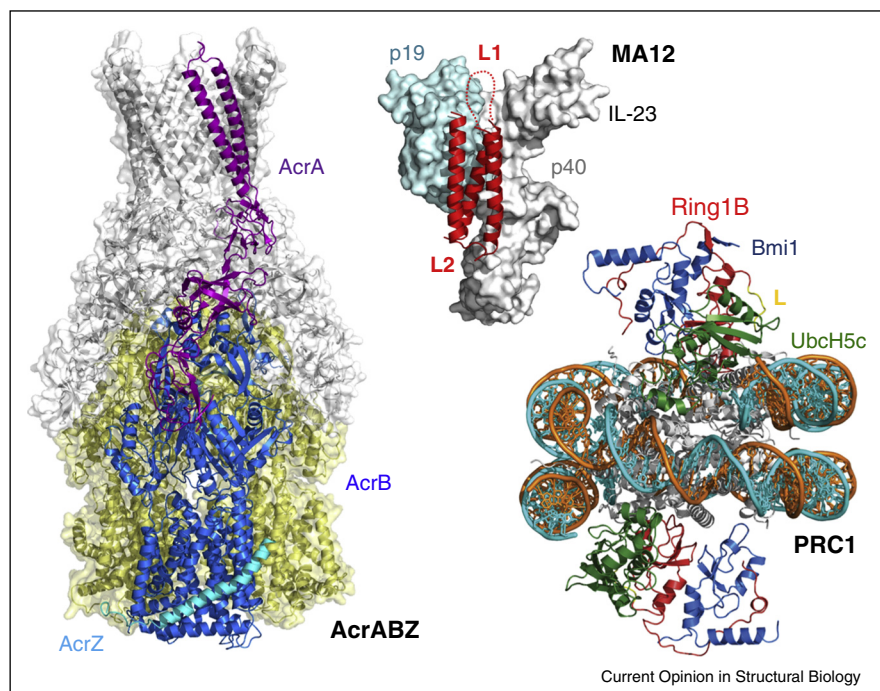
Linking individual proteins into polypeptide chains. Proteins engineered into single polypeptide chains were used to obtain suitable sample for structure determination of AcrABZ complex [8[··]], an alphabody (MA12) to neutralize human interleukin Il-23 (p19 and p40) [13] and the first structure of a histone-modifying enzyme, the Polycomb Repressive Complex (PRC) 1 ubiquitylation module, bound to a nucleosome [12[··]]. Linker amino acid segments are marked (L, L1, L2).

AcrA-AcrZ, respectively, each conjoined by extended glycine/serine rich linkers. TolC was co-expressed with the AcrA-AcrZ fusion. This strategy resulted in a reconstituted complex with the correct stoichiometry enabling structure determination by hybrid methods. Fitting of crystal coordinates into EM densities allowed to model AcrBZ/AcrA interactions (Figure 3) as well as the holo-complex containing TolC.

Single-chain engineering likewise enabled the development of alphabodies, a novel scaffold representing a promising alternative to antibodies for various biomedical and biotechnological applications [13]. Alphabodies comprise *in silico* designed short individual alpha-helical protein segments that are then conjoined by covalent

linkers into single-chain antiparallel coiled-coils that are highly stable and suitable for affinity maturation. The crystal structure of a complex with human interleukin (IL-23) revealed the structural basis of IL-23 antagonism by the alphabody, MA12 (Figure 3).

Covalent linking of subunits of a protein complex was required to obtain the first crystal structure of a histone-modifying enzyme complex bound to a nucleosome core particle [12**]. Polycomb repressor complex (PRC) 1, an essential regulator of cell fate, comprises an activity to ubiquitylate nucleosomal histone H2A at residue K119. PRC1 uses its E3 ubiquitin ligase subunits, Ring1B and Bmi1, together with an E2 ubiquitin-conjugating enzyme, UbcH5c for this purpose. The E2–E3 complex was

Figure 4
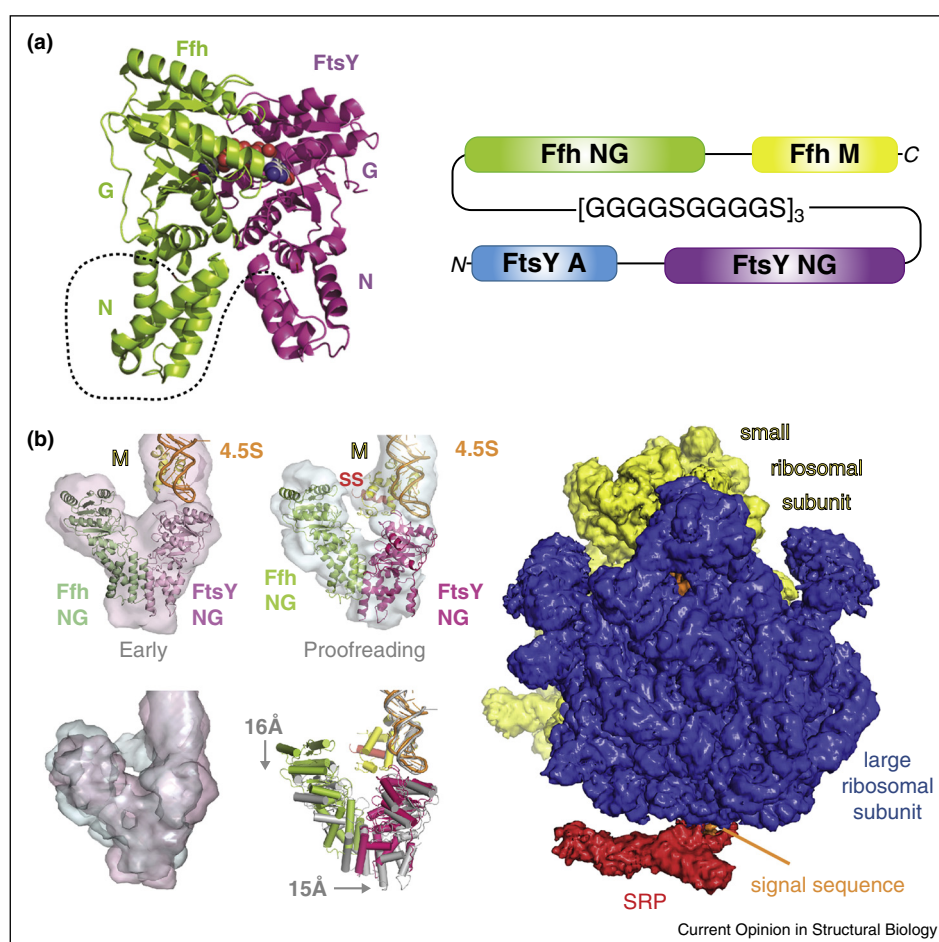
SRP pathway revealed by single-chain protein engineering. **(a)** *E. coli* signal recognition particle (SRP) protein subunit Ffh (green) and the SRP receptor FtsY (purple) were covalently linked into a single polypeptide chain (left). NG domains are marked. The dashed line indicates the extended polypeptide linker. Bound nucleotide is shown (spheres). The single polypeptide construct is illustrated on the right, detailing the arrangement of FtsY A (blue) and NG (purple) domains, the 30 amino acid glycine/serine rich linker connecting FtsY with Ffh, and the Ffh NG (green) and M (yellow) domains. N and C-termini are marked. **(b)** Structures of SRP/FtsY bound to a translating ribosome were elucidated by electron cryo-microscopy (color coding as in panel a). 4.5S RNA is colored in orange. EM density and fitted models of the 'early' and 'proofreading' stages are shown (left, top). Substantial rearrangements are observed as illustrated by the overlay of the EM densities and the model coordinates (left, bottom) (adapted from [10•]). The cryo-EM structure of the ribosome-SRP-FtsY co-translational targeting complex in the closed state is shown on the right (adapted from [11•]).

produced using an engineered single-chain fusion of Ring1B to UbcH5c to overcome the low affinity and salt-sensitive E2–E3 interaction. The resulting single-chain construct was co-expressed with Bmi1 to produce the trimeric complex which was purified and reconstituted with nucleosome core particles, and crystallized [12$^{••}$]. The structure showed two copies of PRC1 E2–E3 complex bound to one nucleosome, revealing intricate interactions (Figure 3).

### SRP pathway revealed by single-chain protein engineering

Around one third of the proteins in living cells are delivered to the plasma membrane. This is carried out by a universally conserved, complex mechanism involving ribosomes that are translating mRNA into membrane-bound nascent polypeptide chains, the signal recognition particle, SRP and the SRP receptor, FtsY [9,40,41]. Snapshots of this elaborate process were obtained by single particle electron cryo-microscopy and biochemical analysis [9,10$^{•}$,11$^{•}$,42–44]. These studies revealed SRP binding to ribosome nascent chain complexes [42], followed by the 'early' [44], 'proofreading' [10$^{•}$] and 'closed' [11$^{•}$] states upon FtsY binding and GTP hydrolysis. Successful structure determination critically relied on stabilizing SRP binding to FtsY. This was achieved by covalently linking the SRP subunit Ffh with FtsY into a single polypeptide chain with virtually wild-type activity (Figure 4).

Polyprotein single-molecule structural biology

Transient unfolding and refolding of proteins can be an essential feature of protein structure space in living organisms, for example in the translocation of proteins into and across cellular membranes or the muscle stroke. The use of atomic force microscopy (AFM) has emerged as a powerful technique to probe protein structure, enabling analysis of the mechanical stability and folding pathway of protein specimens at the single-molecule level. By means of a stretching force, applied through a microscopic cantilever to a biological target fixed to a support and recorded by the deflection of a laser beam, the analyzed protein is unfolded to an extended state (Figure 5).

Historically, polyproteins were used in single-molecule AFM to measure unique mechanical fingerprint profiles of a protein response to the applied mechanical force [14,45$^{•}$]. The muscle protein titin is a prominent example [46–48]. Titin consists of several hundred repeated protein domains including fibronectin and immunoglobulin-like folds. The use of polyproteins in single-molecule AFM results in considerable improvement of the statistical evaluation of singular domains within the polyprotein chain [14]. This can be exploited in the analysis of homomeric polyproteins that are constructed genetically or by chemical fusion reaction from identical copies of the same protein species of choice. Moreover, in addition to
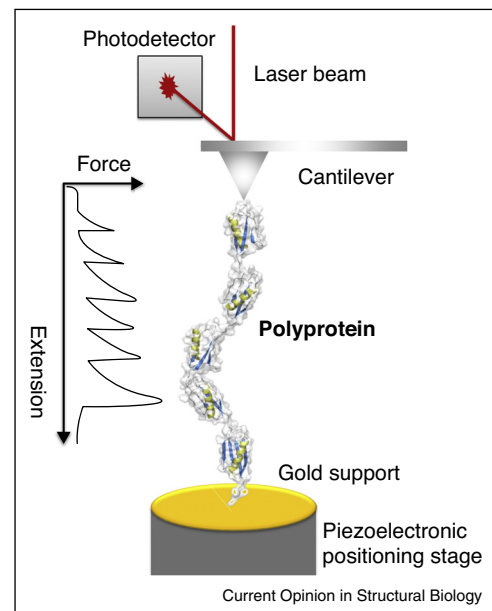
Polyprotein single-molecule structural biology. The setup of atomic force microscopy of polyproteins is shown in a schematic representation (adapted from [14]). The polyprotein is tethered to the gold support resting on the piezoelectronic positioning stage (bottom) on one end, and the tip of a cantilever made of silicon nitride on the other. A laser beam is focused on the back of the cantilever (top). The cantilever is displaced by the force that acts on the polyprotein chain resulting in change of the deflection of the laser beam, recorded by a photodetector. Increasing force causes each domain of the polyprotein to unfold, resulting in characteristic spikes in a force/extension diagram (inset).

providing a clear fingerprint, polyproteins also have the advantage that a larger number of events can be recorded per experiment as compared to only one event if monomeric proteins are used.

For these reasons polyproteins emerge as work-horses of single-molecule structural biology by AFM. Numerous polyproteins have been analyzed for their mechanical properties by using this technique, including poly-I27, derived from the I-band region of Titin, oligo-calmodulin, poly-ubiquitin, polyproteins made of the virulence factor GB1 of *Peptostreptococcus magnus* [46–54] and others, providing unique insights into biological folding/unfolding mechanisms. The availability of a large and growing number of well characterized homomeric specimens furthermore enables now the construction of chimeric polyproteins as a tool to study mechanically uncharacterized proteins, by using the unique fingerprints of the known protein unit as a reference [14,45$^{•}$].

Conclusions

Natural and synthetic polyproteins are at the core of contemporary structural biology. Analysis of viral polyproteins

and the architectural consequences of their processing during maturation not only furthers our understanding of viral mechanisms but provides important clues for drug design to combat viral diseases. Synthetic polyproteins are emerging as invaluable tools to accelerate research by unlocking hitherto inaccessible proteins for high resolution structure determination. Artificial polyproteins obtained by singe-chain protein engineering approaches are instrumental to overcome sample production bottlenecks and provide novel means to illuminate biological mechanisms, including folding and unfolding properties at the single-molecule level. We anticipate a major increase in the use of polyproteins in structural biology as valuable tools to tackle large and complex biological systems in the future.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Al-Tawfiq JA, Zumla A, Memish ZA: **Coronaviruses: severe acute respiratory syndrome coronavirus and Middle East respiratory syndrome coronavirus in travelers**. *Emerg Infect Dis* 2014, **20**:1562-1564.

2. Hilgenfeld R: **From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design**. *FEBS J* 2014, **281**:4085-4096.

3. Lucas S, Nelson AM: **HIV and the spectrum of human disease**. *J Pathol* 2015, **235**:229-241.

4. Vijayachandran LS, Viola C, Garzoni F, Trowitzsch S, Bieniossek C, Chaillet M, Schaffitzel C, Busso D, Romier C, Poterszman A *et al.*: **Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells**. *J Struct Biol* 2011, **175**:198-208.

5. Nie Y, Bellon-Echeverria I, Trowitzsch S, Bieniossek C, Berger I:
• **Multiprotein complex production in insect cells by using polyproteins**. *Methods Mol Biol* 2014, **1091**:131-141.
Detailed description of the methodology and protocols for expressing protein complexes from synthetic polyproteins in the MultiBac baculo-virus insect cell expression system.

6. Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crépin T, Hart D,
•• Lunardi T, Nanao M, Ruigrok RW, Cusack S: **Structural insight into cap-snatching and RNA synthesis by influenza polymerase**. *Nature* 2014, **516**:361-366.

First crystal structure of the influenza polymerase bound to cognate viral RNA substrate, providing unprecedented insights into cap-snatching and RNA synthesis. Influenza polymerase, inaccessible to structural analysis for decades since its discovery, was produced here for the first time successfully from a synthetic polyprotein construct.

7. Pflug A, Guilligay D, Reich S, Cusack S: **Structure of influenza A polymerase bound to the viral RNA promoter**. *Nature* 2014, **516**:355-360.

8. Du D, Wang Z, James NR, Voss JE, Klimont E, Ohene-Agyei T,
•• Venter H, Chiu W, Luisi BF: **Structure of the AcrAB-TolC multidrug efflux pump**. *Nature* 2014, **509**:512-515.
Structure determination of the bacterial AcrABZ-TolC multidrug efflux pump by hybrid methods combining X-ray crystallography and electron cryo-microscopy. Elaborate single-chain protein engineering was a pre-requisite for successful sample preparation of the complete efflux pump.

9. Knoops K, Schoehn G, Schaffitzel C: **Cryo-electron microscopy of ribosomal complexes in cotranslational folding, targeting, and translocation**. *Wiley Interdiscip Rev RNA* 2012, **3**:429-441.

10. von Loeffelholz O, Knoops K, Ariosa A, Zhang X, Karuppasamy M,
• Huard K, Schoehn G, Berger I, Shan SO, Schaffitzel C: **Structural basis of signal sequence surveillance and selection by the SRP-FtsY complex**. *Nat Struct Mol Biol* 2013, **20**:604-610.
Structure of the '*proof-reading*' step in co-translational targeting, revealed by electron cryo-microscopy and biochemical analysis of a reconstituted assembly comprising ribosome nascent chain complex, SRP and SRP receptor, FtsY. The SRP protein subunit Ffh and FtsY were provided as an engineered protein fusion conjoined by an extended glycine/serine rich linker. The resulting protein fusion displayed wild-type activity.

11. von Loeffelholz O, Jiang Q, Ariosa A, Karuppasamy M, Huard K,
• Berger I, Shan SO, Schaffitzel C: **Ribosome-SRP-FtsY cotranslational targeting complex in the closed state**. *Proc Natl Acad Sci U S A* 2015, **112**:3943-3948.
High-resolution electron cryo-microscopy study revealing the structure and mechanism of the '*closed*' step in bacterial co-translational targeting. The emerging signal sequence bound to the Ffh M-domain was observed in near-atomic detail.

12. McGinty RK, Henrici RC, Tan S: **Crystal structure of the PRC1
•• ubiquitylation module bound to the nucleosome**. *Nature* 2014, **514**:591-596.
3.0 Å crystal structure determination of the first histone-modifying enzyme complex, the PRC1 deubiquitylation module, bound to a nucleo-some. Single-chain engineering was required to stabilize the enzyme complex for crystallization.

13. Desmet J, Verstraete K, Bloch Y, Lorent E, Wen Y, Devreese B, Vandenbroucke K, Loverix S, Hettmann T, Deroo S *et al.*: **Structural basis of IL-23 antagonism by an Alphabody protein scaffold**. *Nat Commun* 2014, **5**:5237.

14. Hoffmann T, Dougan L: **Single molecule force spectroscopy using polyproteins**. *Chem Soc Rev* 2012, **41**:4781-4796.

15. Shin G, Yost SA, Miller MT, Elrod EJ, Grakoui A, Marcotrigiano J: **Structural and functional insights into alphavirus polyprotein processing and pathogenesis**. *Proc Natl Acad Sci U S A* 2012, **109**:16534-16539.

16. Yost SA, Marcotrigiano J: **Viral precursor polyproteins: keys of regulation from replication to maturation**. *Curr Opin Virol* 2013, **3**:137-142.

17. Kostyuchenko VA, Zhang Q, Tan JL, Ng TS, Lok SM: **Immature and mature dengue serotype 1 virus structures provide insight into the maturation process**. *J Virol* 2013, **87**:7700-7707.

18. Miller M: **The early years of retroviral protease crystal structures**. *Biopolymers* 2010, **94**:521-529.

19. Noble CG, Shi PY: **Structural biology of dengue virus enzymes: towards rational design of therapeutics**. *Antiviral Res* 2012, **96**:115-126.

20. Konvalinka J, Kräusslich HG, Müller B: **Retroviral proteases and their roles in virion maturation**. *Virology* 2015. pii: S0042-6822(15)00157-9.

21. Bell NM, Lever AM: **HIV Gag polyprotein: processing and early viral particle assembly**. *Trends Microbiol* 2013, **21**:136-144.

22. Briggs JA, Kräusslich HG: **The molecular architecture of HIV**. *J Mol Biol* 2011, **410**:491-500.

23. Sundquist WI, Kräusslich HG: **HIV-1 assembly, budding, and maturation**. *Cold Spring Harb Perspect Med* 2012, **2**:a006924.

24. Bharat TA, Davey NE, Ulbrich P, Riches JD, de Marco A, Rumlova M, Sachse C, Ruml T, Briggs JA: **Structure of the immature retroviral capsid at 8 Å resolution by cryo-electron microscopy**. *Nature* 2012, **487**:385-389.

25. de Marco A, Heuser AM, Glass B, Kräusslich HG, Müller B, Briggs JA: **Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity**. *J Virol* 2012, **86**:13708-13716.

26. Bharat TA, Castillo Menendez LR, Hagen WJ, Lux V, Igonet S, Schorb M, Schur FK, Kräusslich HG, Briggs JA: **Cryo-electron microscopy of tubular arrays of HIV-1 Gag resolves structures essential for immature virus assembly**. *Proc Natl Acad Sci U S A* 2014, **111**:8233-8238.

27. Schur FK, Hagen WJ, Rumlová M, Ruml T, Müller B,
•• Kräusslich HG, Briggs JA: **Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution**. *Nature* 2015, **517**:505-508.
The structure of the immature HIV-1 capsid in intact virus particles formed by HIV Gag precursor polyprotein was elucidated by electron cryo-tomography and sub-tomogram averaging methods, revealing unexpected structural plasticity.

28. Kennedy MW: **The polyprotein lipid binding proteins of nematodes**. *Biochim Biophys Acta* 2000, **1476**:149-164.

29. McAleer MA, Irvine AD: **The multifunctional role of filaggrin in allergic skin disease**. *J Allergy Clin Immunol* 2013, **131**:280-291.

30. Harding CR, Aho S, Bosko CA: **Filaggrin — revisited**. *Int J Cosmet Sci* 2013, **35**:412-423.

31. Meenan NA, Ball G, Bromek K, Uhrín D, Cooper A, Kennedy MW, Smith BO: **Solution structure of a repeated unit of the ABA-1 nematode polyprotein allergen of Ascaris reveals a novel fold and two discrete lipid-binding sites**. *PLoS Negl Trop Dis* 2011, **5**:e1040.

32. Smith JL, Skiniotis G, Sherman DH: **Architecture of the polyketide synthase module: surprises from electron cryo-microscopy**. *Curr Opin Struct Biol* 2015, **31**:9-19.

33. Bukhari HS, Jakob RP, Maier T: **Evolutionary origins of the multienzyme architecture of giant fungal fatty acid synthase**. *Structure* 2014, **22**:1775-1785.

34. Tran TH, Hsiao YS, Jo J, Chou CY, Dietrich LE, Walz T, Tong L: **Structure and function of a single-chain, multi-domain long-chain acyl-CoA carboxylase**. *Nature* 2015, **518**:120-124.

35. Eisfeld AJ, Neumann G, Kawaoka Y: **At the centre: influenza A virus ribonucleoproteins**. *Nat Rev Microbiol* 2015, **13**:28-41.

36. Ortín J, Martín-Benito J: **The RNA synthesis machinery of negative-stranded RNA viruses**. *Virology* 2015. pii: S0042-6822(15)00154-3.

37. Bieniossek C, Imasaki T, Takagi Y, Berger I: **MultiBac: expanding the research toolbox for multiprotein complexes**. *Trends Biochem Sci* 2012, **37**:49-57.

38. Thorsen TS, Matt R, Weis WI, Kobilka BK: **Modified T4 lysozyme fusion proteins facilitate G protein-coupled receptor crystallogenesis**. *Structure* 2014, **22**:1657-1664.

39. Yin J, Mobarec JC, Kolb P, Rosenbaum DM: **Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant**. *Nature* 2015, **519**:247-250.

40. Saraogi I, Shan SO: **Co-translational protein targeting to the bacterial membrane**. *Biochim Biophys Acta* 2014, **1843**:1433-1441.

41. Zhang X, Shan SO: **Fidelity of cotranslational protein targeting by the signal recognition particle**. *Annu Rev Biophys* 2014, **43**:381-408.

42. Schaffitzel C, Oswald M, Berger I, Ishikawa T, Abrahams JP, Koerten HK, Koning RI, Ban N: **Structure of the *E. coli* signal recognition particle bound to a translating ribosome**. *Nature* 2006, **444**:503-506.

43. Zhang X, Schaffitzel C, Ban N, Shan SO: **Multiple conformational switches in a GTPase complex control co-translational protein targeting**. *Proc Natl Acad Sci U S A* 2009, **106**:1754-1759.

44. Estrozi LF, Boehringer D, Shan SO, Ban N, Schaffitzel C: **Cryo-EM structure of the *E. coli* translating ribosome in complex with SRP and its receptor**. *Nat Struct Mol Biol* 2011, **18**:88-90.

45. Scholl ZN, Li Q, Marszalek PE: **Single molecule mechanical
• manipulation for studying biological properties of proteins, DNA, and sugars**. *Wiley Interdiscip Rev Nanomed Biotechnol* 2014, **6**:211-229.
Authoritative overview article presenting the state-of-the-art of single-molecule structure analysis of biological samples (proteins, nucleic acid, sugars) by using atomic force microscopy.

46. Krüger M, Linke WA: **The giant protein titin: a regulatory node that integrates myocyte signaling pathways**. *J Biol Chem* 2011, **286**:9905-9912.

47. Bang ML, Centner T, Fornoff F, Geach AJ, Gotthardt M, McNabb M, Witt CC, Labeit D, Gregorio CC, Granzier H, Labeit S: **The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system**. *Circ Res* 2001, **89**:1065-1072.

48. Williams PM, Fowler SB, Best RB, Toca-Herrera JL, Scott KA, Steward A, Clarke J: **Hidden complexity in the mechanical properties of titin**. *Nature* 2003, **422**:446-449.

49. Valbuena A, Oroz J, Vera AM, Gimeno A, Gómez-Herrero J, Carrión-Vázquez M: **Quasi-simultaneous imaging/pulling analysis of single polyprotein molecules by atomic force microscopy**. *Rev Sci Instrum* 2007, **78**:113707.

50. Carrion-Vazquez M, Oberhauser AF, Fisher TE, Marszalek PE, Li H, Fernandez JM: **Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering**. *Prog Biophys Mol Biol* 2000, **74**:63-91.

51. Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM: **The mechanical stability of ubiquitin is linkage dependent**. *Nat Struct Biol* 2003, **10**:738-743.

52. He C, Genchev GZ, Lu H, Li H: **Mechanically untying a protein slipknot: multiple pathways revealed by force spectroscopy and steered molecular dynamics simulations**. *J Am Chem Soc* 2012, **134**:10428-10435.

53. Cao Y, Li H: **Polyprotein of GB1 is an ideal artificial elastomeric protein**. *Nat Mater* 2007, **6**:109-114.

54. Li YD, Lamour G, Gsponer J, Zheng P, Li H: **The molecular mechanism underlying mechanical anisotropy of the protein GB1**. *Biophys J* 2012, **103**:2361-2368.