A Universal Chemical Method for Rational Design of Protein-based Nanoreactors

Mullapudi Mohan Reddy¹, Punita Bathla², and Britto S. Sandanaraj^{1,2*}

Department of Chemistry¹ & Department of Biology²,

Indian Institute of Science Education and Research - Pune

Abstract

Self-assembly of a monomeric protease to form a multi-subunit protein complex “proteasome” enables targeted protein degradation in living cells. The naturally occurring proteasomes serve as an inspiration and blueprint for the design of artificial protein-based nanoreactors. Here we disclose a general chemical strategy for the design of proteasome-like nanoreactors. Micelle-assisted protein labeling (MAPLab) technology along with the N-terminal bioconjugation strategy is utilized for the synthesis of a well-defined monodisperse self-assembling semi-synthetic protease. The designer protein is programmed to self-assemble into a proteasome-like nanostructure which preserves the functional properties of native protease.

The proteasome machinery which is present in all three domains of life is the finest example for self-assembled protein-based nanostructure that carries out the essential function such as targeted protein degradation.¹ In bacterial proteasome, the protease activity is performed by the *N*-terminal threonine residue of each subunit. The β -subunit of 14-mer is arranged in such a way that the *N*-terminal active residue is positioned inside the central cavity of the hollow cylindrical 20-S core particle to perform proteolysis of target substrates.² This highly complex and most sophisticated nanoreactor is the result of billion years of evolution and therefore it is highly challenging to design an artificial proteasome with such complexity. Nevertheless, these naturally occurring nanomachines serve as a blueprint for the design of primitive proteasomes.

Encapsulation of proteases inside protein-based nanocompartments results in compartmentalization.³ This strategy is widely used in the living system for controlling substrate specificity and preventing unwanted off-target reactions.⁴ In an elegant study, Hilvert and coworkers demonstrated the encapsulation of a protease inside cage-forming lumazine synthase and achieved artificial substrate selectivity.⁵ This top-down approach is also exploited by other groups for the design of various protein-based nanoreactors.⁶ Although extremely powerful, these methods rely on the engineering of naturally occurring protein cages through directed evolution and are therefore restricted to a few scaffolds. In contrast to the top-down strategy, the bottom-up approaches provide opportunities to design protein assemblies with an extraordinary diversity. In this aspect, computational protein design has proven to be a mature technology over the past decade.⁷ However, most of the reports are based on the design of static protein assemblies,⁸ and to our knowledge there are very few reports on synthetic protein nanoassemblies that can exhibit protease activity.⁹

Chemical technologies are emerging as a complementary method for the design of protein assemblies.¹⁰ Recently, our group invented a new method named “micelle-assisted protein labeling (MAPLab) technology (Figure 1a).¹¹ This technology was used for the design of a diverse set of self-assembling semi-synthetic proteins which includes protein amphiphiles^{11a}, protein-dendron bioconjugates^{11b}, protein-peptide bioconjugates^{11c}, photo-responsive protein amphiphiles,^{11a} and photo-sensitive protein-dendron bioconjugates^{11d}. We have demonstrated the successful design of monodisperse protein nanoparticles with a controlled oligomeric state, tunable size, and surface charges.^{11a} We also demonstrated the design of a multi-responsive

protein nanoparticle that can disassemble upon exposure to different stimuli.^{11a, d} However, in our previous method, the increased nucleophilicity of the active site residue of protein was exploited for site-specific bioconjugation and therefore the resultant protein assembly is catalytically inactive. Besides, the reported method is only applicable to the serine protease family. Although there are hundreds of enzymes that belong to this category¹², however, the number of proteins is far less compared to the vast number of proteins that exist in Nature. Therefore, it is important to develop a general chemical methodology that would work for the vast majority of the proteins to make self-assembling semi-synthetic proteins without affecting the function of the target protein. However, to the best of our knowledge, there are no chemical methods available that fulfill all of the above criteria.

During our research, we came across an exciting report from Francis and coworkers. They reported a new chemical methodology for site-specific modification of the *N*-terminal residue of a protein (Figure 1b).¹³ Unlike other bioconjugation methods, this method is versatile and can be used to label a large number of proteins irrespective of their size, shape, surface charge, molecular weight, and secondary, tertiary, and quaternary structures. Although a terrific method, one of the severe limitations of this methodology is that it is used only for bioconjugation of “hydrophilic payloads” which includes imaging agents, drugs, and polymers to make hydrophilic protein conjugates which find applications in the area of targeted drug delivery¹⁴ and biomaterials applications.¹⁵ Although extremely useful, this method cannot be directly used for site-specific bioconjugation of hydrophobic synthetic molecules or macromolecules.

Inspired by this work, we envisioned using this method along with MAPLab technology for the design of functional self-assembling semi-synthetic proteins (FSSSPs). Accordingly, we designed our target probe **8**. This probe contains three structural motifs: (i) 2-pyridine carboxyaldehyde as a site-specific *N*-terminal protein reactive group¹³ (ii) octaethylene linker as a hydrophilic linker, (iii) Octadeca benzyl group as a hydrophobic tail (Scheme 1). The choice of the linker and hydrophobic tail were based on our previous knowledge.^{11a} The synthesis of the target molecule was achieved in seven steps with an overall yield of 9.8%. In brief, intermediate **3** was synthesized following a procedure reported by Francis and coworkers.¹³ Similarly, intermediates **4** and **5** were made following synthetic procedures reported by our group.^{11a} Copper-catalyzed Huisgen [3+ 2] cycloaddition of compounds **4** and **5** yielded intermediate **6**.^{11a}

The obtained intermediate **6** was reacted with piperazine in the presence of THF to get compound **7**. Finally, compound **7** upon treatment with compound **3** in the presence of K₂CO₃ in acetonitrile yielded the amphiphilic probe **8** (Scheme 1, see SI for details).

Considering the amphiphilic nature of the probe molecule, we tested the solubility of probe **8** in an aqueous medium. Unfortunately, probe **8** was not soluble presumably due to the presence of a long hydrophobic tail. Next, we tested whether the probe could be solubilized in the presence of neutral surfactant Triton-X-100. To our delight, the probe was soluble in the presence of Triton-X-100 because of mixed-micelle formation in which the protein-reactive group 2-pyridine carboxaldehyde is presumably presented outside the micellar surface.^{11a-d} Next, we carried out a bioconjugation reaction (Figure 1b) of probe **8** using a MAPLab technology pioneered by our group (Figure 1a).¹¹ As a test case; we have attempted a bioconjugation reaction with a small globular protein lysozyme (abbreviated as LYS hereafter) in an aqueous medium at pH 7.2 room temperature for 12 h. The extent of bioconjugation reaction was monitored using a MALDI-ToF-MS. To our delight, we observed a new peak at high molecular weight corresponding to the facially amphiphilic bioconjugate LYS-OEG-C18 (Figure 2a). Note that this result indeed suggests that the 2-pyridine carboxaldehyde is present on the exterior surface of a mixed-micelle; otherwise, this reaction would not have been viable. Excited by these results, we wanted to investigate whether this reaction could be used for different categories of proteins. Accordingly, we chose three more proteins; a protease (chymotrypsin – abbreviated as CHY), a carrier protein (bovine serum albumin abbreviated as BSA), and a green fluorescent protein (GFP). All four proteins have different sizes, shapes, surface charges, secondary/tertiary structures and therefore modestly represent the structural and functional diversity of naturally occurring proteins.¹³ Bioconjugation of all three proteins was carried out similarly as described above and the progress of the reaction was monitored using the mass spec. To our delight, all of them are equally reactive as evident from mass spec results (Figure 2b-2d). As expected, we did not observe any double-labeling which is consistent with the published report.¹³

Although we were successful in a site-specification bioconjugation of a highly hydrophobic synthetic molecule to a globular protein, the next challenge is the purification and analytical characterization of newly designed protein conjugate. These proteins are extremely challenging to purify because of their amphiphilic nature and hence prone to non-specific aggregation.

Previously, we had reported an easy method for purification of a diverse set of self-assembling semi-synthetic proteins.¹¹ We have adopted the same method for purification. In brief, the Triton-X-100 was removed by ion-exchange chromatography followed by the separation of protein conjugate from native protein using size-exclusion chromatography (SEC). We were delighted to see a single peak for all the bioconjugates manifesting their monodisperse character in mass spec characterization (Figures S2 and S4).

Next, we set out to test whether these bioconjugates can self-assemble into monodisperse protein assemblies. Accordingly, we carried out analytical SEC studies (Figure S7 and S8). We were pleased to see all the protein bioconjugates self-assemble into protein complexes of bigger size as evident from SEC results (Figure 3a-3d). For example, BSA-OEG-C18 oligomerizes to form a 700 kDa protein complex whereas CHY-OEG-C18 and GFP-OEG-C18 formed 400 and 330 kDa protein complexes, respectively. Interestingly, the smallest protein bioconjugate LYS-OEG-C18 also self-assembles to form a trimer of 50 kDa protein nanoreactor (Figure 3a-3d, Table 1). These results are consistent with the previous results from our group that the oligomerization of globular proteins heavily depends on surface charge provided the linker length and hydrophobic tail remain constant.^{11a} This is indeed true in this study as well; LYS, ($M_w = 14.3$ kDa, $D_h = 2$ nm, and $pI = 11.35$) one of the smallest highly cationic globular protein forms only trimer. However BSA ($M_w = 66.6$ kDa, $D_h = 4$ nm, $pI = 4.7$) self-assemble into 700 kDa decamer (Figure 3a-3d, Table 1). Excited by these results, we wanted to check the hydrodynamic size (D_h) of these protein complexes. Accordingly, we carried out dynamic light scattering studies for BSA-OEG-C18 and CHY-OEG-18 bioconjugates. The hydrodynamic sizes of BSA-OEG-C18 and CHY-OEG-18 are 13 and 12 nm, respectively and these results match with SEC results of this study (Figure 3e, Table 1). Besides, SEC profiles and DLS results of these bioconjugates closely matches with the results obtained for other protein amphiphiles made using active-site bioconjugation technique.^{11a}

The most important objective of this study is to design a protein-based nanoreactor that can mimic the function of a naturally occurring proteasome. It is possible that during self-assembly the active-site of each subunit of CHY-OEG-C18 protein complexes can be sterically hindered and hence diminishes the catalytic activity of the protein by reducing the accessibility of the active-site by the substrate (Figure 4). To check that possibility, we carried out the enzymatic

assay of native CHY and CHY-OEG-C18 in parallel. Note that native CHY exists as a monomer whereas CHY-OEG-C18 exists as a decamer (from both SEC and DLS results). For the enzymatic assay, the concentrations of both CHY and CHY-OEG-C18 were kept at 80 μ M. We chose this concentration because dynamic light scattering studies reveal, CHY-OEG-C18 exists as a 12 nm proteasome-like nanoreactor at 80 μ M. The enzyme kinetic results suggest that the rate of hydrolysis of SPNA substrate is comparable between the native CHY and the CHY-OEG-C18 (Figure 4b, 4c). This is a very interesting result considering the catalytic activity of native CHY is not compromised when it oligomerizes into a decamer (Figure 4a). In a way, this process mimics the self-assembly of protease into proteasome which contains 14 subunits of proteases.²

In summary, we have demonstrated that monomeric CHY can be programmed to oligomerize into monodisperse decamer without compromising on the protease activity whereas LYS can be assembled into a trimer. We have shown here that (i) N-terminal bioconjugation strategy along with MAPLab technology could be used for the design of programmable protein-based nanoreactor; (ii) Both N-terminal bioconjugation and self-assembly of protein bioconjugate does not affect the activity of the protein; (iii) the utility of this method with a diverse set of proteins and finally; (iv) the size and oligomeric state of protein complexes can be precisely tuned by choosing an appropriate linker and a hydrophobic tail. The success of this approach depends on the power of two previous technologies i.e. the MAPLab technology¹¹ and the N-terminal bioconjugation method.¹³ Considering, this is a platform technology, one can envisage using any protease of interest to build semi-synthetic proteasomes. Although the designed protein-based nanoreactor is considered as an amateur compared to naturally occurring proteasomes. This study lays the foundation for the design of semi-synthetic protein-based nanomachines with increased complexity and sophistication. This kind of study also would shed some light on our understanding of the oligomerization of natural proteins. The ability to design the protein-based nanoreactors using a general chemical method opens up tremendous opportunities for the use of these nanoreactors in different applications such as biocatalysis, targeted drug delivery, and *in vivo* imaging.

Figure 1. a. Schematic presentation of MAPLab technology. b. Protein bioconjugation reaction: custom-designed probe site-specifically reacts with *N*-terminal residue of the target protein to yield functional self-assembling semi-synthetic proteins.

Scheme 1: Multi-step synthesis of the target probe.

Figure 2: MALDI-ToF MS results **a.** native LYS and a reaction mixture containing the starting material (native LYS) and the product LYS-OEG-C18 **b.** native CHY/reaction mixture **c.** native GFP/reaction mixture **d.** native BSA/reaction mixture. In each case, the successful bioconjugation is evident from appearance of single peak at a higher molecular weight corresponding to the molecular weight of the particular protein bioconjugate.

Figure 3. a. SEC chromatogram of bioconjugation reaction mixtures; a. LYS-OEG-C18 b. CHY-OEG-C18 c. BSA-OEG-C18. Semi-synthetic protein bioconjugates self-assemble into larger protein complex and therefore elutes first whereas native protein which exist as a monomeric proteins elutes later d. SEC data of purified CHY-OEG-C18 and BSA-OEG-C18 e. DLS data of purified CHY-OEG-C18 and BSA-OEG-C18

Sl. No	Native Proteins/Protein Bioconjugate	Elution Vol SEC (mL)	Mol Wt SEC (kDa)	Oligomeric State (mer)
1	BSA Native	7.4	66	1
2	BSA-OEG-C18-1T	13	700	10
3	CHY Native	19.0	25	1
4	CHY-OEG-C18-1T	14.6	400	16
5	GFP Native	18.7	30	1
6	GFP-OEG-C18-1T	14	330	10
7	LYS Native	20.9	14	1
8	LYS-OEG-C18-1T	17.9	50	3

Table 1. Oligomeric state of native and self-assembling semi-synthetic protein bioconjugates. Molecular weights of native proteins and their corresponding self-assembling semi-synthetic protein complexes were calculated using standard calibration curve (Figure S7 and S8).

Figure 4. Functional assay. **a.** Conversion of monomeric chymotrypsin (CHY) into proteasome-like nanoreactor using a MAPLab technology. The active site serine residue of chymotrypsin is highlighted in the figure. Conversion of CHY to CHY-OEG-C18 leads to programmed oligomerization to decamer. **b.** Schematic representation of hydrolysis of SPNA substrate by CHY. **c.** Kinetic assay CHY *V_s* CHY-OEG-C18. The rate hydrolysis of SPNA substrate is found to be similar for both monomeric CHY and decamer CHY-OEG-C18. Concentrations of CHY and CHY-OEG-C18 were kept at 80μM. **d.** CHY (PDB accession: 1YPH) structure showing active site serine residue (yellow) with respect to *N*-terminus residue (green).

References

1. Kish-Trier, E., & Hill, C. P. Structural biology of the proteasome. *Annu. Rev. Biophys.*, **2013**, 42, 29-49.
2. (a) Zühl F, Seemüller E, Golbik R, Baumeister W. Dissecting the assembly pathway of the 20S proteasome. *FEBS Lett.* **1997**, 418, 189-194.
(b) Brannigan JA, Dodson G, Duggleby HJ, Moody PC, Smith JL, Tomchick DR, Murzin AG. A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* **1995**, 378, 416-419.
3. Azuma Y., Hilvert D. Enzyme encapsulation in an engineered lumazine synthase protein cage. *Methods Mol. Biol.* **2018**, 1798, 39-55.
4. Yeates, T. O.; Kerfeld, C. A.; Heinhorst, S.; Cannon, G. C.; Shively, J. M. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.* **2008**, 6, 681-691.
5. (a) Worsdorfer, B.; Woycechowsky, K. J.; Hilvert, D. Directed evolution of a protein container. *Science* **2011**, 331, 589-592.
(b) Worsdorfer, B.; Pianowski, Z.; Hilvert, D. Efficient in vitro encapsulation of protein cargo by an engineered protein container. *J. Am. Chem. Soc.* **2012**, 134, 909-911.
(c) Seebeck, F. P.; Woycechowsky, K. J.; Zhuang, W.; Rabe, J. P.; Hilvert, D. A simple tagging system for protein encapsulation. *J. Am. Chem. Soc.* **2006**, 128, 4516-4517.
6. Das, S., Zhao, L., Elofson, K., & Finn, M. G. Enzyme Stabilization by Virus-Like Particles. *Biochemistry*, **2020**, 59, 2870-2881.
7. Huang, P. S., Boyken, S. E., & Baker, D. The coming of age of de novo protein design. *Nature*, **2016**, 537, 320-327.
8. Bale, Jacob B; Gonen, Shane; Liu, Yuxi; Sheffler, William; Ellis, Daniel; Thomas, Chantz; Cascio, Duilio; Yeates, Todd O; Gonen, Tamir; King, Neil P; Baker, David Accurate design of megadalton-scale two-component icosahedral protein complexes *Science*, **2016**, 353, 389-394.
9. Song, W.J. and Tezcan, F.A., A designed supramolecular protein assembly with in vivo enzymatic activity. *Science*, **2014**, 346,1525-1528.
10. Luo, Q., Hou, C., Bai, Y., Wang, R. and Liu, J., Protein assembly: versatile approaches to construct highly ordered nanostructures. *Chem. Rev.* **2016**, 116, 13571-13632.

11. (a) Sandanaraj, B. S.; Reddy, M.; Bhandari, P.; Sugam, K.; Aswal, V. K. Rational Design of Supramolecular Dynamic Protein Assemblies Using a Micelle-Assisted Activity-based Protein Labelling Technology. *Chem. - Eur. J.* **2018**, 24, 16085-16096.
(b) Sandanaraj, B. S, Bhandari, P. V, Reddy, M. M, Lohote, A. B, Sahoo, B. Design, Synthesis and Self-Assembly Studies of Suite of Monodisperse Facially Amphiphilic Protein-Dendron Conjugates. *ChemBioChem* **2020**, 21, 408-416.
(c) Sandanaraj, B. S.; Reddy, M. M.; Jagan, K.; Bhandari, P. J. Rational Design of Semi-Synthetic Protein Complexes with the Defined Oligomeric State. *ChemistrySelect.* **2019**, 4, 6397-6402.
(d) Bhandari, P. B., Sandanaraj, B. S. Programmed and Sequential Disassembly of Multi-responsive Supramolecular Protein Nanoassemblies: A Detailed Mechanistic Investigation *ChemBioChem* **2020**, Accepted <https://doi.org/10.1002/cbic.202000581>
(e) Bhandari, P. B., Reddy, M., Jagan, K. Sandanaraj, B. S. Expedient Chemical Synthesis of Selfassembling Semi-synthetic Proteins. *J. Org. Chem.* **2020**, Under Revision.
12. Simon, G. M., & Cravatt, B. F. Activity-based proteomics of enzyme superfamilies: serine hydrolases as a case study. *J. Biol. Chem.* **2010**, 285, 11051-11055.
13. MacDonald, J.I., Munch, H.K., Moore, T. and Francis, M.B., One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **2015**, 11, 326-331.
14. Sangsuwan, R., Tachachartvanich, P., & Francis, M. B. Cytosolic delivery of proteins using amphiphilic polymers with 2-pyridinecarboxaldehyde groups for site-selective attachment. *J. Am. Chem. Soc.* **2019**, 141, 2376-2383.
15. Koo, B., Dolan, N. S., Wucherer, K., Munch, H. K., & Francis, M. B. Site-Selective Protein Immobilization on Polymeric Supports through N-Terminal Imidazolidinone Formation. *Biomacromolecules* **2019**, 20, 3933-3939.

Acknowledgments

This work was supported by DST-SEB Early Career Award to B.S.S. (ECR/2015/000253) and DBT grant (BT/PRI1450/BRB/I0/1370/2015)