CHEMISTRY
A European Journal

A Journal of
ChemPubSoc Europe

Accepted Article

Title: Rational Design of Supramolecular Dynamic Protein Assemblies Using a Micelle-Assisted Activity-based Protein Labeling Technology

Authors: Britto Sandanaraj, Mohan Reddy, Pavankumar Bhandari, Sugam Kumar, and Vinod Aswal

To be cited as: *Chem. Eur. J.* 10.1002/chem.201802824

Link to VoR: http://dx.doi.org/10.1002/chem.201802824

Supported by
ACES

WILEY-VCH

10.1002/chem.201802824

Rational Design of Supramolecular Dynamic Protein Assemblies Using a Micelle-Assisted Activity-based Protein Labeling Technology

Britto S Sandanaraj[1#], Mullapudi Mohan Reddy[1*] and Pavankumar Janardhan Bhandari[1*], Sugam Kumar[2] and Vinod K. Aswal[2]

[1]Department of Chemistry & Biology, Indian Institute of Science Education and Research (IISER), Pune, 411 008, India

[2]Solid State Physics Division, Bhabha Atomic Research Centre (BARC), Mumbai, 400085, India

* These authors contributed equally to this work

e-mail: sandanaraj.britto@iiserpune.ac.in

Abstract

Self-assembly of proteins into higher order super-structures is ubiquitous in biological systems. Genetic methods comprising both computational and rational design strategies are emerging as powerful methods for design of synthetic protein complexes with high accuracy and fidelity. Although useful, most of the reported protein complexes lack dynamic behavior that may limit their potential applications. On the contrary, protein engineering using chemical strategies offers excellent possibilities for the design of protein complexes with stimuli-responsive functions and adaptive behavior. However, designs based on chemical strategies are not accurate and therefore yield polydisperse samples that are difficult to characterize. Here, we describe simple design principles for the construction of protein complexes through supramolecular chemical strategy. Micelle-assisted activity-based protein labeling technology has been developed to synthesize library of facially amphiphilic synthetic proteins, which self-assemble to make protein complexes through hydrophobic interaction. The proposed methodology is amenable for the synthesis of protein complex libraries with molecular weights and dimensions comparable to naturally occurring protein cages. The designed protein complexes display very rich structural diversity, oligomeric states, sizes and surface charges that can be engineered through macromolecular design. The broad utility of this method is demonstrated by design of most sophisticated stimuli-responsive systems that can be

programmed to assemble/disassemble in a reversible/irreversible fashion using pH or light as a trigger.

Introduction

Design of protein super-structure through bottom-up approach has gained enormous interest in the recent years[1-5]. Genetic engineering[6-8] and supramolecular chemical strategies[9-12] are two complementary technologies that are widely used for the design strategy. The structures of protein complexes made through genetic methods have been determined to a near-atomic resolution using X-ray crystallography[13] and cryo-electron microscopy (cryo-EM)[14] . In addition, small-angle X-ray scattering (SAXS) technique has also been used to determine low-resolution structures[15,16]. Despite these advantages, genetic methods are prone to high failure rate and often yield structures not intended by the design[16-18]. Most importantly, a majority of the designs produced only static structures; integration of dynamic behavior has yet to be explored. In addition, the standard genetic method is restricted to a small set of building blocks - 20 canonical amino acids - that are used for the design purpose.

Natural protein complexes evolved to perform advanced functions *in vivo*. They achieve this remarkable feat through several reversible/irreversible post-translational modifications, thus converting the static structure into dynamic one and *vice versa*[19]. Hence, access to more building blocks (chemical entities) is essential to realize dynamic/adaptive behavior. While the incorporation of several unnatural amino acids onto protein through modified genetic method was demonstrated long back[20], to date there are only few reports on the use of unnatural amino acids for design of *de novo* protein complex[21]. Alternatively, protein engineering using chemical strategies provide opportunities for the design of stimuli-responsive protein complexes because of access to innumerable building blocks[22,23]. Despite this advantage, most of the designs yield polydisperse samples and therefore lack detailed biophysical characterization. Hence, the underlying design rules are still elusive. Here, we report design and development of micelle-assisted activity-based protein labeling technology and its utility in the synthesis of libraries of protein complexes with defined size, surface charge and oligomeric state. In addition, the same method is utilized for design of protein complex with stimuli-responsive function/adaptive behavior.

Results

Macromolecular design

The inspiration for our design came from structural, biophysical and self-assembly studies of naturally occurring facially amphiphilic hydrophobin proteins[24] (Fig. 1a). In particular, a recent report of the high-resolution structure of a bacterial hydrophobin reveals that it exists as a decamer[25] where the hydrophobic face of the protein is shielded away from the water molecules by forming a dense hydrophobic core and hydrophilic face is presented outside, which mimics protein cages[26]. This study gave us a clue for the "minimal design" of facially amphiphilic semi-synthetic protein. The macromolecular design has three core structural elements: (i) hydrophilic globular protein, (ii) flexible hydrophilic linker, and (iii) hydrophobic tail (Fig. 1a). We envisioned that the designed protein would self-assemble into complex architecture, primarily driven by strong hydrophobic interaction, which is one of the major driving force for the formation of supramolecular protein assembly in natural[27] and artificial systems[28]. The basic tenet of our hypothesis is that the attributes of a protein complex (size, oligomeric state, surface charge and shape) would be dictated by choice of a core structural element of macromolecular design[29].

The modular strategy for synthesis of libraries of facially amphiphilic semi-synthetic protein is shown in Fig. 1f (Supplementary Results, Supplementary Scheme. 22). Activity-based protein labelling method[30] was chosen for following reasons: (i) site-specific modification of an active-site residue of serine proteases; (ii) fast reaction kinetics and (iii) serine proteases encompasses a large set of proteins and therefore provide opportunity to make large library (library I, Fig. 1c, Supplementary Fig. 1a). In addition, monodisperse oligoethylene glycol (defined repeating units) is used as spacer moiety because of its water solubility and protein repellent properties and most importantly, spacer length could be systematically engineered to generate a unique library of protein conjugates (library II, Fig. 1d, Supplementary Fig. 1b). Finally, tail hydrophobicity can be systematically varied by modifying tail length/branching utilizing click chemistry strategy[31] (library III, Fig. 1e, Supplementary Fig. 1c)

As expected, the solubility of the probe (FP-OEG-C12-3T, Supplementary Fig. 2a) posed a challenge in terms of choosing an optimum condition to carry out protein labelling reaction in 100% aqueous medium. Several reaction conditions were attempted but none of them was

successful. Considering the amphiphilic nature of the probe, we hypothesized that it could be solubilized in the presence of a small-molecule micelle. A reasonable assumption is that the amphiphilic probe would make a mixed micelle in the presence of an another suitable surfactant (Fig. 1b), where the tail component of a probe would be part of the hydrophobic micellar interior and hydrophilic end (oligoethylene glycol with a terminal fluorophosphonate group) would be available for active-site labelling of a serine protease[32] (Fig. 1b). To test this hypothesis, extensive surfactant screening was carried out. Triton X-100 was found to be best, as it satisfies all the requirements for protein labelling reaction. As a test case, micelle-assisted labelling of trypsin (serine protease) by FP-OEG-C12-3T was first attempted (Fig. 1b and 1f). The reaction was carried out in the presence of 10X critical micelle concentration (CMC) of triton X-100 at room temperature pH 7.4. The extent of protein modification was monitored using MALDI-ToF at different time points. After 12 h, a major peak for trypsin conjugate (Try-OEG-C12-3T) at 24,453 Da and a minor peak at 23,272 Da for native trypsin were observed (Fig. 2a). Motivated by these results, we set out to test whether this method could be widely applicable to other serine proteases as well (chymotrypsin, proteinase K, and subtilisin) to further validate the method (Fig. 1c). FP-OEG-C12-3T labeled all the serine proteases as evident from MALDI-ToF results (Fig. 2a). Encouraged by these results, we focused our efforts on the design and synthesis of 4 additional AABPs (AABP library I, Supplementary Fig. 2a, Supplementary Scheme. 22) having different linker length (di-, tetra-, dodeca- and cetyl-ethylene glycol). All the probes labelled trypsin except FP-DEG-C12-3T because of sub-optimal linker length (Fig. 2b, Supplementary Fig. 2a). Finally, to demonstrate the systematic variation of hydrophobic tail, 7 new AABPs were synthesized (AABP library II, Supplementary Fig. 2b, Fig. 1e, Supplementary Scheme. 22) with varying alkyl chain length (hexyl, dodecyl, and octadecyl abbreviated as C6/C12/C18) and branching (1-tail/2-tails/3-tails abbreviated as 1T/2T/3T). All the probes including a most hydrophobic probe FP-OEG-C18-3T (Fig. 2c, Supplementary Fig. 2b) were reactive as evident from MALDI-ToF data.

The next major challenge in this study is the scale-up and purification of a protein conjugate as it resembles natural hydrophobin/membrane proteins (Fig. 1a) in terms of hydropathy index[24]. Synthesis of a Try-OEG-C12-3T conjugate was performed in large scale (200 mg scale). Triton X-100 and unreacted probe in the reaction mixture were successfully removed using ion-exchange chromatography (IEX) method (Fig. 2d). Since Try-OEG-C12-3T

is facially amphiphilic, we hypothesized that increasing ionic strength of the solution would promote self-sorting of Try-OEG-C12-3T through strong hydrophobic interaction but not the native trypsin, which would aid in purification by the size-exclusion chromatography (SEC) method[33]. As expected, elution of protein conjugate Try-OEG-C12-3T (1 M NaCl phosphate buffer pH 7.4) in void volume suggests formation of higher-order complex, while monomeric native trypsin eluted at 65 mL in SEC experiment (Fig. 2e). Finally, NaCl was removed by standard desalting method (Fig. 2f). After three-step purification, a single peak for the purified conjugate at 24,453 Da with a very high purity was obtained with an overall yield of 20% (Fig. 1g). Most importantly, the same method was used for synthesis and purification of all the protein conjugates in the library I to III except for Try-OEG-C18-3T (Fig. 1b, 1f, 2d-2f, 2g-2i, Supplementary Scheme. 1-16).

Programmed self-assembly

In order to test the self-assembling ability of designed protein conjugate (Fig. 1g-i), detailed self-assembly studies were carried out using a variety of techniques which include analytical SEC, dynamic light scattering (DLS), small-angle neutron scattering (SANS), size-exclusion chromatography coupled with multi-angle laser light scattering (SEC-MALS) and zeta potential measurements. First, the effect of protein structure (Fig. 1c, Supplementary Fig. 7), size (Supplementary Fig. 4, Supplementary Table. 2a) and surface charge (Fig. 3m, Supplementary Fig. 6) on self-assembly was studied. For this purpose, we focused our attention on protein library I (Fig. 1c and 1g, Supplementary Fig. 1a). All the protein conjugates in this library have an identical linker (OEG) and a hydrophobic tail (C12-3T) except the protein is systematically varied in the macromolecular design. Analytical SEC studies revealed that all four conjugates eluted at lower elution volumes (11.5 to 10.9 mL) indicative of large protein complexes (Fig. 3a, Supplementary Fig. 3 and Supplementary Table. 1). In order to further understand various attributes of protein complexes such as the radius of gyration (R_g), oligomeric state and shape (σ), detailed SANS studies were carried out[34]. Both trypsin and chymotrypsin complexes exist as a 10-mer with molecular mass of 245 kDa whereas proteinase K conjugate self-assemble into a 22-mer with molecular mass of 661 kDa (Fig. 3d, Supplementary Fig. 4 and Supplementary Table. 2b). However, SANS experiment of subtilisin complex failed as sample precipitated during the course of SANS measurement (12 h). Nevertheless, it was evident from

the SEC and DLS results that subtilisin conjugate is also capable of forming higher-order protein complex. The R_g of protein complexes were obtained from Guinier region of SANS data. The R_g of trypsin, chymotrypsin, and proteinase K complexes are 41 Å, 40 Å and 53 Å, respectively (Table. 1, Supplementary Table. 2b). The hydrodynamic radii (R_h) were determined by DLS studies (Fig. 3g, Supplementary Fig. 5 and Supplementary Table. 3). R_h data showed an interesting trend similar to R_g data (Table. 1, Supplementary Table. 3). The shape factor σ (R_g/R_h) were obtained from SANS and DLS data. The obtained σ values of all of the protein complexes are in the range of 0.60-0.79 (Table. 1, Supplementary Table. 3), indicating that the designed protein complexes exist as a spherical object similar to both naturally occurring[26,35] and synthetic protein cages[7,14]. To confirm SANS results, the molecular mass of Try-OEG-C12-3T was verified using SEC-MALS and value of 276 kDa (Fig. 3j) obtained was in agreement with SANS results (245 kDa, Table. 1). Zeta potential measurements were carried out to investigate the surface charge of designed protein complexes (Supplementary Data Fig. 6 and Table. 1). Both trypsin and chymotrypsin complexes are highly cationic (+16.5 mV and +8.7 mV), proteinase K complex is slightly anionic (-3.2 mV) and subtilisin complex is close to neutral (-0.7 mV).

In order to understand the effect of linker length on protein self-assembly, we focused our attention on protein library II (Fig. 1d and 1h, Supplementary Fig. 1b). All the protein complexes are monodisperse as evident from SEC and SANS results (Fig. 3b and 3e, Supplementary Fig. 3 and 4). As predicted, R_g of protein complex increased with increase in spacer length. The size of Try-TEG-C12-3T is 32 Å whereas Try-CEG-C12-3T is 49 Å (Table. 1). SANS studies also revealed that molecular mass of Try-TEG-C12-3T with the shortest linker (TEG) is 243 kDa whereas Try-CEG-C12-3T with the longest linker (CEG) is 372 kDa (Fig. 3e, Supplementary Fig. 4 and Supplementary Table. 2b). DLS results also showed a similar trend (Fig. 3h, Supplementary Fig. 5 and Table. 1). Molecular mass of Try-CEG-C12-3T was independently verified by SEC-MALS. The obtained value (400 kDa) matches with SANS results (Fig. 3k, Table. 1). As expected, all the protein complexes are cationic and zeta potential values are between +4.6 to +16.5 mV (Supplementary Fig. 6 and Table. 1).

Finally, the influence of tail hydrophobicity on protein conjugate self-assembly was systematically studied (protein library III, Fig. 1e and 1i, Supplementary Fig. 1c). As predicted,

negative design Try-OEG-C6-3T with short hydrophobic tail predominantly exists in monomeric form at 200 μM concentration (Supplementary Fig. 3). However, Try-OEG-C6-3T self-assembled to trimer at higher concentration (800 μM) as evident from SANS and DLS results (Supplementary Fig. 4 and 5, Table.1 and Supplementary Table. 2b and 3). Similarly, Try-OEG-C12-1T predominantly exists as monomer during SEC experiment, whereas same conjugate behaved as higher-order complex when analyzed by SANS and DLS experiments. This is primarily attributed to result of dilution effect during SEC experiment (Supplementary Fig. 4 and 5, Table.1 and Supplementary Table. 2b and 3). As expected, protein conjugates with sufficient tail length and hydrophobicity (Try-OEG-C12-2T/3T and Try-OEG-C18-2T) self-assemble into large protein complexes (Fig. 3c, 3f and 3i, Table.1, Supplementary Fig. 3, 4 and 5, Supplementary Table. 1, 2b and 3). Remarkably, protein conjugates with just 1 hydrophobic tail, Try-OEG-C18-1T, self-assemble into large protein complex as evident from SEC studies (Fig. 3c, Supplementary Fig. 3). To further corroborate SANS results, the molecular mass of Try-OEG-C12-2T was independently verified using SEC-MALS and the result obtained (261 kDa) matches with SANS results (Fig. 3l, Table. 1)

In order to probe the secondary and tertiary structure of designed protein complexes, detailed biophysical studies were carried out using fluorescence[36] and CD spectroscopy[37]. Fluorescence studies reveal there was no shift in tryptophan emission maximum which indicates the 3D structure of a protein is preserved in the complex state (Supplementary Fig. 7a-7c). However, CD indicates a small perturbation in the secondary structure of some conjugates (Supplementary Fig. 7d-7f). A detailed investigation is underway and the results will be reported in due course.

Design of stimuli-responsive supramolecular protein complex

In order to test whether change in pH has any effect on self-assembly of designed protein conjugate[7,14,35], SEC studies of Chy-OEG-C12-3T were carried out at different pH. The elution volume of protein complex was shifted to lower side by 0.4 mL at pH 10. This result indicates the size of protein complex enlarged with increase in pH. Surprisingly, decreasing solution pH to 5 resulted in the shrinkage of the size with respect to dimension of protein complex at pH 7.4 (Supplementary Fig. 8).

In order to demonstrate the versatility of our method, photo-responsive semi-synthetic protein was designed (Fig. 4a, Supplementary Fig. 1d and 2c). Design principle is same as

previous design except a photo-responsive group (2-nitrobenzyl derivative) is installed between octaethylene glycol (hydrophilic spacer) and a hydrophobic tail (C12-2T) (Fig. 4a, Supplementary Fig. 2c). Our hypothesis is upon photo-irradiation; cleavage of a 2-nitrobenzyl group[38] would lead to a separation of hydrophobic domain from the rest of protein conjugate (hydrophilic globular domain) that would result in dis-assembly of a protein complex into constitutive monomers (Fig. 4a), as the assembly is kept intact by strong hydrophobic interaction. Next, time-dependent photo-chemical reactions were carried out on the photo-responsive protein conjugate (50 μM) and the samples were analyzed by analytical SEC. As expected, at time zero, the protein conjugate eluted at 11.5 mL indicative of higher-order complex (8-mer). However, at intermediate time intervals (2 mins to 45 mins), two peaks were observed, first peak at 11.5 mL and the second peak at 17.0 mL corresponding to a protein complex and a cleaved protein conjugate, respectively (Fig. 4b). The concentration of protein complex gradually decreased with a concomitant increase in concentration of a cleaved monomeric protein as function of sample exposure time to UV light (Fig. 4b). In parallel, detailed MALDI-ToF analyses were carried out for reaction mixtures at different time points (Supplementary Fig. 9). At time zero, a single peak at 24,362 Da for the full conjugate Try-OEG-NB-C12-3T was observed (Fig. 4c). At 2 minutes, two peaks were seen, one major peak for the full conjugate and a very small peak for a cleaved protein Try-OEG-NH_2 at 23,723 Da (Fig. 4c). With increase in UV light exposure, an appreciable decrease in peak height for the full conjugate and concomitant increase for a cleaved protein were observed. At 90 mins, complete conversion of full conjugate into a cleaved protein was achieved (Fig. 4c and Supplementary Fig. 9). The peak at 10.5 mL did not vanish completely even after prolonged irradiation time (90 mins). MALDI-ToF results revealed that sample did not contain trace amount of full conjugate or a cleaved protein. The width of the peak at 11.5 mL was slightly increased and elution volume moved to lower side (11.3 mL), which suggests that the released amphiphilic small molecule (2-nitrosobenzaldehyde derivative) forms a micelle-like structure. Attempt to characterize this compound by ^{1}H-NMR and mass spec was not successful. We have also carried out the control experiment by exposing Try-OEG-NB-C12-3T to the visible light, as expected there was no disassembly (Supplementary Fig. X). Similarly, protein complex Try-OEG-C12-3T is stable upon exposure to UV light. These control experiments prove that the disassembly of Try-OEG-NB-C12-3T protein complex is because of selective cleavage of nitro-benzyl group.

Discussion

Accurate design of supramolecular protein complex through bottom-up approach is at the forefront of bio-molecular engineering efforts[6-8]. Protein engineering utilizing chemical technologies offer remarkable advantages over genetic method as they are not restricted to standard 20 canonical amino acids[9-12]. However, most of the reported chemical methods yield polydisperse sample, restricted to few proteins and lacks modular approach[9]. Our results demonstrate that intelligent rational design based on chemical method can yield supramolecular protein complex with narrow polydispersity. The protein self-assembly is mediated by strong favorable interaction between hydrophobic tails opposed by electrostatic repulsion between protein head group containing point charges (presence of acidic and basic amino acid at physiological pH and ionic strength)[39]. This mechanism is further supported by the fact that protein conjugate with short hydrophobic tail (Try-OEG-C6-2T) failed to self-assemble. This is because electrostatic repulsion between proteins would be much higher compared to weak hydrophobic interaction in the assembled state, which would be energetically unfavorable[39]. In addition, oligomeric state of ProK-OEG-C12-3T is quiet different compared to Try-OEG-C12-3T or Chy-OEG-C12-3T. This result can be explained based on the surface charge of proteins. Proteinase K surface charge is close to neutral and hence the complex can accommodate more subunits without much electrostatic repulsion. The most striking features of our method are (i) simple design, (ii) high success rate, (iii) rapid generation of library of supramolecular protein complexes with unique structural diversity, and (iv) the size, oligomeric state and surface charge of protein complex can be precisely engineered through macromolecular design.

Accurate design of photo-responsive monodisperse protein complexes is a fascinating idea. However, to date, there are no reports based on this concept. This kind of design is very difficult to achieve through standard genetic method as 20 naturally occurring amino acids have not evolved to respond to light[1,2]. Amber-codon suppression method could be used, however, that method is cumbersome and engineering-intensive[20]. Our results demonstrate that accurate design of most sophisticated photo-responsive molecular machine can be achieved through chemical method. The programmed dis-assembly of protein complex is achieved upon exposure to light. These results also emphasize the proposed mechanism that favorable interaction between the hydrophobic tails is the major driving force for protein conjugate to self-assemble.

The cleavage of hydrophobic tail upon photo-irradiation leads to complex destabilization because of loss of strong hydrophobic interactions.

The simple design principle described here makes it easy to extend this methodology to large number of enzyme families and proteins, for example amphiphilic activity-based probe equipped with appropriate reactive groups[40] can be used to label large number to enzymes (cysteine, aspartic and metallo proteases, etc) using similar approach described here. Similarly, this methodology can be extended for labeling of N-terminal amino acid of any protein[41]. In addition, modular fashion of our synthetic strategy provides opportunity to introduce chemically diverse linkers to make stimuli-responsive protein complexes which can respond to ionic strength, temperature, enzyme, pH, redox, photo and electro-trigger. Most importantly, hydrophobic tail of a protein conjugate can be systematically varied using monodisperse linear polymers and dendrimers. These modifications provide opportunity for custom-design of protein complexes in the size range 20-100 nm and molecular weights in mega-dalton range. Therefore, this method has immense potential to push the boundaries of bio-molecular engineering efforts[42-46]. We firmly believe this work would inspire genetic methods (both computational and rational approach) for construction of stimuli-responsive protein complexes using unnatural amino acids as building blocks through bottom-up approach. The ability to engineer small molecule coupled with technology to make synthetic protein complex opens up new avenues for accurate design of "smart protein complex". This technology would have huge impact in the area of vaccine design[47], targeted drug delivery[48], *in vivo* diagnostics[49] and synthetic biology[50].

References

1 Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of *de novo* protein design. *Nature* **537**, 320-327 (2016).

2 Yeates, T. O. Geometric principles for designing highly symmetric self-assembling protein nanomaterials. *Annu. Rev. Biophys.* **46**, 23-42 (2017).

3 Salgado, E. N., Radford, R. J. & Tezcan, F. A. Metal-directed protein self-assembly. *Acc. Chem. Res.* **43**, 661-672 (2010).

4 Reynhout, I. C., Cornelissen, J. J. L. M. & Nolte, R. J. M Synthesis of polymer-biohybrids: from small to giant surfactants. *Acc. Chem. Res.* **42**, 681-692 (2009).

5 Luo, Q., Hou, C., Bai, Y., Wang, R. & Liu, J. Protein Assembly: versatile approaches to construct highly ordered nanostructures. *Chem. Rev.* **116**, 13571-13632 (2016).

6 Lai, Y. T., Tsai, K. L., Sawaya, M. R., Asturias, F. J. & Yeates, T. O. Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. *J. Am. Chem. Soc.* **135**, 7738-7743 (2013).

7 King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103-108 (2014).

8 Wörsdörfer, B., Woycechowsky, K. J. & Hilvert, D. Directed evolution of a protein container. *Science* **331**, 589-592 (2011).

9 Velonia, K., Rowan, A. E. & Nolte, R. J. Lipase polystyrene giant amphiphiles. *J. Am. Chem. Soc.* **124**, 4224-4225 (2002).

10 Heredia, K. L. *et al.* In situ preparation of protein-"smart" polymer conjugates with retention of bioactivity. *J. Am. Chem. Soc.* **127**, 16955-16960 (2005).

11 Biswas, S. *et al.* A tubular biocontainer: metal ion-induced 1D assembly of a molecularly engineered chaperonin. *J. Am. Chem. Soc.* **131**, 7556-7557 (2009).

12 Sim, S., Niwa, T., Taguchi, H. & Aida, T. Supramolecular nanotube of chaperonin groel: length control for cellular uptake using single-ring groel mutant as end-capper. *J. Am. Chem. Soc.* **138**, 11152-11155 (2016).

13 King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171-1174 (2012).

14 Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389-394 (2016).

15 Lai, Y. T. *et al.* Designing and defining dynamic protein cage nanoassemblies in solution. *Sci. Adv.* **2**, e1501855 (2016).

16 Kobayashi, N. *et al.* Self-assembling nano-architectures created from a protein nano-building block using an intermolecularly folded dimeric de novo protein. *J. Am. Chem. Soc.* **137**, 11285-11293 (2015).

17 Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl Acad. Sci. USA* **98**, 2217-2221 (2001).

18 Sciore, A. *et al.* Flexible, symmetry-directed approach to assembling protein cages. *Proc. Natl Acad. Sci. USA* 8681–8686 (2016).

19 Bowman, G. D. & Poirier, M. G. Post-translational modifications of histones that influence nucleosome dynamics. *Chem. rev.* **115**, 2274-2295 (2014).

20 Cornish, V. W., Mendel, D. & Schultz, P. G. Probing protein structure and function with an expanded genetic code. *Angew. Chem. Int. Edn. Engl.* **34**, 621-633 (1995).

21 Mills, J. H. *et al.* Computational design of a homotrimeric metalloprotein with a trisbipyridyl core. *Proc. Natl Acad. Sci. USA* **113**, 15012–15017 (2016).

22 Lavigueur, C. *et al.* Thermoresponsive giant biohybrid amphiphiles. *Polym. Chem.* **2**, 333-340 (2011).

23 Xia, Y., Tang, S. & Olsen, B. D. Site-specific conjugation of RAFT polymers to proteins via expressed protein ligation. *Chem. Commun.* **49**, 2566-2568 (2013).

24 Ren, Q., Kwan, A. H. & Sunde, M. Two forms and two faces, multiple states and multiple uses: properties and applications of the self□assembling fungal hydrophobins. *Biopolymers* **100**, 601-612 (2013).

25 Hobley, L. *et al.* BslA is a self-assembling bacterial hydrophobin that coats the *Bacillus subtilis* biofilm. *Proc. Natl Acad. Sci. USA* **110**, 13600-13605 (2013).

26 Theil, E. C. Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. *Annu. Rev. Biochem.* **56**, 289-315 (1987).

27 Simons, K., Helenius, A., Leonard, K., Sarvas, M. & Gething, M. J. Formation of protein micelles from amphiphilic membrane proteins. *Proc. Natl Acad. Sci. USA* 75, 5306–5310 (1978).

28 Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136-139 (2016).

29 Israelachvili, J. N. *Intermolecular and surface forces*. (Academic press, 2011).

30 Liu, Y., Patricelli, M. P. & Cravatt, B. F. Activity-based protein profiling: the serine hydrolases. *Proc. Natl Acad. Sci. USA* **96**, 14694-14699 (1999).

31 Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise huisgen cycloaddition process: copper (I)□catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Edn. Engl.* **114**, 2708-2711 (2002).

32 Bader, B. *et al.* Bioorganic synthesis of lipid-modified proteins for the study of signal transduction. *Nature* **403**, 223-226 (2000).

33 Ceres, P. & Zlotnick, A. Weak protein− protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **41**, 11525-11531 (2002).

34 Jacques, D. A. & Trewhella, J. Small□angle scattering for structural biology-Expanding the frontier while avoiding the pitfalls. *Protein Sci.* **19**, 642-657 (2010).

35 Speir, J. A., Munshi, S., Wang, G., Baker, T. S. & Johnson, J. E. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* **3**, 63-78 (1995).

36 Desie, G., Boens, N. & Schryver F. C. D. Study of the time-resolved tryptophan fluorescence of crystalline. alpha.-chymotrypsin. *Biochemistry* **25**, 8301-8308 (1986).

37 Provencher, S. W. & Gloeckner, J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33-37 (1981).

38 Cameron, J. F. & Frechet, J. M. Photogeneration of organic bases from o-nitrobenzyl-derived carbamates. *J. Am. Chem. Soc.* **113**, 4303-4313 (1991).

39 Kegel, W. K. & van der Schoot, P. Competing hydrophobic and screened-coulomb interactions in hepatitis B virus capsid assembly. *Biophys. J.* **86**, 3905-3913 (2004).

40 Evans, M. J. & Cravatt, B. F. Mechanism-based profiling of enzyme families. *Chem. Rev.* **106**, 3279-3301 (2006).

41 Rosen, C. B. & Francis, M. B. Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol.* **13**, 1094–1101 (2017).

42 Wagenbauer, K. F., Sigl, C. & Dietz, H. Gigadalton-scale shape-programmable DNA assemblies. *Nature* **552**, 78-83 (2017).

43 Ljubetič, A. *et al.* Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. *Nature Biotechnol.* **35**, 1094–1101 (2017).

44 Mou, Y., Yu, J. Y., Wannier, T. M., Guo, C. L. & Mayo, S. L. Computational design of co-assembling protein-DNA nanowires. *Nature* **525**, 230-233 (2015).

45 Mosayebi, M. *et al.* Beyond icosahedral symmetry in packings of proteins in spherical shells. *Proc. Natl Acad. Sci. USA* **114**, 9014-9019 (2017).

46 Garcia-Seisdedos, H., Empereur-Mot, C., Elad, N. & Levy, E. D. Proteins evolve on the edge of supramolecular self-assembly. *Nature* **548,** 244–247 (2017).

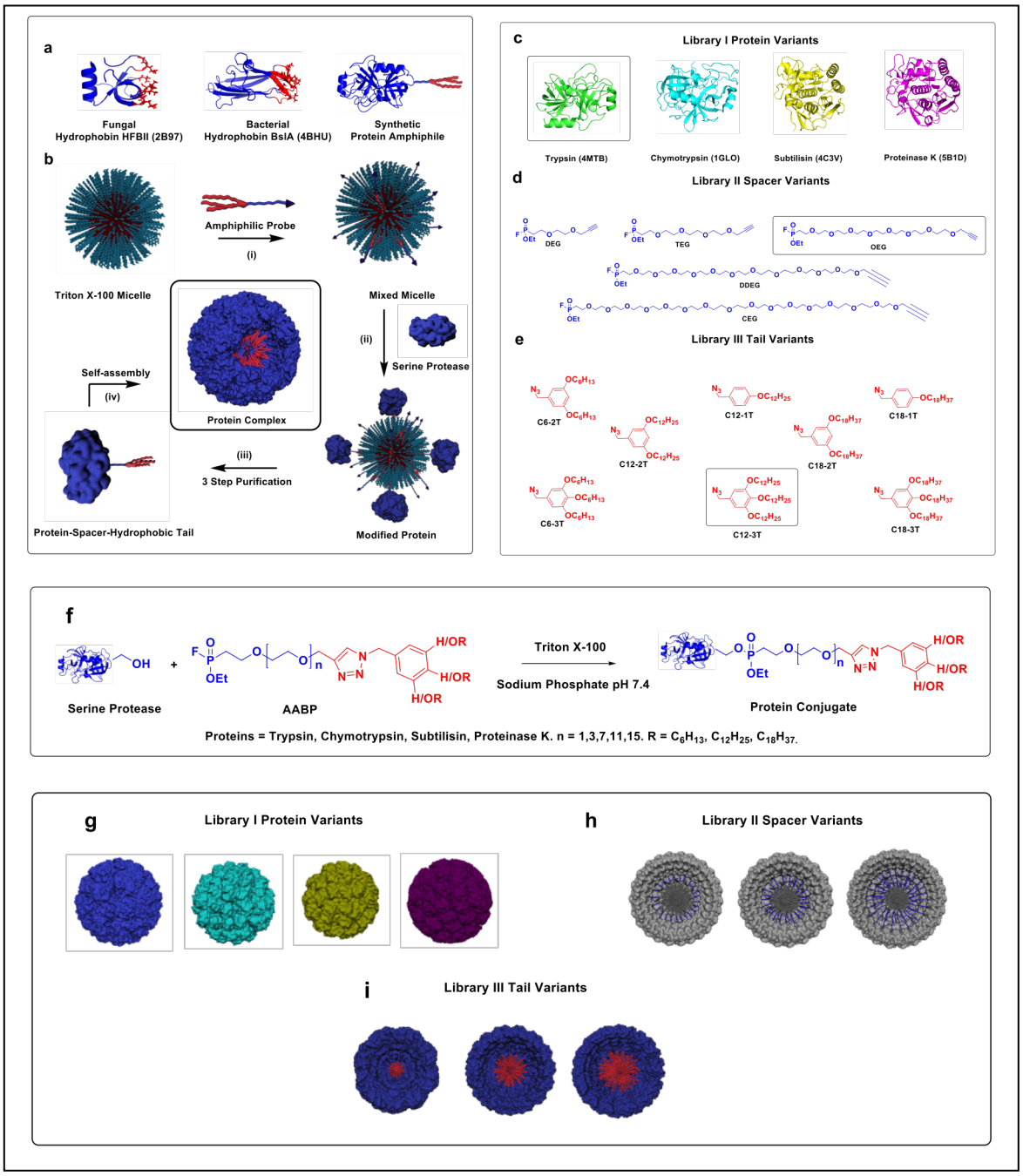
47 Smith, M. L. *et al.* Modified tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. *Virology* **348**, 475-488 (2006).

48 Ma, Y., Nolte, R. J. & Cornelissen, J. J. Virus-based nanocarriers for drug delivery. *Adv. Drug Deliv. Rev.* **64**, 811-825 (2012).

49 Weissleder, R., Nahrendorf, M. & Pittet, M. J. Imaging macrophages with nanoparticles. *Nat. Mater.* **13**, 125-138 (2014).

50 Smanski, M. J. *et al.* Synthetic biology to access and expand nature's chemical diversity. *Nat. Rev. Microbiol.* **14**, 135-149 (2016).

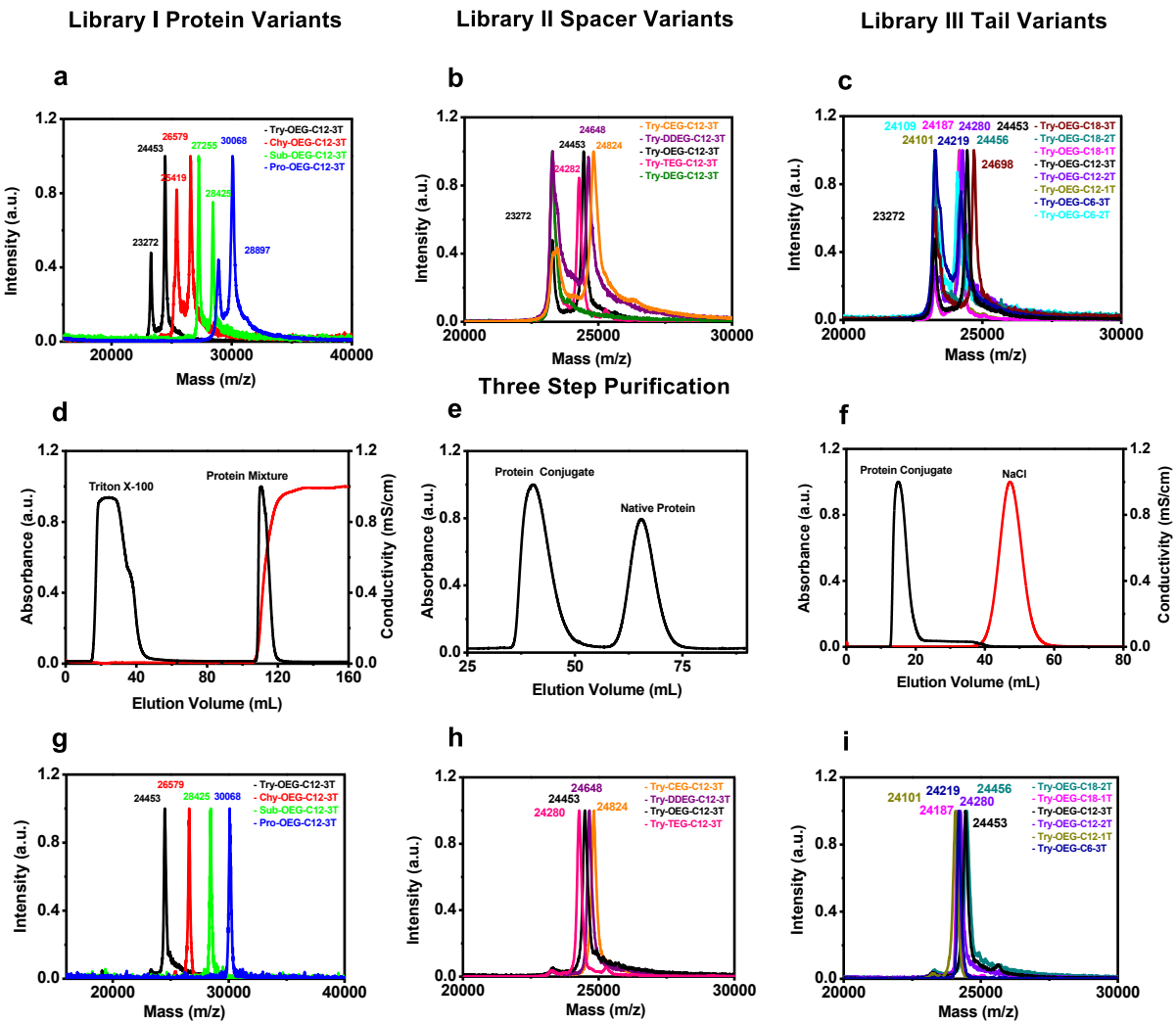
10.1002/chem.201802824



10.1002/chem.201802824

Figure 1 | Overview of the technology. a. Crystal structure of fungal and bacterial hydrophobins (PDB accession number 2B97 and 4BHU) and designed synthetic amphiphilic protein - consists of globular water soluble protein and hydrophobic tail connected by hydrophilic linker (hydrophilic and hydrophobic faces are shown in blue and red, respectively). **b.** Summary of chemical methodology to make synthetic amphiphilic proteins. i. The amphiphilic probe was solubilised in the presence of triton X-100, ii. Micelle-assisted protein labelling rection. iii. Three-step purification to remove triton X-100 and the unreacted protein using IEX, SEC and standard desalting. iv. Programmed protein self-assembly**. c.** Crystal structures of serine proteases used in this study (trypsin, chymotrypsin, subtilisin, proteinase K - PDB accession number 4MTB, 1GLO, 4C3V, 5B1D respectively) **d, e.** Chemical structures of hydrophilic linker and hydrophobic tails. **f.** General scheme for making libraries of synthetic amphiphilic proteins. **g, h, i.** Cartoon representation of protein complexes. **g.** Protein variants (Library I). **h.** Spacer variants (Library II). **i.** Tail variants (Library III).

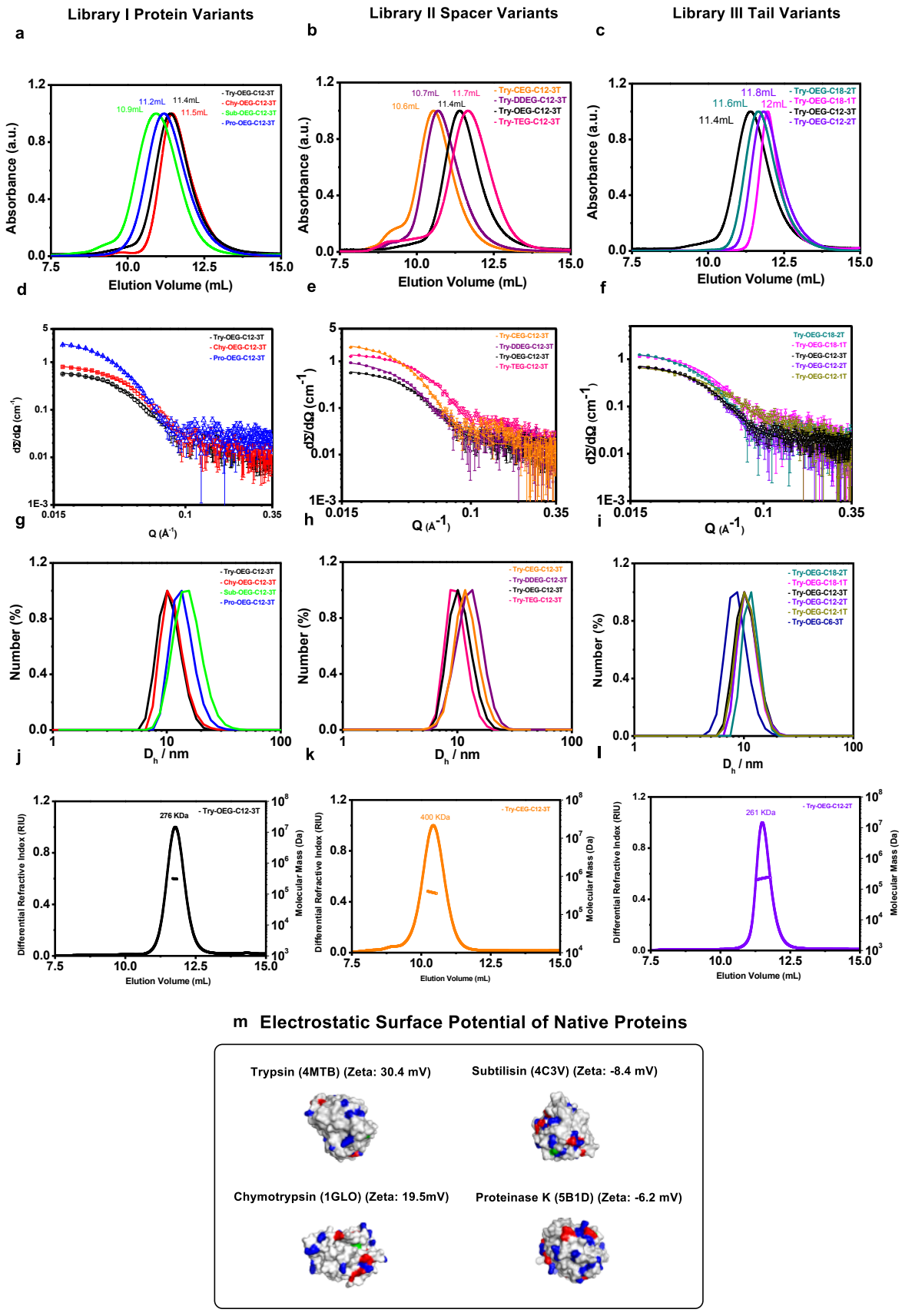
10.1002/chem.201802824



10.1002/chem.201802824

Figure 2 | MALDI-ToF characterization of amphiphilic proteins before and after purification. a, g. Molecular weights of protein variants (Library I). **b, h.** spacer variants (Library II) and **c, i.** tail variants (Library III) before and after purification respectively. To determine molecular weights and to monitor the extent of protein modification, the samples were directly withdrawn from reaction mixture using a pipette. To check the purity of protein conjugate, after size exclusion, 98 μL of the sample from fractions containing the modified proteins, were mixed with 2 μL of triton X-100 (2% or 100 times the CMC) in a separate microcentrifuge tube and vortexed for 4 h. After that, 1:1:1 ratio of protein sample, 2% TFA and final matrix mixture were mixed and monitored using MALDI-ToF MS. **d, e, f.** Represents three-step purification for separating protein conjugate from reaction mixtures. **d.** Representative IEX chromatogram showing the separation of of triton X-100 form protein mixture (native and modified protein). **e.** Representative SEC chromatogram showing the separation of native and modified proteins. The modified protein appears larger in size as evident from elution volume, because of complex formation, **f.** Representative desalting chromatogram, shows the removal of salt from the purified conjugates.

10.1002/chem.201802824



10.1002/chem.201802824

Figure 3 | Self-assembly studies of protein complexes. a, b, c. SEC data - The molecular weights of the protein complexes were calculated from elution volumes of standard proteins. SEC runs were performed in Superdex-200 10/300 GL in 50 mM sodium phosphate pH 7.4, 1 M NaCl with 0.25 mL/min as the flow rate. The concentration and volume of protein samples were 5 mg/mL and 0.5 mL respectively. **d, e, f.** SANS data - For SANS measurements, the samples (5 mg/mL) were dissolved in 50 mM sodium phosphate pH 7.4 prepared in D_2O, in order to minimize the incoherent scattering and to increase contrast. The temperature was kept fixed at 30 □C during the measurements. **g, h, i.** DLS data - Measurements were carried out at 5 mg/mL concentration in 50 mM sodium phosphate pH 7.4. 1 mL of sample was taken in disposable polystyrene cells and the mean size of the complexes was measured at 90º scattering angle. **j, k, l.** SEC-MALS Data - Molecular weight determination of protein complexes using SEC-MALS was performed on a Superdex-200 10/300 GL column connected to an Agilent HPLC system equipped with the 18-angle light scattering detector and a refractive index detector. 100 µL of the protein samples at concentrations (5 mg/mL) were injected, and the molecular weight was calculated using ASTRA software. **m.** Electrostatic surface potential of native proteins. Negatively charged residues like Asp and Glu are colored red. Positively charged residues like Arg and Lys are colored blue. His residues are colored green and active site serine alone is represented in magenta (site of conjugation).

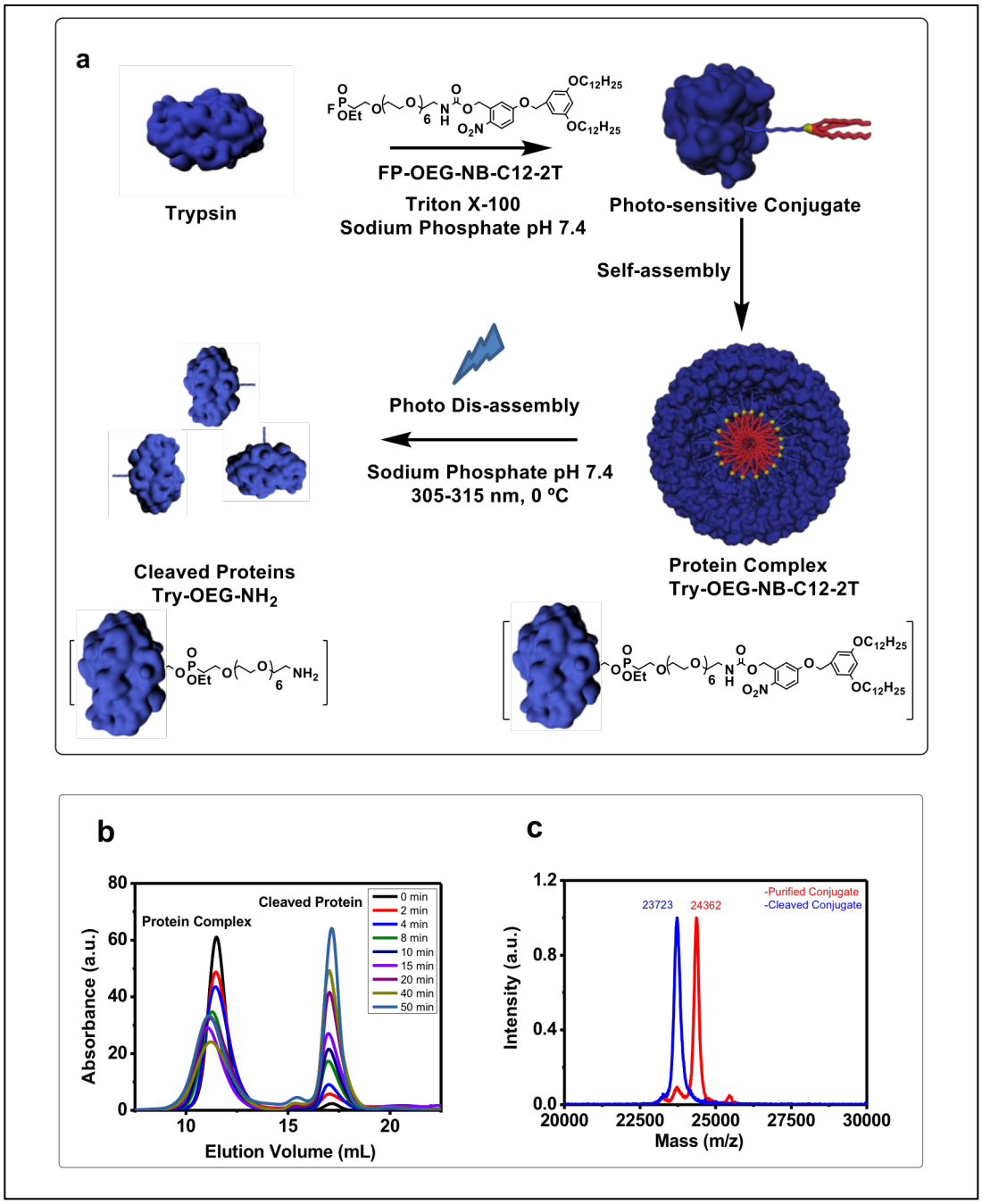
10.1002/chem.201802824

Table 1 | Summary of bio-physical data of protein complexes

Protein Conjugate	Elution Vol SEC (mL)	Mol Wt - SEC (kDa)	Mol Wt SANS (kDa)	Oligomeric state SANS (mer)	R_g from P(r) SANS (Å)	R_h DLS (nm)	Shape Factor ($\sigma = R_g/R_h$)	Zeta potential (mV)
Library I Protein Variants								
Try-OEG-C12-3T	**11.4**	**276**	**245**	**10**	**41**	**5.7**	**0.72**	**16.5**
Chy-OEG-C12-3T	**11.5**	**258**	**266**	**10**	**40**	**5.2**	**0.77**	**8.7**
Pro-OEG-C12-3T	**11.2**	**300**	**661**	**22**	**53**	**7.0**	**0.76**	**-3.2**
Sub-OEG-C12-3T	**10.9**	**339**	**----**	**----**	**---**	**8.8**	**----**	**-0.7**
Library II Spacer Variants								
Try-TEG-C12-3T	**11.7**	**241**	**243**	**10**	**32**	**5.3**	**0.60**	**10.3**
***Try-OEG-C12-3T**	**11.4**	**276**	**245**	**10**	**41**	**5.7**	**0.72**	**16.5**
Try-DDEG-C12-3T	**10.7**	**375**	**320**	**13**	**49**	**7.3**	**0.67**	**9.8**
Try-CEG-C12-3T	**10.6**	**388**	**372**	**15**	**49**	**6.7**	**0.73**	**4.6**
Library III Tail Variants								
Try-OEG-C6-3T	**17.3**	**23**	**73**	**3**	**35**	**4.2**	**-----**	**6.5**
Try-OEG-C12-1T	**17.3**	**23**	**289**	**12**	**37**	**5.0**	**0.74**	**5.5**
Try-OEG-C12-2T	**11.8**	**225**	**267**	**11**	**42**	**5.5**	**0.76**	**7.6**
***Try-OEG-C12-3T**	**11.4**	**276**	**245**	**10**	**41**	**5.7**	**0.72**	**16.5**
Try-OEG-C18-1T	**12.0**	**210**	**411**	**17**	**43**	**5.5**	**0.78**	**7.1**
Try-OEG-C18-2T	**11.7**	**241**	**440**	**18**	**47**	**5.9**	**0.79**	**12.6**

* Included again for the sake of comparison.

10.1002/chem.201802824



10.1002/chem.201802824

Figure 4 | Programmed dis-assembly of protein complex. a. Schematic representation of irreversible dis-assembly of protein complex upon photo-irradiation. The photo-sensitive conjugate was dissolved in 50 mM sodium phosphate pH 7.4 in a beaker and the temperature was maintained using ice bath during entire reaction. The sample was exposed to the UV lamp (range: 305-315 nm). Samples were withdrawn at different time points using a pipette and analysis of the same was carried out using SEC and MALDI-ToF MS. **b.** SEC chromatogram of time-dependent photo-chemical irreversible dis-assembly. **c.** MALDI-ToF data of protein conjugate before and after cleavage.

METHODS

Selection of proteins used in this work

We chose 4 serine proteins for our study *i.e*. trypsin (23 kDa), chymotrypsin (25 kDa), subtilisn (27 kDa) and proteinase K (29 kDa). Proteins were purchased from commercial vendors and molecular weights were determined prior to modification to track the molecular weight changes accurately.

Matrix preparation and molecular weight determination

Apart from molecular weight determination of native proteins, all stages of protein modification and purification were also followed by MALDI-ToF MS. The samples were analysed in Linear High Mass mode in AB Sciex 4800 plus MALDI-ToF/ToF analyzer with 4000 Series Explorer as software. Mass was scanned between 10,000 Da and 40,000 Da with focus mass at 25,000-29,000 Da depending on the protein analysed.

4 mg of DHAP and 4.5 mg of DAHC were weighed separately in microcentrifuge tubes. 150 µL of ethanol and 200 µL of milli-Q water were added to each tube, respectively. Both the solutions were sonicated using bath sonicator for 1 minute and vortexed for another minute. Then 50 µL of DAHC aqueous solution was transferred to DHAP solution and the resulting solution was vortexed for one more minute to yield matrix mixture. The sample preparation was carried out in a 500 µL microcentrifuge tube, 1:1:1 ratio of protein sample, 2% trifluoroacetic acid (TFA) and matrix mixture were mixed and vortexed for 1 minute. The samples were then kept undisturbed. When the crystallization observed in centrifuge tubes, 0.5 µL of the samples were spotted on the MALDI plate and air dried for 15 minutes. The plate was then loaded and fired to get accurate molecular weight both in +1 and +2 states. 100 µM protein concentration was found to be optimum for MALDI-ToF MS analysis.

Protein modification and monitoring

Protein modification was carried out at the concentration of 100 µM, which was found to be optimum for MALDI-ToF MS monitoring. Triton X-100 was used to solubilize the AABPs at concentrations 100 times (20 mM) more than CMC or 2% of total volume of reaction mixture. Typically, for test reactions, final volume of reaction mixture is 1 mL. Proteins were weighed (2.3 to 2.9 mg, depending on protein) in microcentrifuge tubes and 500 µL 50 mM sodium phosphate pH 7.4 was added and mixed gently with a pipette to make 200 µM solutions. Then AABPs (1 or 2 equivalents) were weighed in a different microcentrifuge tube, followed by

addition of 20 µL triton X-100 and 480 µL of 50 mM sodium phosphate pH 7.4 and vortexed for 15 minutes. When the AABP solution becomes clear, the protein solution was added into AABP solution to get 100 µM (1 mL) protein solutions and allowed to react for 24 h on rotospin at 20 rpm at 25 ºC. For probes FP-OEG-C18-2T and FP-OEG-C18-3T, 10 equivalents were weighed in a microcentrifuge tube and 4% triton X-100 was added and sonicated till the mixture became homogenous (please note that we used bath sonicator to solubilize these two AABPs). Protein modifications were carried out in falcon tubes at 200 mg scale following the linear scale up of above mentioned procedure for understanding the self-assembly behavior.

To monitor the extent of protein modification, the samples were directly withdrawn from reaction mixture using a pipette, and analyzed. In brief, 2 µL of reaction mixture was mixed with 2 µL of 2% TFA and 2 µL of matrix mixture as previously stated, vortexed and spotted on MALDI-ToF MS plate. Please note that the matrix preparation and sample preparation procedures remained same here.

Purification of modified proteins

All the conjugates were purified by three-step purification *i.e* IEX, SEC and desalting, performed using either Aktaprime or Aktaprime plus or Akta Explorer or Akta Pure. IEX was performed to remove triton X-100 using either SP sepharose or Q sepharose resins (GE) depending on isoelectric point (pI) and surface charges of proteins (Try: 30.4, Chy: 19.5, Pro K: -8.4, Sub: -6.2). For example, to purify the reaction mixture of trypsin or chymotrypsin, we used SP sepharose, a cation-exchange resin at pH 7.4 and to purify subtilisin and proteinase K we used Q sepharose, an anion-exchange resin at pH 10.

During cation-exchange chromatography, the column was pre-equilibrated with the same buffer (50 mM sodium phosphate pH 7.4) which was used for modification and then sample was injected followed by post injection equilibration for at least 2 Column Volumes (CVs) or until the complete removal of triton X-100 for large scale reactions. The elution of native protein and its corresponding conjugate together as mixture was later achieved using 50 mM sodium phosphate pH 7.4, 1 M NaCl as elution buffer.

In the case of anion-exchange chromatography, the protein reaction mixtures were buffer-exchanged first to 50 mM tris base pH 10 (since the buffer used for modification was 50 mM sodium phosphate pH 7.4). Then the anion-exchange column was pre-equilibrated using same

buffer (50 mM tris base pH 10) which was used for buffer-exchange and then sample was injected followed by post injection equilibration for at least 2 CVs or until the complete removal of triton X-100 for large scale reactions. The elution of the native protein and its conjugate together as mixture was later achieved using 50 mM tris base pH 10, 1 M NaCl as elution buffer. The obtained IEX fractions were subjected to SEC immediately to remove the native proteins from the protein conjugates. For the separation of native protein from protein conjugate, 50 mM sodium phosphate pH 7.4, 1 M NaCl was used as buffer using either Sephacryl S-100 HR 16/60 or Sephacryl S-200 HR 16/60 or Sephacryl S-300 HR 16/60. The NaCl was later removed by Sephadex-G25 desalting column. The desalted fractions were quickly lyophilized and later dissolved in required buffer when needed.

Monitoring of IEX, SEC and desalted fractions were carried out using the same procedure mentioned for native proteins and reaction mixtures except for the addition of triton X-100 in protein fractions. In short, 98 µL of samples from each fraction were mixed with 2 µL of triton X-100 (2% or 100 times the CMC) in a separate microcentrifuge tube and vortexed for 4 h. The samples were then analysed using the same procedure as mentioned above. Please note that the matrix and sample preparation procedure remained same.

Size exclusion-based molecular weight determination

To determine the molecular weights of soluble protein complexes, we performed SEC in Superdex-200 10/300 GL (GE healthcare). For this purpose, we chose standard proteins (GE Healthcare) *i.e*. blue dextran (2000 kDa), to determine the void volume of this column, thyroglobulin (660 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa) and trypsin (23 kDa).

A series of size exclusion runs were performed with these proteins in 50 mM sodium phosphate pH 7.4, 1 M NaCl with 0.25 mL/min as the flow rate. All the proteins were dissolved individually in milli Q water at 3 mg/mL and 0.5 mL was injected. With the obtained elution volumes, the partition coefficients (K_{av}) for standard proteins were calculated according to standard protocol (GE Healthcare) using the formula $K_{av} = V_e-V_o / V_c-V_o$ where, V_o = column void volume (8.5 mL, from the elution volume of blue dextran), V_e = elution volume of particular sample, and V_c = geometric CV (23.5 mL). Calibration curve was plotted for K_{av} of standards against relative molecular weights (M_r) of standard proteins. ($r^2 = 0.99$, please refer to the calibration curve). Now, to determine the molecular weights of soluble protein complexes,

the protein conjugates were dissolved individually in milli Q water at 5 mg/mL and 0.5 mL was injected.

Dynamic light scattering

Analysis of protein complexes was performed using DLS using Zetasizer Nano 2590 (Malvern, UK). Protein samples (5 mg/mL) were prepared in 50 mM sodium phosphate pH 7.4. *1 mL of sample was taken in disposable polystyrene cells and then* the mean size of the complexes was measured at 90° scattering angle.

SANS measurements

SANS experiments were carried out at SANS spectrometer operating at Dhruva Reactor, Bhabha Atomic Research Centre (BARC), Mumbai, India[51]. The facility employs a neutron velocity selector to provide a monochromatic neutron beam with a mean wavelength (λ) of 5.2 Å and wavelength spread (Δλ/λ) approximately 15%. The scattered neutrons were detected using a one dimensional ^{3}He position sensitive detector (PSD) in an angular range of 0.5-15° corresponding to wave vector transfer ($Q = 4\pi\sin(\theta/2)/\lambda$, where θ is scattering angle) range of 0.015-0.3 Å^{-1}. The measured SANS data were corrected for the background, the empty cell contribution, and the transmission. The corrected data were presented on an absolute scale using the standard protocols. For SANS measurements, the samples (5 mg/mL) were prepared in sodium phosphate pH 7.4 buffer in D_2O in order to minimize the incoherent scattering and to increase contrast. The temperature was maintained 30°C during the measurements.

SANS Analysis

SANS is a very promising technique to investigate self assembled structures in nanometer length scales. In SANS experiment, one usually measures the differential scattering cross section ($d\Sigma/d\Omega$) per unit volume as a function of wave vector transfer (Q). For monodispersed particles in a medium, it can be expressed as[52]

$$\frac{d\Sigma}{d\Omega} = n\,P(Q)\,S(Q) + B \qquad (1)$$

where n is the particle number density. $P(Q)$ is the intra-particle structure factor (square of the form factor) and S(Q) is the inter-particle structure factor. $P(Q)$ provides the specific size and shape of the scatterer whereas $S(Q)$ decides the spatial arrangement of the particles and thereby gives the information about the inter-particle interaction. For a dilute system, S(Q) may be

approximated to unity. *B* is a constant term representing the incoherent background arising mainly from the hydrogen atoms present in the sample.

P(*Q*) for spherical particles of radius *R* is given by[53]

$$P(Q) = \frac{16\pi^2}{9}\left(\rho_p - \rho_s\right)^2 R^6 \left[3\frac{sin(QR) - (QR)cos(QR)}{(QR)^3}\right]^2 \quad (2)$$

where ρ_p and ρ_s are the scattering length densities of particle and solvent respectively.

For prolate ellipsoidal particles, P(Q) can be expressed as[53]

$$P(Q) = \frac{16\pi^2}{9}\left(\rho_p - \rho_s\right)^2 \left(ab^2\right)^2 \int_0^1 \left[F(Q,\mu)\right]^2 d\mu \quad (3)$$

where the functions are given by

$$F(Q,\mu) = \frac{3(\sin x - x\cos x)}{x^3} \text{ and } x = Q[a^2\mu^2 + b^2(1-\mu^2)]^{1/2}$$

where *a* and *b*=*c* are, respectively semi-major and semi-minor axes of prolate ellipsoid (*a*>*b*=*c*). The variable μ is the cosine of the angle between the directions of *a* and *Q*.

The following expression provides *P*(*Q*) for a spherical core-shell structure[53]

$$P(Q) = \frac{16\pi^2}{9}[(R+dR)^3\left(\rho_{shell} - \rho_s\right)\frac{sinQ(R+dR) - Q(R+dR)cosQ(R+dR)}{Q^3(R+dR)^3} - R^3\left(\rho_{shell} - \rho_{core}\right)\frac{sinQR - QRcosQR}{Q^3R^3}]^2$$

where R is core radius and dR is shell thickness. ρ_{core} and ρ_{shell} represents the scattering length densities of the core and shell, respectively.

Throughout the data analysis corrections were made for instrumental smearing. The calculated scattering profiles were smeared by the appropriate resolution function to compare with the measured data. The parameters in the analysis were optimized by means of a nonlinear least-square fitting program[54].

SEC-MALS

SEC-MALS analyses of the protein complexes were performed on a Superdex-200 10/300 GL column (GE Healthcare) connected to an Agilent HPLC system equipped with the 18-angle light scattering detector (Wyatt Dawn HELIOS II) and a refractive index detector (Wyatt Optilab TrEX). The system was calibrated with BSA at a concentration of 2 mg/mL; 100 μL of the protein conjugates at concentrations 5 mg/mL were injected, and the molecular weights were calculated using ASTRA software (Wyatt Technologies).

Fluorescence Spectroscopy

The fluorescence of native proteins and protein conjugates were measured using Xenon arc lamp on Fluorolog spectrofluorimeter (HORIBA, Jobin-Yvon). Protein samples (1 mg/mL) were prepared in 50 mM sodium phosphate pH 7.4. Measurements were carried out at tryptophan specific excitation wavelength of 295 nm in a 10 mm path length cuvette. The samples remained clear during the meseaurement.

Circular Dichroism (CD) Spectroscopy

The CD spectra of native proteins and protein conjugates were recorded on Jasco J-815 spectrometer. Protein samples (0.25 mg/mL) were prepared in 50 mM sodium phosphate pH 7.4 and measurements were performed in a 1 mm path length cuvette. The spectra (background corrected) were obtained at scanning speed of 50 nm/minute within range of 195-250 nm. Mean residue elasticity was converted from accumulated spectra of five measurements for each sample.

Zeta potential measurement

The zeta potential for native proteins and protein conjugates was measured using Zetasizer Nano 2590 (Malvern, UK). Protein samples (1 mg/mL) were prepared in milli-Q water.

Programmed Dis-assemby of photo-sensitive supramolecular protein complex

The photo-sensitive protein conjugate was dissolved in 50 mM sodium phosphate pH 7.4 in a 50 mL beaker and temperature was maintained at 0° C using ice bath over the entire reaction. The sample was exposed to the Sankyo Denki G8T5E UVB linear lamp (8 W output) which emits ultraviolet rays between 280 to 360 nm (peak at 305-315 nm). Samples were withdrawn at different time points using a pipette and analysis of the same was carried out using SEC and MALDI-ToF MS.

10.1002/chem.201802824

References

51. Aswal, V. & Goyal, P. Small-angle neutron scattering diffractometer at Dhruva reactor. *Curr. Sci.* **79**, 947-953 (2000).
52. Svergun, D.I. & Koch, M. H. J. Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* 66, 1735–1782 (2003).
53. Pedersen, J. S. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Adv. Colloid Interface Sci.* 70, 171–210 (1997).
54. Chen, S.H. & Lin, T. L. 16. Colloidal Solutions. *Methods Exp. Phys.* **23**, 489-543 (1987).

Acknowledgments

This work was supported by IISER-Pune Start-up Grant, DST-SERB Early Career Award to BSS (ECR/2015/000253) and DBT grant We thank Dr. Radha Chauhan at NCCS-Pune for giving us access to SEC-MALS instrument. We thank Dr.Amrita Hazra @ IISER Pune for access to FPLC system. We also than Mr. Jegannadha Rao, for help with synthesis of few intermediate compounds.

Author Contributions

BSS conceived the idea and designed the research. MMR and PJB designed the experimental approach under the guidance of B.S.S. MMR performed protein bio-conjugation reaction and complete analytical characterization. MMR performed all biophysical experiments except SANS. PJB synthesized and characterized all the chemical probes. BSS, MMR and PJB analyzed complete data set except SANS data. SK and VKS performed SANS experiment and analyzed the SANS data. BSS wrote the paper.

Author Information

Reprints and permission information is available at www.nature.com/reprints. Correspondence and request for materials should be addressed to BSS (sandanaraj.britto@iiserpune.ac.in, sandanaraj.britto@gmail.com)