

A semisynthetic protein nanoreactor for single-molecule chemistry

Joongoo Lee

Wolfson College



A thesis submitted to the University of Oxford

for the degree of

Doctor of Philosophy

Major Subject: Chemical Biology

May 2015

Declaration

The work described in this thesis was carried out between October 2011 and May 2015 in the Chemistry Research Laboratory, University of Oxford, UK under the supervision of Professor Hagan Bayley. All the work described is entirely my work, unless where otherwise acknowledged in the text. No work in this thesis has been submitted previously for any other degree at the University of Oxford, or any other universities.

Joongoo Lee

May, 2015

For my father, in heaven

Acknowledgement

I would like to thank the following people for their contribution to this work.

Prof. Hagan Bayley, for his guidance and support during my graduate study and for enabling my development as a scientist. This work was accomplished by his rigour in all details, from experimental procedures to the use of language in writings.

All the other members of the Bayley group, both past and present, for having contributed to making my time in the lab both stimulating and enjoyable.

Prof. Carol Robinson and Prof. Stefan Howorka, my examiners, for their time to read this thesis.

My mother and parents-in-law, who are responsible for my better qualities.

Kyungmi, for sacrificing everything for family, and for being Songhee's beloved mom.

I was mainly supported by a studentship from European Research Council. But I was also in part indebted to the Korean government and Mogam science foundation for awarding scholarships.

Abstract

The covalent chemistry of individual reactants bound within a protein nanopore can be monitored by observing the ionic current flow through the pore, which acts as a nanoreactor responding to bond-making and bond-breaking events. However, chemistry investigated in this way has been largely confined to the reactions of thiolates, presented by the side chains of cysteine residues. The introduction of unnatural amino acids would provide a large variety of reactive side chains with which additional single-molecule chemistry could be investigated.

An efficient method to incorporate unnatural amino acid is semisynthesis, which allows site-specific modification with a chemically-defined functional group. However, relatively little work has been done on engineered membrane proteins. This deficiency stems from attributes inherent to proteins that interact with lipid bilayer, notably the poor solubility in aqueous buffer.

In the present work, four different derivatives α -hemolysin (α HL) monomer were obtained either by two- or three-way native chemical ligation. The semisynthetic α HL monomers were successfully refolded to heptameric pores and used as nanoreactors to study single-molecule chemistry. The semisynthetic pores show similar biophysical properties to native α HL pores obtained from an *in vitro* transcription and translation technique. Interestingly, when α HL pores with one semisynthetic subunit containing a terminal alkyne group were used to study Cu(I)-catalyzed azide-alkyne cycloaddition, a long-lived intermediate in the reaction was directly observed.

Table of Contents

| | |
|---|----|
| Declaration | 1 |
| Acknowledgement..... | 3 |
| Abstract | 4 |
| Table of Contents | 5 |
| List of Figures | 8 |
| List of Tables | 11 |
| Abbreviations | 12 |
| Chapter 1. Introduction | 15 |
| 1.1. Structure of α HL | 15 |
| 1.2. Prepore intermediate in assembly | 16 |
| 1.3. Stability of α HL pore | 17 |
| 1.4. Preparation of heptameric α HL pores | 17 |
| 1.5. Electrical recording for single-channel on planar lipid bilayers | 18 |
| 1.5.1. Equipment..... | 18 |
| 1.5.2. Chambers | 18 |
| 1.5.3. Electrodes..... | 19 |
| 1.6. The nanoreactor approach to single-molecule chemistry | 20 |
| 1.6.1. Measuring insertion of pores | 21 |
| 1.6.2. Interpreting single-channel recording data | 21 |
| 1.6.3. Previous work using the nanoreactor approach..... | 23 |
| 1.6.4. Further discussion on the α HL nanoreactor..... | 26 |
| Conclusion..... | 27 |
| References | 29 |
| Chapter 2. Building a protein pore by native chemical ligation | 32 |
| 2.1. Introduction | 32 |
| 2.1.1. Semisynthesis..... | 32 |

| | |
|---|-----|
| 2.1.2. Ligation sites..... | 34 |
| 2.1.3. Generation of α thioesters..... | 34 |
| 2.1.4. Generation of N-terminal cysteine..... | 36 |
| 2.1.5. Studies of semisynthesis on other membrane proteins | 36 |
| 2.2. Results and discussion..... | 40 |
| 2.2.1. Overview of building protein pores by native chemical ligation..... | 40 |
| 2.2.2. α HL pores with a full-length semisynthetic monomer (SM_{FL}) | 41 |
| 2.2.3. Semisynthetic truncated barrel mutant (TBM) pores | 64 |
| 2.3. Conclusion | 82 |
| 2.4. Materials and methods | 83 |
| 2.4.1. Assembly of WT α HL pore | 83 |
| 2.4.2. Assembly of a truncated pore..... | 93 |
| 2.4.3. General Analytical LC-MS..... | 96 |
| 2.4.4. General methods required for SCR..... | 96 |
| 2.4.5. Protein structure modelling..... | 98 |
| References | 99 |
| Chapter 3. A semisynthetic protein nanoreactor for single-molecule chemistry | 106 |
| 3.1. Introduction | 106 |
| 3.2. Results and discussion..... | 107 |
| 3.2.1. Preparation of Polypeptide Fragments..... | 107 |
| 3.2.2. Native Chemical Ligation | 111 |
| 3.2.3. Refolding of semisynthetic α HL monomers to heptameric pores..... | 116 |
| 3.2.4. Comparison of electrical properties of WT to semisynthetic pores..... | 122 |
| 3.2.5. CuAAC reaction in the semisynthetic pore | 125 |
| 3.3. Conclusion | 137 |
| 3.4. Materials and methods | 139 |
| 3.4.1. Preparation of plasmids and polypeptides..... | 139 |

| | |
|--|-----|
| 3.4.2. Peptide synthesis..... | 139 |
| 3.4.3. NCL | 140 |
| 3.4.4. Ionic Current Recordings and Single-Molecule Chemistry..... | 141 |
| 3.4.5. Analytical HPLC..... | 141 |
| 3.4.6. Peptide purification | 142 |
| 3.4.7. Analytical LC-MS | 142 |
| 3.4.8. Protein Purification..... | 142 |
| 3.4.9. Renaturation of the synthetic monomer..... | 143 |
| 3.4.10. Heptameric α HL pores..... | 143 |
| 3.4.11. Hemolysis Assays..... | 144 |
| 3.4.12. Limited proteolysis | 145 |
| 3.4.13. Single channel recording and data analysis | 145 |
| 3.4.14. β CD kinetics..... | 146 |
| 3.4.15. Molecular dynamics simulations..... | 146 |
| 3.4.16. MALDI analysis..... | 147 |
| 3.4.17. CuAAC kinetics in solution | 147 |
| Reference | 149 |
| Appendix..... | 155 |
| DNA and amino acid sequence | 155 |
| Amino acid sequence of the semisynthetic monomers | 161 |
| Modeller script..... | 162 |
| Alignment file (e.g. TBMd6.ali)..... | 163 |

List of Figures

| | |
|--|----|
| Figure 1.1. Structure of the α HL protein pore (PDB: 7AHL)..... | 16 |
| Figure 1.2. Proposed mechanism for the assembly of α HL pore | 17 |
| Figure 1.3. Custom-built vertical chamber and bilayer formation | 19 |
| Figure 1.4. Ag/AgCl electrodes | 20 |
| Figure 1.5. Kinetic analysis of single-channel recording data | 22 |
| Figure 1.6. A pioneering experiment in the single-molecule chemistry | 23 |
| Figure 1.7. Single-molecule reactions UQ (1) and QU(D), d1-1 with the thiolate of α HL pore | 24 |
| Figure 1.8. An example of reversible single-molecule chemistry studying photoisomerization of azobenzene..... | 26 |
| Figure 2.1. Schematic representation of NCL and expressed protein ligation (EPL).... | 33 |
| Figure 2.2. Schematic representation of three α HL pores built using the semisynthetic approach via expressed protein ligation (EPL)..... | 38 |
| Figure 2.3. Preparation of plasmid encoding NTF _{FL} - α thioester (1-126) (pTXB3-NTF _{FL}) | 42 |
| Figure 2.4. PCR-amplified DNA for NTF _{FL} -thioester | 43 |
| Figure 2.5. Purification of NTF _{FL} | 43 |
| Figure 2.6. Schematic representation of intein-mediated cleavage by a thiol to prepare NTF[Ala ¹ -Gly ¹²⁶]- α thioester (NTF _{FL}) | 44 |
| Figure 2.7. Characterization of purified NTF _{FL} - α thioester | 44 |
| Figure 2.8 Preparation of the CTF (127-293)-D ₈ H ₆ plasmid (pT7-SC1-CDH) | 46 |
| Figure 2.9. Schematic representation of direct preparation of CTF (127-293) from <i>E. coli</i> | 47 |
| Figure 2.10. Purification of CTF visualised on SDS-PAGE. CTF was purified using a His ₆ -tag at the C-terminus in a gravity column..... | 48 |
| Figure 2.11. Characterization with LC-MS of the purified CTF | 49 |
| Figure 2.12. Preparation and purification of a full length α HL monomer (SM _{FL})..... | 50 |
| Figure 2.13. Hemolytic assay of the synthetic monomer | 51 |
| Figure 2.14. Preparation and purification of WT α HL-H6 for quantitative analysis of hemolytic activity..... | 52 |

| | |
|---|----|
| Figure 2.15. WT _{7-n} SM _n Heteroheptameric pores | 54 |
| Figure 2.16. Two digestion sites of proteinase K on α HL monomer..... | 55 |
| Figure 2.17. Limited proteolysis for SM ₇ pores..... | 56 |
| Figure 2.18. Multichannel insertions of the synthetic pores | 57 |
| Figure 2.19. Count histogram of ionic currents corresponding to single synthetic pore insertions | 58 |
| Figure 2.20. Single-channel IV curves of WT ₆ (SM _{FL}) ₁ and WT α HL pores..... | 59 |
| Figure 2.21. Binding of β CD into the pore containing a single synthetic monomer (WT ₆ SM ₁)..... | 60 |
| Figure 2.22. Dwell time histogram for β CD binding events at different concentration of β CD | 61 |
| Figure 2.23. Kinetic analysis of β CD binding in the synthetic pore | 62 |
| Figure 2.24. Single-channel recordings of the α HL pore in the presence of PEG-OPSS (5 k Da)..... | 64 |
| Figure 2.25. Truncated barrel mutants and sequences in the β strand..... | 65 |
| Figure 2.26. Design of NCL to prepare the two different truncated synthetic monomers | 66 |
| Figure 2.27. Preparation of the NTF _{Δ6} plasmids by homologous recombination | 67 |
| Figure 2.28. PCR products to encode NTF _{Δ6} | 68 |
| Figure 2.29. Preparation and characterization of NTF _{Δ6-3M} | 69 |
| Figure 2.30. Preparation and characterization of NTF _{Δ6-3F} | 70 |
| Figure 2.31. Characterization of NTF by LC-MS | 70 |
| Figure 2.32. Preparation of the pT7-CTF _{Δ6} -DH plasmid | 71 |
| Figure 2.33. PCR products to encode CTF _{Δ6} | 72 |
| Figure 2.34. Purification of overexpressed CTF _{Δ6} by a Ni-NTA affinity column..... | 73 |
| Figure 2.35. Characterization of CTF _{Δ6} by LC-MS..... | 73 |
| Figure 2.36. Purification of TSMs produced by NCL..... | 74 |
| Figure 2.37. Characterization of TSM _{Δ6-3M} and TSM _{Δ6-3F} | 75 |
| Figure 2.38. Hemolysis assays for WT α HL and TBM Δ 6 prepared by IVTT, and truncated synthetic monomers (TSM _{Δ6-3M} and TSM _{Δ6-3F}) | 76 |

| | |
|--|-----|
| Figure 2.39. Assembly to homoheptamer from truncated synthetic monomer | 77 |
| Figure 2.40. Single channel I-V curve of TBM Δ 6 from IVTT and truncated synthetic pores | 78 |
| Figure 2.41. Kinetic analysis of β CD binding to homoheptameric (TSM Δ 6-3M) $_7$ pore | 79 |
| Figure 2.42. Representative single-channel trace of cyclodextrins binding to homoheptameric (TSM Δ 6-3F) $_7$ pore | 80 |
| Figure 2.43. The Dwell time histogram for β CD binding events in (TSM Δ 6-3F) $_7$ in the presence of 10 μ M β CD | 81 |
| Figure 3.1. Schematic representation for Preparation of an α HL polypeptide containing an unnatural amino acid by semisynthesis | 108 |
| Figure 3.2. Synthesis of SP-Nbz | 110 |
| Figure 3.3. Characterization of SP-Nbz | 111 |
| Figure 3.4. The first NCL reaction and characterization of the product | 112 |
| Figure 3.5. Deprotection of Thz114 after the first NCL reaction | 113 |
| Figure 3.6. The ligation reaction between NTF and the first ligation product (second ligation) | 114 |
| Figure 3.7. Characterization of the full-length synthetic monomer | 115 |
| Figure 3.8. Deconvoluted mass spectrum of the synthetic α HL monomer (SM) | 115 |
| Figure 3.9. Hemolysis assays | 116 |
| Figure 3.10. Three repetitive hemolysis assays | 117 |
| Figure 3.11. SDS-PAGE analysis of WT and synthetic α HL | 118 |
| Figure 3.12. Heteroheptameric pores | 119 |
| Figure 3.13. Amino acid sequence and proteinase K sites in the synthetic α HL monomer (SM) | 120 |
| Figure 3.14. Limited proteolysis with proteinase K of SM in solution and in the presence of DPhPC liposomes (1 mg mL $^{-1}$) | 121 |
| Figure 3.15. Structure of the WT $_6$ SM $_1$ α HL pore | 122 |
| Figure 3.16. Electrical properties of the WT $_6$ SM $_1$ α HL pore | 123 |
| Figure 3.17. Dependence of τ_{off} and τ_{on} on the concentration of β CD (trans) for WT $_6$ SM $_1$ pores | 124 |
| Figure 3.18. Binding of β CD in the M113F and M113N mutant pores | 124 |

| | |
|---|-----|
| Figure 3.19. MALDI-TOF mass spectra of azide-PEG-biotins | 125 |
| Figure 3.20. Currents observed after the addition of an azide substrate and catalyst to a single WT α HL pore (WT ₇) | 126 |
| Figure 3.21. Binding of APB700 to the WT ₆ SM ₁ pore in the presence of 20 mM Cu(I) | 127 |
| Figure 3.22. Currents observed after the addition of an azide substrate and catalyst to a single WT ₆ SM ₁ α HL pore | 128 |
| Figure 3.23. A click reaction in a WT ₆ SM ₁ α HL nanoreactor..... | 130 |
| Figure 3.24. I-V curves after the click reaction. a) Reaction with APB400 (n = 3) | 132 |
| Figure 3.25. The mean lifetime of the reaction intermediate..... | 132 |
| Figure 3.26. The mean reaction time | 133 |
| Figure 3.27. Characterization of the reagents used to determine the rate constant for the CuAAC reaction in solution | 135 |
| Figure 3.28. Monitoring the CuAAC reaction in solution | 136 |
| Figure 3.29. Analysis of the CuAAC kinetics | 137 |

List of Tables

| | |
|---|-----|
| Table 2.1. Semisynthetic strategies for generating polypeptide fragments containing α thioester | 35 |
| Table 2.2. Semisynthetic strategies for generating polypeptide fragments containing N-Cys | 36 |
| Table 3.1. The mean current decreases (δI_{b1} and δI_{b2}) and the root mean square noises (I _{rms}) during the formation of intermediate with the two different substrates..... | 131 |

Abbreviations

| | |
|------------------|--|
| aa | amino acids |
| amp | ampicillin |
| Bis-Tris propane | 1,3-bis(tris(hydroxymethyl)methylamino)propane |
| Boc | tert-butyloxycarbonyl |
| BSA | bovine serum albumin |
| CBD | chitin binding domain |
| CTF | C-terminal fragment |
| β CD | β -cyclodextrin |
| DCM | dichloromethane |
| DIEA | N,N-diisopropylethylamine |
| DMF | N,N-dimethylformamide |
| DNA | deoxyribonucleic acid |
| DPhPC | 1,2-diphytanoyl-sn-glycero-3-phosphocholine |
| DTT | dithiothreitol |
| EDTA | ethylenediaminetetraacetic acid |
| EPL | expressed protein ligation |
| ESI | electrospray ionization |
| FL | full-length |
| Fmoc | fluorenylmethyloxycarbonyl |
| Fmoc-Dbz | 3-(Fmoc)-4-diaminobenzoic acid |
| GuHCl | guanidine hydrochloride |
| α HL | α -hemolysin |

| | |
|--------|--|
| HBTU | N,N,N',N'-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate, O-(benzotriazol-1-yl)-N,N,N',N'- tetramethyluronium hexafluorophosphate |
| HOBt | N-hydroxybenzotriazole |
| HPLC | high performance liquid chromatography |
| IVTT | <i>in vitro</i> transcription and translation |
| IPTG | isopropyl- β -D-thiogalactopyranoside |
| LB | Luria-Bertani (Broth) medium |
| LC-MS | liquid chromatography-mass spectrometry |
| MBSA | 150 mM NaCl, pH 7.4, containing 1 mg mL ⁻¹ BSA |
| MESNa | sodium 2-sulfanylethanesulfonate |
| MOPS | 3-(4-morpholinyl)-1-propanesulfonic acid |
| MPAA | 4-mercaptophenylacetic acid |
| MWCO | molecular weight cut-off |
| NCL | native chemical ligation |
| Ni-NTA | nickel-nitrilotriacetic acid |
| NMP | N-methylpyrrolidone |
| NMR | nuclear magnetic resonance |
| NTF | N-terminal fragment |
| OD | optical density |
| OPSS | orthopyridyl disulfide |
| PCR | polymerase chain reaction |
| PEG | polyethylene glycol |
| PMSF | phenylmethylsulfonyl fluoride |
| PTFE | polytetrafluoroethylene |
| rRBCs | rabbit red blood cells |
| rRBCm | rabbit red blood cell membrane |

| | |
|----------|---|
| SCR | single-channel recording |
| SDS | sodium dodecyl sulfate |
| SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SEC | size exclusion column |
| SM | synthetic monomer |
| SP | synthetic pore |
| SPPS | solid-phase peptide synthesis |
| TBE | Tris-Boric acid-EDTA buffer |
| TBM | truncated barrel mutant |
| TSM | truncated synthetic monomer |
| TCEP | tris(2-carboxylethyl)phosphine |
| TE8 | 20 mM Tris.HCl, 0.1 mM EDTA, pH8.0 |
| TFA | trifluoroacetic acid |
| TGS | Tris-Glycine-SDS Buffer |
| Thz | thiazolidine |
| TIS | triisopropylsilane |
| Tris | 2-amino-2-hydroxymethyl-propane-1,3-diol |
| WT | wild type |

Chapter 1. Introduction

α -Hemolysin (α HL) is a bacterial pore-forming toxin secreted from *Staphylococcus aureus*. α HL is secreted as water-soluble monomer and oligomerizes on target membranes to form homoheptameric pores. The pore formation causes permeabilization of the bilayer to ions, water and low molecular weight molecules (< 2 kDa), leading to lysis of red blood cells¹. The heptameric α HL has been engineered for a variety of applications in biotechnology such as stochastic sensing², DNA sequencing³, and as a nanoreactor⁴ for monitoring chemical reactions at the single-molecule level. This chapter introduces the fundamental characteristics and unique features of α HL as a nanoreactor. A brief introduction to the single-molecule techniques utilized in this study is also presented.

1.1. Structure of α HL

The crystal structure of the α HL protein⁵ at 1.9 Å resolution reveals the presence of heptameric oligomers. The oligomers are composed of seven α HL subunits each containing 293 amino acids (33.2 kDa). The assembly imparts to the protein a mushroom-like shape, with a ~50 Å β -barrel stem protruding from the cap domain through the lipid bilayer into the cell's interior (Figure. 1.1 a and b). The cap of the protein conceals a large vestibule connected to the cell's exterior through a large opening at the top of the cap. The narrowest (1.4 nm in diameter) part of the channel is at the midpoint of the lumen, where the β barrel connects to the vestibule (Figure. 1.1 a and b).

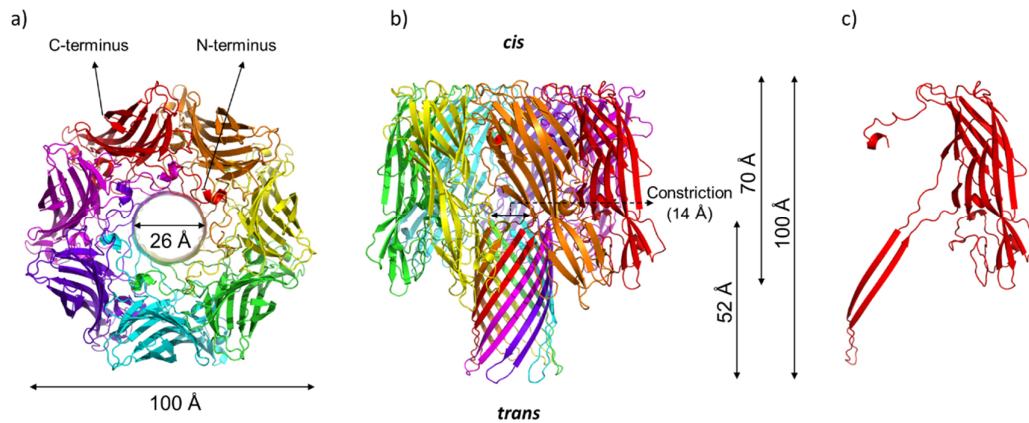


Figure 1.1. Structure of the α HL protein pore (PDB: 7AHL). Structure of α HL protein (PDB: 7AHL). Top (a) and side (b) view of the protein pore. The N- and C-termini are labeled in the top view. The structure is visualized in Pymol and each subunit is illustrated in a different color. (c) An individual subunit from the heptameric pore.

1.2. Prepore intermediate in assembly

The β -barrel pore-forming toxins are thought to assemble via a prepore intermediate⁶⁻⁸. The water-soluble monomer (α_1 , state 1 in Figure 1.2) first binds onto the lipid membrane to form membrane-associated α HL monomers (α_1^* , state 2) that organize to form nonlytic membrane-bound heptameric prepores (α_7^* , state 3). Finally, the stem domain inserts into the membrane to yield functional pores (α_7 , state 4). This process can occur on an artificial lipid bilayer, for example, that formed by 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC). The pore formation can be observed more rapidly on rabbit red blood cells (rRBCs), which is due to a receptor-mediated process, increasing the affinity of α HL to the cell membrane. The prepore mechanism will be discussed again in detail with experimental results in Chapter 2 and 3.

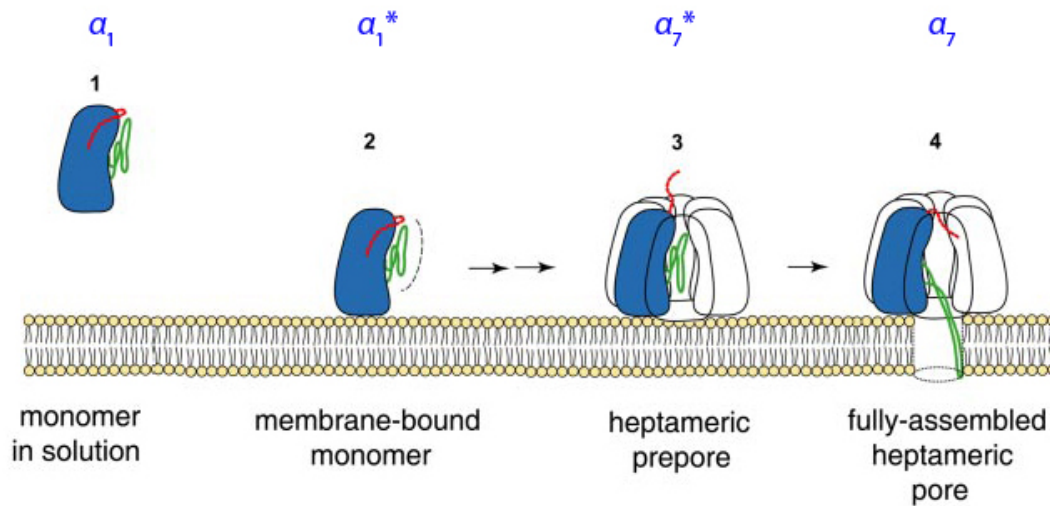


Figure 1.2. Proposed mechanism for the assembly of α HL pore. The water-soluble monomer (α_1 , state 1) bind to membranes (α_1^* , state 2). Oligomerization of α_1^* yields the nonlytic membrane-bound heptameric prepore (α_7^* , state 3). The functional pore is formed by insertion of the stem domain into the membrane. Reprinted with permission from J. Biol. Chem.⁶, copyright (2006).

1.3. Stability of α HL pore

The α HL pore is stable and remains open over a wide range of ionic strengths (from 0.001 to 4 M KCl), temperature⁹ and pH (pH3-10)¹⁰. The open state is of long duration, and lasts for hours, except at extremes of pH and transmembrane potential¹¹.¹². The α HL pore is also stable to sodium dodecyl sulfate (SDS)¹³, a common detergent used in gel electrophoresis. This enables direct purification of functional heptameric pores from SDS gels for further experiments.

1.4. Preparation of heptameric α HL pores

Although heptameric α HL pores can be obtained directly from α HL monomers for single-channel recording experiments, pre-assembled homoheptamers are preferred as they give more consistent results than monomers.

Typically, α HL protein is expressed by *in vitro* transcription and translation (IVTT). α HL heptamers used for the bilayer experiments are then formed by incubating monomers with rabbit red blood cell membranes (rRBCm) and further purified by SDS-PAGE¹⁴. The concentration of heptamers after purification is typically $\sim 1\text{ ng } \mu\text{L}^{-1}$ ¹⁵.

1.5. Electrical recording for single-channel on planar lipid bilayers

Protein nanopores employed for single-channel recordings (SCR) are reconstituted into artificial lipid bilayers that mimic biological membrane. Single-channel recording experiments presented in this thesis are carried out with planar lipid bilayer with the use of sensitive equipment^{16, 17} to measure the ionic currents (pA scale) passing through a single pore. The following sections cover the specific apparatus and methods used in our setting to perform SCR.

1.5.1. Equipment

SCR measurements are carried out using the commercial Axopatch 200B patch-clamp amplifier, with a 1440A 16-bit digitizer (Molecular Devices). Experiments are conducted in the voltage-clamp mode, where the current through the pore is measured at a fixed potential.

Experiments are performed in a Faraday cage to avoid unnecessary electrical noise from external sources during data acquisition. The enclosure, a steel box ($\sim 50 \times 50 \times 50$ cm in dimensions) with rigid side walls (> 1 mm thick), is grounded via the common ground in the patch-clamp amplifier. External noise is further reduced by placing the Faraday cage on pneumatic antivibration tables or support with partially inflated soft rubber tubing¹⁵ (see section 2.4.4. for detailed procedure).

1.5.2. Chambers

The opening of the pore in the cap domain is referred to as the *cis* side, whilst the opening at the end of the β -barrel is referred to as the *trans* side (Figure 1.1b). When the protein crosses the lipid bilayer and inserts into the membrane, the meaning of the terminology is used in the same effect. *Cis* is the chamber to which protein is added. *Cis* is kept at ground and the transmembrane potential is given as the potential on the *trans* side¹⁵.

The chamber is a custom-built vertical-bilayer device where two 500 μ L compartments (*cis* and *trans*) are separated by a 25 μ m-thick polytetrafluoroethylene film. The only path between the two compartments is a central aperture in the PTFE film across which the bilayer is formed (Figure 1.3) by Montal-Mueller method¹⁸. The aperture is typically 50-100 μ m in diameter and is produced by ‘zapping’ the films with a custom high-voltage spark generator (producing a spark ~15 mm long at a frequency of ~ 1Hz) (see section 2.4.4. for detailed procedure).

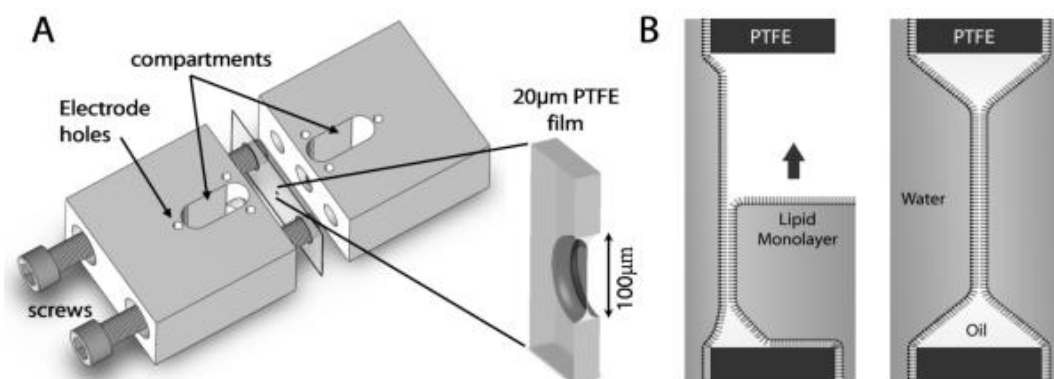


Figure 1.3. Custom-built vertical chamber and bilayer formation. a) The two halves ($\sim 3 \times 3 \times 1.5$ cm in dimension, each) are fastened with screws and sealed for water-tightness. b) The bilayer is formed by the Montal-Mueller method. Reprinted with permission from Elsevier¹⁵, copyright (2010).

1.5.3. Electrodes

The working and ground electrode connect the *trans* and *cis* compartments to the recording equipment, respectively. Ag/AgCl electrodes are used due to their long-term chemical stability, reversible electrochemical behavior, predictable junction potentials and their low-noise electrical performance. The electrodes are prepared by immersing Ag wire overnight in ~30 % hypochlorite (NaClO) solution¹⁹ (see section 2.4.4. for detailed procedure).

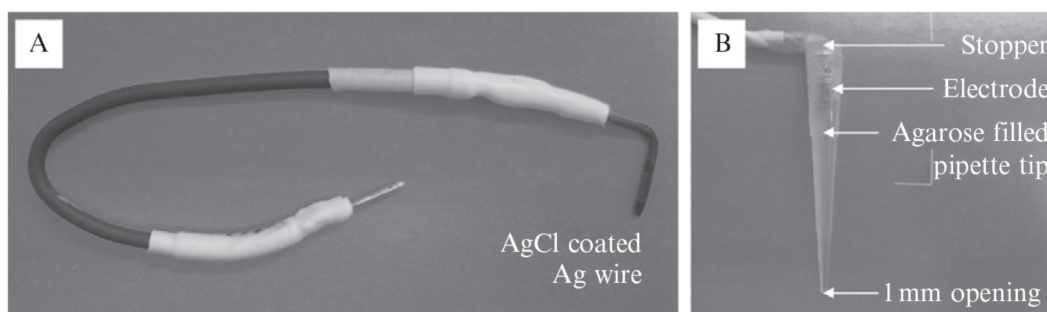


Figure 1.4. Ag/AgCl electrodes. a) The electrode is soldered onto standard (Sn/Cu alloy) electrical wire. b) The bare electrode is incorporated into a 200 μ L pipette tip filled with agarose (3 % w/v, 3 M KCl). Reprinted with permission from Elsevier¹⁵, copyright (2010).

1.6. The nanoreactor approach to single-molecule chemistry

For the single-channel experiments, a single α HL pore is located in an artificial lipid bilayer that separates two compartments containing aqueous electrolyte. An electrical potential is applied across the bilayer and the flow of ions through the α HL pore is monitored. Molecules of interest are investigated by analyzing the associated modulation of the ionic current when the molecules pass through the pore or interact with the lumen. This allows to obtain the kinetics of formation (association), dissociation, and, in turn, equilibrium constant of the molecules interacting with the pore under a given potential, which has been utilized to study chemical kinetics where current

modulation generated by a comparable size of molecules or intermediates bound to α HL pore can be differentiated.

To explain the nanoreactor approach using the α HL pore, interpretation of SCR data, early works employed for single-molecule chemistry, and pros and cons of the nanoreactor are discussed in the following sections (1.6.1-1.6.4)

1.6.1. Measuring insertion of pores

There are a number of ways to determine the insertion of pores^{20, 21}. One of the most straightforward methods is to wait until a single conductance is observed under a positive potential (> 100 mV) after the addition of preassembled α HL heptameric pores (Section 1.1.4). Typically ~ 0.5 μ L of SDS-solubilized heptamer (~ 1 ng μ L⁻¹) is added to the *cis* compartment and the solution in the compartment is then stirred gently. When a single channel is produced, a conductance state (~ 1 nS, e.g. 100 pA at +100 mV in 1 M KCl) is observed (mostly within 10-20 min). The amount of free protein in the *cis* compartment is then reduced by perfusion using fresh buffer (see experimental details).

1.6.2. Interpreting single-channel recording data

The ionic current flowing through the transmembrane under a fixed potential is modulated depending on analyte (A) binding to the pore (P) or chemical reactions taking place inside the pore. The time-current trace in Figure 1.5a is representative of a reversible binding of A to P.

The binding events are dependent on the concentration of A, e.g. the number of binding events (the number of low conductance states) increases in a certain time, which lowers the duration of open pore conductance. Thus, the mean inter-event duration ($\bar{\tau}_{on}$, the mean value of all the open pore states on the trace) is inversely proportional to the concentration of the species.

The mean dwell time ($\bar{\tau}_{off}$, the mean duration of low P-A states) and the current amplitude of the events can provide a specific characteristic of the analyte. Because the lifetime ($\bar{\tau}_{off}$) is independent on the concentration of A, $\bar{\tau}_{off}$ of a single molecule bound to the pore is equal to the reciprocal of the sum of unimolecular rate constants of reactions leading away from the pore. From this fact, one can derive the following equations, $\bar{\tau}_{on} = 1/k_{on}[A]$ and (Eq. 1) and $\bar{\tau}_{off} = 1/k_{off}$ (Eq. 2)¹⁷.

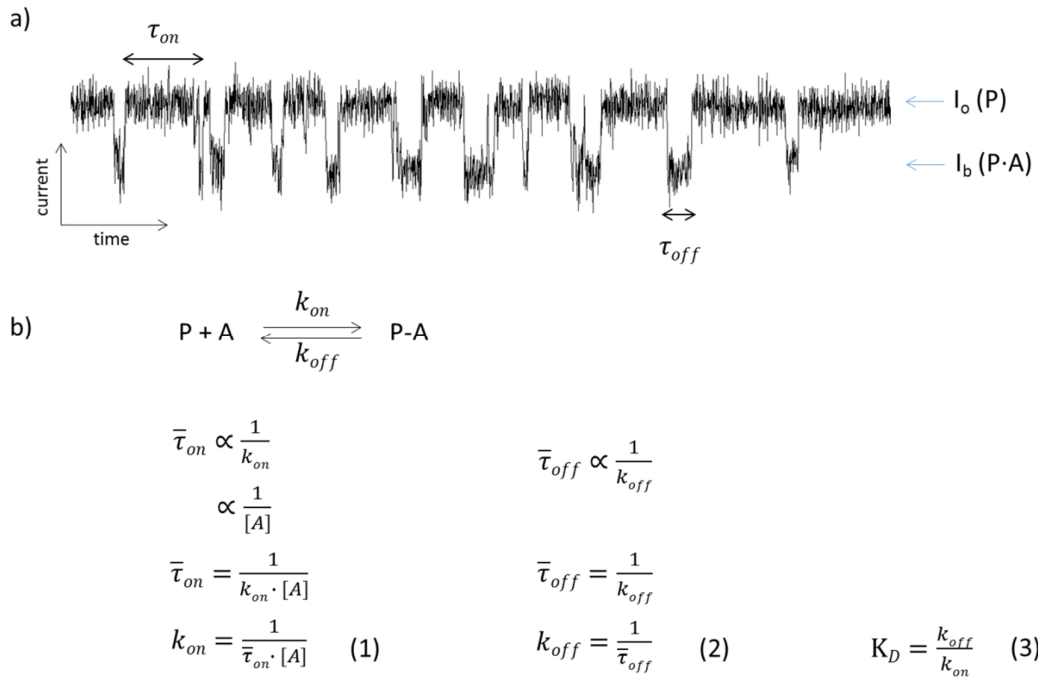


Figure 1.5. Kinetic analysis of single-channel recording data. (a) An actual time-current trace for reversible binding of an analyte (A) to a pore (P) under a positive potential. b) Equations are derived from the relationship of concentration of analyte and binding to the pore. (k_{on} : association rate constant, k_{off} : dissociation rate constant, $\bar{\tau}_{on}$: mean open pore time, $\bar{\tau}_{off}$: mean duration of low conductance state, i.e. mean time of analyte binding)

Values of $1/\bar{\tau}_{on}$ shows linear dependence on $[A]$ meaning the binding of A to P follows the first-order kinetics, whereas $1/\bar{\tau}_{off}$ is independent of $[A]$ meaning the

breakdown of P-A to P meets the zeroth-order kinetics. Thus, the formation of P-A (α HL pore-bound analyte) is a bimolecular reaction, while the reverse reaction is unimolecular. Using the two rate constants (k_{on} and k_{off}) obtained from analysis of a current trace, the dissociation constant (K_d) is determined Eq. (3).

1.6.3. Previous work using the nanoreactor approach

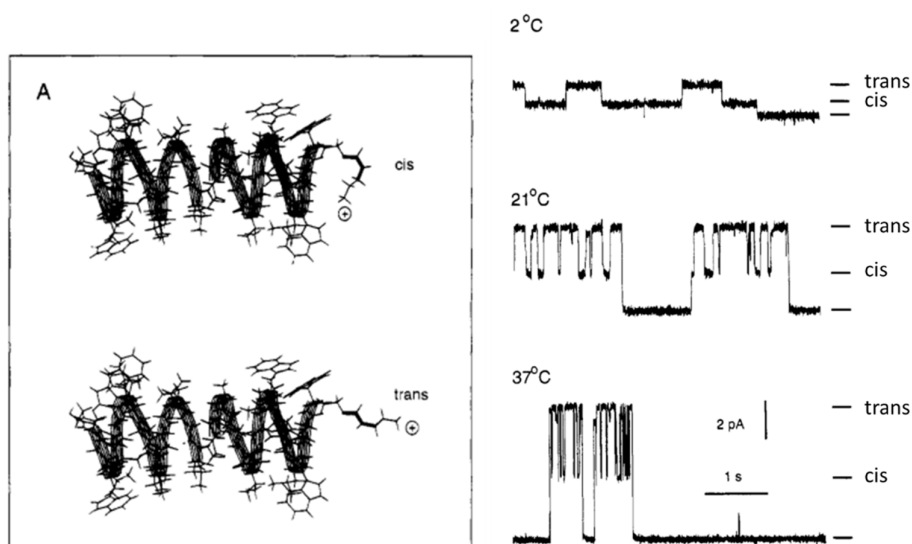


Figure 1.6. A pioneering experiment in the single-molecule chemistry. The isomerization of a covalently attached carbamate on the gramicidin A channel was observed in the single-molecule level. Reprinted with permission from American Chemical Society²², copyright (1995).

The ability to measure the movement of ions through individual channels in planar lipid bilayers was developed in the 1960s²³ and brought to fruition when currents were recorded flowing through single channels of known chemical structure²². In Figure 1.6, a carbamate was covalently attached to the dimeric gramicidin A channel. Under an applied potential *cis-trans* isomerization of the carbamate bond was reflected in a modulation of the single channel conductance. From an analysis over the temperature

range 2°C to 37°C, standard enthalpies and entropies, and activation enthalpies and entropies for the isomerization were obtained.

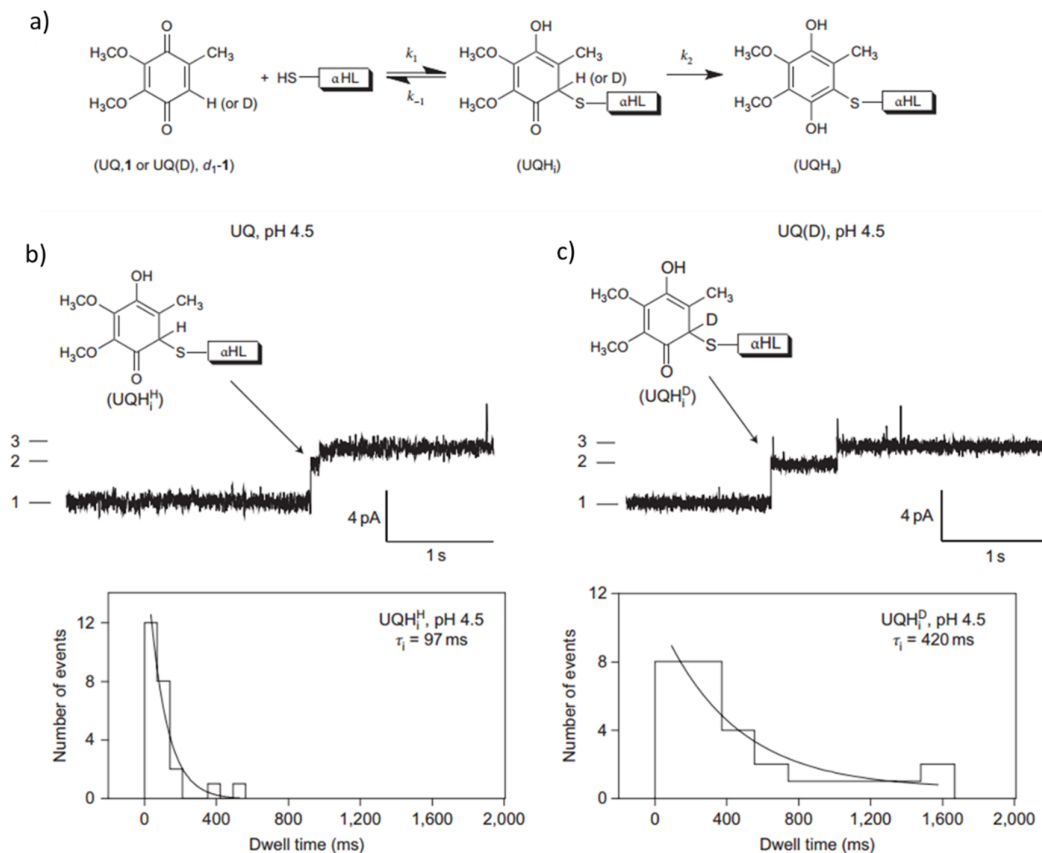


Figure 1.7. Single-molecule reactions UQ (1) and QU(D), d1-1 with the thiolate of α HL pore.

a) The Michael addition reaction of a quinone with the thiolate in the protein nanoreactor. The thiol attacks UQ(1) at C6 to form a tetrahedral intermediate (UQH_i), which yields subsequently UQH_a. b) and c) The lifetimes of UQH_i^H and UQH_i^D were collected and fitted to an exponential curve. The mean lifetimes of UQH_i^D was found to be longer than UQH_i^H. (2 M KCl, 50 mM sodium acetate, 100 μ M EDTA, pH 4.5, and the applied potential: -50 mV) Reprinted with permission from Nature²⁴, copyright (2010)

As bond making and breaking in individual reactant molecules tethered within a single pore can be observed as current modulation, reaction intermediates can be

identified by this nanoreactor approach if the lifetime of intermediates falls within a detection limit of nanopore. A good example of single-molecule chemistry observed in α HL pore is the reaction of quinone with a thiol on the pore to form a substituted hydroquinone by reductive 1,4-Michael addition (Figure 1.7a)²⁴. Interestingly, the analysis of the lifetimes of UQH_i (the exponential fits in Figure 1.7b and c) reveals that the deuterated intermediate (UQH_i^{D}) has a longer duration (420 ms) than the protonated (97 ms) intermediate (UQH_i^{H}). This is remarkable because a primary hydrogen-deuterium isotope effect could be measured straightforward for a particular reaction step at the single-molecule level, which cannot be in an ensemble measurement.

The examination of irreversible bond formation at the single-molecule level can be rewarding, but it is tedious and time-consuming because each of the experiments reveals only a single value of the lifetime for each event, thus, requiring many repeats to obtain a reliable statistic value.

Reversible covalent chemistry^{25, 26} would help in this sense because multiple events would be observed in each experiment which allows the rapid accumulation of reliable mean lifetime. Azobenzenes are photochromic dyes, which are converted from the *trans* to the *cis* form and vice versa by irradiation with near UV light. The photoisomerization was studied recently using the single-channel recording technique described above. An azobenzene molecule attached to the thiol group of Cys at position 117 was used in the study²⁵ (Figure 1.8a). The current passing through the pore is measured at a constant potential and the bilayer is irradiated with light of appropriate wavelength. When the bilayer is illuminated at 330 nm, transitions between the initial level and a low conductance level are observed, which corresponded to the photoisomerizations of the azobenzene molecule. It is interesting that the isomerization (to the *trans* or the *cis* form) of azobenzene can be observed (Figure 1.8b) with light of a fixed wavelength, which cannot be identified in bulk solution.

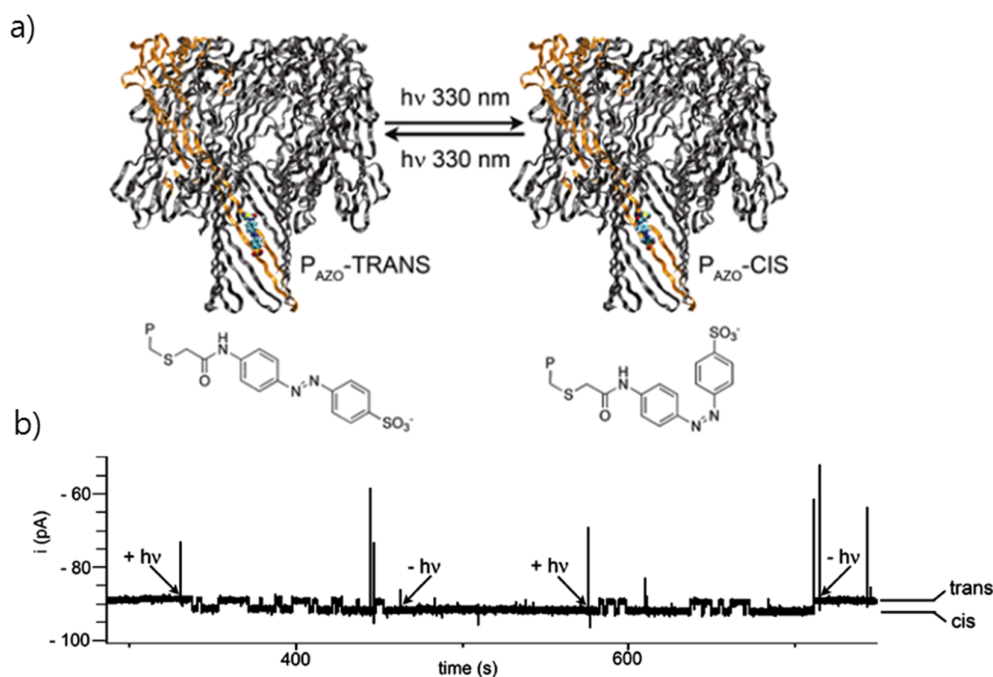


Figure 1.8. An example of reversible single-molecule chemistry studying photoisomerization of azobenzene. a) An α HL pore containing an azobenzene group at position 117 on the sulfhydryl group of cysteine and the molecular structure of the *trans* and the *cis* forms of azobenzene. b) *Trans-cis* isomerization in a single pore in a planar lipid bilayer (10 mM Tris-HCl, pH 8.5, 2 M KCl, 100 μ M EDTA, and the applied potential: -50 mV). Reprinted with permission from American Chemical Society²⁵, Copyright (2006)

1.6.4. Further discussion on the α HL nanoreactor

α HL pore has several advantages as an almost ideal single-molecule nanoreactor. i) α HL is very stable against a wide range of pH and temperature²⁷. ii) Covalent^{4, 24, 26, 28, 29} and noncovalent binding³⁰⁻³² of single-molecule can be monitored in a readout of ionic current modulation, i.e. no labeled fluorophore is required to detect single-molecule reaction. From the readout, information of single-molecule chemistry can be provided, which would not be readily accessible from ensemble experiments.

However, several limitations and a few issues must be considered. i) No organic solvents can be used. Organic solvents disturb and break the lipid bilayer, thus, reactants must be water-soluble. ii) Some reactants (e.g. DTT) may destabilize the bilayers when it is used in high concentration (> 20 mM, observed result). iii) Reactants must react with the desired functional group attached on inner surface of α HL pore. In other words, they must bind to the pore to make changes in ionic flow. iv) To detect the single-molecule reaction within a reasonable time scale, the reaction rate constants should be 10^{-3} to 10^3 s^{-1} (reaction lifetime of 1 ms – 1000 s) for first-order reactions; 10^{-1} to 10^5 $\text{M}^{-1}\cdot\text{s}^{-1}$ for second-order reactions (calculated when 10 mM reactant is used). v) The single-molecule chemistry is largely confined to the reaction of thioate (i.e. deprotonated Cys sidechain) and of groups that can be attached to Cys residues. An important option is the incorporation of unnatural amino acid, which would provide a wide selection of reactive functional group on sidechains. The experiments to resolve the limitation stated on point (v) will be discussed in Chapter 3.

Conclusion

Nanoreactor approach using α HL pore has been widely applied to various studies especially for single-molecule chemistry. In comparison to other single-molecule techniques, the nanoreactor can directly provide evidence of reaction intermediate at the single-molecule level, which cannot be monitored in the ensemble measurements. However, the scope of chemistry that can be studied within the nanoreactor should be expanded since the reactions that have been studied so far are confined to the use of thiolates. Incorporation of unnatural amino acids will be an important option that may allow us to modify proteins with a variety of chemically defined functional groups.

In this sense, I will discuss the semisynthetic approach to engineer α HL and explain the biophysical characteristics of semisynthetic α HL pores in Chapter 2. In Chapter 3, a nanoreactor containing an unnatural amino acid will be discussed and used to demonstrate a click reaction (Cu-catalyzed azide alkyne cycloaddition, CuAAC).

References

1. Krasilnikov, O.V., Sabirov, R.Z., Ternovsky, V.I., Merzliak, P.G. & Tashmukhamedov, B.A. The structure of *Staphylococcus aureus* alpha-toxin-induced ionic channel. *Gen Physiol Biophys* **7**, 467-473 (1988).
2. Bayley, H. & Cremer, P.S. Stochastic sensors inspired by biology. *Nature* **413**, 226-230 (2001).
3. Branton, D. et al. The potential and challenges of nanopore sequencing. *Nat Biotechnol* **26**, 1146-1153 (2008).
4. Bayley, H., Luchian, T., Shin, S.-H. & Steffensen, M. in *Single Molecules and Nanotechnology*, Vol. 12. (eds. R. Rigler & H. Vogel) 251-277 (Springer Berlin Heidelberg, 2008).
5. Song, L. et al. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859-1866 (1996).
6. Jayasinghe, L., Miles, G. & Bayley, H. Role of the amino latch of staphylococcal alpha-hemolysin in pore formation: a co-operative interaction between the N terminus and position 217. *J Biol Chem* **281**, 2195-2204 (2006).
7. Degiacomi, M.T. et al. Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. *Nat Chem Biol* **9**, 623-629 (2013).
8. Yamashita, D. et al. Molecular basis of transmembrane beta-barrel formation of staphylococcal pore-forming toxins. *Nat Commun* **5**, 4897 (2014).
9. Kang, X.F., Gu, L.Q., Cheley, S. & Bayley, H. Single protein pores containing molecular adapters at high temperatures. *Angew Chem Int Ed Engl* **44**, 1495-1499 (2005).
10. Gu, L.Q. & Bayley, H. Interaction of the non-covalent molecular adapter, beta-cyclodextrin, with the staphylococcal alpha-hemolysin pore. *Biophys J* **80**, 494a-494a (2001).
11. Korcev, Y.E. et al. *Staphylococcus aureus* alpha toxin induced pores: Channel-like behavior in lipid bilayers and patch clamped cells. *J Membrane Biol* **144**, 185-185 (1995).
12. Menestrina, G. Ionic channels formed by *Staphylococcus aureus* alpha toxin: voltage-dependent Inhibition by divalent and trivalent cations. *J Membrane Biol* **90**, 177-190 (1986).
13. Japrun, D., Henricus, M., Li, Q.H., Maglia, G. & Bayley, H. Urea facilitates the translocation of single-stranded DNA and RNA through the alpha-hemolysin nanopore. *Biophys J* **98**, 1856-1863 (2010).

14. Cheley, S., Braha, G., Lu, X.F., Conlan, S. & Bayley, H. A functional protein pore with a "retro" transmembrane domain. *Protein Sci* **8**, 1257-1267 (1999).
15. Maglia, G., Heron, A.J., Stoddart, D., Japrun, D. & Bayley, H. Analysis of single nucleic acid molecules with protein nanopores. *Methods Enzymol* **475**, 591-623 (2010).
16. Moczydlowski, E. in *Ion Channel Reconstitution*. (ed. C. Miller) 75-113 (Springer US, 1986).
17. Sakmann, B. & Neher, E. *Single-channel Recording*. (Springer Science+Business Media, 2009).
18. Montal, M. & Mueller, P. Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. *Proc Natl Acad Sci U S A* **69**, 3561-3566 (1972).
19. Krnjevic, K. Microelectrode methods for intracellular recording and iontophoresis. *Trends in Neurosciences* **5**, 133.
20. Gillis, K. in *Single-Channel Recording*. (eds. B. Sakmann & E. Neher) 155-198 (Springer US, 1995).
21. Kado, R.T. Membrane area and electrical capacitance. *Methods in Enzymology* **221**, 273-299 (1993).
22. Jaikaran, D.C.J. & Woolley, G.A. Characterization of thermal isomerization at the single molecule level. *The Journal of Physical Chemistry* **99**, 13352-13355 (1995).
23. Hladky, S.B. & Haydon, D.A. Discreteness of conductance change in bimolecular lipid membranes in the presence of certain antibiotics. *Nature* **225**, 451-453 (1970).
24. Lu, S., Li, W.-W., Rotem, D., Mikhailova, E. & Bayley, H. A primary hydrogen–deuterium isotope effect observed at the single-molecule level. *Nat Chem* **2**, 921-928 (2010).
25. Ludwig, S. & Bayley, H. Photoisomerization of an individual azobenzene molecule in water: an on-off switch triggered by light at a fixed wavelength. *J Am Chem Soc* **128**, 12404-12405 (2006).
26. Shin, S.H., Luchian, T., Cheley, S., Braha, O. & Bayley, H. Kinetics of a reversible covalent-bond-forming reaction observed at the single-molecule level. *Angew Chem Int Ed* **41**, 3707-3709 (2002).

27. Kang, X.F., Gu, L.Q., Cheley, S. & Bayley, H. Single protein pores containing molecular adapters at high temperatures. *Angew Chem Int Ed* **44**, 1495-1499 (2005).
28. Luchian, T., Shin, S.H. & Bayley, H. Single-molecule covalent chemistry with spatially separated reactants. *Angew Chem Int Ed* **42**, 3766-3771 (2003).
29. Choi, L.S. & Bayley, H. S-Nitrosothiol Chemistry at the Single-Molecule Level. *Angew Chem Int Ed* **51**, 7972-7976 (2012).
30. Banerjee, A. et al. Molecular bases of cyclodextrin adapter interactions with engineered protein nanopores. *Proc Natl Acad Sci U S A* **107**, 8165-8170 (2010).
31. Hammerstein, A.F., Shin, S.H. & Bayley, H. Single-molecule kinetics of two-step divalent cation chelation. *Angew Chem Int Ed* **49**, 5085-5090 (2010).
32. Movileanu, L., Cheley, S. & Bayley, H. Partitioning of individual flexible polymers into a nanoscopic protein pore. *Biophys J* **85**, 897-910 (2003).

Chapter 2. Building a protein pore by native chemical ligation

2.1. Introduction

An efficient method to expand the study of single-molecule chemistry with the α HL nanoreactor would be to have the nanoreactor engineered with unnatural amino acids. A semisynthetic method may be a useful approach because it enables us to site-specifically modify proteins by direct incorporation of an unnatural amino acid into the protein. This chapter will discuss how to apply the semisynthetic method to an integral membrane protein and the biophysical properties of semisynthetic α HL pores used as nanoreactors.

2.1.1. Semisynthesis

Expressed protein ligation (EPL) is a synthetic approach that allows the assembly of a target protein from unprotected polypeptide building blocks. The techniques are derived from a class of auto-processing proteins containing inteins¹. EPL is an extension of the immensely successful native chemical ligation (NCL)^{2, 3} in which two unprotected synthetic peptides are chemically ligated together. In EPL, one or more of the peptides is of *in vivo* origin, but the actual ligation step is an *in vitro* chemical process that can be performed under a wide range of reaction conditions⁴.

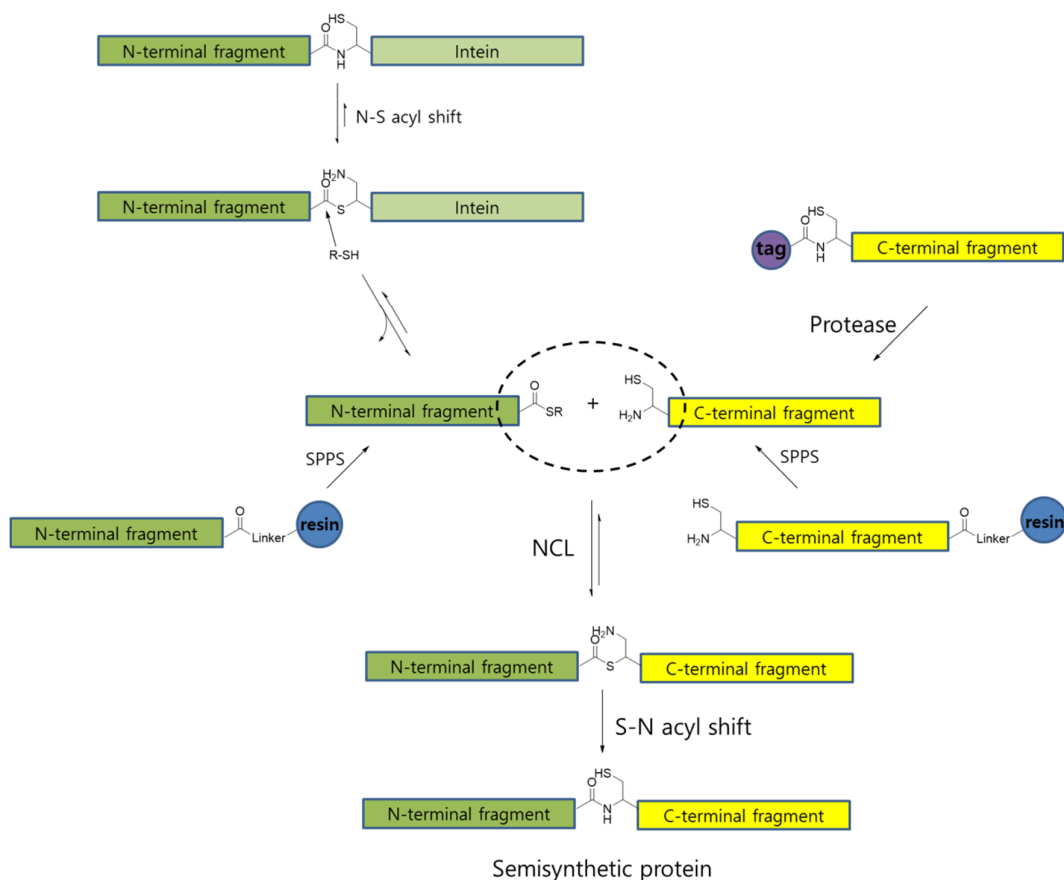


Figure 2.1. Schematic representation of NCL and expressed protein ligation (EPL). In NCL (dotted circle), the initial reactant contains a α thioester that rearranges by intramolecular nucleophilic attack to give a native peptide (amide) bond at the ligation site. EPL is a semisynthetic approach using NCL in which synthetic or recombinant polypeptides are regioselectively ligated together. An N-terminal fragment (NTF) expressed as a fusion protein with an intein can be cleaved from the intein by thiolysis to give a α thioester. A C-terminal fragment (CTF) containing N-Cys can be made recombinantly by masking the cysteine with a tag that can be removed later. Protein building blocks can also be efficiently synthesized by SPPS on a solid resin containing a linker when the number of amino acids is less than 50.

In the simplest case, EPL allows the synthesis of a target protein from two polypeptide fragments. People have shown that the coupling of two polypeptide

fragments allows introduction of isotopically labeled recombinant protein fragments for NMR spectroscopy studies^{5, 6} and incorporation of peptide probes⁷ within the flanking regions of a protein. The two-way coupling is also used to study the renaturation ability of channel proteins after the synthesis in denaturing conditions⁸. Moreover, this approach has proven very useful for the production of cytotoxic proteins from non-toxic fragments or non-canonical amino acids^{9, 10}.

More sophisticated approaches are used to ligate together three polypeptide fragments in a regioselective fashion, thereby allowing the incorporation of specifically modified species¹¹⁻¹³.

2.1.2. Ligation sites

As EPL is an extension of NCL, the ligation product will have a cysteine residue at the ligation junction (Figure 2.1). Thus, at least one Cys (either it is used from a native sequence or generated by mutagenesis) must be present at the N-terminus of the CTF. Another functional group required at the ligation site will be an α thioester at the C-terminus of the NTF, which acts as a good leaving group to form a native amide bond by nucleophilic attack of the Cys side chain.

2.1.3. Generation of α thioesters

Recombinant protein α thioesters can be generated by thiolysis of the corresponding intein fusion¹⁴. Several modified inteins are commonly used for this purpose and many inteins are commercially available for *E. coli* expression vectors. The *Mycobacterium xenopi* DNA gyrase A (Mxe GyrA) is commonly used because it is relatively small (198 amino acids) and can be cleaved efficiently with a variety of thiols to give α thioesters¹⁵⁻¹⁷. Also, it can be efficiently refolded, which is especially important since it allows Mxe GyrA fusions to be recovered from bacterial inclusion

bodies¹⁸. This useful characteristics of Mxe GyrA enables to produce hydrophobic proteins or overexpressed protein in a high temperature (37°C).

Peptide α thioesters can also be prepared synthetically by solid-phase peptide synthesis (SPPS). Several solid-phase methods and commercially available linkers are available for the chemical synthesis of peptide α thioesters¹⁹⁻²⁴. The most conventional way of SPPS is the use of tert-butoxycarbonyl (Boc) protected amino acids to elongate peptides in which acid-based deprotectors (e.g. 95 % of trifluoroacetic acid) are used repetitively to allow free amines to couple with carboxylic acids of next amino acid. α Thioesters remain stable during peptide elongation to the acid-deprotectors. In contrast, SPPS uses 9-fluorenylmethoxycarbonyl (Fmoc) protected amino acids and requires the use of base (e.g. 20 % piperidine), which renders this approach incompatible with α thioester groups. However, Fmoc-SPPS is more commonly used rather than Boc-SPPS because of safety issue caused by hydrogenfluoride (HF) gas, used to cleave peptide from the resin. Several techniques developed to allow the Fmoc synthesis with α thioesters are introduced below (Table 2.1).

Table 2.1. Semisynthetic strategies for generating polypeptide fragments containing α thioester¹⁷

| Polypeptides | Preparation methods | Comments |
|---------------------------------|--|--|
| Recombinant α thioesters | Mxe GyrA inteins ¹⁶ | Commercially available, widely used. |
| | Sortase ²⁵ | Extra amino acid sequences attached after digestion. |
| Synthetic α thioesters | Mercaptopropionamide linkers ^{26, 27} | Only applicable to Boc-SPPS due to incompatibility in basic condition |
| | N-acylbenzimidazolinone (Nbz) ²⁴ | Only applicable to Fmoc-SPPS. Isomeric acylurea may be produced (Figure 3.2) as the linker contains diamino benzene. |

2.1.4. Generation of N-terminal cysteine

Several methods have been developed to generate recombinant proteins bearing an N-terminal cysteine (N-Cys). All the methods rely on removal of a leader sequence from the N-terminus of the polypeptide. The leader sequences can be the initiating methionine, or can be one of several protease recognition sequences (Table 2.2). In chemical synthesis, N-Cys peptides can be prepared relatively easily by routine SPPS using either Boc or Fmoc chemistry^{7, 19-24}.

Table 2.2. Semisynthetic strategies for generating polypeptide fragments containing N-Cys¹⁷

| Polypeptides | Preparation methods | Comments |
|--|---|---|
| Recombinant N-Cys | Leader sequence N-Met ²⁸ | Depends on endogenous methionyl aminopeptidase; N-Cys may be masked <i>in vivo</i> to form a thiazolidine structure (Figure 2.9) due to reaction with pyruvate. |
| | Factor Xa protease ²⁹ | Cleaves at the C-terminal end of the IEGR sequence. |
| Synthetic N-Cys | TEV protease ³⁰ | Recognizes E-Xa-Xb-Y-Xc-Q-(G/S/C) sequences. ENLYFQC is most commonly used to generate N-Cys. |
| | Direct synthesis | Can be prepared either by Fmoc or Boc-SPPS. |
| Synthetic peptide containing N-Cys and α thioester together | Direct synthesis | N-Cys needs to be protected ³¹ to prevent from cyclization. |

2.1.5. Studies of semisynthesis on other membrane proteins

While EPL has been used mainly in investigations of soluble proteins, in recent years it has been increasingly used to study structural and functional characteristic of integral membrane proteins, such as MscL³², KcsA^{18, 33-35} and NaK³⁶. Clayton *et al.*

achieved a preparation of a mechanosensitive channel, MscL, by using chemical ligation to connect sequential three fragments produced by SPPS. A group led by Tom Muir synthesized an ionic channel protein, KcsA, by a semisynthetic approach and successfully refolded the synthetic protein. Furthermore, an unnatural amino acid, was introduced to KcsA^{34, 35} into the semisynthetic pores. In NaK proteins, the semisynthetic strategy revealed that the acidic residue in the selectivity filter is necessary for optimal ion flux through the channel. However, access to integral membrane proteins via the semisynthetic approach has been impeded due to their inherent poor solubility. For example, a polypeptide segment containing a membrane region of α HL has usually been found to precipitate out under neutral conditions (unpublished result), although the full-length α HL monomer is water-soluble. Denaturing conditions including 8 M urea or 6 M guanidine hydrochloride (Gu-HCl) in aqueous buffer may improve the solubility of the segments. However, the denaturant used to resolve the solubility problem may inhibit the activity of proteases that are required for generation of polypeptide fragments bearing an N-Cys. Yet, another hurdle in the preparation of integral proteins is the final folding, which must be carried out to obtain a functional protein.

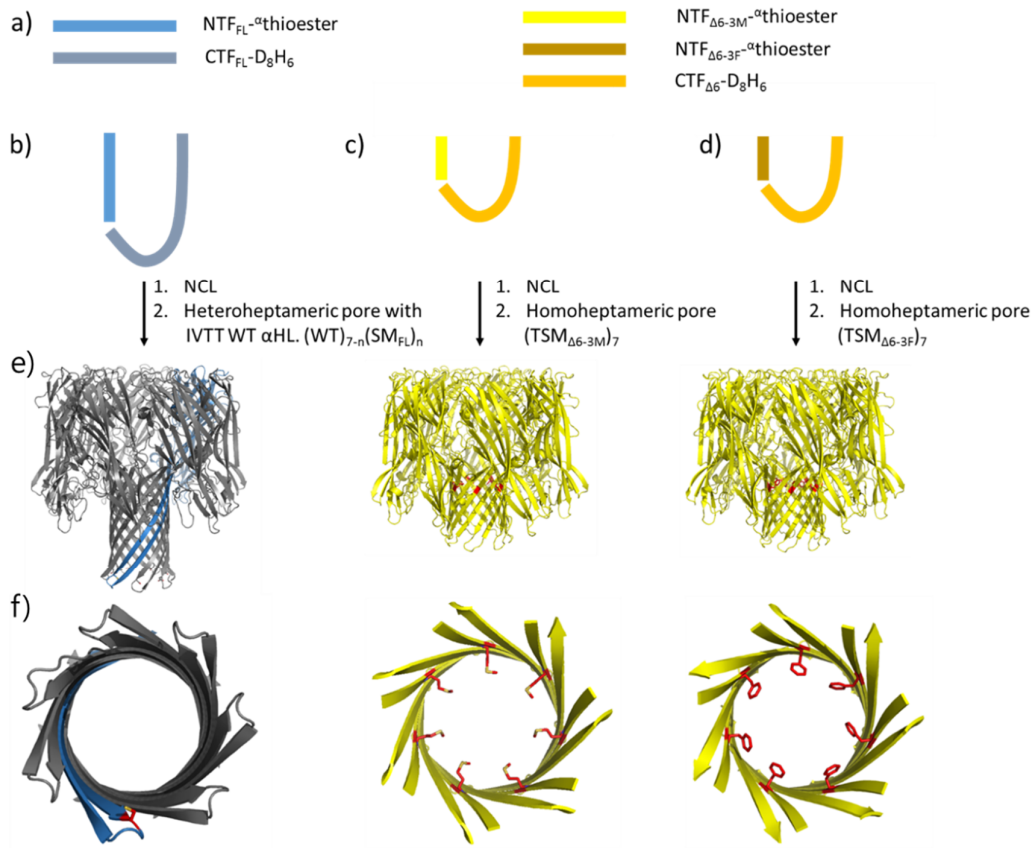


Figure 2.2. Schematic representation of three α HL pores built using the semisynthetic approach via expressed protein ligation (EPL). a) N-terminal fragments and C-terminal fragments are produced from *E. coli*. Two polypeptide fragments required to produce α HL derivatives are purified and coupled to full length α HL monomer via NCL (Figure 2.1b and c). b) $[\text{NTF}_{\text{FL}}]\text{-}\alpha\text{thioester}$ (sky blue) and $\text{CTF}_{\text{FL}}\text{-D}_8\text{H}_6$ (grey) are coupled to produce a full length α HL monomer (SM_{FL}). c-d) $[\text{NTF}_{\Delta 6-3\text{M}}]\text{-}\alpha\text{thioester}$ (light yellow) and $[\text{NTF}_{\Delta 6-3\text{F}}]\text{-}\alpha\text{thioester}$ (dark yellow) were separately coupled with $\text{CTF}_{\Delta 6}\text{-D}_8\text{H}_6$ (yellow) to give truncated mutant synthetic monomers (TSM). e) The synthetic monomers are renatured and oligomerized to functional heptameric pores, $\text{WT}_6(\text{SM}_{\text{FL}})_1$, $(\text{TSM}_{\Delta 6-3\text{M}})_7$, and $(\text{TSM}_{\Delta 6-3\text{F}})_7$, respectively. f) Bottom view for $\text{WT}_6(\text{SM}_{\text{FL}})_1$ and top views for $(\text{TSM}_{\Delta 6-3\text{M}})_7$, and $(\text{TSM}_{\Delta 6-3\text{F}})_7$. Cys¹²⁷, Met¹¹³, and Phe¹¹³ are depicted in red. Images were prepared using Modeller and Pymol.

Herein, three different synthetic α HL pores possessing natural amino acids³⁷ are investigated. α HL monomers were prepared by two-way ligation, i.e. two recombinant polypeptides expressed from *E. coli* were coupled via NCL. The ligation products were purified, renatured, and used to form homo- and hetero-heptameric pores. Subsequently, the functional and structural characteristics of the heptameric pores were analyzed, demonstrating that the semisynthetic pores are almost identical to α HL pores produced by IVTT.

In the first part of this chapter, synthesis of a full length α HL monomer (SM_{FL}) and its assembly to a heteroheptameric α HL pore ($WT_6(SM_{FL})_1$) containing a synthetic subunit are investigated. The functional properties of the semisynthetic pore as a nanoreactor are also discussed. In the second part of the chapter, the synthesis of two different truncated monomers, $TSM_{\Delta 6-3M}$ and $TSM_{\Delta 6-3F}$, which contain different amino acids residues (Met or Phe) at residue 113 is presented. Also, assembly to homoheptameric pores, $(TSM_{\Delta 6-3M})_7$ and $(TSM_{\Delta 6-3F})_7$, and the biological and electrophysiological characteristics of the pores are investigated.

2.2. Results and discussion

2.2.1. Overview of building protein pores by native chemical ligation

We chose to make a full length wild type monomer (SM_{FL}) and two truncated barrel mutants³⁸ ($TSM_{\Delta 6-3M}$ and $TSM_{\Delta 6-3F}$) that would provide unique characteristics when the synthetic α HL monomers are assembled to the corresponding pores. SM_{FL} was chosen for direct comparison of the synthetic pore to wild type α HL pore prepared by IVTT. The two truncated barrel mutants were chosen because the short β barrels are also able to form stable and functional pores, which rearranges the lipid bilayer to a toroidal shape³⁹. Interestingly, a mutant (M113F) of the truncated pore showed a strong interaction with an adaptor molecule which binds permanently inside the lumen of pore. A two-way ligation strategy was used to produce the three different derivatives of α HL monomer, where the different N-terminal fragments and C-terminal fragments were coupled (Figure 2.2) by NCL.

Three polypeptides referred to as N-terminal fragments (NTF_{FL} , $NTF_{\Delta 6-3M}$, and $NTF_{\Delta 6-3F}$) were produced by thiolysis of a fusion protein (NTF -Mxe GyrA intein-CBD, *Mycobacterium xenopi* DNA gyrase A (Mxe GyrA) intein-chitin binding domain (CBD)) after expression in *E. coli*. as shown in Figure 2.3⁴⁰. $NTF_{FL}[Ala^1-Gly^{126}]$ - α thioester was generated for assembly of a full-length monomer of α HL and $NTF_{\Delta 6-3M}[Ala^1-Met^{113}]$ - α thioester and $NTF_{\Delta 6-3F}[Ala^1-Phe^{113}]$ - α thioester were obtained in the same way for assembly of truncated mutants.

Two different polypeptides referred to as C-terminal fragments (CTF_{FL} and $CTF_{\Delta 6}$) were also prepared after expression in *E. coli*. $CTF_{FL}[Cys^{127}-Asn^{293}]$ - D_8H_6 for the synthesis of a full-length α HL monomer and $CTF_{\Delta 6}[Cys^{114}-Asn^{293}, \Delta 114-293$ and $\Delta 120-125]$ - D_8H_6 for truncated mutants were generated by expressing proteins from the genes beginning with Met-Cys codon (Figure 2.9)⁴¹. The additional 14 amino acids, D_8H_6 , at

the C-terminus of CTF were used for purification. The His₆ (H₆) tag was used to purify the CTFs by use of Ni-NTA affinity column and the Asp₈ (D₈) tag was used to directly purify heteroheptamers having different stoichiometry of the synthetic monomers from SDS gel as the heteroheptamers having more synthetic subunits move faster on gel due to the more negatively charged D₈ tag.

The purified NTFs and CTFs were coupled together by NCL under a denaturing condition at room temperature to give a full-length α HL monomer or two truncated monomers. The synthetic monomers (yields: <50 %) were purified by gel filtration and renatured to either heteroheptameric or homoheptameric pores in the presence or absence of lipids at room temperature. The synthetic monomers were then characterized by mass spectrometry and the synthetic pores were further analyzed by using single-channel recordings to compare the properties with the α HL pores prepared by IVTT.

2.2.2. α HL pores with a full-length semisynthetic monomer (SM_{FL})

2.2.2.1. Preparation, purification and characterization of NTF_{FL}

Recombinant NTF- α thioester was prepared by thiolysis of a fusion protein containing Mxe GyrA intein-CBD (*Mycobacterium xenopi* DNA gyrase A intein-chitin binding domain) derived from the plasmid pTXB3 (NEB) (Figure 2.3). The DNA encoding residues 1-126 of α HL were PCR-amplified (Figure 2.4) from a plasmid (pT7-SC1) bearing a full-length α HL gene and cloned into pTXB3.

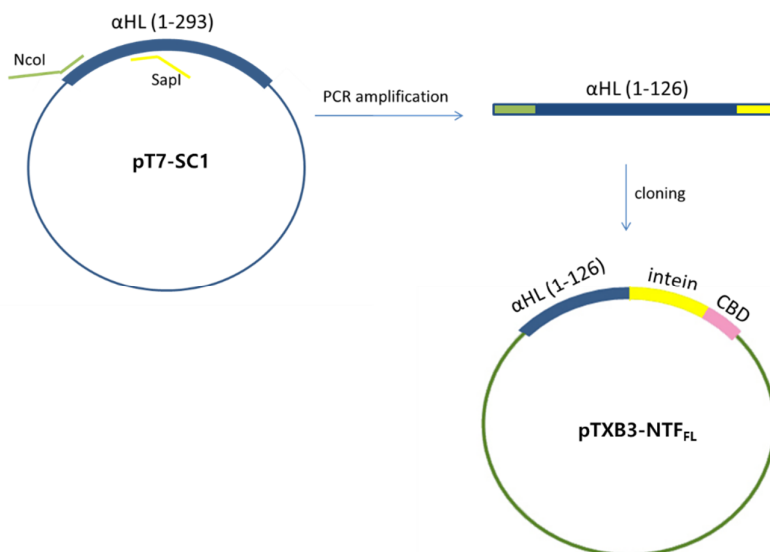


Figure 2.3. Preparation of plasmid encoding NTF_{FL}-αthioester (1-126) (pTXB3-NTF_{FL}). The NTF gene was PCR amplified from pT7-SC1 (blue) containing the WT αHL gene with the two primers (green and yellow) which bear restriction enzyme sites (*NcoI* and *SapI*) at flanking sequence. Subsequently, the PCR amplicon was cloned into a commercial vector (pTXB3, NEB, green) encoding a fusion protein with an intein and a chitin binding domain.

After expression at 37 °C, high levels of the fusion protein were found within inclusion bodies in the *E. coli* host. Following cell lysis and centrifugation, most of the pellet could be solubilized in denaturing buffer containing 8 M urea (Figure 2.5). To allow binding to chitin beads, the concentration of the urea was reduced to 1 M. Unbound proteins were washed off the column and the bound fusion protein was treated with 2-mercaptoethane sulfonate (MESNa) under gravity flow, yielding the NTF[Ala¹-Gly¹²⁶]-αthioester (or NTF_{FL}, Figure 2.6) in high purity as determined by SDS-PAGE and LC-MS ([M+H]⁺, calc.14150 obs.14150). (Figure 2.7 and see section 2.4.1.1. for detailed experimental procedure).

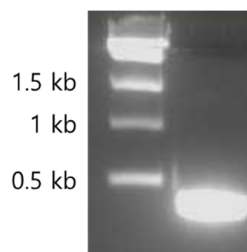


Figure 2.4. PCR-amplified DNA for NTF_{FL}-thioester. Lane 1: molecular DNA marker; lane 2: PCR-amplified DNA product. The band containing the product was gel-purified and cloned into pTXB3 (NEB).

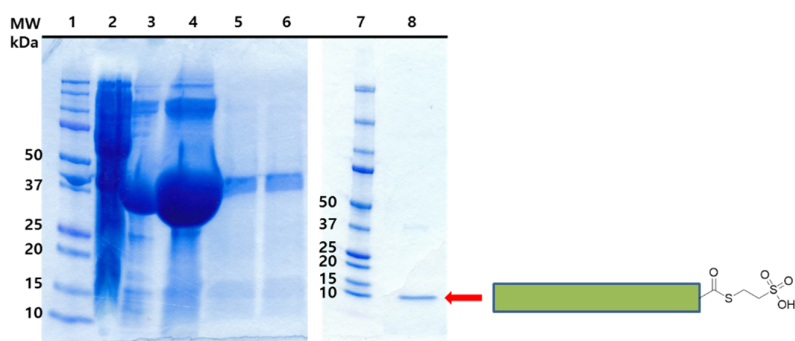


Figure 2.5. Purification of NTF_{FL}. An NTF_{FL}- α thioester was obtained by on-column cleavage from a chitin affinity column with MESNa. The preparation was monitored by using SDS-PAGE (Any kD, and 4-15% gel, Bio-Rad). Lanes 1 and 7: molecular mass markers; lane 2: cell lysate cell debris before separation of supernatant and pellet); lane 3: cell extract supernatant; lane 4: cell extract pellet; lane 5: flow-through from the chitin affinity column; lane 6: wash; lane 8: NTF_{FL} (red arrow, 14.1 kDa) eluted from the column under gravity flow.

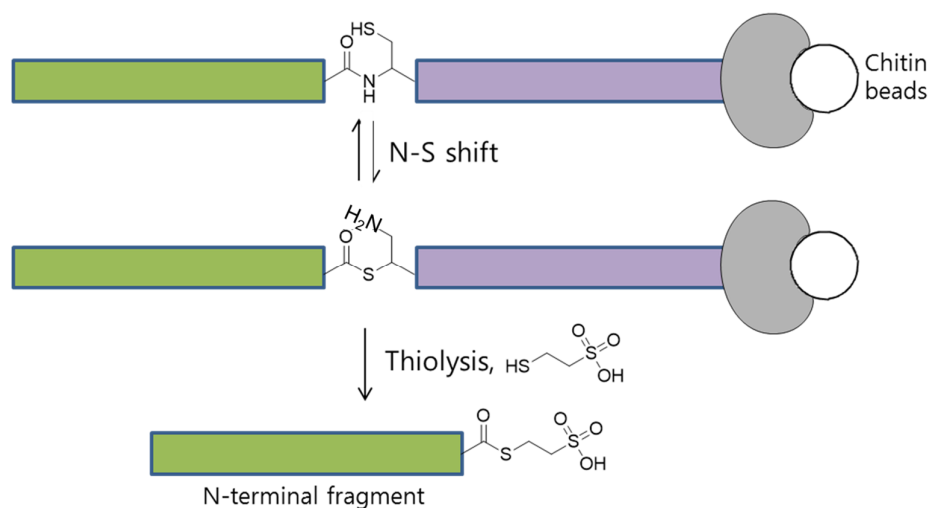


Figure 2.6. Schematic representation of intein-mediated cleavage by a thiol to prepare NTF[Ala¹-Gly¹²⁶]-αthioester (NTF_{FL}). A linear thioester intermediate at the N-terminus of the intein is formed by an N-S acyl rearrangement at Cys. Incubation with a thiol compound, 2-mercaptoethanesulfonic acid (MESNa) results in cleavage of peptide bond generating the linkage of MESNa to the C-terminus of the target protein by a thioester bond. The light purple and grey domain represents the *Mxe* GyrA intein and chitin binding domain, respectively. The white circle represents chitin beads used in a gravity column.

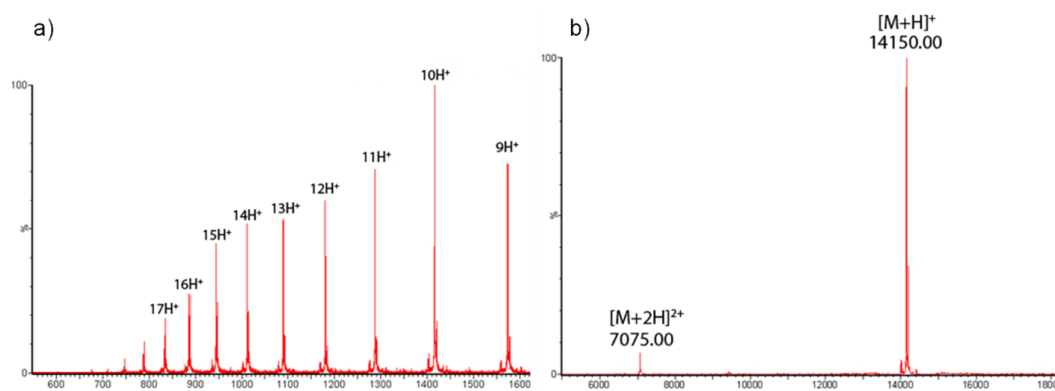


Figure 2.7. Characterization of purified NTF_{FL}-αthioester. a) ESI-MS spectrum of NTF_{FL}-αthioester. b) Deconvoluted mass spectrum of NTF_{FL}-αthioester ([M+H]⁺ = 14150 Da (obs), 14150 (calcd.)).

2.2.2.2. Preparation, purification and characterization of CTF

Current approaches to generate N-terminal Cys proteins involve cleavage of a precursor fusion protein with a site-specific protease to expose an internal cysteine residue as the new N-terminus^{35, 42, 43}. With these methods, CTF bearing an N-Cys has been produced from a fusion protein. For example, fusion proteins containing a recognition site of a tobacco etch virus (TEV) protease or a small ubiquitin-related modifier (SUMO) protease recognition site on their upstream are digested to generate N-Cys. These methods have found to be effective for generating N-terminal cysteine residues^{7, 44-46}. However, it is not appropriate to apply these methods to the proteins obtained from inclusion bodies because the denaturing condition used to purify the proteins would disrupt the enzymatic activity of the protease. Also, an additional step is required to purify the product away from the protease and the cleaved recognition domains.

Thus, a preferable approach to the production of N-terminal Cys proteins would be to take advantage of endogenous methionine aminopeptidase⁴⁷ to cleave the initiating methionine from recombinantly expressed proteins containing cysteine as the second residue. Met is often removed cotranslationally before the completion of synthesis or folding of polypeptide chains⁴⁸⁻⁵⁰.

The catalytic activity of methionine aminopeptidase varies depending on the identity of the adjacent residue to the initiating methionine⁵¹. When the penultimate amino acid bears a small side chain, the initiating methionine is excised efficiently, whereas, when the amino acid has a large side chain, the methionine is not cleaved. For example, the initiator methionine is completely removed when it precedes Val, Gly, Ala, Ser, Cys, and Thr⁵². As a proof, the observed mass of NTF in Figure 2.7 shows that N-terminal Met bearing Ala on the next amino acid residue was found to be cleaved (in Figure 2.7).

To make a plasmid for CTF_{FL} preparation, the codons for residues 1-126 in pT7-SC1 were replaced with codons for Met-Cys by PCR. The residues 1-126 were replaced by the first homologous recombination (Figure 2.8a) and the D₈H₆ tail was introduced at the C-terminus by the next round of homologous recombination. (Figure 2.8b)

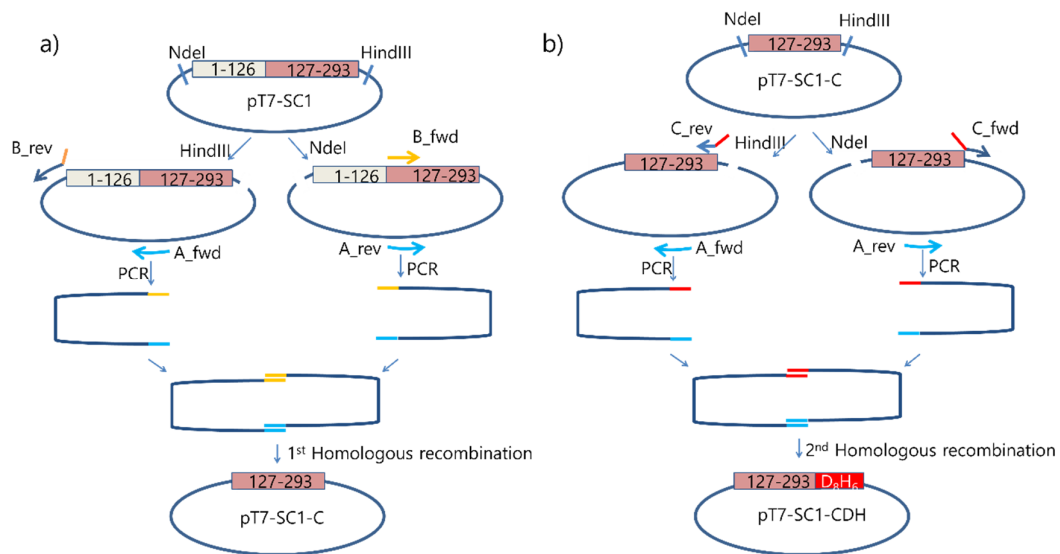


Figure 2.8 Preparation of the CTF (127-293)-D8H6 plasmid (pT7-SC1-CDH). a, b) Preparation of plasmids containing genes for CTF by homologous recombination. a) The plasmid pT7-SC1 containing the gene for WT α HL was separately treated with two restriction enzymes (*NdeI* and *HindIII*). The A_fwd and B_rev primers were used to obtain the left-hand fragment by PCR template of the *HindIII*-digested plasmid, while B_fwd and A_rev were used to obtain the right-hand fragment from the *NdeI*-digested plasmid. The two fragments, which have homologous ends, were mixed and recombined *in vivo* to generate a circular plasmid (pT7-SC1-C) containing a CTF gene initiated with Met-Cys codons. b) A second homologous recombination with A_fwd/C_rev and C_fwd/A_rev primers yielded the plasmid pT7-SC1-CDH containing the gene for CTF with a C-terminal D₈H₆ tail.

After the initiating methionine residue is cleaved from a protein expressed in *E. coli*, however, the N-terminal cysteine residue undergoes a condensation reaction with pyruvic acid, which is an abundant metabolite in *E. coli* to generate a thiazolidine product. This *in vivo* chemical modification of the N-terminal amine can be chemically reversed with methoxyamine hydrochloride (MeONH₂·HCl) at low pH. (pH ~3.5)⁴¹ (Figure 2.9).

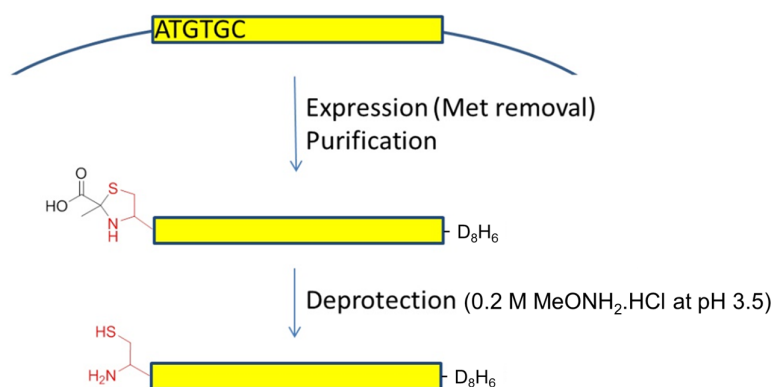


Figure 2.9. Schematic representation of direct preparation of CTF (127-293) from *E. coli*.

The cysteine residue for NCL was placed as a penultimate residue. The methionine was removed during post-translation modification and the cysteine exposed at the N-terminus reacted with one of the most abundant metabolites in the cell, pyruvate. The pyruvate group was removed by 0.2 M MeONH₂·HCl in pH 3.5.

Under the same conditions used for the preparation of NTF, the C-terminal fragment was expressed in *E. coli* and obtained within the inclusion bodies. The CTF was purified by the use of a Ni-NTA affinity column by using His₆ tag placed at the C-terminus (Figure 2.9 and 2.10). The purified product was analyzed with SDS-PAGE and characterized with LC-MS. As observed in the preparation of NTF, the N-terminal Met was found to be cleaved. The unmasked Cys residue had undergone condensation with pyruvic acid, an abundant metabolite, yielding a species +70 Da over in molecular mass (Figure 2.11a). Free N-Cys was regenerated by demasking the

thiazolidine ring (Thz) with 0.2 M methoxylamine hydrochloride for 3 h at room temperature. The purified CTF was characterized by LC-MS ($[M+H]^+$, calc: 20973, obs: 20973, Figure 2.11 b).

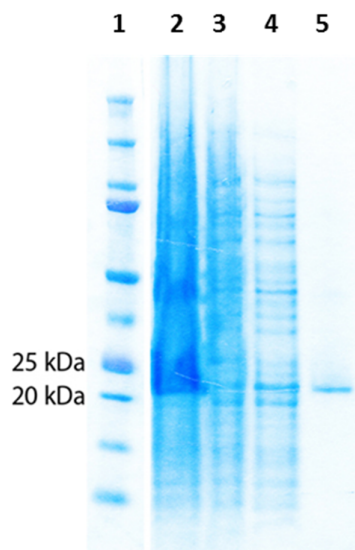


Figure 2.10. Purification of CTF visualised on SDS-PAGE. CTF was purified using a His₆-tag at the C-terminus in a gravity column. Lane 1: molecular marker; Lane 2: Cell lysate; Lane 3: flow-through after the pellet dissolved in the same buffer was applied into the column; Lane 4: wash Lane 5; eluted target peptide.

2.2.2.3. NCL

The NCL reaction was conducted using 50 μ L of \sim 0.3 mM NTF and 50 μ L of \sim 0.2 mM CTF in the NCL buffer (6 M Gu.HCl, 0.2 M sodium phosphate (pH 7.0), 5 mM TCEP, and 200 mM MPAA). The reactants were mixed with 400 μ L of the NCL buffer and the reacting solution was concentrated in 50 μ L using a centrifugal filter (Amicon, MWCO 3k) at 15,000 g for 10 min. The mixture was allowed to react overnight at room temperature. The resulting crude product was directly analysed with SDS-PAGE (Figure 2.12a).

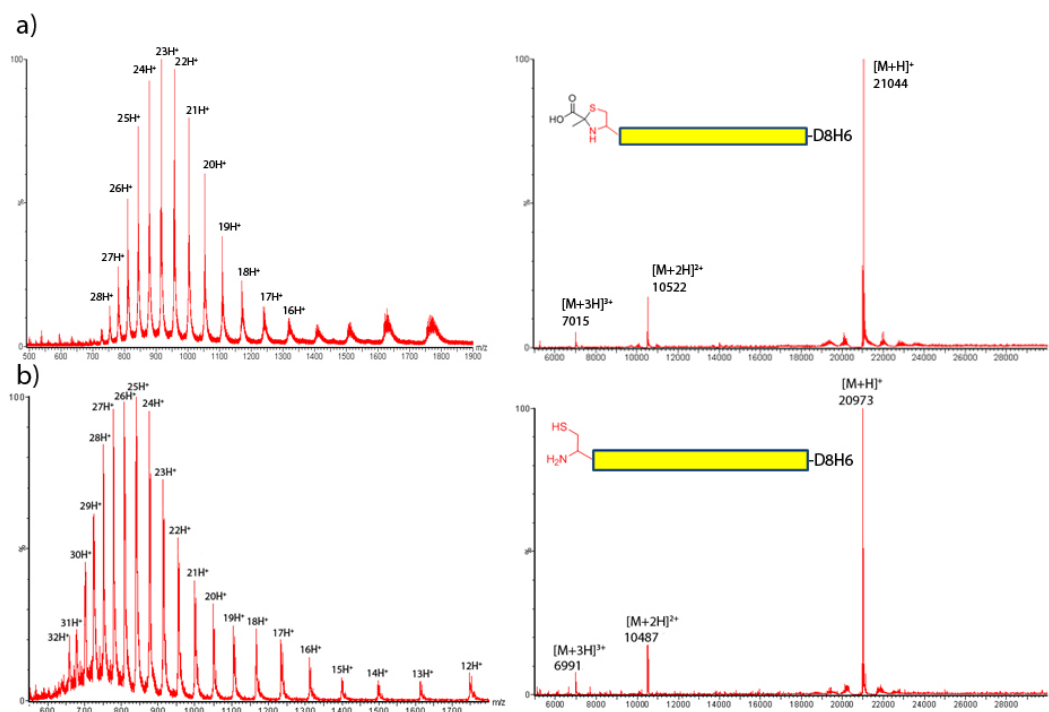


Figure 2.11. Characterization with LC-MS of the purified CTF. (a) CTF with pyruvate-protected N-Cys ($[M+H]^+$, calc: 21043, obs: 21044). (b) CTF after removal of pyruvate to generate a N-terminal cysteine ($[M+H]^+$, calc: 20973, obs: 20973).

2.2.2.4. Purification and characterization of the synthetic monomer

The synthetic monomer was purified with a size exclusion column under denaturing conditions (8 M urea) to prevent oligomerization of the synthetic monomer to a heptameric pore. The fractions collected from the first peak in the gel filtration chromatogram (Figure 2.12b) had the mass expected of the synthetic monomer (calc: 34981, obs: 34983) (Figure 2.12c). The intensity of the polypeptide band in a SDS gel (Figure 2.12a) was quantified (inset in Figure 2.12) by using ImageJ (NIH), which gave a yield of ~ 48 %.

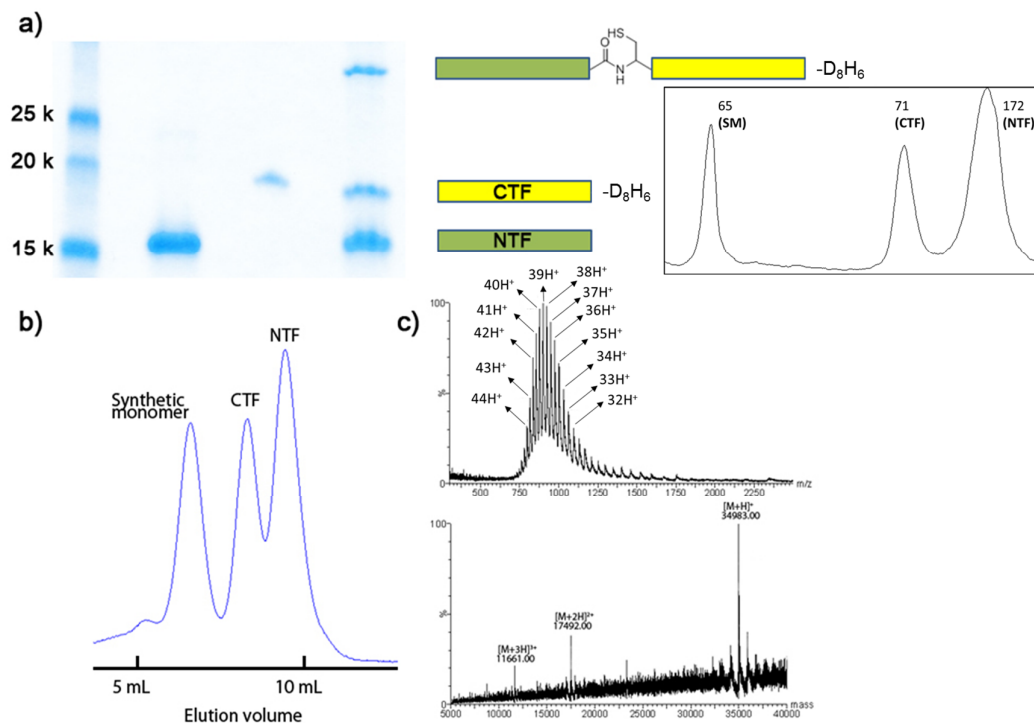


Figure 2.12. Preparation and purification of a full length α HL monomer (SM_{FL}). a) SDS-PAGE analysis for crude NCL product. Lane 1: molecular marker, Lane 2: NTF, Lane 3: CTF, Lane 4: crude product after the ligation reaction. A new band with a corresponding mass to the α HL monomer (~35 kDa) has been generated. b) Purification of the synthetic monomer with gel filtration. c) LC-MS analysis of the purified synthetic monomer.

2.2.2.5. Hemolysis assays for SM_{FL}

Hemolytic assay was conducted to verify whether the synthetic monomer regains a functional activity upon the synthesis condition and purification condition using the denaturants (6 M Gu.HCl and 8 M urea, respectively). To renature the purified synthetic monomer dissolved in 8 M buffer urea, the protein in the 8 M urea was diluted to contain 1 M urea by the addition of buffer including no urea. The diluted protein was then concentrated to the original volume to have the same concentration. The hemolytic activity was analyzed by collecting light scattering data points every 30 s

over a period of 100 min at 595 nm. Active α HL monomers are able to assemble to form a heptameric pore on the lipid bilayer of rabbit red blood cells (rRBCs) and lyse the cells. As the amount of cells lysed increases, the light scattering decreases, which results in a sigmoidal decrease in the apparent light scattering (Figure 2.13).

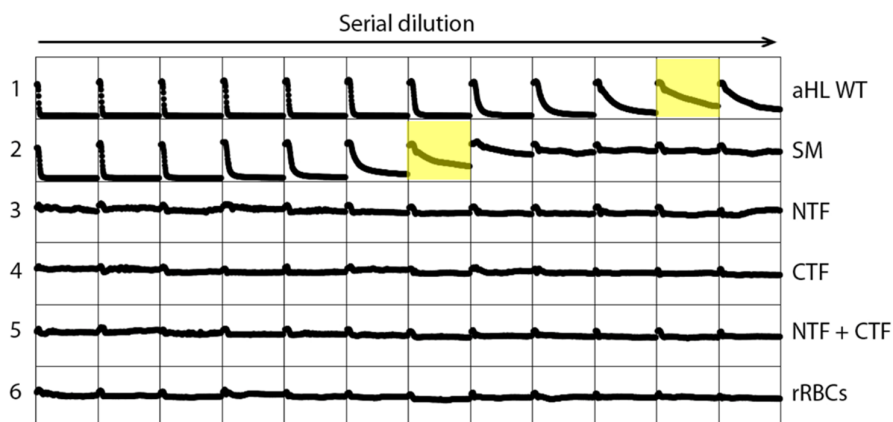


Figure 2.13. Hemolytic assay of the synthetic monomer. The synthetic monomer regains hemolytic activity after synthesis and purification under denaturing conditions. Each well in each lane was serially diluted (2-fold) from the preceding lane. WT α HL monomer, synthetic monomer, NTF, and CTF were added to each lane as follows. Lane 1: WT α HL monomers (1 μ L, 2.17 mg mL⁻¹), Lane 2: purified synthetic monomers (1 μ L, 0.59 mg mL⁻¹), Lane 3: NTF (1 μ L, 0.83 mg mL⁻¹), Lane 4: CTF (1 μ L, 0.64 mg mL⁻¹), Lane 5: NTF (1 μ L, 0.83 mg mL⁻¹) and CTF (1 μ L, 0.64 mg mL⁻¹) were added at the same well. Lane 6: rRBCs only. The specific hemolytic activity was obtained from the amount in the wells in light yellow.

In order to obtain the efficiency of refolding of the synthetic monomer quantitatively, the specific hemolytic activity (HC_{50} , amount of protein that gives a 50 % of lysis in a given time at room temperature) was calculated and compared to WT α HL monomer. As it is difficult to measure the concentration of IVTT product due to the very low yield of IVTT (~ 5 ng μ L⁻¹), WT α HL produced in *E. coli* was used (Figure 2.14).

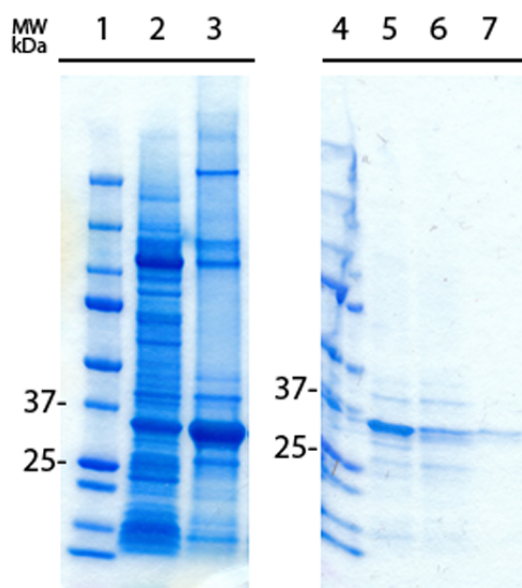


Figure 2.14. Preparation and purification of WT α HL-H6 for quantitative analysis of hemolytic activity. Lane 1 and 4: molecular makers; lane 2: supernatant of cell lysate; lane 3: pellet of cell lysate; lane 5: flow-through from the Ni-NTA affinity column after the pellet was applied; lane 6: column washings; lane 7: elution of recombinant α HL-H6 monomer.

The concentration of protein was obtained by measuring the absorbance at 280 nm by a spectrophotometer (ND-1000, Nanodrop) (Section 2.4.1.6, Table. 2.1). A known amount of each protein was then added to the first well and serially diluted by 2-fold at each well. The wells highlighted in light yellow in Figure 2.10 were chosen as the amount required to give 50% of lysis in 30 min and HC_{50} for the synthetic monomer was found to be similar to that of WT α HL (literature value⁵³: 31 ng mL⁻¹, WT α HL: 21 ng mL⁻¹, and SM: 93 ng mL⁻¹). This similar specific hemolytic activity of the synthetic monomer to lyse the rRBCs demonstrated that the synthetic monomer prepared by NCL of two recombinant polypeptides was correctly refolded to a heptameric pore and was able to regain its functionality after being unfolded in denaturants such as 8 M urea and 6 M Gu.HCl.

2.2.2.6. Heptameric pores with synthetic monomer

2.2.2.6.1. Oligomerization to heteroheptamer (WT₆SM₁)

Since it is known that α HL can easily oligomerize to a heptameric pore in the absence of a membrane³⁹, the synthetic monomer was never exposed to non-denaturing condition in any synthetic process. Moreover, previous studies⁵⁴ have shown that α HL IVTT monomer denatured in 1-2 M urea does not affect its refolding ability and hemolytic activity. Thus, 10 μ L of the synthetic monomer (0.59 mg mL⁻¹) in 8 M urea was added to 70 μ L of TE8 which diluted the concentration of urea to 1 M and the synthetic monomer to 75 ng μ L⁻¹. The diluted synthetic monomer was directly mixed with WT α HL expressed by IVTT. The ratios of mixture varied from 7:0 to 1:6 (WT:SM) in the basis of the assumption that ~250 ng of α HL is expressed in 50 μ L reaction scale (5 ng μ L⁻¹). The mixtures were incubated with rRBCm loaded into 5 % gel after being mixed with the same volume of loading dye. The gel was vacuum-dried, exposed to an autoradiography film, and developed to visualize the proteins.

Attributed to the synthetic monomer that contains octa-aspartate (D₈) tag at the C-terminus, heptameric pores assembled with different numbers of synthetic monomer subunits exhibit different mobility on a SDS gel. For example, Lane 1 in Figure 2.15 shows the homoheptameric pore (WT₇) formed with no subunits containing a D₈ tag, which has the slowest mobility of the protein. From Lane 2 to 5, as heteroheptameric pores were assembled with more synthetic monomers, the mobility on the gel increased compared to others. Each of the heteroheptamers was assigned by counting the number of WT subunit from IVTT. The characterization of one of the homoheptamers will be discussed in the section 2.2.2.7.

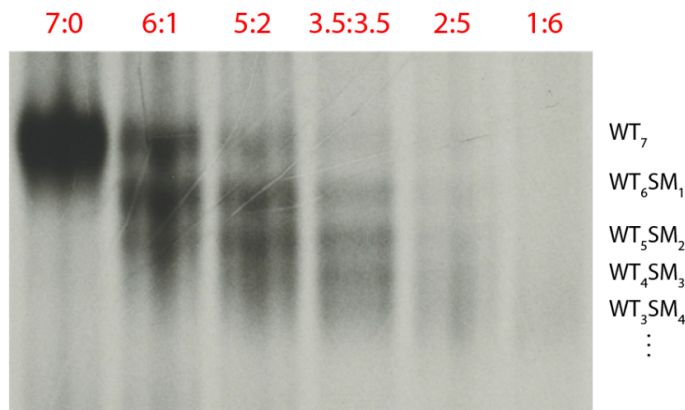


Figure 2.15. WT_{7-n}SM_n Heteroheptameric pores. The heptameric pores were formed with different stoichiometries (WT₇, WT₆SM₁, WT₅SM₂, WT₄SM₃, and WT₃SM₄ were shown) with the synthetic monomer and WT α HL expressed with IVTT. By varying the mixture ratios from 7:0 to 1:6 (WT:SM), various heptameric pores with different stoichiometry of subunit (WT_{7-n}SM_n, n=0-7) were obtained.

2.2.2.6.2. Oligomerization to homoheptamer (SM₇)

The homoheptamer (SM₇) can be prepared by adding a membrane to trigger the assembly of monomer. However, if rRBCm is used for oligomerizing the monomer, it may be unclear to analyze the protein bands in a stained gel because unknown proteins contained in rRBCs are also stained. To avoid this problem, liposomes made of DPhPC (1 mg mL⁻¹) (section 2.4.1.10 for preparation of liposome) were used to oligomerize the monomer instead of rRBCm.

2.2.2.6.3. Limited proteolysis

To determine the conformations of the synthetic monomer to a heptameric synthetic pore and to verify the synthetic monomer refolds through a 'prepore' intermediate state, a proteolysis experiment was carried out with proteinase K, which has previously been shown to cleave α HL in solution at two different sites (Figure 2.16)

The mechanism of membrane proteins forming a pore has been explained as a prepore formation^{53, 57, 58}. Particularly, for assembly of α HL to a heptameric pore, two intermediates have been proposed. i) a membrane-bound monomer (**2** in Figure 1.2) and ii) an impermeable oligomeric prepore (**3** in Figure 1.2). As depicted, while the monomer assembles to a heptamer, a conformational change occurs and the glycine-rich loops (109-147, red in Figure 2.16) are buried in the lipid bilayer.

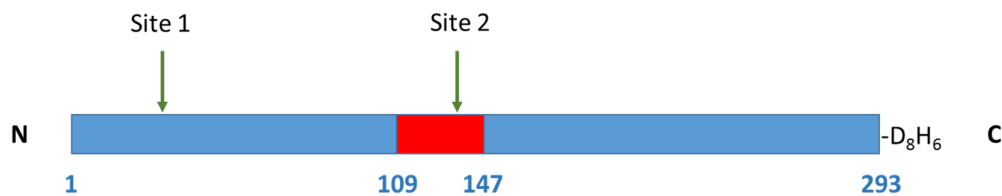


Figure 2.16. Two digestion sites of proteinase K on α HL monomer. The green arrows are predicted proteinase K digestion sites in the SM monomer⁵⁸ in solution. On binding to a bilayer, the sites in the transmembrane region (red) become occluded in the prepore structure⁵⁹. In the fully assembled pore, all the proteinase K sites are occluded.

When the monomer was treated by proteinase K, it was digested to 16 and 18 kDa fragments (Lane 2, orange and green arrows in Figure 2.17). In the presence of liposome, most of the monomers are fully protected by lipid bilayer (Lane 4). The proteinase K only cleaves a short N-terminal domain, therefore, a digested fragment missing the short N-terminal domain is visualized on SDS gel when the heptameric prepore is disassembled into monomer (small yellow arrow, Lane 4 in Figure 2.17).

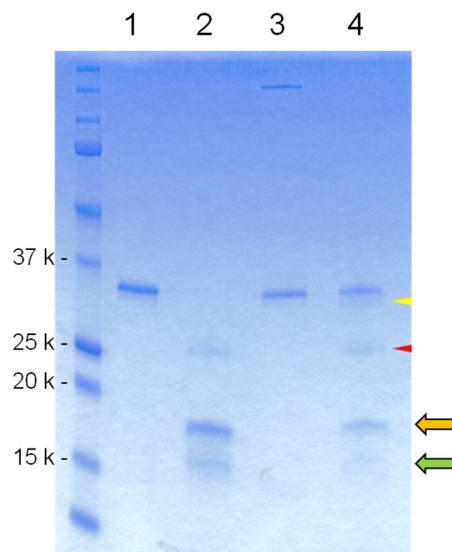


Figure 2.17. Limited proteolysis for SM₇ pores. Lane 1: the purified synthetic monomer. Lane 2: the synthetic monomer treated with proteinase K (small red arrow). The synthetic monomer was digested with proteinase K at both site 1 and 2. The synthetic monomer cleaved at the red site produces two fragments (large orange and green arrow). The other small fragment cleaved at the N-terminal site is not seen on the gel due to its low molecular weight. Lane 3: The synthetic monomer is oligomerized with DPhPC vesicles and a heptameric pore is produced at the top. Lane 4: The heptameric pore treated with proteinase K is disassembled to a monomer by heating. The central proteolytic site remains intact. Due to the conformation of prepore, the proteolysis is only allowed at the N-terminal site and this results in generating a protein (small yellow arrow) losing a short N-terminal peptide. (Figure 2.16b and c)

2.2.2.7. Single-channel recordings of semisynthetic pores

2.2.2.7.1. Inserting WT₆SM₁ pores

Once a stable bilayer was formed, 0.5 μL of aliquoted protein ($\sim 1 \text{ ng } \mu\text{L}^{-1}$) was added to the *cis* compartment and the insertions of the synthetic pores were directly observed under an applied potential which shows a step-wise change in the current (Figure 2.18). As the insertion of synthetic pores continued, the conductance state was

shifted in a staircase manner for each insertion. Each of the current steps was analyzed with an event histogram (Figure 2.19) and used to calculate the average conductance of the WT₆SM₁ pores. The conductance calculated as current/voltage was 0.79 ± 0.09 nS ($n=130$), which is similar to the conductance (~ 0.80 nS) of WT α HL pores prepared by IVTT under the same SCR condition (in 1 M KCl, Tris (pH 7.0) at +50 mV).

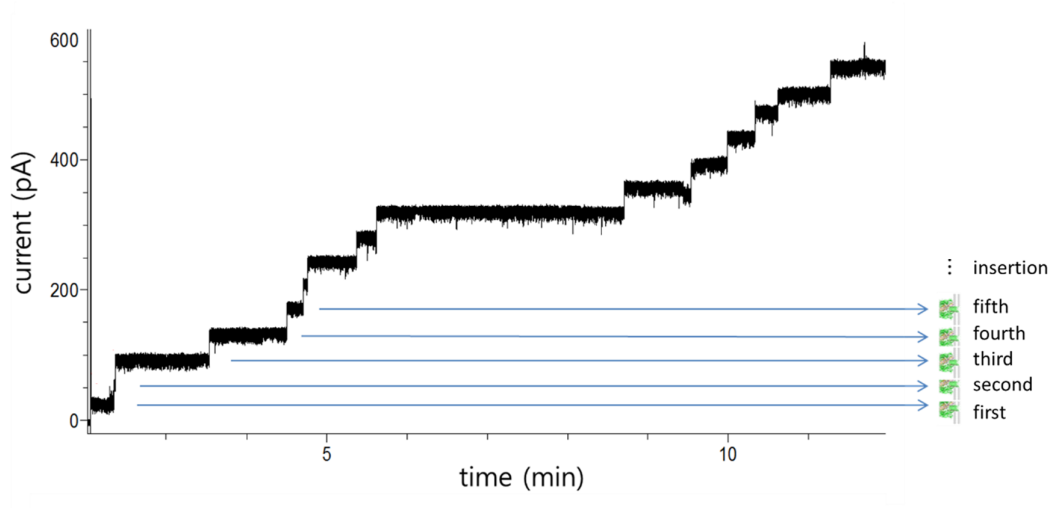


Figure 2.18. Multichannel insertions of the synthetic pores. The conductance state changes in stepwise at each of insertions of the pore into the bilayer under a given potential (+50 mV). (Sampling 25 kHz, Filtering 5 kHz)

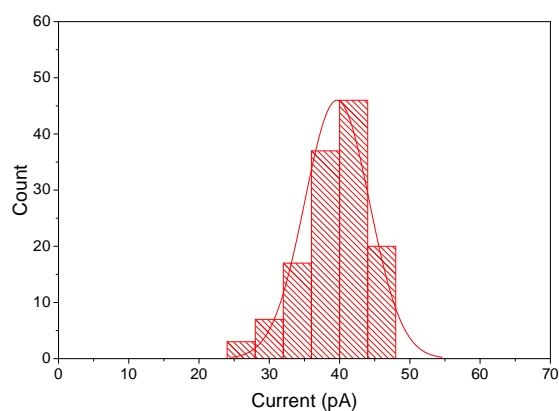


Figure 2.19. Count histogram of ionic currents corresponding to single synthetic pore insertions. The mean current and conductance was calculated as 39.6 ± 4.6 pA ($n=130$) at +50 mV and 0.79 nS (1 M KCl, 20 mM Tris (pH 7.0)), respectively.

2.2.2.7.2. I-V curve of individual WT₆SM₁ pores

The episodic stimulation protocol was used to produce I-V (current-voltage) curve for the synthetic pore. While the voltage was changed from -100 mV to +100 mV with 10 mV increase, the current signal was acquired for 0.5 s at each given voltage. (Figure 2.20). The mean currents acquired for 0.5 s were obtained from 5 individual pores and fitted in a curve. Due to the rectification⁶⁰ of α HL pore, the ionic flow at higher negative potential shows smaller current than the current at positive potential.

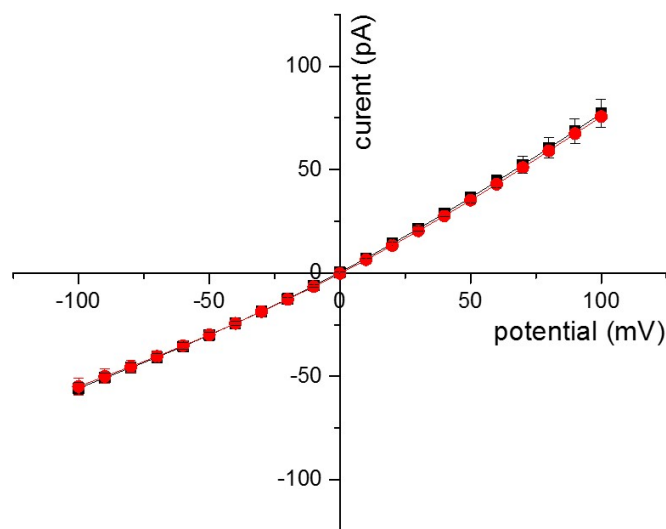


Figure 2.20. Single-channel IV curves of WT₆(SM_{FL})₁ and WT αHL pores. WT₆(SM_{FL})₁ (●), (red, n=4) and WT αHL pores (■) prepared by IVTT (black, mean values, n=3) show similar conductance under the applied potential (-100 - +100 mV).

2.2.2.7.3. Kinetics of β-cyclodextrin binding to WT₆(SM_{FL})₁ pores

β-cyclodextrin (βCD) is known as a noncovalent molecular adapter⁶¹, which resides in the lumen of αHL for several hundred microseconds. The molecule acts as a channel blocker when it becomes lodged in the lumen of the αHL pore, thus, the investigation of βCD binding within the pore may provide evidence that the synthetic monomer folds properly to form a rigid lumen structure composed of 14 strands in a β-sheet.

Once a single pore had inserted, the *cis* compartment (500 μL) was perfused by manual pipetting to reduce the protein concentration. This step consists of removing the protein-containing buffer and replacing it with a protein-free buffer (section 2.4.4.2. for detail experimental protocol), thus, reducing the probability of subsequent protein insertions.

After the perfusion, β CD was added to the *trans* compartment at +150 mV and the expected host-guest interactions caused by binding of a cyclodextrin within the lumen of the α HL pore was observed (Figure 2.21). The binding was shown as a transient reduction in current flowing through the pore. The increase of β CD concentration (25 to 100 μ M) in the *trans* compartment resulted in an increase of binding events, which indicated that the binding event was generated by the presence of β CD.

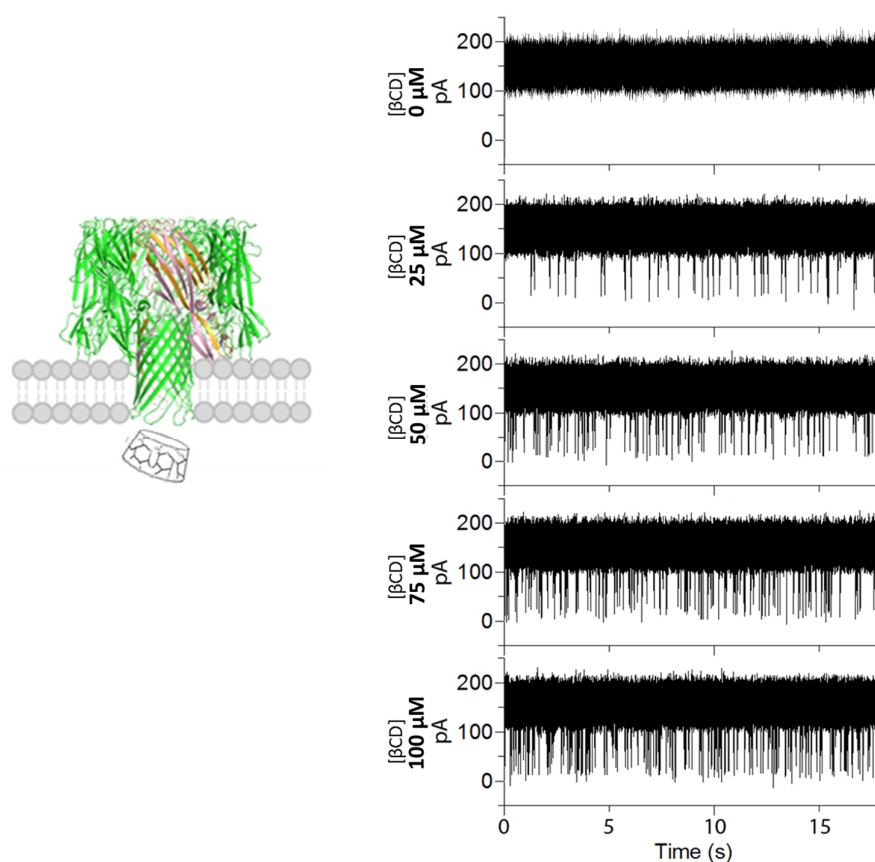


Figure 2.21. Binding of β CD into the pore containing a single synthetic monomer (WT₆SM₁). The pore does not exhibit any binding events before the addition of β CD in the *trans* side, whereas increase in β CD concentration from 25 to 100 μ M results in an increase in event frequency, indicating that β CD binds to the inside pore and the ionic current transiently is blocked.

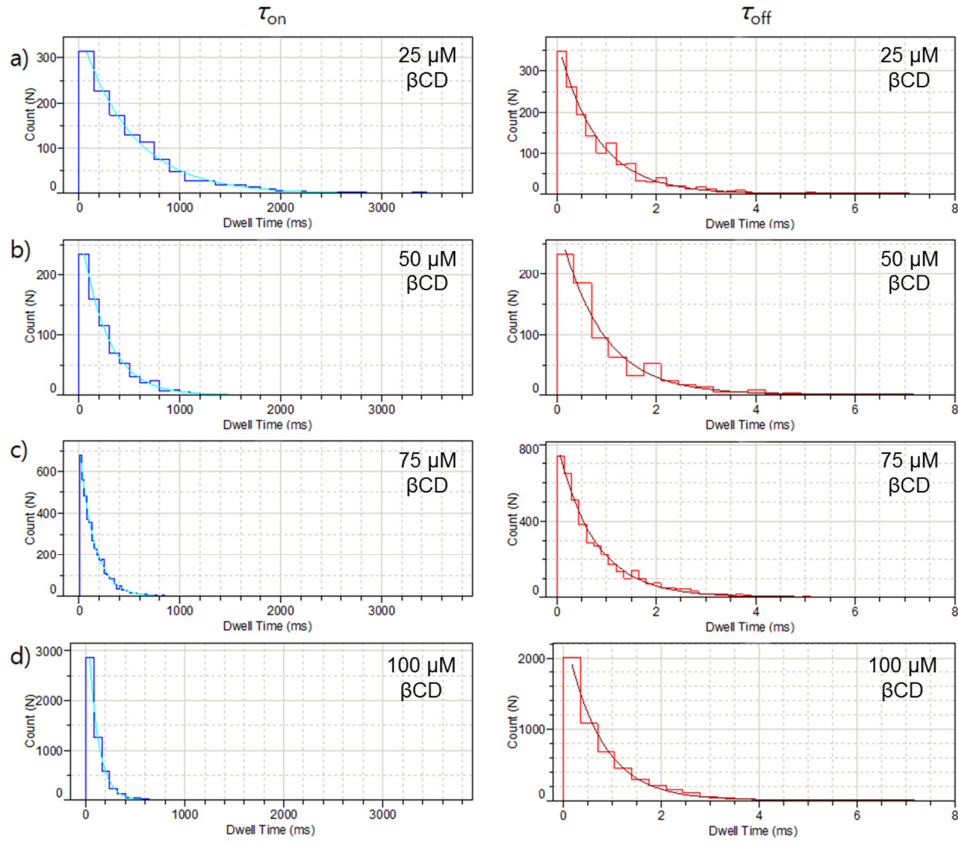


Figure 2.22. Dwell time histogram for β CD binding events at different concentration of β CD. The blue and red histograms represent the durations of open and partially closed pores with β CD, respectively. From the mean value of the exponential fit, the mean τ_{on} and the mean τ_{off} can be obtained. The dwell time of β CD at given concentration is obtained by fitting the histogram to a single exponential function. As the concentration of β CD increases from 25 to 100 μ M by 25 μ M (from a to d), τ_{on} (in blue curve) decreases while τ_{off} (in red curve) remains constant. τ_{on} and τ_{off} obtained are a) 373 ± 7 and 0.79 ± 0.03 ms b) 259 ± 6 and 0.86 ± 0.05 ms c) 145 ± 2 and 0.74 ± 0.01 d) 99 ± 0.8 and 0.72 ± 0.02 ms. The amplifier's internal low-pass Bessel filter was set at 5 kHz. Data were acquired at a sampling rate of 25 kHz.

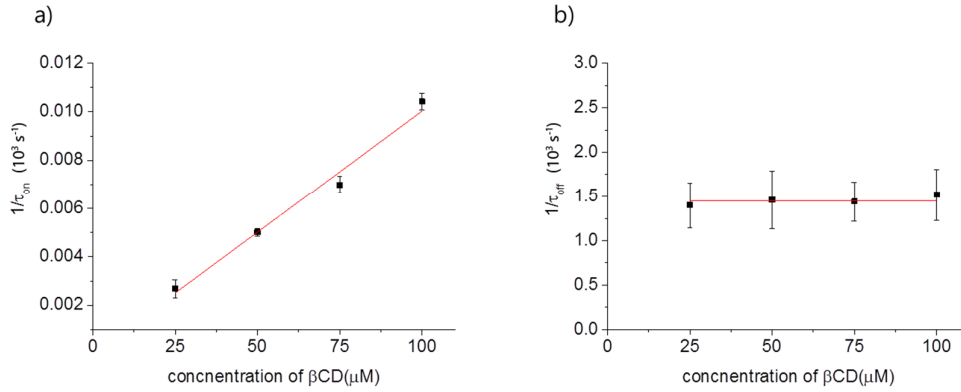


Figure 2.23. Kinetic analysis of βCD binding in the synthetic pore. The association (k_{off}), dissociation rate constant (k_{on}) were determined from plots of $1/\tau_{on}$ and $1/\tau_{off}$ versus $[\beta\text{CD}]$ using linear fit tools in Origin 9.1. Dissociation constants (K_d) were obtained from the rate constants and error (s.d.) propagated from the errors in k_{on} and k_{off} from three independent experiments ($Z = X/Y$, $\Delta Z = Z\sqrt{(\Delta X/X)^2 + (\Delta Y/Y)^2}$). $k_{off} = 1.45 \pm 0.02 \times 10^3 \text{ s}^{-1}$, $k_{on} = 1.00 \pm 0.02 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$, and $K_d = 1.45 \pm 0.04 \times 10^{-2} \text{ M}$. The analysis was carried out under 1 M KCl, 20 mM Tris (pH 7.0) at 150 mV.

To determine the kinetic constants from the βCD binding measurements, data was acquired from three independent experiments, with more than 1000 events per each experiment. τ_{on} and τ_{off} values for βCD were obtained at +150 mV with varying concentration (25, 50, 75, and 100 μM βCD) to generate dwell-time histograms (Figure 2.22) that were fitted to single exponentials by the Levenberg-Marquardt procedure⁶². Each data set from individual experiment yielded similar τ_{on} and τ_{off} values, indicating the stationary kinetics prevailed. Kinetic constants were calculated by using $k_{off} = 1 / \tau_{off}$, $k_{on} = 1 / \tau_{on} [\beta\text{CD}]$, and $K_d = k_{off} / k_{on}$, where $[\beta\text{CD}]$ was the concentration of βCD (Figure 2.23, see section 1.2.7 and Figure 1.6 for derivation of the equation). Values for unitary conductance are: $k_{off} = 1.45 \pm 0.02 \times 10^3 \text{ s}^{-1}$; $k_{on} = 1.00 \pm 0.02 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$; and $K_d = 1.45 \pm 0.04 \times 10^{-2} \text{ M}$. Similar values were obtained previously for the WT αHL pore⁶³.

2.2.2.7.4. Chemistry of thiolate

As a result of NCL, the ligation product will have a cysteine residue at the ligation junction. Because cysteine is the second least common of the 20 amino acids in proteins (the first least common amino acid: Trp)⁶⁴ and α HL contains no native cysteines, introducing a non-native cysteine into α HL was carried out in the basis of previous study⁶⁵. Asp127 was chosen for mutation to Cys, because the cysteine mutation at residue 127 has been found to be tolerated⁶⁵ to the function of α HL and the side chain is directed inside the α HL pore, which may be applicable to covalent chemistry.

The existence of a native amide bond at the ligation site, generated after the synthesis was verified by thiolate chemistry. In the absence of PEG-OPSS (Figure 2.24b), the open state had a long duration, and lasted without any significant binding events. However, the addition of PEG-OPSS (1 mM) to the *trans* side of the bilayer produced reversible partial channel blockades (Figure 2.24c). A few seconds after addition of the reagent, without stirring, the current became noisy with short spikes (level 0), which most likely represented noncovalent interactions between the pore and substrate. Subsequently, the current dropped from level 0 to level 1 (Figure 2.24c); the magnitude of the drop was 86 %, which is 23 % larger than the amplitude of the blockades by free PEG-0.94 k reacting with the sulfhydryl group at position 117 within the lumen of the transmembrane domain of the channel, but similar to the extent of current drop when MePEG-OPSS(6.0 kDa) was added to the pore⁶⁶. Interestingly, after ~ 2 min, a new downward shift in the single-channel current occurred to level 2 (Figure. 2.24d) and the current repeatedly changes from level 1 to level 2 and vice versa (Figure. 24e). This result was presumably due to the large size of PEG bound at the loop perturbing the ionic flow by clogging and opening the pore reversibly. However, when 5 mM DTT was added to the chambers (both the *cis* and the *trans* compartment),

the pore opened back restoring the current to the level 0, as the PEG chain was cleaved by DTT from the Cys-127 residue.

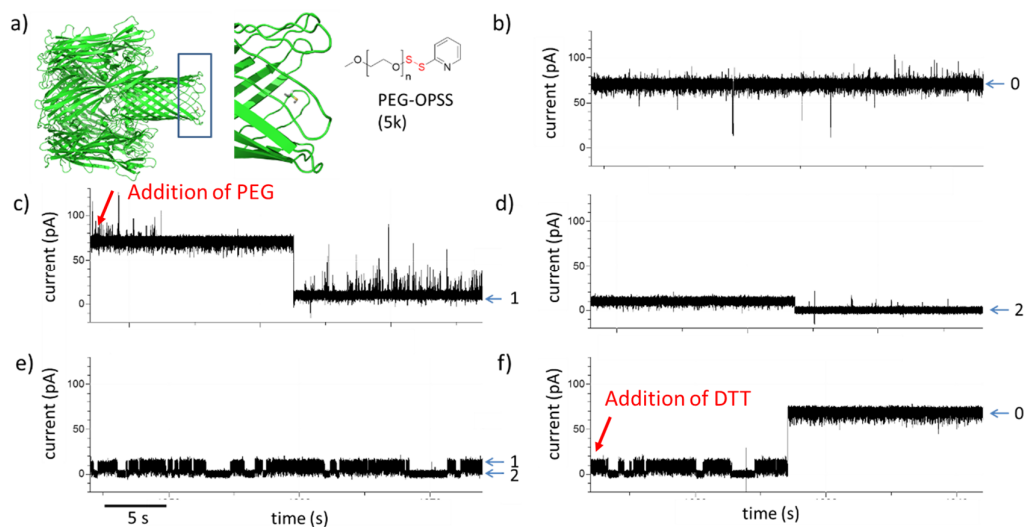


Figure 2.24. Single-channel recordings of the α HL pore in the presence of PEG-OPSS (5 k Da). a) Representation of the synthetic pore (WT₆SM₁) containing a free cysteine at the junction of NCL. The free thiolate reacts with substrate by forming a disulfide bond with PEG-OPSS. b) Addition of PEG-OPSS (1 mM) into open pore. c) The ionic current dropped to level 1 as PEG bound to the synthetic pore. d) Subsequent current decrease to level 2 was attributed to complete clogging at the bottom of pore by attached PEG. e) The current level continuously changed from level 1 and level 2 presumably because of fluctuation of the attached PEG at the bottom of the pore. f) The current level comes back to open state (level 0) after the addition of 5-fold DTT to the pore. All measurements were carried out at potential of +100 mV and in 1 M KCl, 20 mM Tris, pH 7.5, at room temperature.

2.2.3. Semisynthetic truncated barrel mutant (TBM) pores

TBM Δ 6 is a version of the truncated barrel mutant pores of α HL containing a shorter lumen structure than WT. For TBM Δ 6, 12 amino acids (F120-T125 and G133-A138) from each β strand (Figure 2.25) are removed³⁸. The truncated barrel mutant

has attracted interest as it can form a conductive channel despite the length of β barrel not being long enough to span the entire lipid bilayer. Previous studies have proposed the stabilization of the truncated pores through toroidal characteristics of the membrane³⁸.

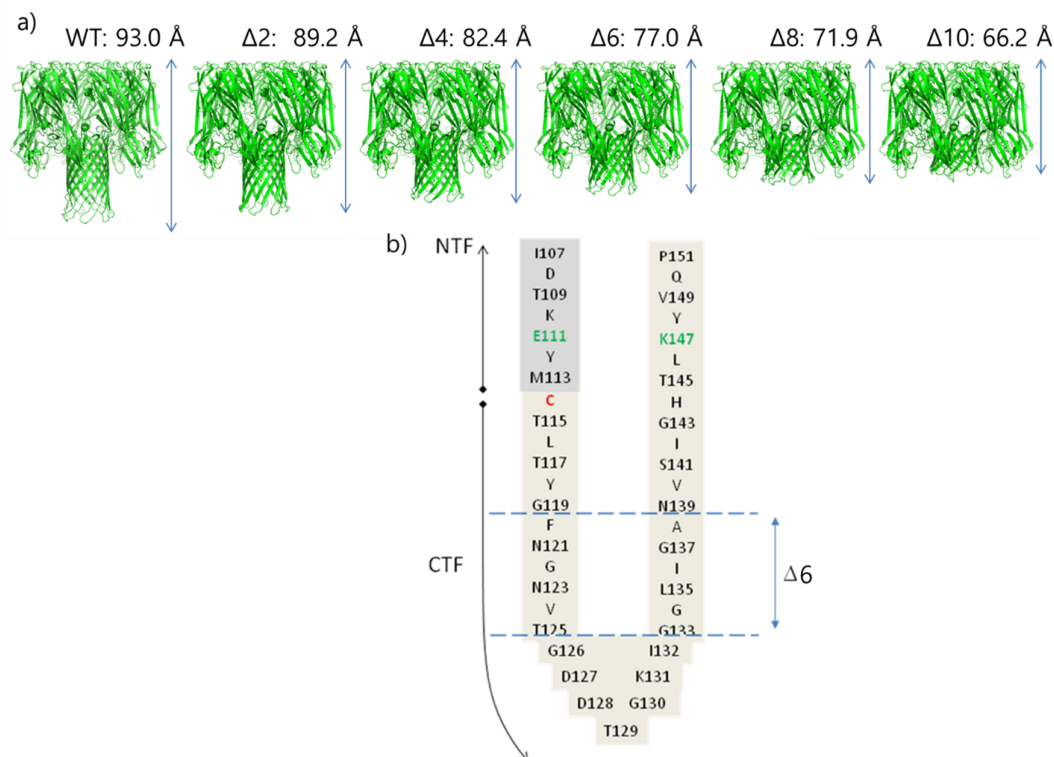


Figure 2.25. Truncated barrel mutants and sequences in the β strand. a) Cut-through representations of the WT and TBM pores ($\Delta 2$ - $\Delta 10$) after homologous modeling of the crystal structure of the α HL pore (PDB: 7AHL) using MODELLER. b) Schematic representation of the β barrel of truncated α HL (TBM $\Delta 6$). Each β strand lacks 12 amino acid residues between positions 120-125 and 133-138 (represented with blue dotted line).

The nature of the truncated α HL pores allows a defined dimension in the β -barrel which can accommodate diverse adaptors such as β CD and $\text{am}_7\beta$ CD as they bind to the wild type α HL pores. Another unique feature of TBM $\Delta 6$ is that the mutant TBM $\Delta 6$ /M113F, in particular, binds $\text{am}_7\beta$ CD very tightly although most of the truncated

pores are expected to be similar to that of the full-length WT pores^{38, 67, 68}. Thus, this experiment would provide a conclusive result on the rigid lumen structure produced by correct refolding of the synthetic monomer.

To confirm the structure of lumen of truncated semisynthetic α HL pores, particularly for homoheptameric pores, we analyzed the kinetics of several different β CD by use of two different semisynthetic mutant pores. The two truncated synthetic mutants (TSM) are i) TSM $_{\Delta 6-3M}$ bearing a wildtype amino acid sequence at 113 with Met and ii) TSM $_{\Delta 6-3F}$ bearing a mutation at 113 with Phe of TBM $\Delta 6$. The functional activities were examined by single channel recordings using the homoheptameric synthetic pores, (TSM $_{\Delta 6-3M}$)₇ and (TSM $_{\Delta 6-3F}$)₇.

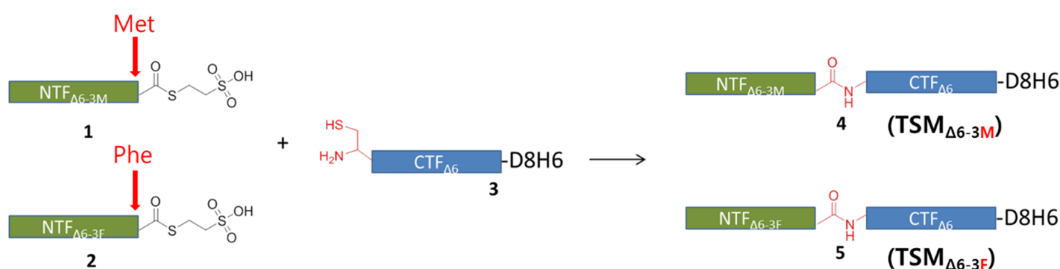


Figure 2.26. Design of NCL to prepare the two different truncated synthetic monomers.

Coupling of NTF $_{\Delta 6-3M}$ (1)/CTF $_{\Delta 6}$ (3) and NTF $_{\Delta 6-3F}$ (2)/CTF $_{\Delta 6}$ (3) yields TSM $_{\Delta 6-3M}$ (4) and TSM $_{\Delta 6-3F}$ (5), respectively.

The residues of NTF $_{\Delta 6}$ and CTF $_{\Delta 6}$ were designed to span 1-113 and 114-293 except the residue 120-125 and 133-138 (denoted as $\Delta 120-125$, $\Delta 133-138$) (Figure. 2.25b). Two different truncated barrel mutants for TBM $\Delta 6$ containing Met and Phe were then synthesized with NCL by coupling the same CTF $_{\Delta 6}$ (3) either with two different NTFs (NTF $_{\Delta 6-3M}$ or NTF $_{\Delta 6-3F}$) containing Met (1) and Phe (2) at 113 residue separately. (Figure 2.26).

2.2.3.1. Preparation of plasmid for NTF_{Δ6-3M} and NTF_{Δ6-3F}

Two additional plasmids encoding NTF_{Δ6-3M} and NTF_{Δ6-3F} residues (1-113) of αHL were prepared by PCR amplification followed by homologous recombination (Figure 2.27).

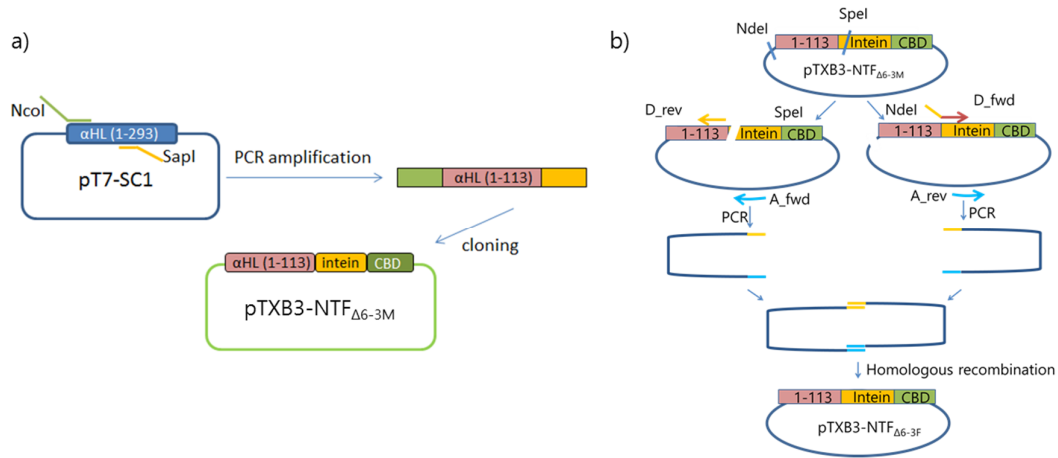


Figure 2.27. Preparation of the NTF_{Δ6} plasmids by homologous recombination. a) Preparation of the plasmid encoding NTF_{Δ6-3M} (αHL residues 1-113). The target sequence (blue) was PCR-amplified from the plasmid (pT7-SC1) bearing a full-length αHL gene. The DNA was cloned into pTXB3, upstream of sequences encoding an intein and a chitin-binding domain. The recombinant plasmid was transformed into *E. coli* and the target protein was overexpressed. b) Preparation of the plasmid encoding NTF_{Δ6-3F}. The A_fwd and D_rev primers were used to obtain the left-hand fragment by PCR template of the SpeI-digested plasmid, while A_rev and D_fwd were used to obtain the right-hand fragment from the NdeI-digested plasmid. The two fragments, which have homologous ends, were mixed and recombined *in vivo* to generate a circular plasmid (pTXB3-NTF_{Δ6-3F}) containing NTF_{Δ6-3M} (1-113) in a fusion protein. The D_fwd and D_rev primers contains TTC (Phe) instead of ATG (Met) on the flanking sequence. (See the section 2.4.1.3 and 2.4.2.1)

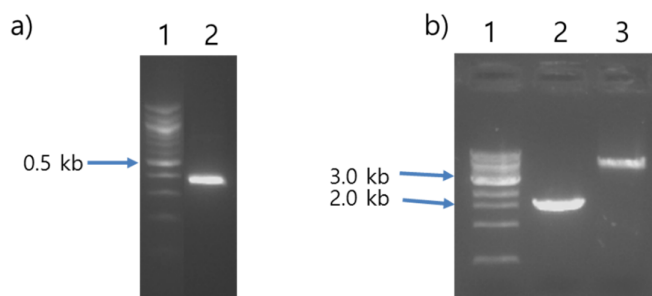


Figure 2.28. PCR products to encode NTF $_{\Delta 6}$. PCR products to give the gene encoding NTF $_{\Delta 6-3M}$ plasmid (a) and the plasmid for NTF $_{\Delta 6-3F}$ (b). a) Lane 1: molecular markers; lane 2: PCR-amplified DNA, which was gel-purified and cloned into pTXB3. b) Lane 1: molecular markers; lane 2 and 3: The two PCR products amplified by A_rev/D_fwd (1.7 kb) and A_fwd/D_rev (5.3 kb), respectively; D_rev and D_fwd contains the codon (TTC) to encode Phe at 113.

The two PCR-amplified products (lane 1 and 2 for the NTF $_{\Delta 6-3M}$ plasmid (Figure 2.28a) and lane 2 and 3 for the NTF $_{\Delta 6-3F}$ plasmid (Figure 2.28b), respectively) were mixed together with competent cells and heat-shocked. The DNA fragments were recombined *in vivo* to generate a circular plasmid encoding the mutated gene.

2.2.3.2. Preparation, purification and characterization of NTF $_{\Delta 6-3M}$ and NTF $_{\Delta 6-3F}$

Both the NTF $_{\Delta 6-3M}$ (1) and NTF $_{\Delta 6-3F}$ (2) in Figure 2.26 were prepared and purified (Figure 2.29 and 2.30) in the same methods described in section 2.2.2.1. The products were also characterized by LC-MS (Figure 2.31).

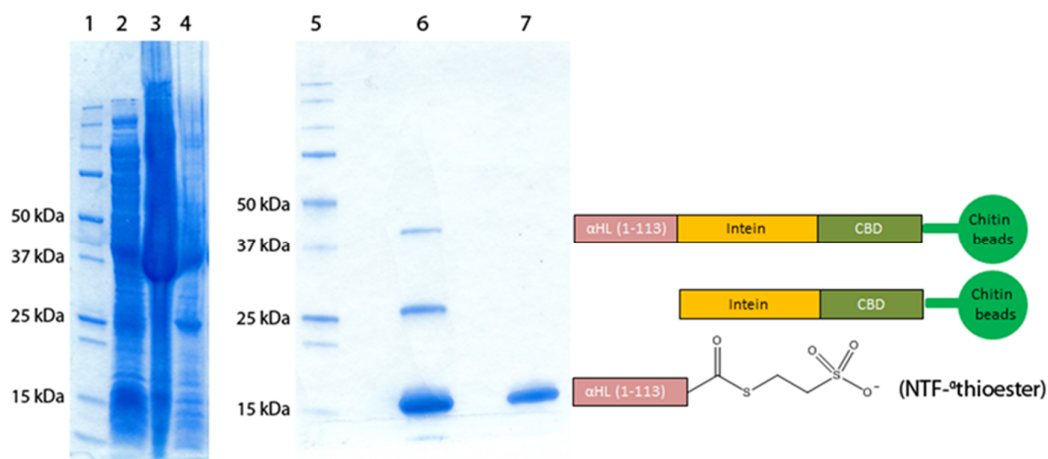


Figure 2.29. Preparation and characterization of NTF_{Δ6-3M}. NTF_{Δ6-3M} was obtained by on-column cleavage from a chitin affinity column with MESNa. The preparation was monitored by using SDS-PAGE (Any kD, Bio-Rad). An NTF- α thioester was obtained by on-column cleavage from a chitin affinity column with MESNa. The preparation was monitored using SDS-PAGE (Any kD, Bio-Rad). Lanes 1 and 5: molecular mass markers; lane 2: cell extract supernatant from *E. coli*; lane 3: cell extract pellet from *E. coli*, in which the overexpressed target protein (40.5 kDa) appears; lane 4: flow-through from the chitin affinity column; lane 6: a small portion of the chitin beads from the column was directly loaded on the gel to monitor the overnight cleavage by MESNa: top band (40.5 kDa), fusion protein; middle band (27.8 kDa), intein-CBD domain; bottom band (12.8 kDa), NTF- α thioester; lane 7: NTF- α thioester eluted from the column under gravity flow.

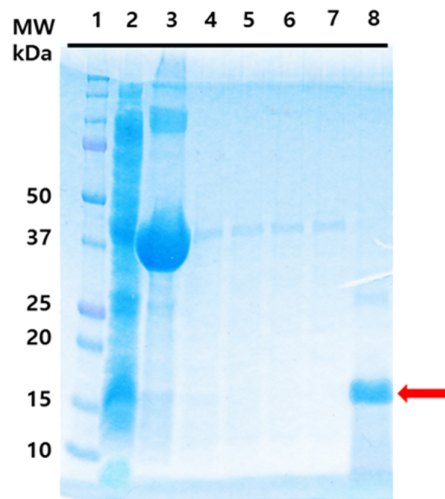


Figure 2.30. Preparation and characterization of NTF $_{\Delta 6-3F}$. NTF $_{\Delta 6-3F}$ was obtained by the same way described in Figure 2.29. Lane 1: molecular mass marker; lane 2: cell extract supernatant; lane 3: cell extract pellet; lane 4: flow-through from the chitin affinity column; lane 5-7: washes; lane 8: NTF $_{\Delta 6-3F}$ (red arrow) eluted from the column under gravity flow.

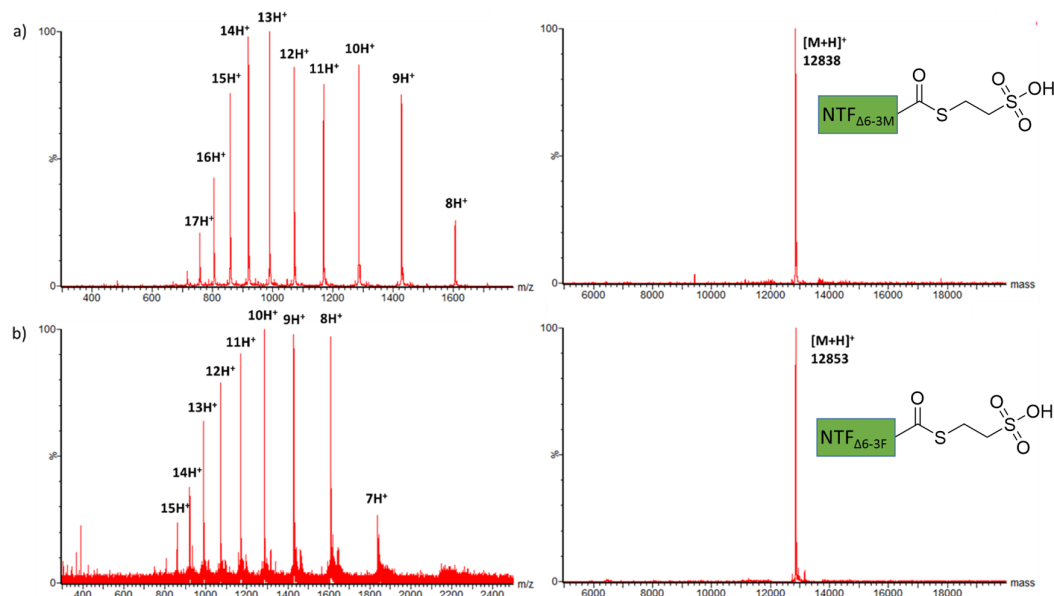


Figure 2.31. Characterization of NTF by LC-MS. ESI-MS spectrum and deconvoluted mass spectrum of a) NTF $_{\Delta 6-3M}$ (calc.12838 obs: 12838) and b) NTF $_{\Delta 6-3F}$ (calc.12853 obs: 12853).

2.2.3.3. Preparation, purification and characterization of CTF_{Δ6}

For preparation of CTF_{Δ6}, the pT7-CTF_{Δ6}-DH plasmid was produced using the same method as described in the preparation of CTF_{FL}, where the codons for residues 1-126 in the pT7 plasmid were replaced with codons for Met-Cys. From the two successive homologous recombination (Figure 2.32), the residues 1-113 were replaced and D8H6 tail introduced at the C-terminus.

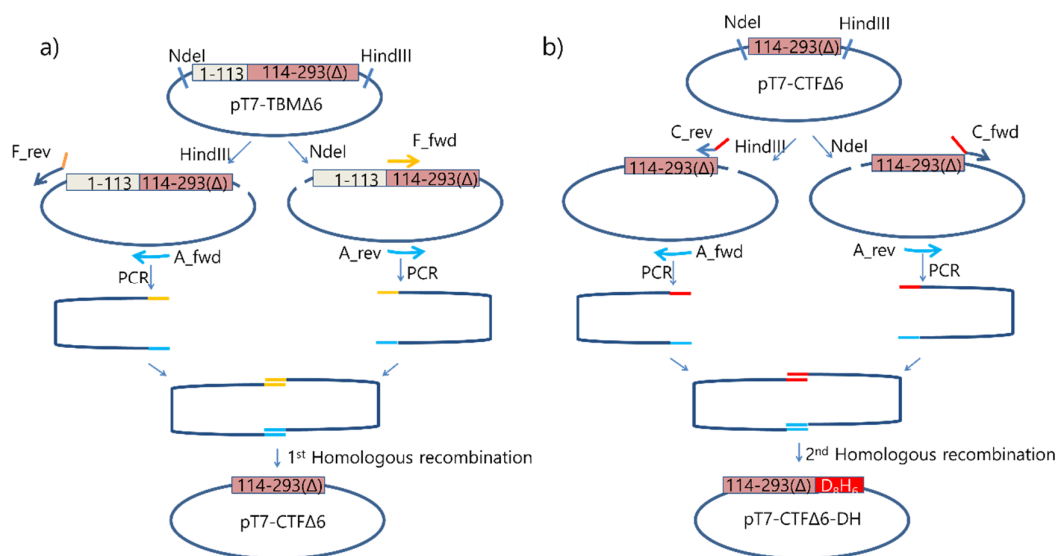


Figure 2.32. Preparation of the pT7-CTF_{Δ6}-DH plasmid. The plasmid containing genes for CTF_{Δ6} were produced by two successive homologous recombinations. a) The pT7-TBMA6 plasmid containing the gene for TBMA6 monomer was separately treated with two restriction enzymes (*NdeI* and *HindIII*). The A_{fwd}/F_{rev} primers were used to obtain the left-hand fragment by PCR template of the *HindIII*-digested plasmid, while F_{fwd}/A_{rev} were used to obtain the right-hand fragment from the *NdeI*-digested plasmid. The two fragments, which have homologous ends, were mixed and recombined *in vivo* to generate a circular plasmid (pT7-CTF_{Δ6}) replacing the residue (1-113) to an upstream of CTF gene initiated with Met-Cys codons. b) A second homologous recombination with A_{fwd}/C_{rev} and C_{fwd}/A_{rev} primers yielded the plasmid pT7-CTF_{Δ6}-DH containing the D8H6 tails at the C-terminus.

The two PCR-amplified products (lane 2 and 3 for the CTF_{Δ6} plasmid (Figure 2.33), respectively) were mixed together with competent cells, heat-shocked, allowed to recombine *in vivo* to generate a circular plasmid (pT7-CTF_{Δ6}) encoding the mutated gene. A successive homologous recombination with the next set of primers yielded the plasmid pT7-CTF_{Δ6}-DH (see section 2.4.2.2. for detailed experimental procedure).

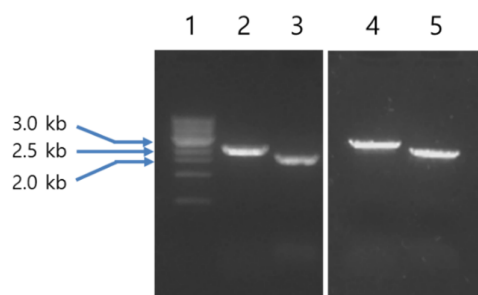


Figure 2.33. PCR products to encode CTF_{Δ6}. PCR products to give the CTF_{Δ6} plasmid. Lane 1: molecular markers; lane 2 and 3: The two PCR products (2.5 and 2.3 kb) amplified by two sets of two primers, A_rev/B_fwd and A_fwd/B_rev (Figure 2.28a), were used to construct pT7-CTF_{Δ6} in the first homologous recombination; lane 4 and 5: The two PCR products amplified by two sets of two primers, C_rev/B_fwd and C_fwd/B_rev (Figure 2.28b), were used to construct pT7-CTF_{Δ6}-DH in the second homologous recombination.

Preparation and characterization of CTF_{Δ6} was carried out in the same procedure as described in section 2.2.3. However, the purification of CTF_{Δ6} was carried out in an automated micropreparative purification system (AKTA) to improve the yield and purity of the product (Figure 2.34). After purification, the fragment was characterized by LC-MS (Figure 2.35).

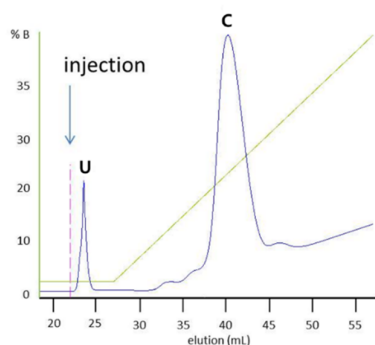


Figure 2.34. Purification of overexpressed CTF Δ_6 by a Ni-NTA affinity column. The cell lysate was injected to the column and the column was washed to remove unbound proteins. The unbound proteins (**U**) was eluted as soon as injected. The buffer (**b**) containing 500 mM imidazole increased in a slow gradient (change of %B: 1 % min⁻¹) and the CTF (**C**) containing His-tag at the C-terminus was eluted when 20-30 % (100-150 mM) of imidazole was added to the column. The fractions for peak **C** were collected and used for characterization. Buffer A: 6M Gu.HCl, 50 mM NaH₂PO₄ (pH 7.5), 250 mM HCl, 5 mM TCEP; Buffer B: 6M Gu.HCl, 50 mM NaH₂PO₄ (pH 7.5), 250 mM HCl, 5 mM TCEP, 500 mM imidazole.

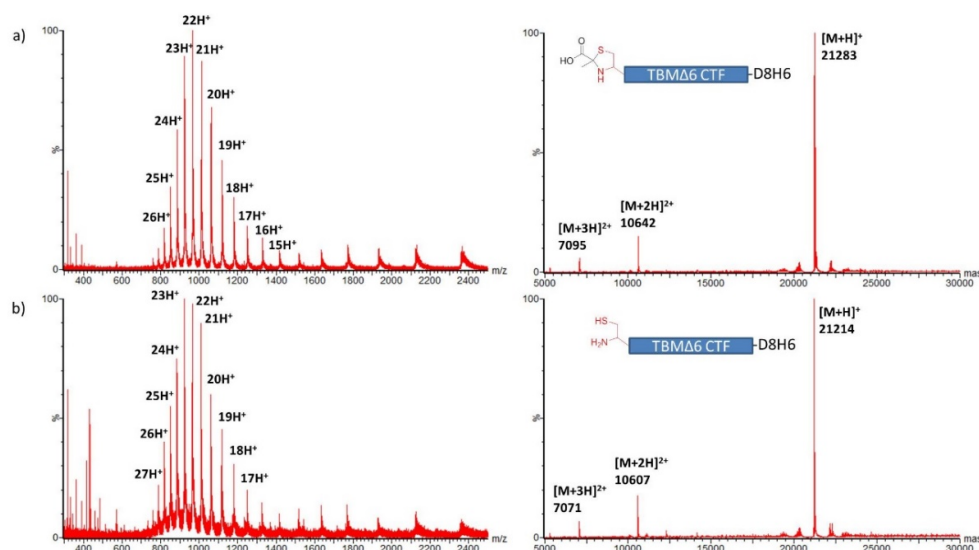


Figure 2.35. Characterization of CTF Δ_6 by LC-MS. a) Mass spectrum of the purified CTF Δ_6 by LC-MS (calc: 21283, obs:21283). (b) N-Cys is protected by a pyruvate, which increases the mass by +72 from the theoretical mass. (b) Thz group was deprotected by MeONH₂·HCl, which corresponded to the calculated mass (calc: 21212, obs: 21214).

2.2.3.4. NCL for preparation of truncated monomers

NTF $_{\Delta 6-3M}$ (1) and NTF $_{\Delta 6-3F}$ (2) were mixed with CTF $_{\Delta 6}$ (3) separately in NCL buffer and allowed to react overnight to generate the truncated synthetic monomers, TSM $_{\Delta 6-3M}$ and TSM $_{\Delta 6-3F}$ (Figure 2.22). The crude product was directly purified by gel filtration (Figure 2.36) and analyzed by LC-MS (Figure 2.37).

2.2.3.5. Purification and characterization of truncated synthetic monomers

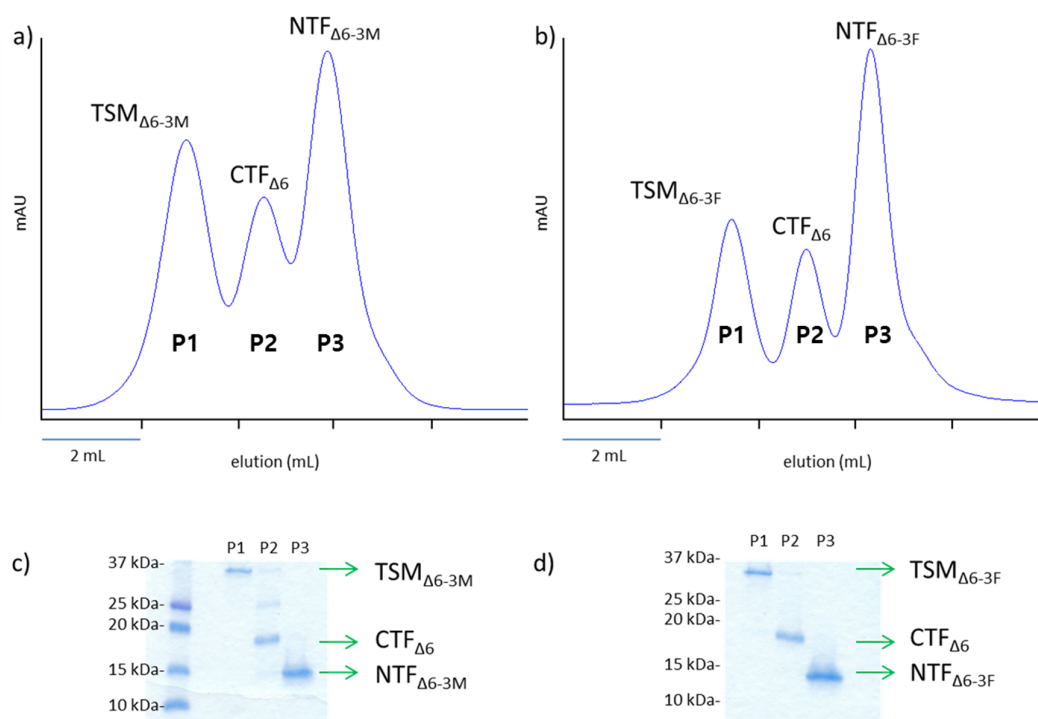


Figure 2.36. Purification of TSMs produced by NCL. Chromatogram of TSM $_{\Delta 6-3M}$ (a) and TSM $_{\Delta 6-3F}$ (b) purification in gel filtration. After the purification, all the fractions were collected and identified by SDS-PAGE. Each of the fractions collected from the first peak corresponds to the expected protein mass (NTF $_{\Delta 6}$ (12.8 kDa), CTF $_{\Delta 6}$ (21.2 kDa), and TSM $_{\Delta 6}$ (33.9 kDa)).

The synthetic monomers were purified in a size exclusion column (SuperdexTM 200 10/300 GL) with the same method described in section 2.2.2.4. The crude product

was loaded to the column and the purification was monitored at 280 nm. Three major peaks were observed and the fractions at each peak of the chromatogram (TSM $_{\Delta 6-3M}$ and TSM $_{\Delta 6-3F}$ for Figure 2.36a and b, respectively) were collected. To ensure the purification result, the fractions at the three peaks were directly loaded to Any kD (Biorad) SDS-gel and stained with Coomassie blue. As shown in Figure 2.36c and d, the fractions collected at the first peak (P1) show a single band which corresponds to the distance that the monomer of truncated barrel mutant migrates. The fractions collected at the other two peaks (P2 and P3) show unreacted CTF (P2) and NTF (P3).

The fractions containing the TSMs were collected and further characterized by LC-MS (Figure 2.37) and matched the theoretical mass.

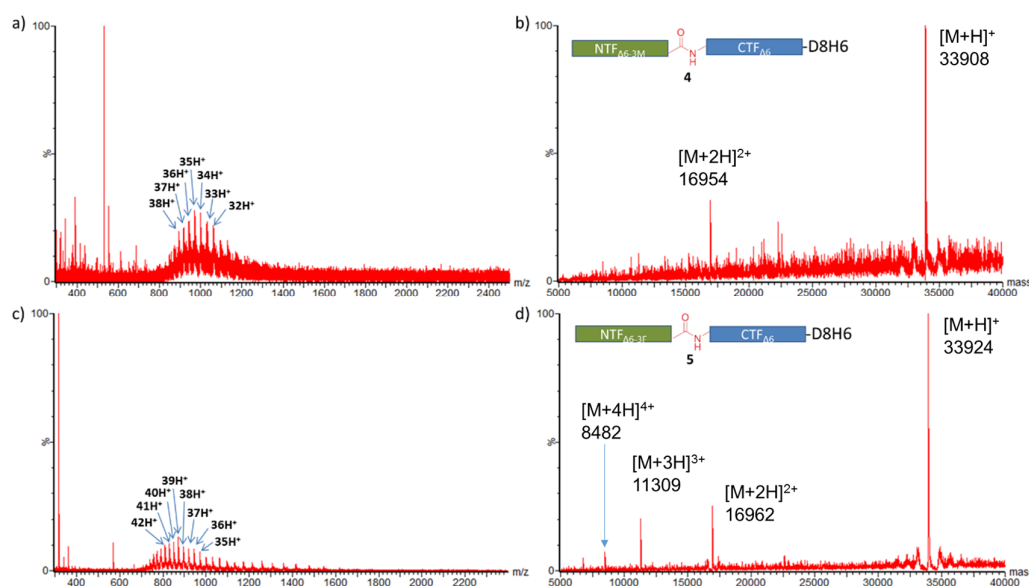


Figure 2.37. Characterization of TSM $_{\Delta 6-3M}$ and TSM $_{\Delta 6-3F}$. Mass spectrum and deconvoluted mass of TSM $_{\Delta 6-3M}$ (**4**) (a and b, calc: 33906, obs: 33908) and TSM $_{\Delta 6-3F}$ (**5**) (c and d, calc: 33923, obs: 33924).

2.2.3.6. Hemolysis assays

TBMΔ6 is known to be hemolytically inactive³⁸ as the length of β-barrel is not long enough to penetrate through the membrane of rabbit's red blood cells and trigger lysis. As shown in Figure 2.38, TSM_{Δ6-3M} and TSM_{Δ6-3F} did not show hemolytic activity, demonstrating semisynthetic monomers have similar functional activity to that of TBMΔ6 pores prepared by IVTT.



Figure 2.38. Hemolysis assays for WT αHL and TBMΔ6 prepared by IVTT, and truncated synthetic monomers (TSM_{Δ6-3M} and TSM_{Δ6-3F}). The decrease in light scattering (OD at 595nm) over time corresponds to lysis of rRBCs due to the activity of pore forming αHL proteins. The hemolytic assay of IVTT αHL monomer and TSM was performed in a 96-well plate, where IVTT αHL, TBMΔ6, and two different TSMs were added to the first column on row 1, 2, 3, and 4 of the plate, respectively, and subsequently serially diluted two-fold in a final volume of 100 μL across each rows. No protein was added to row 5 as a negative control. Only WT αHL monomer prepared by IVTT were found to be hemolytically active, whereas the truncated barrel mutants produced by either IVTT approach or semisynthesis were inactive.

2.2.3.7. Oligomerization to homoheptamer ((TSM_{Δ6-3M})₇ and (TSM_{Δ6-3F})₇)

It has been previously demonstrated that the IVTT product of TBMΔ6 monomer is able to oligomerize into heptameric pores in the absence and presence of lipid membrane³⁸. In the case of the purified TSM_{Δ6-3M} and TSM_{Δ6-3F}, they were found to form homoheptamers regardless of the existence of a lipid membrane (Figure 2.39),

once they were incubated at 37°C. This result suggests that the folding characteristics of TSM $_{\Delta 6-3M}$ and TSM $_{\Delta 6-3F}$ are similar to that of TBM $\Delta 6$ prepared by IVTT. The (TSM $_{\Delta 6-3M}$) $_7$ and (TSM $_{\Delta 6-3F}$) $_7$ homoheptamers were directly obtained by eluting the proteins from SDS gel as described in section 2.4.1.10 and used for single-channel recording.

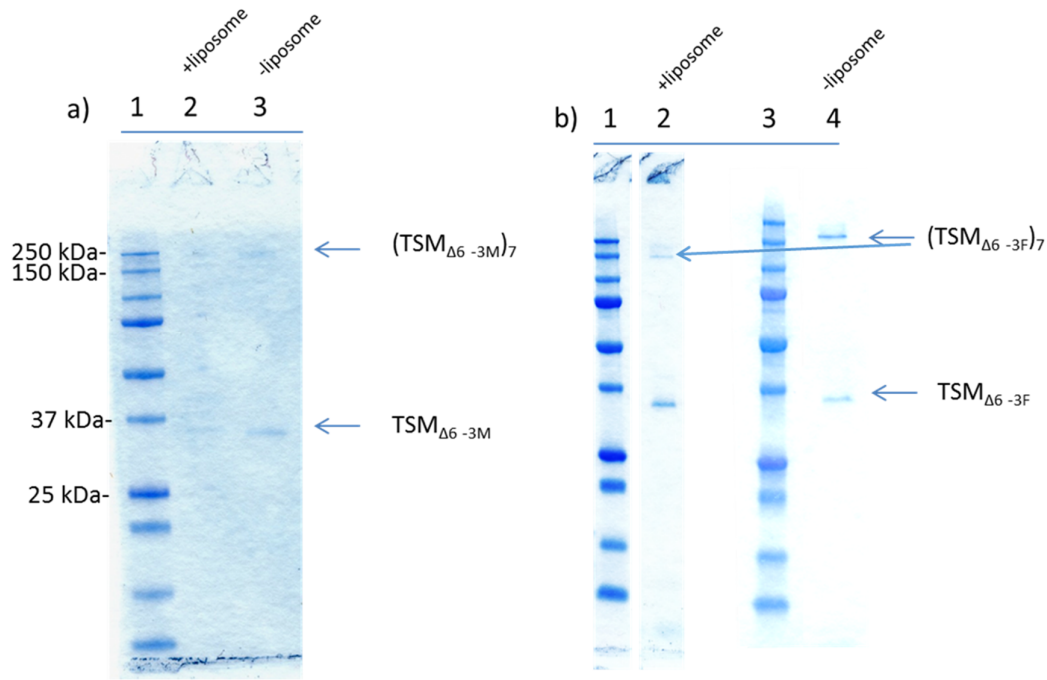


Figure 2.39. Assembly to homoheptamer from truncated synthetic monomer. The truncated synthetic monomers TSM $_{\Delta 6-3M}$ (a) and TSM $_{\Delta 6-3F}$ (b) oligomerize to homoheptamer in the absence and presence of liposome. Lane 1 in a) and Lane 1 and 3 in b) represent a protein molecular marker. The heptameric bands in Lane 2 (Figure 2.39a) are rather weak presumably because the polypeptide sample is diluted after the addition of liposome.

2.2.3.8. Single-channel recordings with truncated synthetic pores

2.2.3.8.1. I-V curves

The semisynthetic homoheptameric pores were characterized by monitoring the magnitude of ionic flow thorough the synthetic pores under applied potentials. The electric characteristics were analyzed in comparison with TBM $\Delta 6$ prepared with IVTT. As shown in Figure 2.40, the I-V curve of (TSM $_{\Delta 6-3M}$) $_7$ and (TSM $_{\Delta 6-3F}$) $_7$ were similar to

that of TBM Δ 6 pores obtained from IVTT, further validating the similarity between semisynthetic and native pores.

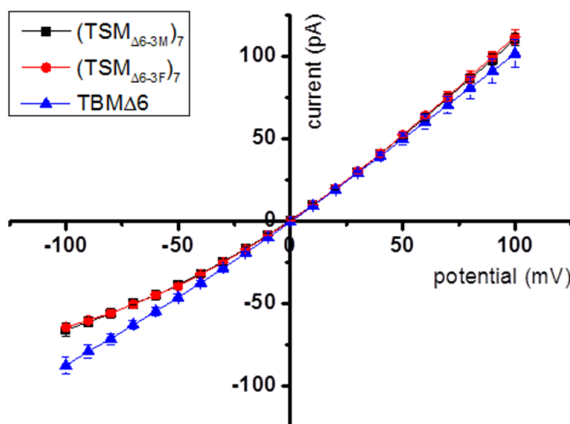


Figure 2.40. Single channel I-V curve of TBM Δ 6 from IVTT and truncated synthetic pores.

IVTT truncated pore (TBM Δ 6 (blue, $n = 3$)) and the two synthetic truncated pores ((TSM Δ_{6-3M}) $_7$ (black, $n = 4$) and (TSM Δ_{6-3F}) $_7$ (red, $n = 3$)) show similar conductance under the applied potential (-100 – 100 mV). TBM Δ 6 containing E111N/K147N (NN) mutations shows larger conductance at the negative potential because a rectification from WT sequences is lost in the NN mutant pore^{69, 70}.

2.2.3.8.2. Binding of β CDs to (TSM Δ_{6-3M}) $_7$ binding

To test the integrity of the barrel in homohepameric (TSM) $_7$ pores, the association and dissociation rate constants of β CD and am $_7\beta$ CD were measured. The binding of β CD to the (TSM Δ_{6-3M}) $_7$ pore showed a similar binding constant ($K_d = 65 \pm 2$ mM) to that of TBM Δ 6 pore ($K_d = 230 \pm 20$ mM)³⁹, suggesting that the rigidity of β barrel was maintained with all the seven semisynthetic monomers (Figure 2.41).

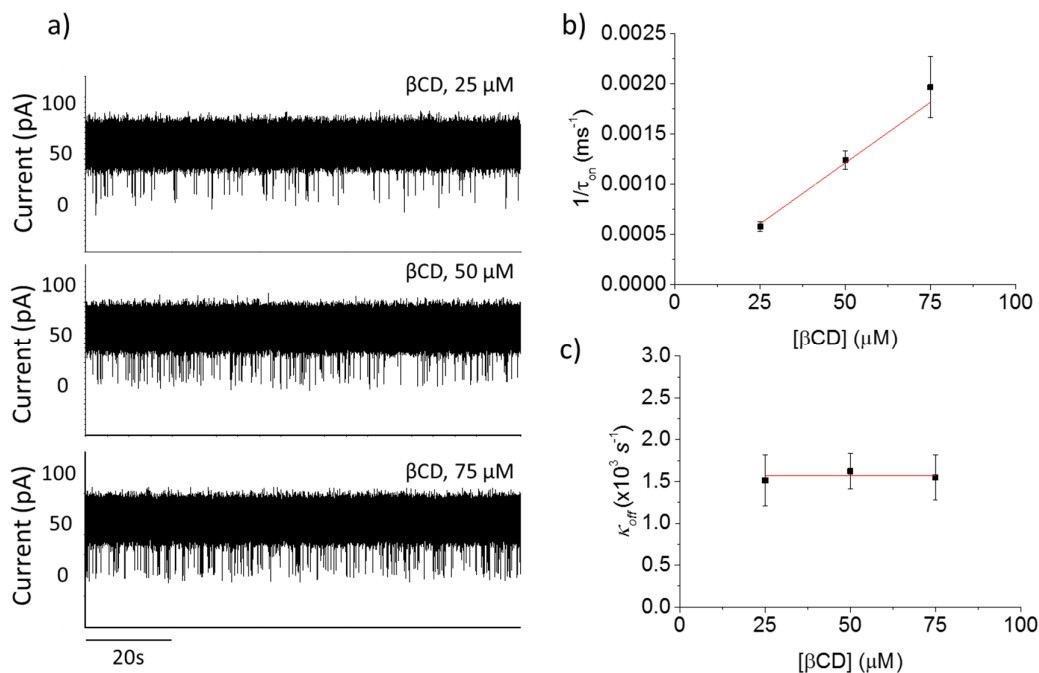


Figure 2.41. Kinetic analysis of β CD binding to homoheptameric (TSM $_{\Delta 6-3M}$) $_7$ pore. k_{off} , k_{on} , and K_d were obtained as a mean value measured from three separate experiments. $k_{off} = 1.6 \pm 0.1 \times 10^3 s^{-1}$, $k_{on} = 2.4 \pm 0.1 \times 10^4 M^{-1} s^{-1}$, and $K_d = 6.5 \pm 0.2 \times 10^{-2} M$. The analysis was carried out under 1 M KCl, 20 mM Tris (pH 7.0) at +50 mV.

2.2.3.8.3. Binding of β CD and am $_7\beta$ CD to (TSM $_{\Delta 6-3F}$) $_7$ pores

The mutation Met113 \rightarrow Phe113 is known to strengthen β CD binding in the WT α HL pore⁶⁸. Particularly, am $_7\beta$ CD previously has shown a permanent binding to a truncated barrel mutant pore (TBM $_{\Delta 6-3F}$)⁷¹. The β CD shows a significantly low dissociation constant ($K_d = 6.1 \pm 0.2 \times 10^{-2} mM$) in the (TSM $_{\Delta 6-3F}$) $_7$ pore (Figure 2.42a) compared to that in (TSM $_{\Delta 6-3M}$) $_7$ ($K_d = 65 mM$) (Figure 2.43). The am $_7\beta$ CD remained bound to the (TSM $_{\Delta 6-3F}$) $_7$ pore for >30 min. (Figure 2.42b). The dissociation constant of am $_7\beta$ CD was not able to be obtained due to a low number of events.

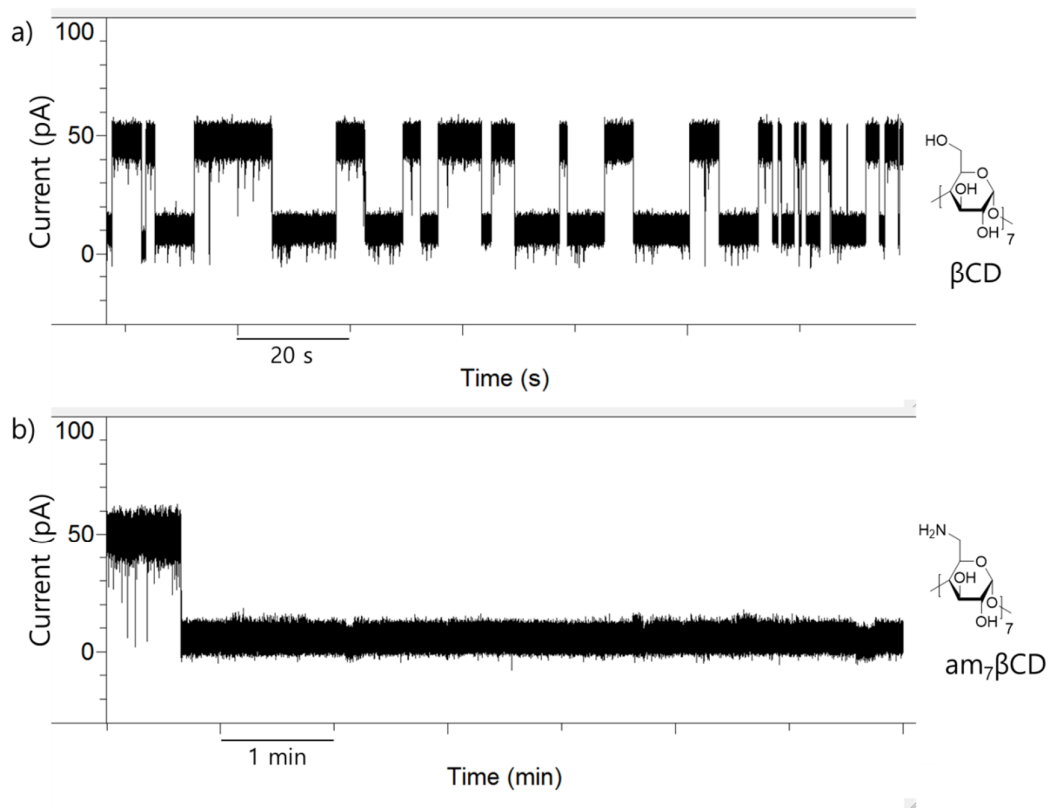


Figure 2.42. Representative single-channel trace of cyclodextrins binding to homoheptameric (TSM_{Δ6-3F})₇ pore. a) Current blockades in the presence of 10 μM of βCD. Condition: 1 M KCl, 20 mM Tris (pH 7.0) at 50 mV. The current was filtered at 5 kHz and sampled at 25 kHz. For display, further post filtering was carried out at 2 kHz (8-pole lowpass Bessel filter). b) Current blockades in the presence of 10 μM of am₇βCD. am₇βCD binds to the pore almost permanently.

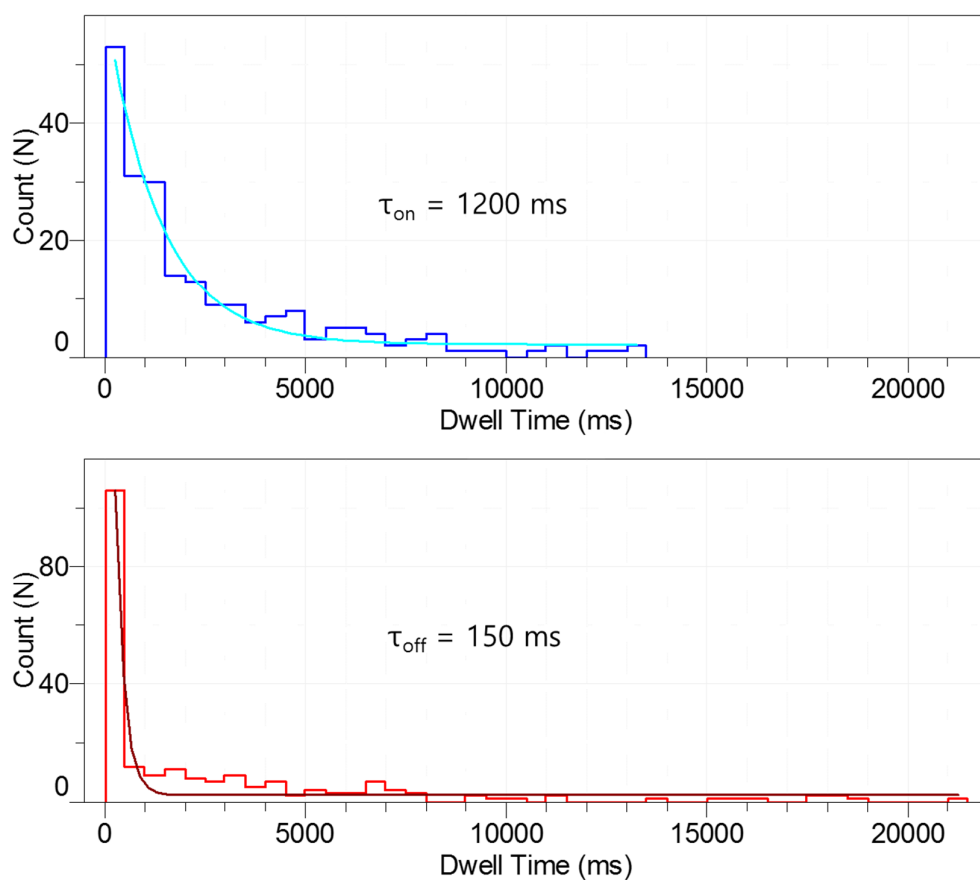


Figure 2.43. The Dwell time histogram for β CD binding events in $(\text{TSM}_{\Delta 6-3\text{F}})_7$ in the presence of $10 \mu\text{M}$ β CD. The dissociation constants was determined from mean values of $1/\tau_{\text{on}} \cdot [\beta\text{CD}]$ or $1/\tau_{\text{off}}$ ($k_{\text{on}} = 7.4 \pm 0.5 \times 10^4 \text{ M}^{-1} \cdot \text{s}^{-1}$, $k_{\text{off}} = 4.5 \pm 0.8 \text{ s}^{-1}$, $K_{\text{d}} = 6.1 \pm 0.2 \times 10^{-2} \text{ mM}$, $n=3$).

2.3. Conclusion

In this chapter, three different semisynthetic nanopores were investigated. A full-length α HL and two truncated mutants of α HL were prepared by NCL and characterized by SDS-PAGE and LC-MS.

The denatured protein monomers were successfully renatured to homo- and hetero-heptameric pores and the function of the pores were verified by hemolytic assays and single-channel recordings. The integrity of the β barrel was proved at the single-molecule level, by observing that the semisynthetic pores accommodated cyclodextrin adaptors with comparable binding kinetics to the native pores prepared by IVTT. The correct refolding was also verified by limited proteolysis, as the full-length synthetic monomer was resistant to proteolysis because the proteolytic site was inaccessible after the monomer was assembled into heptameric pores in the presence of liposome. In addition, the free thiolate at the ligation site was used to monitor disulfide bond formation at the single-molecule level, which demonstrated that synthetic pores also hold potential as nanoreactors.

From the results of hemolytic activity, cyclodextrin binding kinetics, limited proteolysis, and chemistry of sulfhydryl group, it is concluded that the semisynthetic pores show very similar biological and electrophysiological properties to the native pores prepared by IVTT. Furthermore, results presented here open the door to many other applications in engineering of membrane proteins, such as the development of a novel semisynthetic nanoreactor discussed, in chapter 3.

2.4. Materials and methods

2.4.1. Assembly of WT α HL pore

2.4.1.1. Plasmid preparation for generation of N-terminal fragment (NTF)

The gene to encode corresponding NTF was cloned into the pTXB3 vector (NEB), which results in a fusion protein with the intein (198 amino acids) from the *Mycobacterium xenopi* GyrA gene and a chitin binding domain (CBD). The CBD binds to chitin beads (NEB) and it is used for purification of the target protein.

Briefly, the cloning procedures were carried out as follows.

I. *PCR amplification of the target gene (1-126 of α HL):* 2X Phusion high-fidelity

PCR master mix (NEB) was employed to PCR amplify NTF from the pT7 vector containing the wild type gene of α HL by use of a 5'-TATGGATTCAACGGTAATGTTACTGGTTGCATCACGGGAGATGCACTAGT-

3' (forward) and a 5'-GTAACATTACCGTTGAATCCATAAGTTAAAGTACTCATATACTCTTTTGT-3' (reverse) primer. This bears restriction enzyme sites on both end of its flanking sequence. The product (~400 bp, Figure 2.4) was identified by 1 % (w/v) agarose gel electrophoresis in TBE buffer, and the corresponding band was excised and gel-purified with QIAquick gel extraction kit (Qiagen).

Thermocycling conditions were as follows: 3 min of heating at 95 °C followed by 30 cycles of 95 °C for 5 s, 60 °C for 10 s, 72 °C for 1 min and a final elongation step at 72 °C for 10 min. All the polymerase chain reactions shown in this thesis follow the conditions below.

| Reagent | Volume |
|------------|------------------------|
| Template | 1 μ L |
| Primer_fwd | 1 μ L (50 μ M) |
| Primer_rev | 1 μ L (50 μ M) |

| | |
|---------------|-------|
| 2X polymerase | 25 µL |
| water | 22 µL |

- II. *Digestion*: After verification of the desired products by 1% agarose gel electrophoresis, the PCR product was used directly for restriction enzyme digestion without an extra purification step. Restriction enzyme digestion was carried out by incubating 1.5 µl of the amplified chip oligonucleotides and 1 µL of pTXB3 (molar ratio, 3:1) in the same reaction tube with 1 µl *NcoI* and *SapI* (NEB), 2 µl NEB Buffer #1, and 11 µL of nuclease-free water at 37 °C for 3 h. Subsequently, the flanking sequence-cleaved oligonucleotides were purified by PCR purification kit (Qiagen) to remove the restriction enzymes. The genes were eluted with 30 µL of nuclease-free water.
- III. *Ligation*: 3 µL of the purified genes were mixed with 1 µL of T4 DNA ligase (NEB), 2 µL of 10X T4 DNA ligase buffer (NEB), 16.5 µL of nuclease-free water. The mixture was incubated at 16 °C overnight.
- IV. *Transformation*: 5 µL of the reaction mixture was mixed with 20 µL of XL-10 Gold® Ultracompetent cells (Agilent) and incubated on ice for 30 min. The cell/DNA mixture was heat-shocked using a water bath at 42 °C for 30 sec and subsequently returned to ice. The heat-shocked cells were spread onto agar plates containing LB medium and ampicillin (100 µg/mL), and incubated overnight at 37 °C.
- V. *Colony PCR*: To verify and select the plasmid successfully cloned with PCR amplified gene, 16 transformed colonies were picked at random and grown in LB media at 37 °C with shaking for 1 h. 1 µL culture was mixed with Taq 2X PCR master mix (NEB) and was PCR amplified with 5'-TAATACGACTCACTATAGGG-3' (forward) and 5'-ATAAAGTTGCAGGAC

CACTTCTG-3' (reverse) primer. Thermocycling conditions were as follows: 3 min of heating at 95 °C followed by 30 cycles of 95 °C for 5 s, 60 °C for 10 s, 72 °C for 1 min and a final elongation step at 72°C for 10 min. The products were identified by 1 % (w/v) agarose gel electrophoresis and the plasmid with the corresponding band was sent for sequencing.

VI. *Plasmid Extraction and sequencing*: DNA samples was obtained by Qiagen Plasmid Mini Kit and it was sequenced at the Source BioScience LifeSciences centre, located in the Department of Biochemistry, at the University of Oxford.

2.4.1.2. Plasmid preparation for generation of C-terminal fragment (CTF)

The pT7-SC1-CDH plasmid for CTF[Cys¹²⁷-Asn²⁹³]-D₈H₆ expression was obtained by two successive homologous recombinations⁷² beginning with pT7-SC1⁷³, which resulted in the removal of α HL codons 1-126 and the addition of DNA encoding a D₈H₆ tag (Figure 2.8).

1) *Preparation of pT7-SC1-C*: α HL codons 1-126 from pT7-SC1 (containing the WT α HL gene) were replaced by Met-Cys codons (Figure 2.8a). pT7-SC1 was digested with either the *NdeI* or *HindIII* enzymes and each of the two linearized plasmids was used as a PCR template. Two sets of two primers were required to produce a plasmid containing a gene encoding 127-293 residues: 5'-ATAAAGTTGCAGGACCACTTCTG-3' (forward, A_fwd / 5'-CATATGTATATCTCCTTCTTAAAG-3' (reverse, B_rev) was mixed with the plasmid template linearized by *HindIII* and 5'-GATATACATATGTGCGATACCGGCAAAATTGGAAAT-3' (forward, B_fwd) / 5'-CAGAAGTGGTCCTGCAACTTTAT-3' (reverse, A_rev) was mixed with the plasmid template linearized by *NdeI* (Figure 2.8a). The thermocycling conditions were set to: 3 min at 95°C followed by 30 cycles of 95 °C for 5 s, 60 °C for 10 s, 72 °C for 4 min and a final elongation step at 72 °C for 10 min. The size of each product was examined in a 1% (w/v) agarose gel. The two PCR fragments (10 μ L each) were mixed together with *E*.

coli XL-10 Gold Ultracompetent cells (100 μ L, Agilent) and incubated on ice for 30 min. The cells were transformed by heat shock and plated onto ampicillin plates (100 μ g mL⁻¹), which were incubated overnight at 37 °C. A colony was picked and grown overnight in LB medium containing ampicillin (100 μ g mL⁻¹). Plasmid pT7-SC1-C was extracted with a QIAprep miniprep kit (Qiagen) and the sequence of the insert was verified by Sanger sequencing, with the sequencing primers T7F (5'-TAATACGACTCACTATAGGG-3') and T7R (5'-GCTAGTTATTGCTCAGCGG-3') (Source Bioscience).

2) *Preparation of pT7-SC1-CDH*: DNA encoding D₈H₆ tail was added to the end of the CTF gene in pT7-SC1-C (Figure 2.8b) using the same protocol described to prepare pT7-SC1-C except that two other primers were used to yield D₈H₆ tail: 5'-ATAAAGTTGCAGGACCACTTCTG-3' (forward, A_fwd) / 5'-GGTGATCATCGTCGTCA TCATCGTCATCATTTGTCATTTCTTCTTTTTC-3' (reverse, C_rev) and 5'-CGACGATG ATCACCACCATCACCATCATTGATAAGCTTGGATCCGGCTGC-3' (forward, C_fwd) / 5'-CAGAAGTGGTCCTGCAACTTTAT-3' (reverse, A_rev).

2.4.1.3. Preparation of N-terminal fragment (NTF)

The pTXB3-NTF_{FL} vector containing 1-126 residues were transformed into 20 μ L of competent cell (T7 Express Competent *E. coli*, (High Efficiency), NEB) and spread onto an agar plate starter containing 100 mg mL⁻¹ of ampicillin. The plate was incubated at 37 °C overnight. A colony was picked from the plate and transferred into 25 mL of LB media and incubated at 37 °C overnight as a starter culture. The starter was transferred to 500 mL of LB media and the overexpression of NTF was induced by 1 mM IPTG at OD₆₀₀ ~0.6. To take advantage of the poor solubility of membrane segment, the overexpression was conducted at 37 °C, to ensure high expression levels of the protein within inclusion bodies.

The cells in the culture were separated by centrifugation at 5,000 g for 5 min and the supernatant was discarded. The cell pellets (~5 g) were lysed with 25 mL of Bug Buster (Merck Millipore) at 4 °C, followed by sonication for complete lysis. After the lysate was centrifuged at 16,000 for 20 min, the cell pellet was separated from the supernatant (the supernatant was stored for analysis) and resuspended in a denaturing buffer containing 8 M urea, 50 mM phosphate (pH 7.0), 250 mM NaCl. The suspension was re-centrifuged at 16,000 g to remove cell debris which is insoluble even in denaturing buffer. The pellet was discarded and the supernatant was diluted to 1M urea with native buffer containing 50 mM phosphate (pH 7.0), 250 mM NaCl in order for chitin binding domain of the fusion protein to refold.

The chitin bead column was washed with MilliQ water which is 20 times of bed volume to remove 20 % alcohol. Then, the column was equilibrated with 10 times of bed volume of buffer containing 1 M urea, 50 mM phosphate (pH 7.0), 250 mM NaCl. Into the equilibrated column, the diluted cell lysate was added and the flow-through was collected for further analysis. In addition, the column was washed with 20 times of bed volume of the same buffer used for column equilibration and each fraction of washes was also collected for analysis.

~2-bed volume of cleavage buffer (100 mM MESNa (sodium 2-sulfanylethanesulfonate), 1 M urea, 50 mM phosphate (pH 7.0), 250 mM NaCl) was added to the column and ~1-bed volume was flushed through the column to equilibrate the column with the cleavage buffer. The reaction was allowed to proceed overnight. NTF_{FL}-thioester was released from the column by gravity and directly analyzed by SDS-PAGE.

2.4.1.4. Preparation of C-terminal fragment (CTF)

pT7-SC1-CDH was transformed into *E. coli* BL21(DE3) cells (Promega). The cells were plated onto ampicillin (100 µg mL⁻¹) and incubated overnight at 37 °C. A

colony was picked and cultured overnight in LB medium (25 mL) containing the same concentration of ampicillin. This starter culture was added to LB medium (500 mL). At OD₆₀₀ ~0.6, overexpression of CTF was induced by the addition of IPTG (to 100 μ M) and further shaking for 4 h, at 37 °C.

CTF-containing wet cell pellets (~5 g) were lysed with 25 mL of BugBuster® (Merck Millipore) at 4 °C, followed by sonication. The lysate was centrifuged, the supernatant discarded and the cell pellet resuspended in the denaturing buffer (6 M Gu.HCl, 50 mM Na phosphate (pH 7.5), 250 mM NaCl, 5 mM TCEP). The suspension was centrifuged at 16,000 g for 20 min to remove cell debris and the supernatant was loaded onto a Ni-NTA affinity column, which had been equilibrated with the denaturing buffer. After washing the resin with 20 bed volumes of denaturing buffer, the protein was eluted with 6 M Gu.HCl, 50 mM phosphate (pH 7.5), 250 mM NaCl, 5 mM TCEP, containing 250 mM imidazole. Fractions were collected and analyzed by SDS-PAGE. The fraction eluted by imidazole was characterized by LC-MS.

2.4.1.5. Deprotection of N-terminal Cysteine

450 μ L of CTF (2.5 mg mL⁻¹, ~0.12 mM) in the purification buffer (6 M Gu.HCl, 50 mM phosphate (pH 7.5), 250 mM NaCl, 5 mM TCEP, and 250 mM imidazole) was treated with 50 μ L of 4 M MeONH₂.HCl (final concentration 0.4 M). Additional pH adjustment was conducted to make the pH of the reacting solution ~3.5-4.0⁷⁴. The deprotection was carried out overnight at room temperature and the resulting peptide was analyzed by LC-MS without further purification or desalting.

2.4.1.6. NCL for a WT α HL full length monomer

The NCL buffer (6M Gu.HCl, 0.2 M phosphate (pH 6.8), 40 mM TCEP, and 0.2 M MPAA) was prepared, degassed, and sealed under Argon gas before use.

500 μ L of CTF treated with MeONH₂.HCl was concentrated to 100 μ L in a centrifugal filter (Amicon, MWCO 3k) at 15,000 g for 20 min at 20 °C. 400 μ L of NCL

buffer was added and concentrated again at the same condition to reduce the concentration of MeONH₂.HCl and provide the optimal pH (pH 6.8) of NCL. The washes were carried out five times to assume the concentration of MeONH₂.HCl as 128 μ M, which would not affect the ligation reaction.

The concentration was quantified by a spectrophotometer (ND-1000, Nanodrop) with the extinction coefficient and molecular weight for each protein fragments.

Table 2.1. Extinction coefficient and molecular weight used to obtain the concentration from the absorbance at 280 nm.

| proteins | E (M ⁻¹ cm ⁻¹) | MW (Da) |
|----------------|---------------------------------------|---------|
| α HL-H6 | 64,860 | 34,070 |
| NTF-OH | 17,420 | 14,025 |
| CTF-D8H6 | 47,440 | 20,972 |

*Note the extinction coefficient and MW were calculated as a protein with no MESNa at the C-terminus.

To 100 μ L of concentrated CTF (~0.6 mM), 100 μ L (4.6 mg mL⁻¹, ~0.35 mM) of NTF was added and mixed with 300 μ L of NCL buffer in a 500 μ L centrifugal filter (Amicon, MWCO 3k) and centrifuged at 15,000 g for 20 min to concentrate the reacting polypeptides. The flow-through was discarded. 400 μ L of NCL buffer was subsequently added to the reacting mixtures. The mixtures were centrifuged and refilled with NCL buffer four times to reduce the concentration (625-fold) of urea and MESNa used in purification of NTF. The final resulting mixture was concentrated to 100 μ L and allowed to sit at room temperature overnight. The crude reaction mixture was analyzed with SDS-PAGE without further desalting process, which showed a newly generated band, corresponding to the theoretical mass.

2.4.1.7. Purification of the full length α HL synthetic monomer

The purification of the synthetic monomer was carried out by using an automated AKTA FPLC with a size exclusion column (Superdex™ 200 10/300 GL), at room temperature. The column was equilibrated in a denaturing condition using 5 column volume of the purification buffer (8 M urea, 50 mM phosphate, 250 mM NaCl, 5 mM TCEP). The crude product was washed with the same buffer in a centrifugal filter (Amicon, MWCO 3k) by subsequent concentration (15,000 g, 15 min) and diluted to reduce the concentration of MPAA and maintain an identical condition in the purification condition. The sample was manually injected and loaded to the column and the purification was monitored at 280 nm in real-time. The sample was purified at the flow rate (0.7 mL/min) and all the fractions were collected in an automated sampler with 0.5 mL.

2.4.1.8. Hemolysis assays

Rabbit red blood cells were washed repeatedly in MBSA (10 mM 3-[N-morpholino]propane sulfonic acid (MOPS), 150 mM NaCl, pH 7.4, containing 0.1 % bovine serum albumin). The hemolytic activity of synthetic monomers was assayed in a 96-well plate. The protein was added to the first column in the plate and serially diluted two-fold with MBSA such that each well contained a volume of 50 μ L. An equal volume of rabbit red blood cells (~1% in MBSA) was then added to each well, starting with the well which contained the most dilute protein sample. Lysis was recorded for 100 min by monitoring the decrease in light scattered at 595 nm with a microplate reader (Bio-Rad Laboratories Ltd, UK)⁷⁵.

2.4.1.9. Protein expression by *in vitro* transcription and translation

For preparation of heteroheptameric pores with synthetic monomer or control experiment with WT, α HL was also obtained from a cell free *E. coli in vitro* transcription and translation (IVTT) system (*E. coli* T7 S30 Extract System for Circular DNA, Cat.

#L1130, Promega) in the presence of expressed in the presence of [³⁵S]Met (1200 Ci/mmol, 10 mCi/mL, MP Biomedicals). To produce IVTT protein, 20 µL of S30 Ppremix, 5 µL of amino acid mixture (-Met), 15 µL of T7 S30 extract, 1 µL of [³⁵S]Met, 3 to 10 µL of template plasmid (>1 µg) depending on its concentration, and nuclease-free water to make total volume to 50 µL. The IVTT mixture was incubated for 1 h at 37°C and centrifuged at 25,000 g to remove the pellet before use.

2.4.1.10. Preparation of liposomes

A solution of 1,2-diphytanoyl phosphophatidylcholine (DPhPC, Avanti Polar Lipids) in pentane (200 µL, 5 mg mL⁻¹) was dried under high vacuum in a desiccator overnight. The DPhPC was suspended in rehydration buffer (1 mL of 10 mM Tris-HCl, pH 8.0, containing 1 mM EDTA), and extruded through two 0.1 µm polycarbonate membranes to obtain unilamellar liposomes.

2.4.1.11. Protein folding and preparation of heptameric αHL pore

To form heptamers, the purified synthetic monomer in denaturing condition (purification buffer) was washed in a centrifugal filter with TE buffer by repeated concentration and dilution. The supernatant from IVTT product were mixed in various ratios in concentration (7:0, 6:1, 5:2, 3.5:3.5, 2:5, and 1:6) with the denaturants-removed synthetic monomer in the presence of rabbit erythrocyte membranes (rRBCm) as previously described⁷⁶. The mixture was incubated for 1 h at 37 °C and the reaction mixture was centrifuged at 21,000 g. The resulting membrane pellet containing assembled homo- and heteroheptamers was washed and solubilised in a loading buffer and run on a 5 % SDS-polyacrylamide gel overnight at 60 V. The gel was vacuum-dried for 3-4 h onto Whatman 3M filter paper, without heating. The dried gel was exposed to autoradiographic film for 6 hrs and the developed film was used to locate the position of the radio-labelled, heptameric protein in the gel. The protein-containing region of the gel was excised, rehydrated with 400 µL of TE8 and crushed. The

polyacrylamide was removed by centrifuging the suspension at 14,000 g for 10 min at room temperature through a hydrophilic PVDF micro spin column (Generon Ltd.)^{76, 77} Aliquots of the purified protein were stored at -80 °C.

To form homoheptamer ($\text{SM}_{\text{FL}})_7$, a lipid vesicle (DPhPC, 1 mg mL⁻¹)⁷⁸ was used. A solution of DPhPC in pentane (200 μL , 5 mg mL⁻¹) was dried out by N₂ purging and it was further dried under high vacuum in a desiccator overnight. The dried lipid was suspended in rehydration buffer (20 mM TE8) to make 1 mg mL⁻¹ and sonicated for 10 min to disrupt a large multilamellar vesicle (LMV) to a small unilamellar vesicle (SMV). The size of vesicles was then homogenized by extruding the lipid solution with an extruder (Avanti Polar Lipids, Inc.) using a filter paper (Avanti Polar Lipids, Inc., pore size, 100 μm).

1 μL of liposome in 1 M KCl was added to 5 μL (0.3 mg mL⁻¹) of and incubated at 37 °C for 1 hr. The resulting mixture was loaded onto two different lanes on a 7.5 % (Bio-Rad) SDS-PAGE without further washes and run at 100 V for 1.5 hrs. One lane was stained with Coomassie (Expedeon) to visualize the region of protein. The protein-containing region of the gel was excised from the other unstained lane. The gel was crushed in 300 μL of TE8 and the polyacrylamide was removed by the same method used in purification of heteroheptameric pores.

2.4.1.12. Limited Proteolysis of αHL pores with proteinase K

Fresh proteinase K (Sigma Aldrich) solution of 0.1 mg mL⁻¹ was prepared⁷⁶. ($\text{SM}_{\text{FL}})_7$ pores were prepared as described in the previous section. 2 μL of the proteinase K solution was added. Proteolysis was allowed to occur for 10 min at room temperature before the proteinase K was inactivated by the addition of 40 mM PMSF in ethanol (phenylmethylsulfonyl fluoride, a serine protease inhibitor). The proteinase K-treated ($\text{SM}_{\text{FL}})_7$ pores were heated at 95 °C for 15 min in a PCR machine, to denature oligomeric species to monomer⁷⁶.

2.4.2. Assembly of a truncated pore

2.4.2.1. Plasmid preparation of N-terminal fragment (NTF Δ 6)

The plasmid for NTF Δ 6-3M (pTXB3-NTF Δ 6-3M) was produced by the same methods described in section 2.4.1.1. The gene encoding 1-113 residues were PCR-amplified by two primers: 5'-TATGGATTCAACGGTAATGTTACTGGTTGCATCACGGGAGATGCACTAGT-3' (fwd) 5'-CCCTTGGCTCTTCCGCACATATACTCTTTTGTATCAATC-3' (rev) and cloned to pTXB3. For the pTXB3-NTF Δ 6-3F plasmid, homologous recombination was carried out. Briefly, the two PCR-fragments were mixed with *E. coli* to produce a circular DNA plasmid by homologous recombination (section 2.4.1.2.) The primers used to construct NTF Δ 6-3F with PCR amplification are: set 1) 5'-CGATTGATACAAAAGAGTATTTCTGCATCACGGGAGATGCACTAGT-3' (D_fwd) and 5'-CAGAAGTGGTCCTGCAACTTTAT-3' (A_rev), set 2) 5'-ATAAAGTTGCAGGACCACCTTG-3' (A_fwd) and 5'-GAAAATACTCTTTTGTATCAATCGAAT-3' (D_rev) (The underlined sequence represents M113F). The same thermocycling conditions, transformation, and plasmid extraction methods were used as described in section 2.4.1.2.

2.4.2.2. Plasmid preparation of C-terminal fragment (CTF Δ 6)

The plasmid for CTF Δ 6 was produced from by two successive homologous recombinations as described in Figure 2.31. 1) addition of 115-126 (Δ 120-125) to the pT7-SC1-CDH plasmid bearing CTF[127-293]-D8H6 and 2) deletion of 133-138 from the plasmid. For addition and deletion of sequences, the following sets of primers were used for PCR amplification.

Addition of 115-126(Δ 120-125):

set 1) 5'-TGCACTTTAACTTATGGAGGTGATGATACCGGCAAAATTGGAGGC-3' (F_fwd) and 5'-CAGAAGTGGTCCTGCAACTTTAT-3' (A_rev), set 2) 5'-ATAAAGTTGC

AGGACCACTTCTG-3' (A_fwd) and 5'-ATCACCTCCATAAGTTAAAGTGCAACCGCCT ATCTGCTCACGATG-3' (F_rev). The plasmid was transformed into *E. coli* and spread on to the agar plate containing 100 µg mL⁻¹. A colony was picked up and grown in liquid LB media (10 mL). The cell was collected by the centrifugation of the media containing *E. coli* and the plasmid was extracted from the cell pellet by use of Qiagen Plasmid Mini Kit. The plasmid was sequence-verified before it was used for next round homologous recombination.

Deletion of 133-138:

set 1) 5'- CGACGATGATCACCACCATCACCATCATTGATAAGCTTGGATCCGGCT GC -3' (C_fwd) and 5'-CAGAAGTGGTCCTGCAACTTTAT-3' (A_rev), set 2) 5'- ATAAAGTTGCAGGACCACTTCTG-3' (A_fwd) and 5'- CGACGATGATCACCACCATC ACCATCATTGATAAGCTTGGATCCGGCTGC -3' (C_rev). The same thermocycling conditions, transformation, and plasmid extraction methods were used as described in section 2.2.10.2.

2.4.2.3. Preparation of NTF Δ 6 and CTF Δ 6

NTF Δ 6-3M and NTF Δ 6-3F were prepared and purified using the method described in section 2.4.1.3. The purified products with the α thioester were characterized with SDS-PAGE and LC-MS.

CTF Δ 6 was prepared as described in section 2.4.1.4., however, purification of CTF Δ 6 was carried out by using automated AKTA FPLC with a Ni²⁺ sepharose affinity column (HisTrap FF™, 1mL, GE Healthcare Life Sciences). The column was washed with 10 column volumes of distilled water and then equilibrated with 10 column bed volumes of 95:5 % (A:B) of equilibration buffer (*Buffer A* : 6 M Gu.HCl, 50 mM phosphate (pH 7.5), 250 mM NaCl, and 5 mM TCEP; *Buffer B* : 6 M Gu.HCl, 50 mM phosphate (pH 7.5), 250 mM NaCl, 5 mM TCEP, 500 mM imidazole). The cell lysate dissolved in the equilibration buffer was directly injected and unbound proteins (flow-

through) were discarded. The CTF was eluted with elution buffer using a slow linear gradient 5-60 % of B for 1 hr. The eluted protein was collected and characterized with LC-MS.

2.4.2.4. NCL for a truncated pore

Before ligation reaction, CTF was treated with 0.4 M MeONH₂.HCl to deprotect N-Cys from thiazolidine group. The ligation reactions were separately carried out by mixing NTF_{Δ6-3M} and NTF_{Δ6-3F} with CTF_{Δ6} in NCL buffer as described in the section 2.4.1.6. The crude product monomeric TSM_{Δ6-3M} and TSM_{Δ6-3F} were loaded on SDS-PAGE without further purification.

2.4.2.5. Purification of synthetic monomer

The crude product was purified with gel filtration by using the same column (Superdex™ 200 10/300 GL) in denaturing condition (8 M urea) to prevent oligomerization of the synthetic monomer to heptameric pore. The purification was monitored in real-time at 280 nm. All the fractions were collected and the fraction containing the purified synthetic monomers was further characterized with SDS-PAGE and LC-MS.

2.4.2.6. Hemolysis assays

The hemolysis assays for TSM_{Δ6-3M} and TSM_{Δ6-3F} were carried out as described in section 2.4.1.8.

2.4.2.7. Protein folding and preparation of heptameric TBMΔ6

Due to the ability of truncated monomers oligomerizing to a heptameric pore in the absence of a lipid membrane, the oligomerization was conducted both in the presence and absence of membrane. Briefly, the denaturing buffer dissolving purified SM_{Δ6-3M} and SM_{Δ6-3F} was exchanged in a centrifugal filter with TE8 by repeated concentration and dilution. The resulting SM_{Δ6-3M} and SM_{Δ6-3F} were then incubated in 37 °C either in the absence or presence of a lipid membrane (1 μL, DPhPC 1 mg mL⁻¹).

2.4.3. General Analytical LC-MS

All the prepared polypeptides and reaction products were analyzed on a Waters LCT accurate-mass time-of-flight instrument (ESI-TOF MS) by using the positive ion mode. LC-MS employed a Chromolith RP-18e 50 mm × 2 mm column with a linear gradient of 5-100% eluant B in eluant A, over 8 min, with a flow rate of 1 mL min⁻¹ (eluant A: 0.1% formic acid in water; eluant B: 0.1% formic acid in acetonitrile). The mass spectra were generated with MassLynx software (version 4.1, Waters) by combining mass spectra obtained over the major peak in the total ion chromatogram.

2.4.4. General methods required for SCR

2.4.4.1. Preparation for SCR

Electrical recordings was carried out using planar bilayers formed of 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC, Avanti Polar Lipids). The planar lipid bilayer was formed across an aperture (100-150 µm in diameter)⁷⁹ in a 25 µm-thick polytetrafluoroethylene film (Teflon, Goodfellow, Cambridge) by the method of Montal and Mueller⁸⁰. The aperture in Teflon film was generated by a high-voltage spark generator, which produces a spark ~ 15 mm long at a frequency of 1Hz. The high temperature caused by the spark passing through the Teflon film produced a round aperture.

The two half compartments are fastened together with screws, clamping the PTFE film between them, and sealed with silicone glue (Dow Corning) to ensure water-tightness^{79, 81}. To form a planar lipid bilayer, the film around the aperture was treated with a drop (~1 µL) of 1-10 % (v/v) hexadecane in pentane. 500 µL of buffer solution was added to each compartment, followed by a drop (~5 µL) of DPhPC (5 mg mL⁻¹) in pentane on the buffer solution surface. The pentane was allowed to evaporate and a bilayer was formed by lowering and raising the buffer level contacting the film.

The aperture is pretreated with a drop ($\sim 0.5 \mu\text{L}$) of hexadecane in pentane (1 %, v/v). After the pentane has evaporated, the electrodes are connected and both compartments of the chamber are filled with aqueous buffer. A small drop ($\sim 1 \mu\text{L}$) of lipid in pentane solution (e.g., DPhPC, 5 mg mL^{-1}) is then applied directly to the top of the buffer. After the pentane has evaporated, the aqueous buffer is pipetted from the bottom of a compartment to lower the lipid monolayer at the water-air interface past the aperture level. The solution level is then slowly raised back above the aperture. Repeating the procedure 3-4 times in each compartment leads to a bilayer formation, when the electrical current becomes zero.

2.4.4.2. Preparation of electrodes

Ag/AgCl electrodes are prepared from Ag wire, 1.5 mm in diameter (Sigma-Aldrich). The wire was cut to 1 inch length and coated with AgCl by overnight immersion in ~ 30 % hypochlorite (NaClO) solution (Sigma-Aldrich)⁸². Two electrodes were used in the measurement as the working reference electrodes. These electrodes were connected to the *trans* and *cis* compartment of the recording equipment, respectively.

2.4.4.3. Perfusion

Once a successful bilayer was formed, heptameric αHL was added to the *cis* compartment and a potential (+50-200 mV) was applied across the bilayer with Ag/AgCl electrodes until a single channel is inserted into the lipid bilayer. Then, the solution in *cis* compartment was perfused with the same buffer ($200 \mu\text{L} \times 15$ times) to prevent further insertions of additional pores. Electric current recordings were performed at room temperature with a patch clamp amplifier (Axopatch 200B, Molecular Devices). Unless otherwise stated, the electric signal was recorded with a

low-pass Bessel filter (80 dB/decade) with a corner frequency of 5 kHz, and sampled at 25 kHz with a DigiData 1320 A/D converter (Molecular Devices).

2.4.4.4. Cyclodextrin measurements

Aliquots of β CD were added to the *trans* compartment to produce final concentrations of 25, 50, 75 and 100 μ M. For the determination of each set of kinetic constants with (SM_{FL})₇ and (TSM _{Δ 6-3M})₇ pores, three independent experiments were performed and data acquired for 5 min were analysed. Events were collected by 'single-channel searches' with the Clampfit software (version 10.1), and short events (<0.1 ms) were excluded. For kinetic analysis, dwell times were collected for each pore, plotted as unbinned cumulative histograms and fitted to single exponentials. For (TSM _{Δ 6-3F})₇ pores, 10 μ M of β CD and am₇ β CD were used for kinetic analysis due to its high binding affinity.

2.4.4.5. Sulfhydryl-directed chemistry

Water-soluble, sulfhydryl-directed MePEG-OPSS (monomethoxypoly(ethylene glycol)-*o*-pyridyl disulfide) reagents were purchased from Nanocs (ChemQuest, UK). Including the contribution of the OPSS group (142 Da), the mass of the reagent was $M_w = 5000$ Da. The reagent was dissolved in 1M KCl, 20 mM Tris (pH 7.5) to give 100 mM and added to the *trans* side of bilayer to give 1 mM in final concentration. 500 mM DTT was prepared and diluted in *cis* and *trans* compartment to 5 mM final concentration to form a reversible disulfide bond.

2.4.5. Protein structure modelling

The structure was predicted by using MODELLER with the script files provided from <https://salilab.org/modeller/examples/automodel/model-default.py> after modifications relevant to α HL. The modeller was run with a python script and an alignment file (see Appendix) under the command 'mod9.14 model-default.py'.

References

1. Evans, T.C., Benner, J. & Xu, M.Q. The in vitro ligation of bacterially expressed proteins using an intein from *Methanobacterium thermoautotrophicum*. *J Biol Chem* **274**, 3923-3926 (1999).
2. Dawson, P.E., Muir, T.W., Clarklewis, I. & Kent, S.B.H. Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779 (1994).
3. Hackeng, T.M., Griffin, J.H. & Dawson, P.E. Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology. *Proc Natl Acad Sci U S A* **96**, 10068-10073 (1999).
4. Muir, T.W. Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* **72**, 249-289 (2003).
5. Yamazaki, T. et al. Segmental isotope labeling for protein NMR using peptide splicing. *Journal of the American Chemical Society* **120**, 5591-5592 (1998).
6. Xu, R., Ayers, B., Cowburn, D. & Muir, T.W. Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies. *P Natl Acad Sci USA* **96**, 388-393 (1999).
7. Pentelute, B.L., Barker, A.P., Janowiak, B.E., Kent, S.B.H. & Collier, R.J. A semisynthesis platform for investigating structure function relationships in the N-terminal domain of the anthrax lethal factor. *ACS Chem Biol* **5**, 359-364 (2010).
8. Valiyaveetil, F.I., MacKinnon, R. & Muir, T.W. Semisynthesis and folding of the potassium channel KcsA. *J Am Chem Soc* **124**, 9113-9120 (2002).
9. Evans, T.C., Benner, J. & Xu, M.Q. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* **7**, 2256-2264 (1998).
10. Wu, W., Wood, D.W., Belfort, G., Derbyshire, V. & Belfort, M. Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. *Nucleic Acids Res* **30**, 4864-4871 (2002).
11. Otomo, T., Ito, N., Kyogoku, Y. & Yamazaki, T. NMR observation of selected segments in a larger protein: Central-segment isotope labeling through intein-mediated ligation. *Biochemistry* **38**, 16040-16044 (1999).
12. Cotton, G.J., Ayers, B., Xu, R. & Muir, T.W. Insertion of a synthetic peptide into a recombinant protein framework: A protein biosensor. *J Am Chem Soc* **121**, 1100-1101 (1999).
13. Blaschke, U.K., Cotton, G.J. & Muir, T.W. Synthesis of multi-domain proteins using expressed protein ligation: Strategies for segmental isotopic labeling of internal regions. *Tetrahedron* **56**, 9461-9470 (2000).

14. Muir, T.W., Sondhi, D. & Cole, P.A. Expressed protein ligation: A general method for protein engineering. *Proc Natl Acad Sci U S A* **95**, 6705-6710 (1998).
15. Camarero, J.A. et al. Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proc Natl Acad Sci U S A* **99**, 8536-8541 (2002).
16. Southworth, M.W., Amaya, K., Evans, T.C., Xu, M.Q. & Perler, F.B. Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein. *Biotechniques* **27**, 110-+ (1999).
17. Muralidharan, V. & Muir, T.W. Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nature Methods* **3**, 429-438 (2006).
18. Valiyaveetil, F.I., MacKinnon, R. & Muir, T.W. Semisynthesis and folding of the potassium channel KcsA. *J Am Chem Soc* **124**, 9113-9120 (2002).
19. Alsina, J., Yokum, T.S., Albericio, F. & Barany, G. Backbone amide linker (BAL) strategy for N-alpha-9-fluorenylmethoxycarbonyl (Fmoc) solid-phase synthesis of unprotected peptide p-nitroanilides and thioesters. *J Org Chem* **64**, 8761-8769 (1999).
20. Ingenito, R., Bianchi, E., Fattori, D. & Pessi, A. Solid phase synthesis of peptide C-terminal thioesters by Fmoc/t-Bu chemistry. *J Am Chem Soc* **121**, 11369-11374 (1999).
21. Shin, Y. et al. Fmoc-based synthesis of peptide-(alpha)thioesters: Application to the total chemical synthesis of a glycoprotein by native chemical ligation. *J Am Chem Soc* **121**, 11684-11689 (1999).
22. Swinnen, D. & Hilvert, D. Facile, Fmoc-compatible solid-phase synthesis of peptide C-terminal thioesters. *Org Lett* **2**, 2439-2442 (2000).
23. Camarero, J.A., Hackel, B.J., de Yoreo, J.J. & Mitchell, A.R. Fmoc-based synthesis of peptide alpha-thioesters using an aryl hydrazine support. *J Org Chem* **69**, 4145-4151 (2004).
24. Blanco-Canosa, J.B. & Dawson, P.E. An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew Chem Int Ed* **47**, 6851-6855 (2008).
25. Ling, J.J.J., Policarpo, R.L., Rabideau, A.E., Liao, X.L. & Pentelute, B.L. Protein thioester synthesis enabled by sortase. *J Am Chem Soc* **134**, 10749-10752 (2012).

26. Camarero, J.A., Cotton, G.J., Adeva, A. & Muir, T.W. Chemical ligation of unprotected peptides directly from a solid support. *J Pept Res* **51**, 303-316 (1998).
27. Lee, J., Kwon, Y., Pentelute, B.L. & Bang, D. Use of Model Peptide Reactions for the Characterization of Kinetically Controlled Ligation. *Bioconjugate Chem* **22**, 1645-1649 (2011).
28. Camarero, J.A., Fushman, D., Cowburn, D. & Muir, T.W. Peptide chemical ligation inside living cells: In vivo generation of a circular protein domain. *Bioorgan Med Chem* **9**, 2479-2484 (2001).
29. Erlanson, D.A., Chytil, M. & Verdine, G.L. The leucine zipper domain controls the orientation of AP-1 in the NFAT center dot AP-1 center dot DNA complex. *Chem Biol* **3**, 981-991 (1996).
30. Tolbert, T.J. & Wong, C.H. New methods for proteomic research: Preparation of proteins with N-terminal cysteines for labeling and conjugation. *Angew Chem Int Ed* **41**, 2171-2174 (2002).
31. Bang, D., Pentelute, B.L. & Kent, S.B. Kinetically controlled ligation for the convergent chemical synthesis of proteins. *Angew Chem Int Ed Engl* **45**, 3985-3988 (2006).
32. Clayton, D. et al. Total chemical synthesis and electrophysiological characterization of mechanosensitive channels from Escherichia coli and Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **101**, 4764-4769 (2004).
33. Valiyaveetil, F.I., Leonetti, M., Muir, T.W. & MacKinnon, R. Ion selectivity in a semisynthetic K⁺ channel locked in the conductive conformation. *Science* **314**, 1004-1007 (2006).
34. Devaraneni, P.K. et al. Semisynthetic K⁺ channels show that the constricted conformation of the selectivity filter is not the C-type inactivated state. *Proc Natl Acad Sci U S A* **110**, 15698-15703 (2013).
35. Komarov, A.G., Linn, K.M., Devereaux, J.J. & Valiyaveetil, F.I. Modular strategy for the semisynthesis of a K⁺ channel: investigating interactions of the pore helix. *ACS Chem Biol* **4**, 1029-1038 (2009).
36. Linn, K.M., Derebe, M.G., Jiang, Y.X. & Valiyaveetil, F.I. Semisynthesis of NaK, a Na⁺ and K⁺ conducting ion channel. *Biochemistry* **49**, 4450-4456 (2010).
37. Bayley, H., Cheley, S., Harrington, L. & Syeda, R. Wrestling with native chemical ligation. *ACS Chem Biol* **4**, 983-985 (2009).

38. Stoddart, D. et al. Functional truncated membrane pores. *Proc Natl Acad Sci U S A* **111**, 2425-2430 (2014).
39. Stoddart, D. et al. Functional truncated membrane pores. *Proc Natl Acad Sci U S A* **111**, 2425-2430 (2014).
40. Muir, T.W. Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* **72**, 249-289 (2003).
41. Gentle, I.E., De Souza, D.P. & Baca, M. Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjug Chem* **15**, 658-663 (2004).
42. Erlanson, D.A., Chytil, M. & Verdine, G.L. The leucine zipper domain controls the orientation of AP-1 in the NFAT.AP-1.DNA complex. *Chem Biol* **3**, 981-991 (1996).
43. Tolbert, T.J. & Wong, C.H. New methods for proteomic research: preparation of proteins with N-terminal cysteines for labeling and conjugation. *Angew Chem Int Ed Engl* **41**, 2171-2174 (2002).
44. Ling, J.J., Polcarpo, R.L., Rabideau, A.E., Liao, X. & Pentelute, B.L. Protein thioester synthesis enabled by sortase. *J Am Chem Soc* **134**, 10749-10752 (2012).
45. Tolbert, T.J., Franke, D. & Wong, C.H. A new strategy for glycoprotein synthesis: ligation of synthetic glycopeptides with truncated proteins expressed in E-coli as TEV protease cleavable fusion protein. *Bioorgan Med Chem* **13**, 909-915 (2005).
46. Tolbert, T.J. & Wong, C.H. Conjugation of glycopeptide thioesters to expressed protein fragments: semisynthesis of glycosylated interleukin-2. *Methods Mol Biol* **283**, 255-266 (2004).
47. Sherman, F., Stewart, J.W. & Tsunasawa, S. Methionine or not methionine at the beginning of a protein. *BioEssays* **3**, 27-31 (1985).
48. Wilson, D.B. & Dintzis, H.M. Protein chain initiation in rabbit reticulocytes. *Proc Natl Acad Sci U S A* **66**, 1282-1289 (1970).
49. Yoshida, A., Watanabe, S. & Morris, J. Initiation of rabbit hemoglobin synthesis: methionine and formylmethionine at N-terminal. *Proc Natl Acad Sci U S A* **67**, 1600-& (1970).
50. Jackson, R. & Hunter, T. Role of methionine in the initiation of haemoglobin synthesis. *Nature* **227**, 672-676 (1970).

51. Hirel, P.H., Schmitter, J.M., Dessen, P., Fayat, G. & Blanquet, S. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci U S A* **86**, 8247-8251 (1989).
52. Boissel, J.P., Kasper, T.J. & Bunn, H.F. Cotranslational amino-terminal processing of cytosolic proteins. Cell-free expression of site-directed mutants of human hemoglobin. *J Biol Chem* **263**, 8443-8449 (1988).
53. Walker, B., Krishnasastri, M., Zorn, L., Kasianowicz, J. & Bayley, H. Functional expression of the alpha-hemolysin of *Staphylococcus aureus* in intact *Escherichia coli* and in cell lysates. Deletion of five C-terminal amino acids selectively impairs hemolytic activity. *J Biol Chem* **267**, 10902-10909 (1992).
54. Henricus, M. & Bayley, H., Vol. Thesis (Master of Science) (University of Oxford, 2010).
55. Walker, B., Krishnasastri, M., Zorn, L. & Bayley, H. Assembly of the oligomeric membrane pore formed by Staphylococcal alpha-hemolysin examined by truncation mutagenesis. *J Biol Chem* **267**, 21782-21786 (1992).
56. Walker, B., Braha, O., Cheley, S. & Bayley, H. An intermediate in the assembly of a pore-forming protein trapped with a genetically-engineered switch. *Chem Biol* **2**, 99-105 (1995).
57. Walker, B., Braha, O., Cheley, S. & Bayley, H. An intermediate in the assembly of a pore-forming protein trapped with a genetically-engineered switch. *Chem Biol* **2**, 99-105 (1995).
58. Song, L. et al. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859-1866 (1996).
59. Olson, R., Nariya, H., Yokota, K., Kamio, Y. & Gouaux, E. Crystal structure of staphylococcal LukF delineates conformational changes accompanying formation of a transmembrane channel. *Nat Struct Biol* **6**, 134-140 (1999).
60. Maglia, G. et al. Droplet networks with incorporated protein diodes show collective properties. *Nature Nanotechnology* **4**, 437-440 (2009).
61. Gu, L.Q. & Bayley, H. Interaction of the non-covalent molecular adapter, beta-cyclodextrin, with the staphylococcal alpha-hemolysin pore. *Biophys J* **80**, 494a-494a (2001).
62. Gu, L.Q. et al. Reversal of charge selectivity in transmembrane protein pores by using noncovalent molecular adapters. *Proc Natl Acad Sci U S A* **97**, 3959-3964 (2000).

63. Gu, L.Q. & Bayley, H. Interaction of the noncovalent molecular adapter, beta-cyclodextrin, with the staphylococcal alpha-hemolysin pore. *Biophys J* **79**, 1967-1975 (2000).
64. Mccaldon, P. & Argos, P. Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins* **4**, 99-122 (1988).
65. Walker, B. & Bayley, H. Key residues for membrane binding, oligomerization, and pore forming activity of staphylococcal alpha-hemolysin identified by cysteine scanning mutagenesis and targeted chemical modification. *J Biol Chem* **270**, 23065-23071 (1995).
66. Movileanu, L., Cheley, S. & Bayley, H. Partitioning of individual flexible polymers into a nanoscopic protein pore. *Biophys J* **85**, 897-910 (2003).
67. Gu, L.Q., Cheley, S. & Bayley, H. Prolonged residence time of a noncovalent molecular adapter, beta-cyclodextrin, within the lumen of mutant alpha-hemolysin pores. *J Gen Physiol* **118**, 481-494 (2001).
68. Banerjee, A. et al. Molecular bases of cyclodextrin adapter interactions with engineered protein nanopores. *Proc Natl Acad Sci U S A* **107**, 8165-8170 (2010).
69. Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *P Natl Acad Sci USA* **106**, 7702-7707 (2009).
70. Stoddart, D. et al. Nucleobase recognition in ssDNA at the central constriction of the alpha-hemolysin pore. *Nano Letters* **10**, 3633-3637 (2010).
71. Harrington, L., Cheley, S., Alexander, L.T., Knapp, S. & Bayley, H. Stochastic detection of Pim protein kinases reveals electrostatically enhanced association of a peptide substrate. *Proc Natl Acad Sci U S A* **110**, E4417-4426 (2013).
72. Howorka, S. & Bayley, H. Improved protocol for high-throughput cysteine scanning mutagenesis. *Biotechniques* **25**, 764-+ (1998).
73. Miles, G., Cheley, S., Braha, O. & Bayley, H. The staphylococcal leukocidin bicomponent toxin forms large ionic channels. *Biochemistry* **40**, 8514-8522 (2001).
74. Bang, D., Pentelute, B.L. & Kent, S.B.H. Kinetically controlled ligation for the convergent chemical synthesis of proteins. *Angew Chem Int Edit* **45**, 3985-3988 (2006).

75. Cheley, S. et al. Spontaneous oligomerization of a staphylococcal alpha-hemolysin conformationally constrained by removal of residues that form the transmembrane beta-barrel. *Protein Eng* **10**, 1433-1443 (1997).
76. Cheley, S., Braha, G., Lu, X.F., Conlan, S. & Bayley, H. A functional protein pore with a "retro" transmembrane domain. *Protein Sci* **8**, 1257-1267 (1999).
77. Stoddart, D., Maglia, G., Mikhailova, E., Heron, A.J. & Bayley, H. Multiple base-recognition sites in a biological nanopore: two heads are better than one. *Angew Chem Int Ed Engl* **49**, 556-559 (2010).
78. Mantri, S., Sapra, K.T., Cheley, S., Sharp, T.H. & Bayley, H. An engineered dimeric protein pore that spans adjacent lipid bilayers. *Nat Commun* **4**, 1725 (2013).
79. Mayer, M., Kriebel, J.K., Tosteson, M.T. & Whitesides, G.M. Microfabricated teflon membranes for low-noise recordings of ion channels in planar lipid bilayers. *Biophys J* **85**, 2684-2695 (2003).
80. Montal, M. & Mueller, P. Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. *Proc Natl Acad Sci U S A* **69**, 3561-3566 (1972).
81. Maglia, G., Heron, A.J., Stoddart, D., Japrun, D. & Bayley, H. Analysis of single nucleic acid molecules with protein nanopores. *Methods Enzymol* **475**, 591-623 (2010).
82. Krnjevic, K. Microelectrode methods for intracellular recording and iontophoresis. *Trends in Neurosciences* **5**, 133.

Contents here were submitted to *Proc. Natl. Acad. Sci. U S A.* (Lee, J. and Bayley H., 2015)

Chapter 3. A semisynthetic protein nanoreactor for single-molecule chemistry

3.1. Introduction

The staphylococcal α -hemolysin (α HL) pore has been engineered for a variety of applications in biotechnology. For example, the pore can function as a detector for stochastic sensing¹⁻³ and DNA sequencing^{4, 5}, and act as a 'nanoreactor' for single-molecule covalent chemistry⁶. In the nanoreactor approach, individual bond-making and bond-breaking steps can be observed with sub-millisecond time-resolution by monitoring the ionic current carried by the pore under an applied potential. Single-molecule chemistry has attracted interest because it provides insights into fundamental chemical processes that cannot be observed at the ensemble scale⁶⁻⁸. However, chemistry carried out within the α HL nanoreactor has so far been confined to the reactions of thiolates (i.e., deprotonated Cys side chains)⁷⁻¹² and derivatives of the side chains of Cys residues^{13, 14}. The incorporation of unnatural amino acids into the α HL pore would substantially advance single-molecule chemistry, through the incorporation of a large variety of individual side chains as well as multiple substitutions. We have produced α HL polypeptides with unnatural alkyl and aryl amino acids¹⁵ by using *in vitro* chemically aminoacylated tRNAs. However, the method is demanding and often gives poor results, e.g. when more than one amino acid is introduced. *In vivo* methods that use engineered synthetase/tRNA pairs also enable the incorporation of unnatural amino acids^{16, 17}. Two different amino acids^{17, 18}, but so far not more, can be incorporated. However, not all amino acids are successfully handled, and a new one

often requires additional synthetase development. Further, various amino acid side chains are destroyed *in vivo*, often by reduction¹⁸.

Another means to incorporate unnatural amino acids is through NCL¹⁹⁻²¹, wherein the amino acid of interest is introduced in a peptide obtained by chemical synthesis^{22, 23}. The advantage of NCL is that virtually any amino acid or collection of amino acids can be incorporated into the peptide chain, provided the side chain is stable towards the conditions of solid phase peptide synthesis (SPPS) and NCL, or can be protected and deprotected. An extension of NCL that has seen widespread success is expressed protein ligation (EPL)²⁴⁻²⁶, a semisynthetic method in which large segments of the polypeptide chain are bacterially expressed. However, relatively little work has been reported on the semisynthesis of membrane proteins²⁷⁻³⁰. This deficiency can be attributed to the hydrophobic nature of membrane proteins, which are poorly soluble in either folded or unfolded form³¹. For example, a polypeptide segment encompassing the transmembrane region of the α HL pore precipitates in neutral buffer (unpublished observation). The use of a denaturing environment during NCL may solve the solubility problem³². However, it raises the question of whether the semisynthetic protein can be refolded into the native form. Here, we incorporate an unnatural amino acid with a terminal alkyne side chain into the α HL polypeptide and demonstrate the functional activity of heptameric pores made from it. Further, we investigate click chemistry of the alkyne at the single-molecule level by using the semisynthetic pore as a nanoreactor.

3.2. Results and discussion

3.2.1. Preparation of Polypeptide Fragments

We designed polypeptide fragments that would provide for the efficient semisynthesis of a full-length α HL monomer by a three-way ligation strategy. In our design (Figure. 3.1.), a central segment (SP: SP[Thz¹¹⁴-Gly¹²⁶]-N-acyl-

benzimidazolinone(Nbz)) was obtained by chemical synthesis. An N-terminal fragment (NTF: NTF[Ala¹-Met¹¹³]- α thioester) and a C-terminal fragment (CTF: CTF[Cys¹²⁷-Asn²⁹³]-D₈H₆) were obtained after expression in *E. coli*.

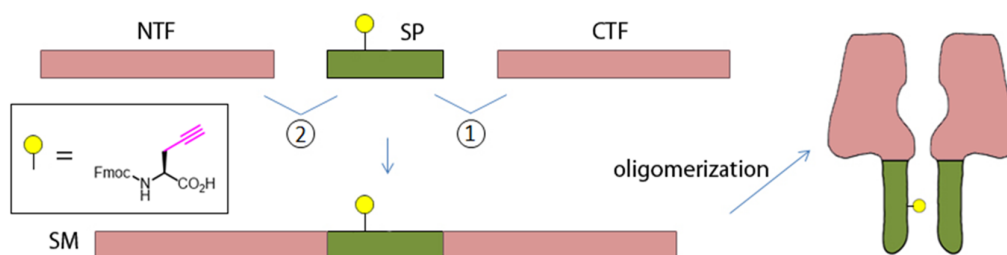


Figure 3.1. Schematic representation for Preparation of an α HL polypeptide containing an unnatural amino acid by semisynthesis. NTF and CTF were prepared by expression in *E. coli*, whereas the central peptide, SP, was produced by SPPS. An unnatural amino acid residue (in this case L-propargylglycine (Prp), inset) with a side chain containing a functional group of interest (yellow ball) was incorporated into the peptide. Two sequential NCL splice the three segments into a full-length α HL polypeptide.

Recombinant NTF[Ala¹-Met¹¹³]- α thioester was produced by thiolysis of a fusion protein, NTF-GyrA intein-CBD (NTF[Ala¹-Met¹¹³]-*Mycobacterium xenopi* DNA gyrase A (Mxe GyrA) intein-chitin binding domain (CBD)) (Figure. 1.13 and 2.5.) derived from the plasmid pTXB3. The DNA encoding residues 1-113 of α HL was PCR-amplified from a plasmid bearing a full-length α HL gene and cloned into pTXB3. After expression at 37°C, high levels of the fusion protein were found within inclusion bodies in the *E. coli* host. Following cell lysis and centrifugation, most of the pellet could be solubilized in denaturing buffer containing 8 M urea. To allow binding to chitin beads, the concentration of the urea was subsequently reduced to 1 M. Unbound proteins were washed off the column and the bound fusion protein was treated with 2-mercaptoethane sulfonate (MESNa) under gravity flow, yielding the NTF[Ala¹-Met¹¹³]-

^athioester in high purity as determined by SDS-PAGE (Figure 2.29a) and LC-MS (Figure 2.30).

We prepared another plasmid (pT7-SC1-CDH, Figure 2.8) for the production of CTF[Cys¹²⁷-Asn²⁹³]-D₈H₆ under control of the T7 promoter. In this plasmid, the codons for residues 1-126 in the pT7-SC1 plasmid, which contains a full-length α HL gene, are replaced with codons for Met-Cys. CTF was again obtained from inclusion bodies, and then purified in 6 M guanidine hydrochloride by use of the His₆ tag at the C terminus. The N-terminal Met was found to be cleaved and the unmasked Cys residue had undergone condensation with pyruvic acid, an abundant metabolite³³. We regenerated free Cys at the N terminus by cleaving the thiazolidine (Thz) with methoxylamine (MeONH₂) at pH 4 (Figure 2.9). The purified CTF was characterized by LC-MS (Figure 2.11).

The central synthetic peptide (SP:Thz¹¹⁴ThrLeu**Prp**TyrGlyPheAsnGlyAsnValThr Gly¹²⁶-Nbz), in which Thr-117 is replaced with L-propargylglycine (**Prp**), was prepared by SPPS³⁴ with Fmoc-protected amino acids (Figure. 3.2). The alkyne side chain of residue 117 was predicted to point into the transmembrane β barrel of the α HL pore³⁵; residue 117 is a sensitive position for the detection of chemical reactions⁶. The peptide as released from the resin contained a C-terminal Nbz group, which can undergo thiolysis, with 4-mercaptophenyl acetic acid (MPAA), to yield a peptide-^aarylthioester³⁴. The HPLC-purified peptide thioester (Figure 3.3a) was characterized by LC-MS (Figure 3.3b).

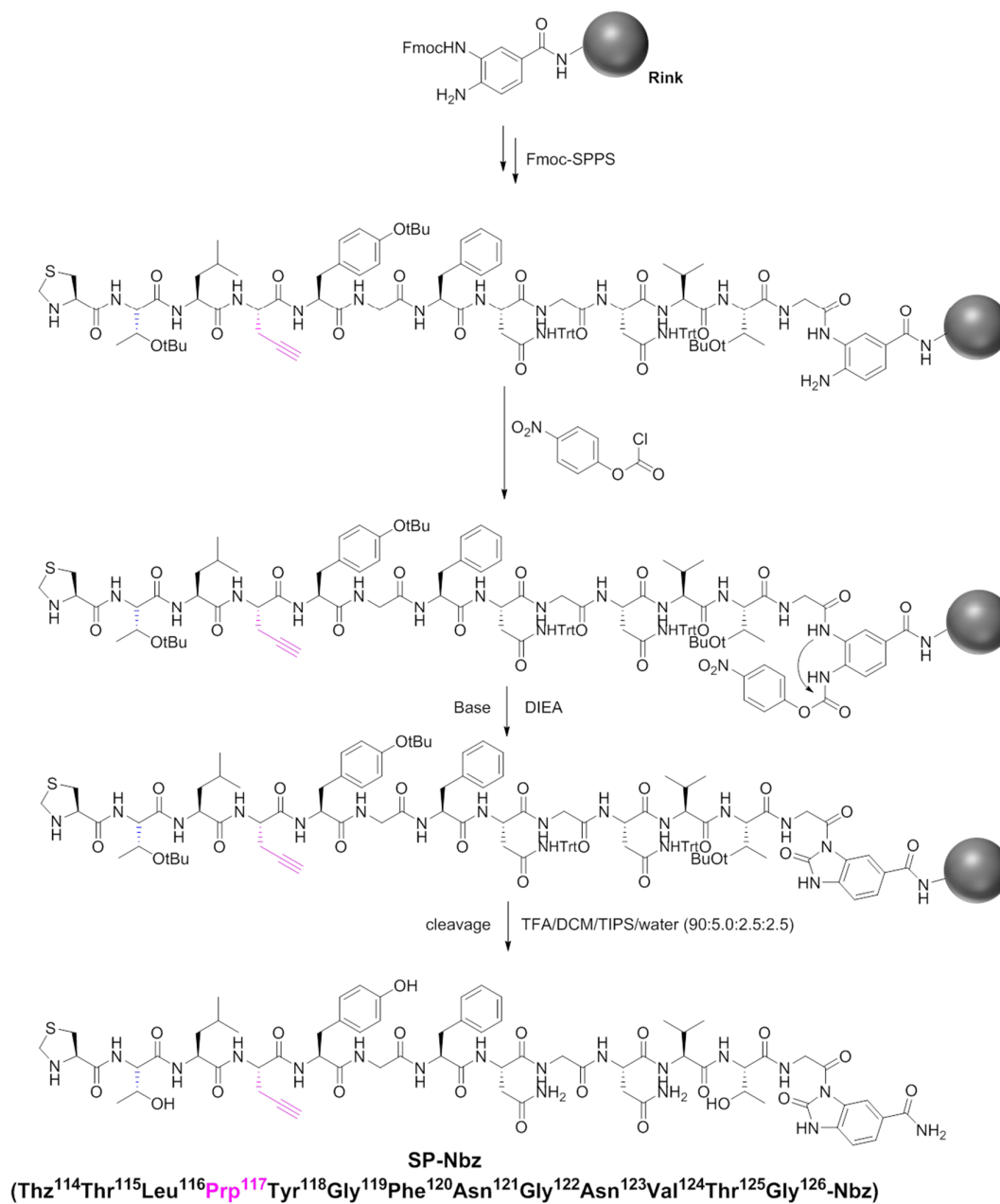


Figure 3.2. Synthesis of SP-Nbz. Synthetic scheme for SP-Nbz bearing an alkyne group (see experimental section 3.4.4 for details)

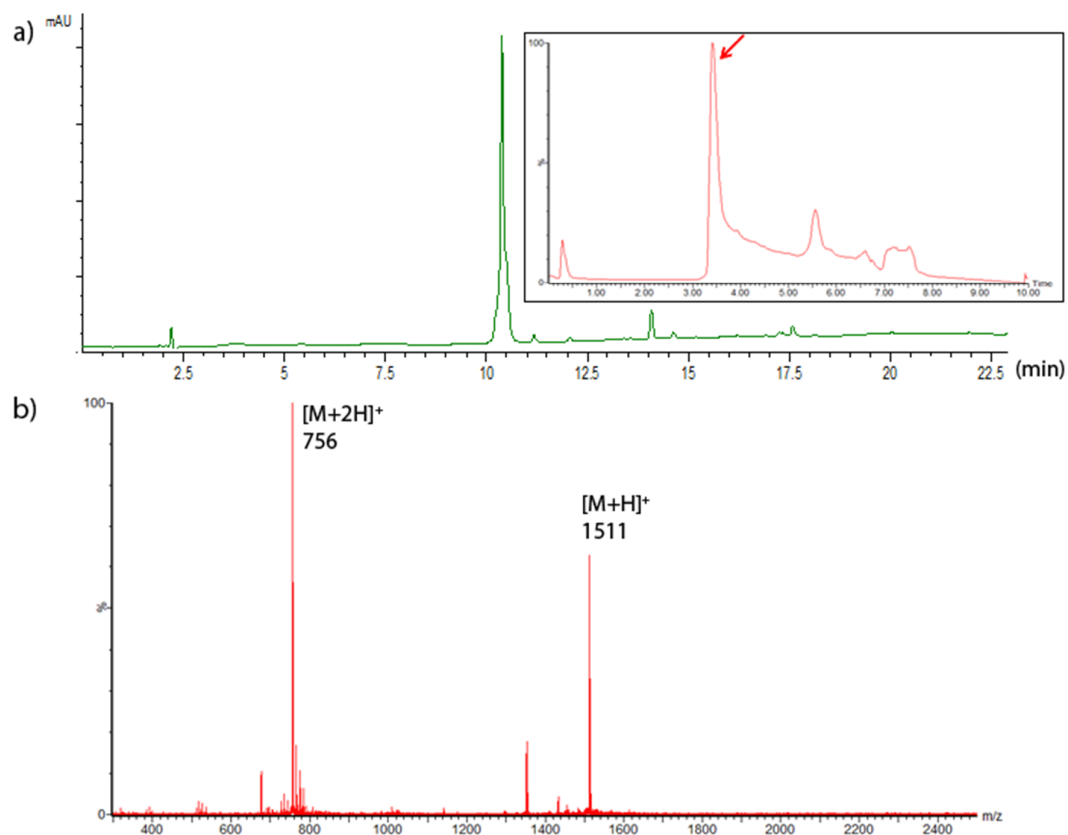


Figure 3.3. Characterization of SP-Nbz. a) The peptide purified by HPLC with a preparative C18 reverse-phase column (Vydac 218TP) was further characterized by analytical LC and LC-MS (inset) by use of C18 analytical columns (Polaris C18 A 5u and Chromolith RP-18e, respectively). The peak (red arrow, inset) was analyzed by combining MS obtained across the peak. b) ESI-MS spectrum ($[M+H]^+ = 1511$ (obs), 1511(calcd.)). The weaker additional peak ($[M+H]^+ = 1353$ (obs), 1353(calcd.)) is attributed to the peptide in which the acylurea has been hydrolyzed. This impurity does not react with the N-terminal Cys of CTF during the first ligation step.

3.2.2. Native Chemical Ligation

We then proceeded to assemble a full-length α HL polypeptide bearing the propargyl group (Figure. 3.1) with two sequential NCL reactions each producing a native peptide bond: 1) ligation of the synthetic peptide to CTF (Figure. 3.4 and 3.5) and 2) ligation of NTF to the first ligation product (Figure 3.6). We purified the final ligation product (Figure 3.6c and 3.6d), a synthetic monomer (SM) of α HL, by gel filtration under denaturing condition (8 M urea) followed by ion-exchange chromatography and characterized it by LC-MS, which yielded a single peak containing material of the theoretical mass (Figure 3.7). The purified synthetic monomer was concentrated under denaturing conditions (8 M urea) to prevent premature assembly into a heptameric pore.

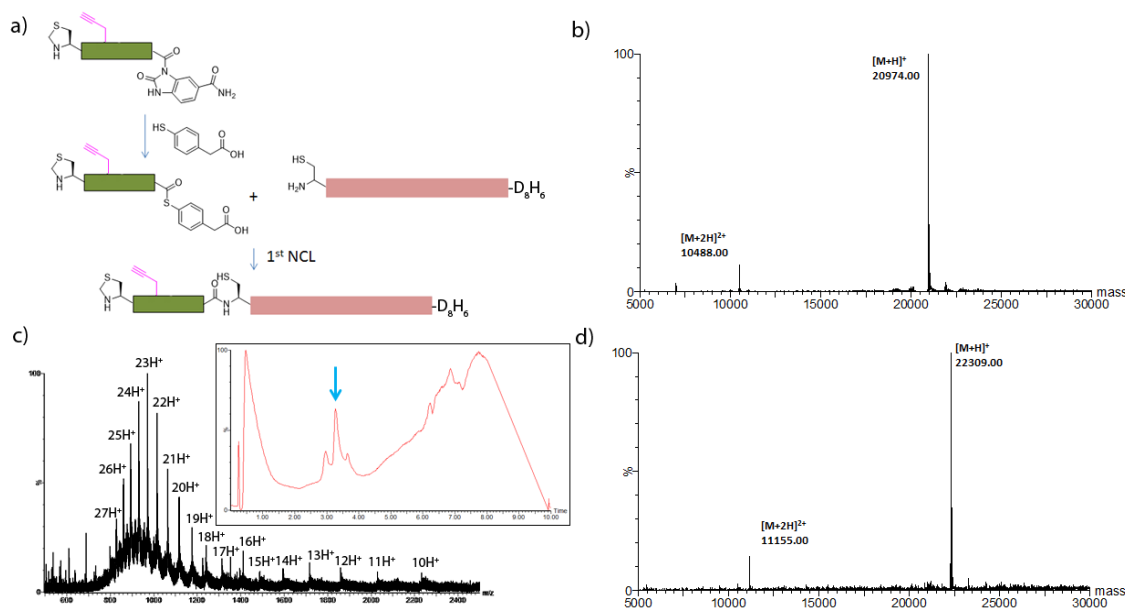


Figure 3.4. The first NCL reaction and characterization of the product. a) Schematic representation of the first ligation reaction between SP-Nbz and CTF. b) Deconvoluted ESI-MS of CTF before the first ligation. c) The inset shows the LC trace of the 1st ligation product. ESI-MS spectrum of the first ligation product obtained over the peak (blue arrow in the inset). d) Deconvoluted ESI-MS of the first ligation product. ([M+H]⁺ = 22309 (obs), 22308(calcd.)).

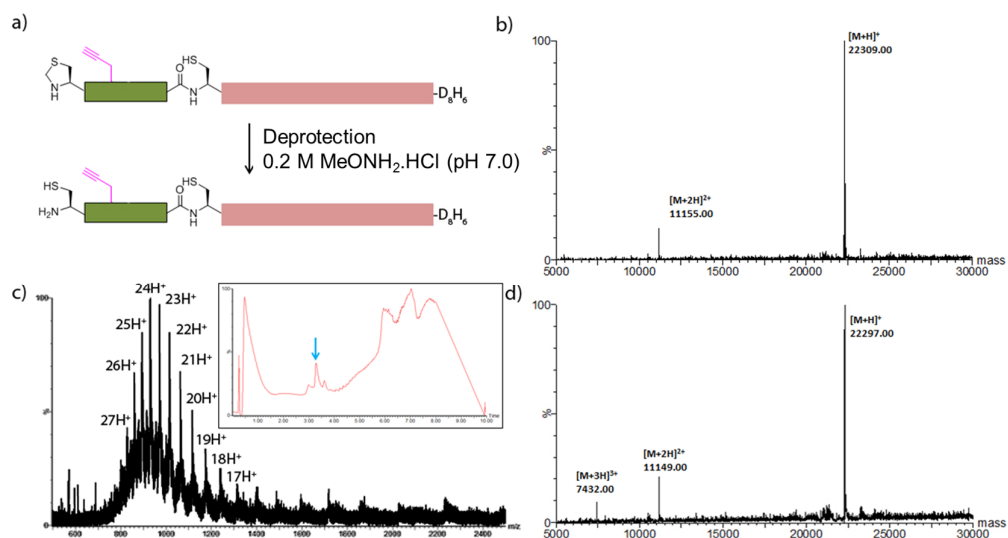


Figure 3.5. Deprotection of Thz114 after the first NCL reaction. **a)** Deprotection of the first ligation product. **b)** Deconvoluted ESI-MS of the first ligation product. **c)** The inset shows the LC trace after deprotection. The ESI-MS was generated by combining mass spectra obtained over the peak (blue arrow). **d)** Deconvoluted ESI-MS of the deprotected ligation product. ($[M+H]^+ = 22297$ (obs), 22297 (calcd.)). There is a decrease of 12 mass units over the first ligation product (Figure. 3.4b).

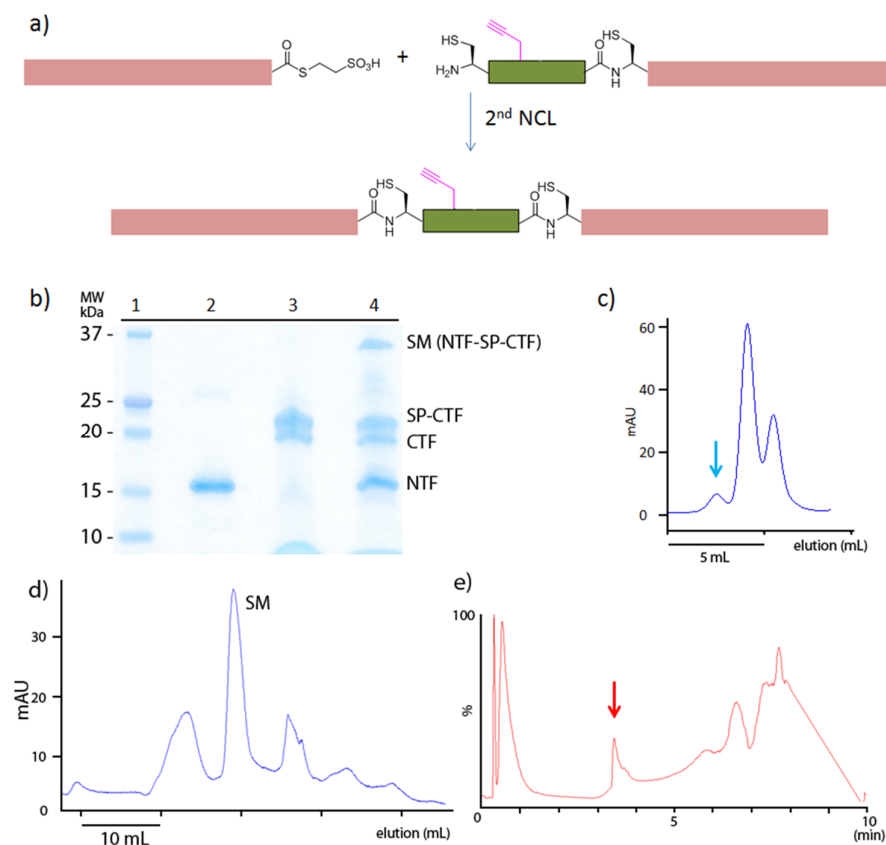


Figure 3.6. The ligation reaction between NTF and the first ligation product (second ligation). a) The ligation was carried out in 6 M Gu.HCl, 200 mM Na phosphate buffer (pH 6.8), 200 mM MPAA, 40 mM TCEP. b) The fragments used for NCL were examined by SDS-PAGE (Any kD, Bio-Rad). Lane 1: molecular mass markers; lane 2: unreacted NTF; lane 3: crude first ligation product; lane 4: crude second ligation product. c) The crude second ligation product was purified by gel filtration under denaturing conditions (8 M urea, 50 mM NaH₂PO₄ (pH 7.0), 5 mM TCEP). The fraction eluted in the first peak (blue arrow) of the column (Superdex™ 200 10/300 GL column) was collected. d) The collected fraction was further purified on an ion-exchange column (HiTrap Q HP, 5 mL). The fractions were collected and SM was identified by LC-MS. e) LC trace for SM (red arrow) on a C18 column.

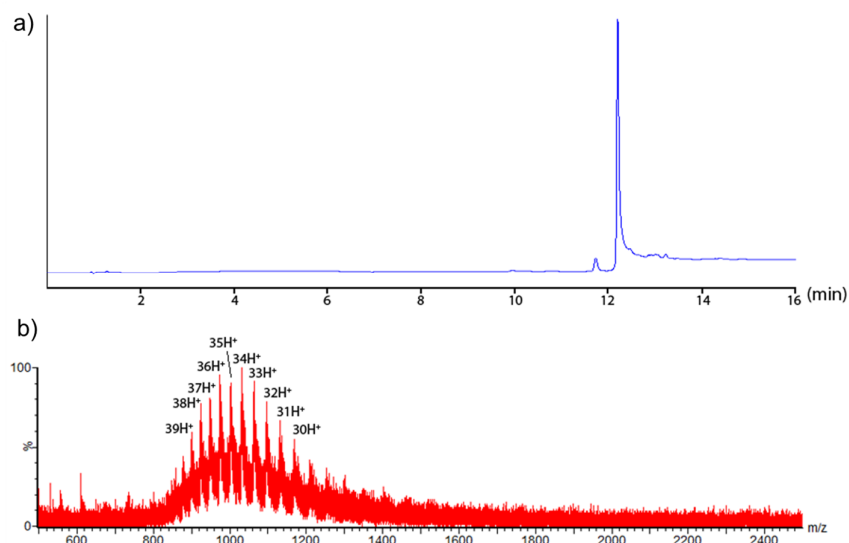


Figure 3.7. Characterization of the full-length synthetic monomer. a) The purity of SM was analysed by HPLC monitored at 280 nm, which yielded a single peak. b) The ESI-MS was generated by combining mass spectra obtained across the peak (red arrow in Figure 3.6e). The deconvoluted ESI-MS is shown in Figure 3.8. ($[M+H]^+ = 34994$ (obs), 34994 (calcd.).

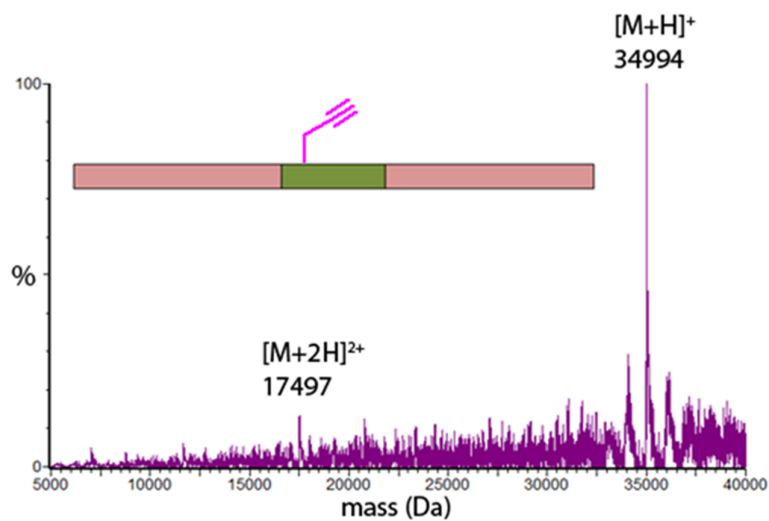


Figure 3.8. Deconvoluted mass spectrum of the synthetic α HL monomer (SM). The synthetic α HL monomer (SM) containing an alkyne group was prepared by NCL, purified, and characterized by LC-MS: ($[M+H]^+ = 34994$ (obs), 34994 Da (calcd.)).

3.2.3. Refolding of semisynthetic α HL monomers to heptameric pores.

We next tested whether SM could be correctly folded into active monomers. WT α HL monomers prepared by *in vitro* transcription and translation (IVTT) and denatured in 8 M urea completely regained hemolytic activity after the urea concentration was reduced to 1 M (data not shown)³⁶. The urea concentration of the SM solution was reduced to ~60 mM by repeated dilution and concentration in a centrifugal filter (Amicon, MWCO 3k). The SM was then examined for: (i) hemolytic activity towards rabbit red blood cells (rRBCs); (ii) heptamer formation on liposomes and red blood cell membranes; (iii) conformational integrity as reflected by limited proteolysis.

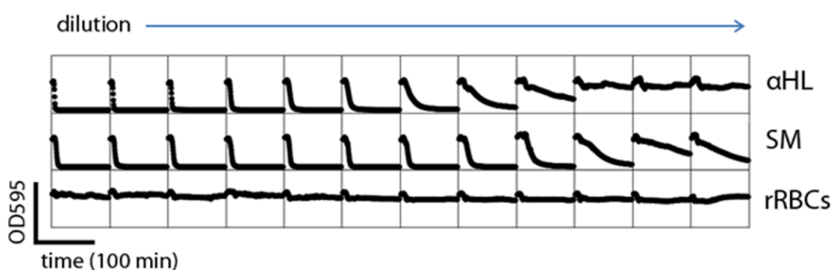


Figure 3.9. Hemolysis assays. Row 1: hemolysis by recombinant WT α HL; row 2: hemolysis by SM; row 3: rabbit red blood cells (rRBCs) with no α HL addition. The decrease in light scattering over time was recorded in a microplate reader. WT (1 μ L, 1.08 mg mL⁻¹) and SM α HL monomer (1 μ L, 4.02 mg mL⁻¹) were added to the first column of row 1 and row 2 of the plate, respectively, and each sample was serially two-fold diluted in a final volume of 100 μ L per well.

The specific hemolytic activity of SM ($HC_{50} = 39$ ng mL⁻¹ ($n = 3$); HC_{50} , the concentration of protein giving 50 % lysis at 100 min, Figure 3.9) was very similar to (Figure. 3.10) that of recombinant WT α HL monomer ($HC_{50} = 35 \pm 12$ ng mL⁻¹ ($n = 3$), previously measured³⁷ as 31 ng mL⁻¹). To visualize the formation of α HL heptamers, we incubated SM for 1 h at 37°C in the presence of liposomes (1 mg mL⁻¹, diphytanoyl

phosphatidylcholine, DPhPC)^{38, 39} and upon SDS-polyacrylamide gel electrophoresis, we observed a new band that migrated with a mobility corresponding to the heptameric state (**Fig. 1c**).

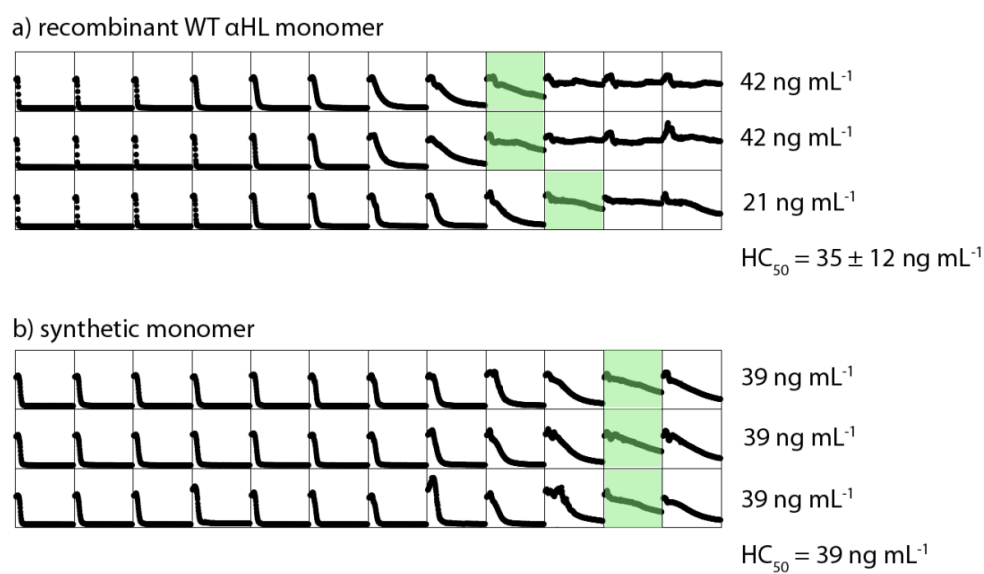


Figure 3.10. Three repetitive hemolysis assays. The concentrations of the proteins that showed 50% cell lysis in 100 min (green boxes) were the specific hemolytic activity values (HC_{50} , right). In each of three experiments (separate rows), the same initial concentrations of α HL protein were used. HC_{50} values for recombinant α HL-H₆ (a, see Figure 2.14 for purification) and synthetic monomer (b) were: $35 \pm 12 \text{ ng mL}^{-1}$ and 39 ng mL^{-1} , respectively. The first rows in a) and b) are shown in Figure. 3.8.

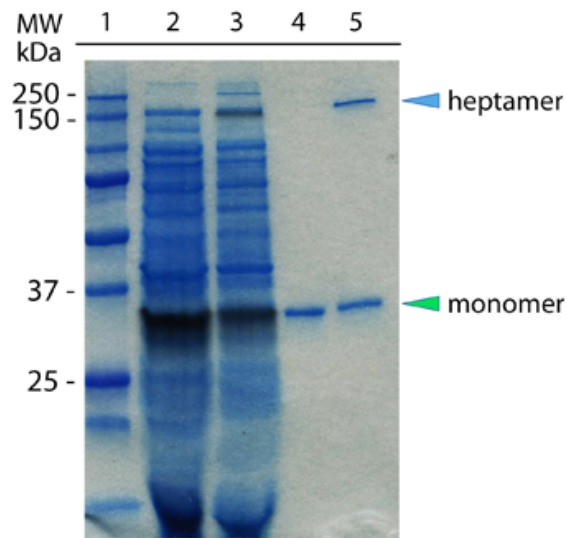


Figure 3.11. SDS-PAGE analysis of WT and synthetic α HL. Lane 1: molecular markers; lane 2: radiolabeled α HL monomer produced by IVTT; lane 3: radiolabeled WT₇ pores assembled in the presence of DPhPC liposomes (1 mg mL⁻¹); lane 4: purified synthetic α HL monomer (SM); lane 5: purified SM after treatment with DPhPC liposomes. SM₇ pores are formed, which co-migrate with the WT₇ pore. An autoradiogram is superimposed on the Coomassie Blue stained gel.

In addition, we incubated SM in different ratios with radiolabeled WT α HL produced by IVTT in the presence of rabbit red blood cell membranes (rRBCm). SM oligomerized to form heteroheptameric states with different stoichiometries (WT_{7-n}SM_n, n = 0-7) (Figure. 3.12). Together, these data show that SM forms functional heptameric pores.

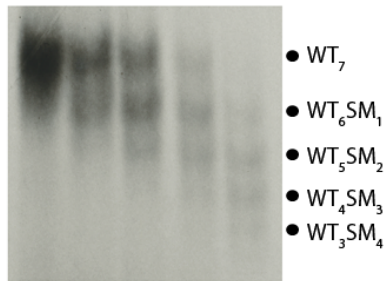


Figure 3.12. Heteroheptameric pores. WT α HL (radiolabelled protein was obtained by IVTT) and SM were mixed in various ratios and incubated at 37°C in the presence of rRBCm. The heptameric pores that formed were separated by SDS-PAGE based on the ability of the oligo-aspartate extension (D₈ tail) at the C terminus of SM to enhance electrophoretic mobility. WT₆SM₁ pores were eluted from the gel for single channel recording.

A limited proteolysis experiment was carried out with proteinase K, to determine the conformational state of the α HL polypeptide in the monomer and the heptamer⁴⁰⁻⁴². We found that the monomer was cleaved into two fragments of 20 kDa and 15 kDa (Figure 3.13), indicating that it is correctly folded (Figure 3.14, lane 4). When the heptamer formed on liposomes was similarly treated, a large fraction of the polypeptide chains was protected from the protease (Figure 3.14, lane 5), indicating that the heptamer was correctly assembled.

10 20 30 40 50 60
 ADSDINIKTG TTDIGSNTTV KTGDLVTYDK ENGMHKKVFY SFIDDKNHNK KLLVIRTKGT
 70 80 90 100 110 120
 IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD YYPRNSIDT **K EYMSTL** **YGF**
 130 140 150 160 170 180
NGNVTGDDTG KIGGLIGANV SIGHTLK YVQ PDFKITLESP TDKKVGWKVI FNNMVNQNWG
 190 200 210 220 230 240
 PYDRDSWNPV YGNQLFMKTR NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK
 250 260 270 280 290 300
 QQTNIDVIYE RVRDDYQLHW TSTNWKGTNT KDKWTD RSSE RYKIDWEKEE MTNDDDDDDD

 DHHHHHH

Figure 3.13. Amino acid sequence and proteinase K sites in the synthetic α HL monomer (SM). The yellow and blue boxes represent the propargylglycine residue and the transmembrane region, respectively. The blue arrows are predicted proteinase K digestion sites in the SM monomer³⁵ in solution. On binding to a bilayer, the sites in the transmembrane region become occluded in the prepore structure⁴². In the fully assembled pore, all the proteinase K sites are occluded. When the pore formed from SM was heat-denatured, most of the subunits had assumed the fully protected conformation.

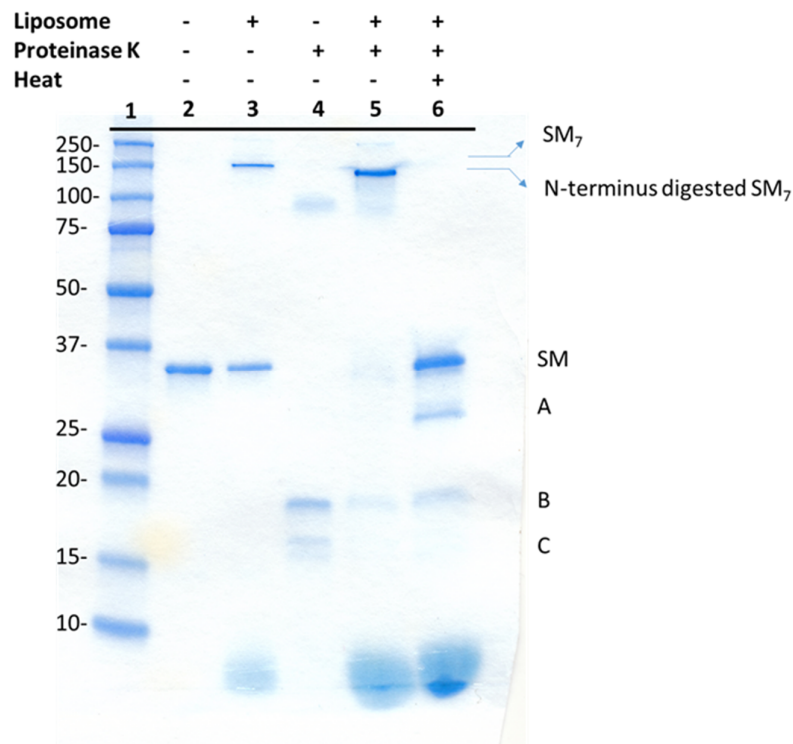


Figure 3.14. Limited proteolysis with proteinase K of SM in solution and in the presence of DPhPC liposomes (1 mg mL^{-1}). Lane 1: molecular markers; lane 2: SM; lane 3: SM treated with DPhPC liposomes; lane 4: SM in solution digested by proteinase K. Two fragments (~ 20 and 15 kDa) were generated (B and C); lane 5: SM treated with DPhPC liposome and further treated with only proteinase K. The oligomeric band shows a smaller size than the SM_7 in lane because only the N-terminal digestion site is open to the proteinase K; lane 6: SM treated with DPhPC liposomes was further treated with proteinase K and then heat-denatured at 95°C to disassemble the heptamers from which the monomer remained intact. The central proteolytic site is inaccessible in SM_7 , and most of the SM polypeptide is undigested. A: fragment generated after cleavage near the N terminus in the pre-pore form of SM_7 ($\sim 32 \text{ kDa}$)⁴¹.

3.2.4. Comparison of electrical properties of WT to semisynthetic pores

We examined the electrical properties of WT₆SM₁ pores (Figure 3.15) by determining the mean conductance under defined conditions (Figure. 3.16a) and measuring IV curves (Figure. 3.16b). The conductance of WT₆SM₁ (0.79 ± 0.07 nS at +50 mV in 1 M KCl, 20 mM Bis-Tris propane, pH 8.5) was similar to that of the WT₇ pore (~ 0.88 nS, same conditions).

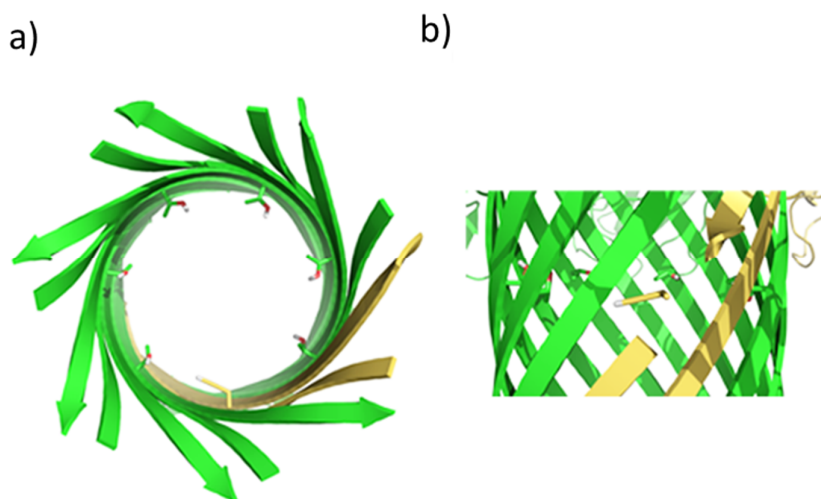


Figure 3.15. Structure of the WT₆SM₁ αHL pore. Top (a) and side view (b) of the β barrel (residues 109-150) of the WT₆SM₁ pore. The alkyne group of the L-propargylglycine in SM is shown as a yellow stick structure, while the Thr-117 residues of the WT subunits are in green. The main chain of 117 residue was removed to clearly show the alkyne sidechain. The crystal structure of αHL (PDB: 7AHL) was used as a template. The structure of WT₆SM₁ was generated by an energy minimization method by using AMBER (version 12, see Materials and methods, section 3.4.1.6).

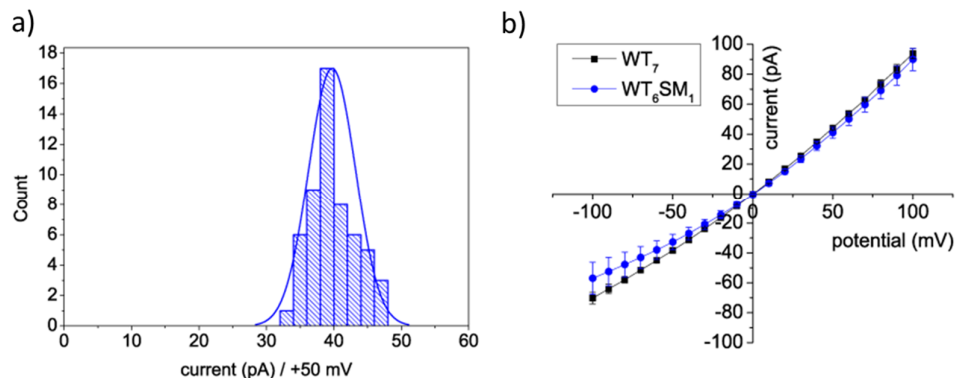


Figure 3.16. Electrical properties of the WT₆SM₁ αHL pore. a) Histogram of unitary conductance values of WT₆SM₁ pores (n = 55). b) IV curves of WT₆SM₁ (mean values, n = 9) and WT αHL pores (mean values, n = 5). The latter were made from WT subunits prepared by IVTT.

We also investigated the binding kinetics of β-cyclodextrin (βCD) with WT₆SM₁ to confirm that the structure of the transmembrane β barrel was intact. βCD is a noncovalent molecular adaptor that acts as a transient channel blocker when it is lodged in the lumen of the αHL pore^{43, 44}. In single-channel recordings, the addition of βCD to the *trans* compartment produced reversible transient, partial blockades of the ionic current. The mean dwell time of βCD in the pore (τ_{off}) did not vary with βCD concentration, whereas the mean lifetime of the unoccupied state (τ_{on}) was inversely proportional to the concentration suggesting a bimolecular interaction (Figure. 3.17). We determined the association rate constant (k_{on}) and dissociation rate constant (k_{off}) of βCD binding from the τ values: $k_{\text{on}} = 1.1 \pm 0.04 \times 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$; $k_{\text{off}} = 1.4 \pm 0.1 \times 10^3 \text{ s}^{-1}$; $K_D = k_{\text{off}}/k_{\text{on}} = 12.7 \pm 0.08 \times 10^{-3} \text{ M}$. Similar values were obtained previously for the WT αHL pore⁴⁴. Therefore, the alkyne residue presented by the SM subunit neither affects the electrical properties of the αHL pore, nor its ability to bind the βCD adaptor (Figure 3.18).

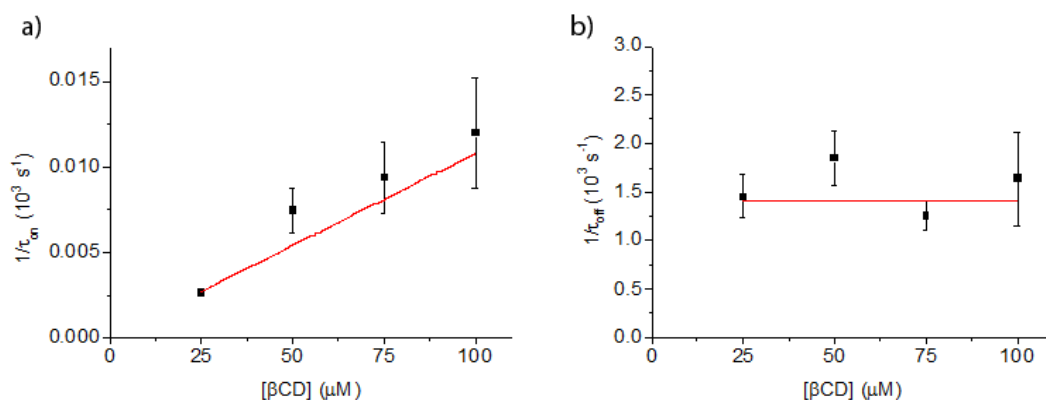


Figure 3.17. Dependence of τ_{off} and τ_{on} on the concentration of βCD (trans) for WT6SM1 pores. Kinetic constants were calculated by using (a) $k_{off} = 1/\tau_{off}$, (b) $k_{on} = 1/\tau_{on}[\beta\text{CD}]$, where [βCD] is the concentration of βCD. k_{on} was obtained from the slope of the linear fit of $1/\tau_{on}$ versus [βCD]. k_{off} was independent of [βCD] and was obtained from the average of the $1/\tau_{off}$ values. $k_{on} = 1.1 \pm 0.3 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$; $k_{off} = 1.4 \pm 0.1 \times 10^3 \text{ s}^{-1}$; $K_D = k_{off} / k_{on} = 1.3 \pm 0.4 \times 10^{-2} \text{ M}$ ($n = 3$).

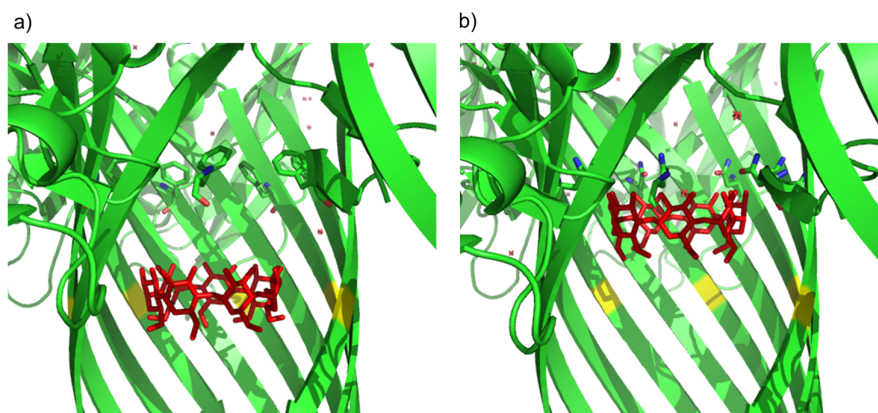


Figure 3.18. Binding of βCD in the M113F and M113N mutant pores . The crystal structure of the M113F (a, PDB: 3M3R) and M113N (b, PDB: 3M4E) mutant pore shows a binding of βCD molecule (red) in the lumen. Note that the binding site of βCD differs due to the different interaction of the side chain at 113¹⁵. The residue at 117 (yellow) where the alkyne group is incorporated does not affect to binding of the βCD.

3.2.5. CuAAC reaction in the semisynthetic pore

We then used the protein nanoreactor approach to investigate click chemistry of the alkyne side chain in WT₆SM₁ at the single-molecule level. The copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC), a click reaction⁴⁵, has been widely studied in materials science⁴⁶, bioconjugation⁴⁷, and drug discovery⁴⁸.

We first examined the unfunctionalized WT₇pore (Figure. 3.20) in the presence of the reagents used for the click reaction, namely the Cu(I) catalyst and azide substrates (Figure. 3.19, APB400: 0.4 kDa azide-PEG-biotin; APB700: 0.7 kDa azide-PEG-biotin).

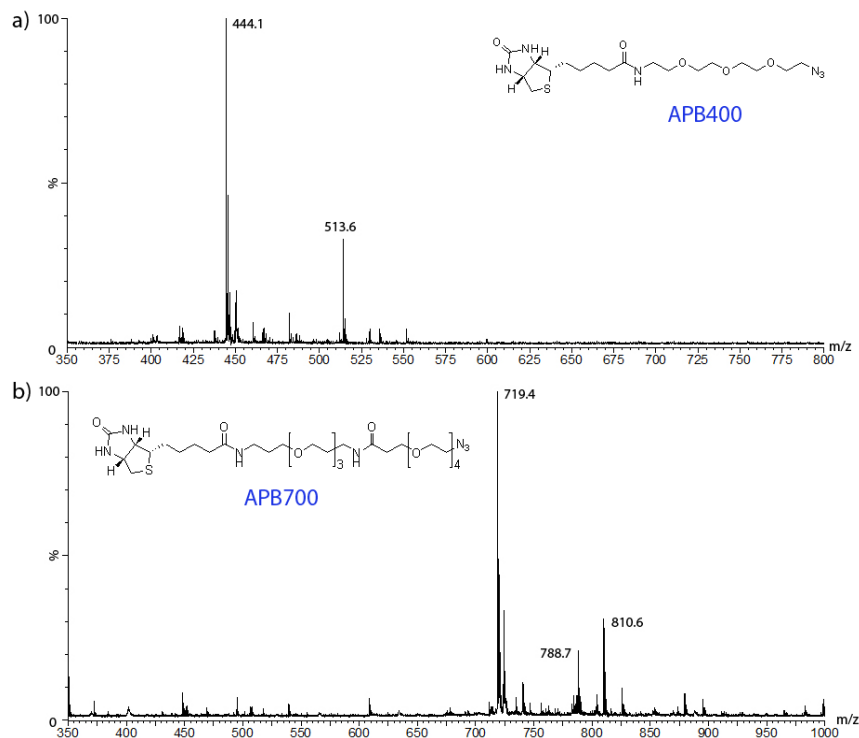


Figure 3.19. MALDI-TOF mass spectra of azide-PEG-biotins. a) azide-PEG₃-biotin (APB400 (Jena Bioscience, Germany)), $[M+H]^+ = 444.1$ (obs), 445.5 Da (calcd.). b) azide-PEG(3+4)-biotin (APB700, AGTC Bioproduct, UK), $[M+H]^+ = 719.4$ (obs), 720.9 Da (calcd.). The materials were analysed as received.

Cu(I) (20 mM) and APB400 (0.5 to 2.0 mM), neither separately nor together, affected the WT₇ pore (Figure. 3.20a-g), whereas APB700, either by itself (Figure. 3.20h-k) or with Cu(I) (Figure. 3.20l), produced reversible events, which increased in frequency with APB700 concentration. These events most likely arise from entry of APB700 into the lumen of the pore for periods sufficient to be detected without covalent attachment⁴⁹.

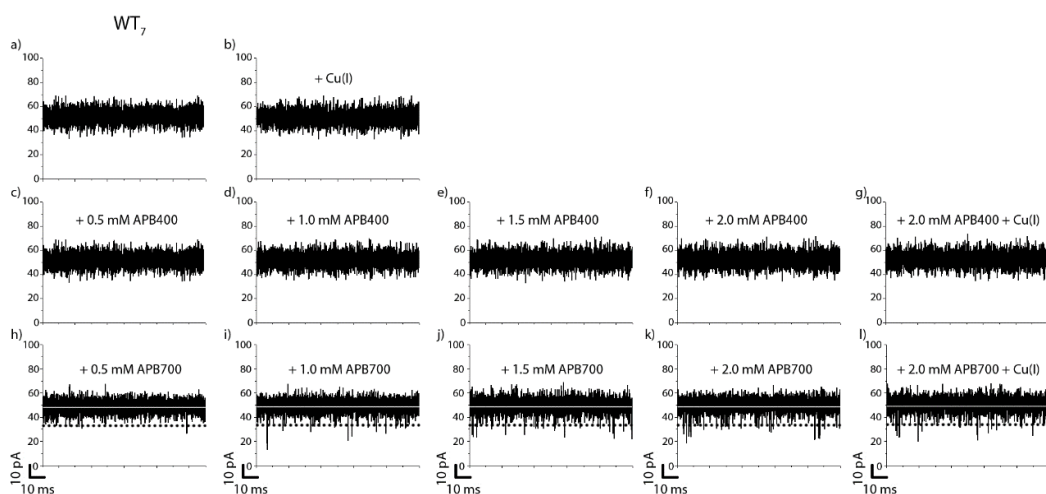


Figure 3.20. Currents observed after the addition of an azide substrate and catalyst to a single WT α HL pore (WT₇). a) Current trace with no substrate and catalyst. b) Current after the addition of 20 mM Cu(I). c-f) The sequential addition of increasing concentrations of APB400 (0.5 to 2 mM). g) The subsequent addition of 20 mM Cu(I). No binding events were seen in 'c' to 'g'. h-k) The sequential addition of increasing concentrations of APB700 (0.5 to 2 mM). Ionic current fluctuations are observed. l) The subsequent addition of 20 mM Cu(I). No change in the trace is seen. The binding kinetics were analyzed by threshold-based event detection in Clampfit (version 10.1). Every spike crossing the dotted line ($\Delta I = -15.0$ pA, from the mean current) was counted as an event (See Figure 3.18). k_{on} , k_{off} , and K_D for these binding events were determined from dwell-time analyses: $k_{on} = 1.3 \pm 0.2 \times 10^5 \text{ M}^{-1}\text{s}^{-1}$; $k_{off} = 8.7 \pm 0.3 \times 10^3 \text{ s}^{-1}$; $K_D = 6.7 \pm 1.0 \times 10^{-2} \text{ M}$ ($n = 3$). Cu(I) was added to both the *cis* and *trans* compartments, while the substrate was added to *trans* side. Measurements were performed in 1 M KCl, 20 mM Bis-Tris propane, 25 μM TCEP, pH 8.5. The applied potential was +50 mV and the filter corner frequency was 10 kHz (sampling: 20 μs).

By examining the concentration dependence of the blockades (APB700, 0.5 mM to 2 mM) in the absence of Cu(I), we obtained: $k_{on} = 1.3 \pm 0.2 \times 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$; $k_{off} = 8.7 \pm 0.3 \times 10^3 \text{ s}^{-1}$; $K_D = k_{off}/k_{on} = 6.7 \pm 1.0 \times 10^{-2} \text{ M}$ (Figure 3.21), similar to the values obtained by Movileanu et al. for a 940 Da PEG. The mean lifetime of the blockades was short ($\sim 100 \text{ } \mu\text{s}$), resulting in a broad distribution of amplitudes under our data acquisition conditions. No permanent current decrease was observed with the WT₇ pore during 3 h of monitoring after the addition of Cu(I), with either of the azides present in both compartments.

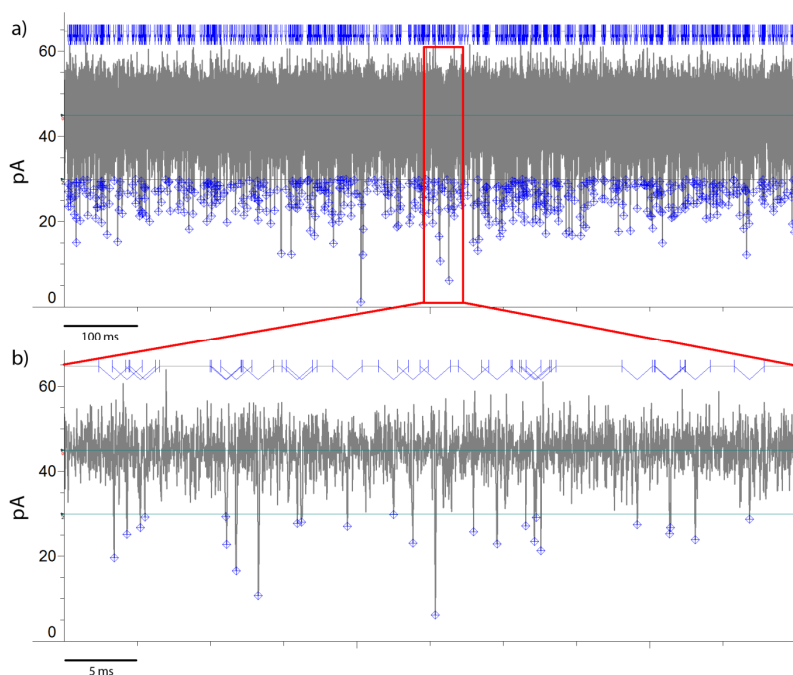


Figure 3.21. Binding of APB700 to the WT₆SM₁ pore in the presence of 20 mM Cu(I). The two green lines were used for a threshold-based search in Clampfit (version 10.1). The events from a 30 s segment were examined to determine the kinetics after each addition of substrate. k_{on} was given by $(n/\text{time})/[\text{APB700}]$, where n is the number of events. k_{off} was calculated as $1/\tau_{off}$, where τ_{off} was the mean value of half-width of the peak.

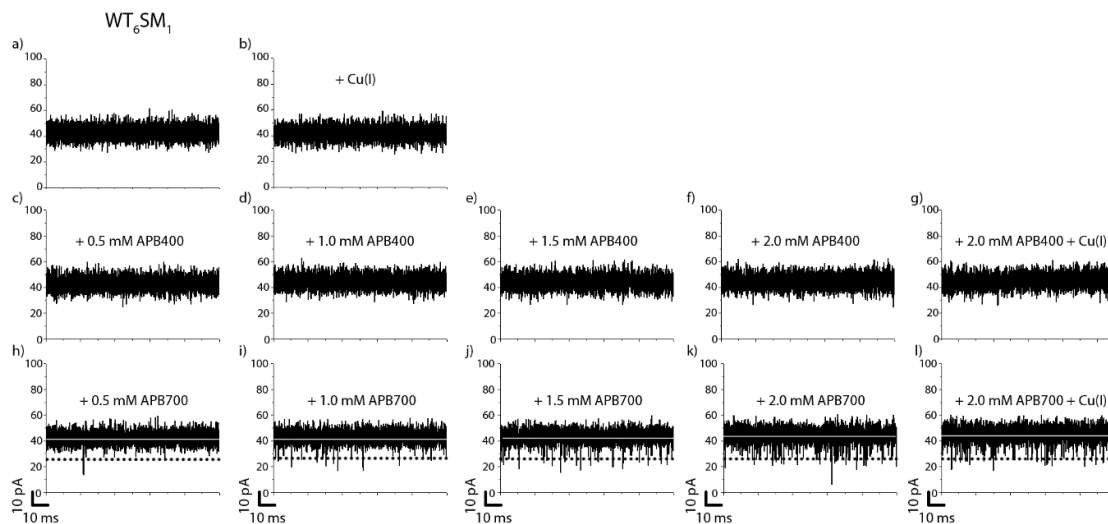


Figure 3.22. Currents observed after the addition of an azide substrate and catalyst to a single WT₆SM₁ α HL pore. a) Current trace with no substrate and catalyst. b) Current after the addition of 20 mM Cu(I). c-f) The sequential addition of increasing concentrations of APB400 (0.5 to 2 mM). g) The subsequent addition of 20 mM Cu(I). No binding events were seen in 'c' to 'g'. h-k) The sequential addition of increasing concentrations of APB700 (0.5 to 2 mM). Ionic current fluctuations are observed. l) The subsequent addition of 20 mM Cu(I). No change in the trace is seen. The binding kinetics were analyzed by threshold-based event detection in Clampfit (version 10.1). Every spike crossing the dotted line ($\Delta I = -15.0$ pA, from the mean current) was counted as an event (See Figure. 3.21). k_{on} , k_{off} , and K_D for these binding events were determined from dwell-time analyses: $k_{on} = 1.6 \pm 0.2 \times 10^5 \text{ M}^{-1}\text{s}^{-1}$; $k_{off} = 5.6 \pm 1.1 \times 10^3 \text{ s}^{-1}$; $K_D = 3.5 \pm 0.8 \times 10^{-2} \text{ M}$ ($n = 3$). Cu(I) was added to both the *cis* and *trans* compartments, while the substrate was added to *trans* side. Measurements were performed in 1 M KCl, 20 mM Bis-Tris propane, 25 μM TCEP, pH 8.5. The applied potential was +50 mV and the filter corner frequency was 10 kHz (sampling: 20 μs).

We then carried out chemistry with single WT₆SM₁ pores. As in the case of the WT₇ pore, Cu(I) added to both the *cis* and *trans* compartments at the same time did not

generate significant changes in the current (Figure. 3.20a and b), suggesting that if a Cu acetylide (I in Figure. 3.21a) is formed it is short-lived. When APB700 was added to the *trans* compartment in the absence of Cu(I), short reversible current blockades characteristic of entry and exit from the pore were seen, which were similar to those observed with the WT₇ pore: $k_{on} = 1.6 \pm 0.2 \times 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$; $k_{off} = 5.6 \pm 1.1 \times 10^3 \text{ s}^{-1}$; $K_D = k_{off}/k_{on} = 3.5 \pm 0.8 \times 10^{-2} \text{ M}$ (Figure. 3.21 c-g and h-l). In the case of APB400, the events were presumably too short to be registered at the data acquisition rate employed⁴⁹.

By contrast with WT₇, the addition of Cu(I) to both compartments in the presence of 2 mM APB400 or APB700 (*trans*) eventually led to an irreversible partial blockade of the WT₆SM₁ pore, suggesting that Cu(I) triggered a reaction between the substrate and the alkyne within the pore. Interestingly, the blockade of the pore proceeded in two steps (Figure. 3.23b-d and Table 3.1). An intermediate state (S_i) was formed first, characterized by a small current decrease ($\delta I_{b1} = I_{S_0} - I_{S_i}$, where S₀ is the unreacted state of the pore), which was subsequently converted to a second state (S_b) associated with a larger current decrease ($\delta I_{b2} = I_{S_i} - I_{S_b}$).

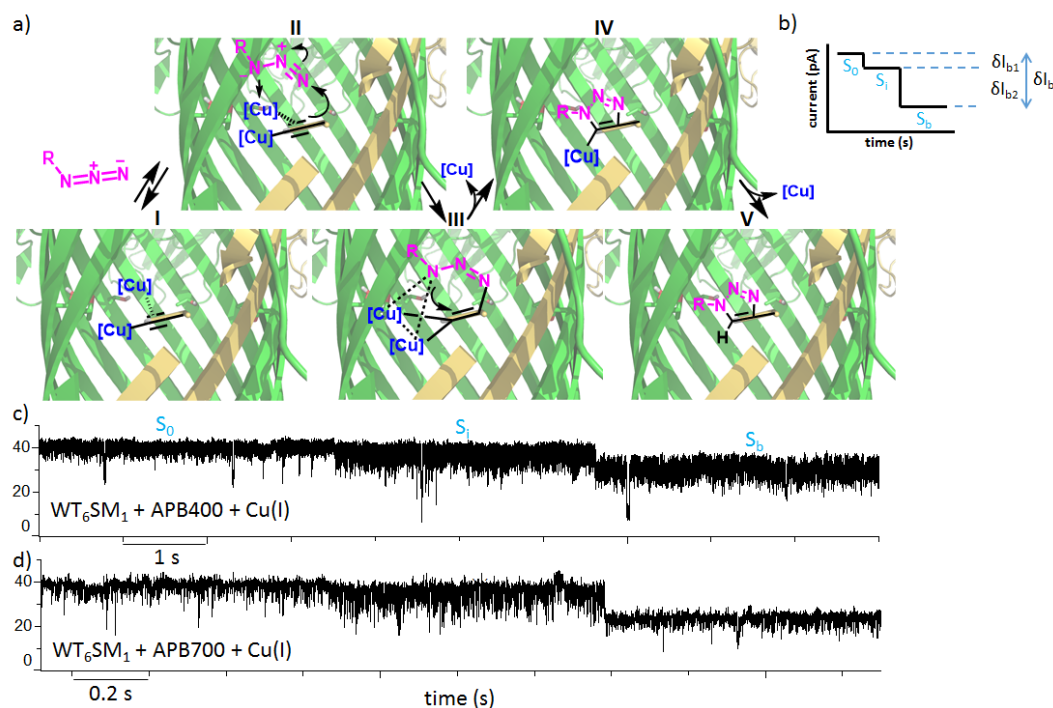


Figure 3.23. A click reaction in a WT₆SM₁ αHL nanoreactor. **a)** A proposed mechanism⁵⁰ for a CuAAC reaction in a single WT₆SM₁ αHL pore. Parts of the β strands have been removed to expose the reactive L-propargylglycine (PrP) residue. **I**, formation of a Cu(I)-acetylide. **II**, reversible binding of an azide substrate. **III**, a dinuclear copper intermediate. **IV**, formation of Cu triazolide intermediate. **V**, the triazole product corresponding to the S_b state. The triazolide **IV** may correspond to the S_i intermediate in the current trace, but this assignment is speculative (see the text). **b)** A schematic of the current traces corresponding to the CuAAC reaction. The unreacted state (S₀) is converted to an intermediate state (S_i) before the pore is permanently, partly blocked (S_b). The magnitudes of the current steps from S₀ to S_i and S_i to S_b are denoted by δI_{b1} and δI_{b2} , respectively ($\delta I_b = \delta I_{b1} + \delta I_{b2}$). **c)** Current recording with azide APB400 as the substrate (2 mM). **d)** Current recording with azide APB700 as the substrate (2 mM). In both cases, the concentration of Cu(I) was 20 mM. Note that the timescales in 'c' and 'd' differ. The buffer was 1 M KCl, 20 mM Bis-Tris propane, 25 μM TCEP, pH 8.5. The recording was made at +50 mV. The current was filtered at 5 kHz and sampled at 25 kHz. For display, further digital filtering was carried out at 1 kHz.

Table 3.1. The mean current decreases (δI_{b1} and δI_{b2}) and the root mean square noises (I_{rms}) during the formation of intermediate with the two different substrates.

| | APB400 (n = 11) | APB700 (n = 10) |
|--------------------|-------------------|-------------------|
| $\% \delta I_{b1}$ | $6.5 \pm 3.4 \%$ | $9.5 \pm 3.6 \%$ |
| $\% \delta I_{b2}$ | $18.9 \pm 3.2 \%$ | $31.1 \pm 3.7 \%$ |
| $\% I_{rms}(S_o)$ | $12.7 \pm 1.6 \%$ | $17.4 \pm 3.2 \%$ |
| $\% I_{rms}(S_i)$ | $16.3 \pm 2.9 \%$ | $22.1 \pm 2.8 \%$ |
| $\% I_{rms}(S_b)$ | $14.1 \pm 2.3 \%$ | $14.7 \pm 3.2 \%$ |

The recordings were carried out at 50 mV in 1 M KCl, 50 mM Bis-Tris propane (pH 8.5), 25 μ M TCEP (n, number of events).

The percent decreases in the open pore current for APB400 (Figure 3.24a) were $\% \delta I_{b1, APB400} = 6.5 \pm 3.4$ and $\% \delta I_{b2, APB400} = 18.9 \pm 3.2 \%$ (n = 11, n represents the number of experiments). For the larger APB700, the percent decreases (Figure 3.24b) were $\% \delta I_{b1, APB700} = 9.5 \pm 3.6$ and $\% \delta I_{b2, APB700} = 31.1 \pm 3.7 \%$ (n = 10). The first step (to form S_i) was irreversible, and the final step yielded a permanent partly blocked state, S_b . The S_i state is relatively noisy (Figure 3.22c, d), which suggests that the intermediate is flexible or undergoes a reversible transformation, such as a conformational change or protonation and deprotonation (for APB400: $I_{rms}(S_o) = 12.7 \pm 1.6 \%$, $I_{rms}(S_i) = 16.3 \pm 2.9 \%$, $I_{rms}(S_b) = 14.1 \pm 2.3 \%$ (n = 11); for APB700: $I_{rms}(S_o) = 17.4 \pm 3.2 \%$, $I_{rms}(S_i) = 22.1 \pm 2.8 \%$, $I_{rms}(S_b) = 14.7 \pm 3.2 \%$ (n = 10), filtering: 5 kHz, sampling: 25 kHz, Table 3.1).

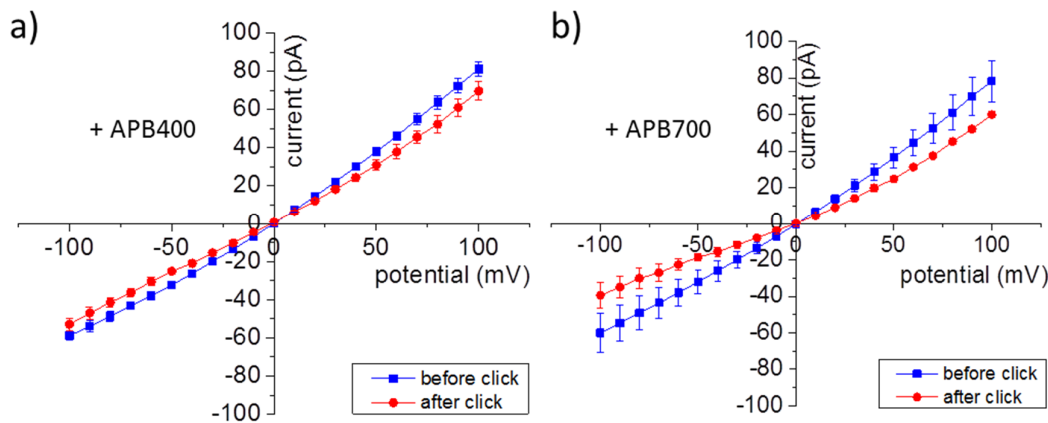


Figure 3.24. I-V curves after the click reaction. a) Reaction with APB400 (n = 3). b) reaction with APB700 (n = 4). The graph shows a current decrease at all applied potentials after the reaction.

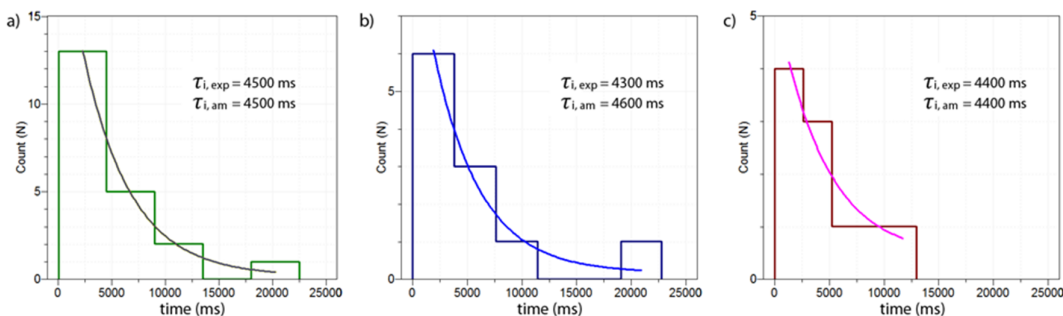


Figure 3.25. The mean lifetime of the reaction intermediate. a) Individual lifetimes (S_i) from all recordings (number of measurements (n) = 21, APB400 and APB700) were plotted and fitted to an exponential function to yield a mean lifetime of 4500 ± 600 ms (correlation coefficient, $r = 0.93$). b) Data from individual recordings with APB400 (n = 11, $r = 0.84$). c) Data from individual recordings with APB700 (n = 10, $r = 0.85$). The lifetimes obtained from the fitted exponential curves (exp) are similar to the values of the arithmetic means of the individual lifetimes (am) ⁵¹.

The mean lifetime of S_i (Figure. 3.25) was $\tau_i = 4.5 \pm 0.6$ s ($k_i = 0.22 \pm 0.03$ s⁻¹) (n = 21, for APB400 (n=11); for APB700 (n=10)) and the mean reaction constant was 0.24

$\text{s}^{-1}\cdot\text{M}^{-1}$ (Figure. 3.26). It is possible that the intermediate is one of the three sequential structures (II-IV, Figure. 3.23a) proposed by Worrell et al.⁵², and based on its lifetime most likely the Cu triazolidine (IV). However, the mechanism of the CuAAC reaction remains contentious, and our assignment is strictly hypothetical. More important is the fact that we observe a long-lived intermediate that must be incorporated into any viable reaction mechanism.

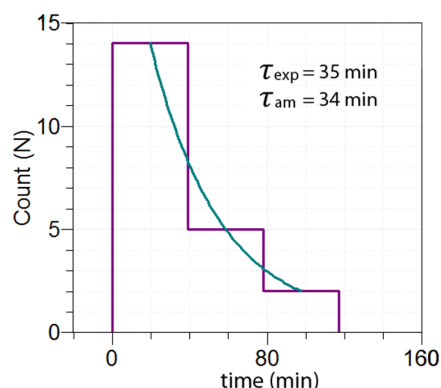


Figure 3.26. The mean reaction time. Individual reaction times from all recordings ($n = 21$) were plotted and fitted to an exponential function to yield a mean reaction time of 35 ± 1 min (correlation coefficient, $r = 0.99$). The reaction time is the time from the beginning of the recording, after mixing of the reagents, to the second blockade step signifying the formation of the triazole product ($S_0 + S_i$ in Figure. 3.23). As expected, the lifetime obtained from the fitted exponential curve (exp) is similar to the value of the arithmetic mean of the individual lifetimes (am)⁵¹. The rate constant was obtained from the mean reaction time and the final concentration of APB used in the reaction ($k = 1/\text{time} \cdot [\text{APB}]$).

Previous work has demonstrated the ability of single αHL pores to display the time-course of bond-making and bond-breaking reactions in individual molecules. Here, we have incorporated an unnatural amino acid bearing an alkyne group into the αHL

pore to expand the range of covalent chemistry that can be examined by the nanoreactor approach.

By NCL, we synthesized a full length α HL monomer including a central synthetic peptide incorporating L-propargylglycine. The alkyne side chain of the amino acid was predicted to project into the lumen of the transmembrane β barrel in the heptameric α HL pore. Our biochemical assays and single-channel current recordings demonstrated that the semisynthetic α HL monomer can be renatured efficiently and included in correctly assembled α HL heptamers. We used pores containing a single alkyne subunit to monitor CuAAC, a click reaction, at the single-molecule level and observed a long-lived (4.5 ± 0.6 s, $n = 21$) intermediate. We also obtained the overall rate constant ($k = 0.24 \pm 0.01$ M⁻¹·s⁻¹) of the click reaction within the pore, which is surprisingly low⁵³.

To compare the rate constant with the reaction in bulk solution, we carried out the CuAAC reaction (Figure. 3.28a) with a water-soluble synthetic tripeptide containing a terminal alkyne (ArgPrpArg, RXR, and Figure. 3.27a) and Fmoc-L- β -azidoalanine (Aza) (Figure. 3.27b). The reaction was monitored by measuring the absorbance of the product (Figure. 3.28b) at 260 nm on a HPLC. The rate constant in bulk solution was 2.1 ± 0.4 M⁻¹·s⁻¹ (Figure. 3.29), while the rate constants for similar reactions in bulk solution are in the range of 10 to 200 M⁻¹·s⁻¹⁵³.

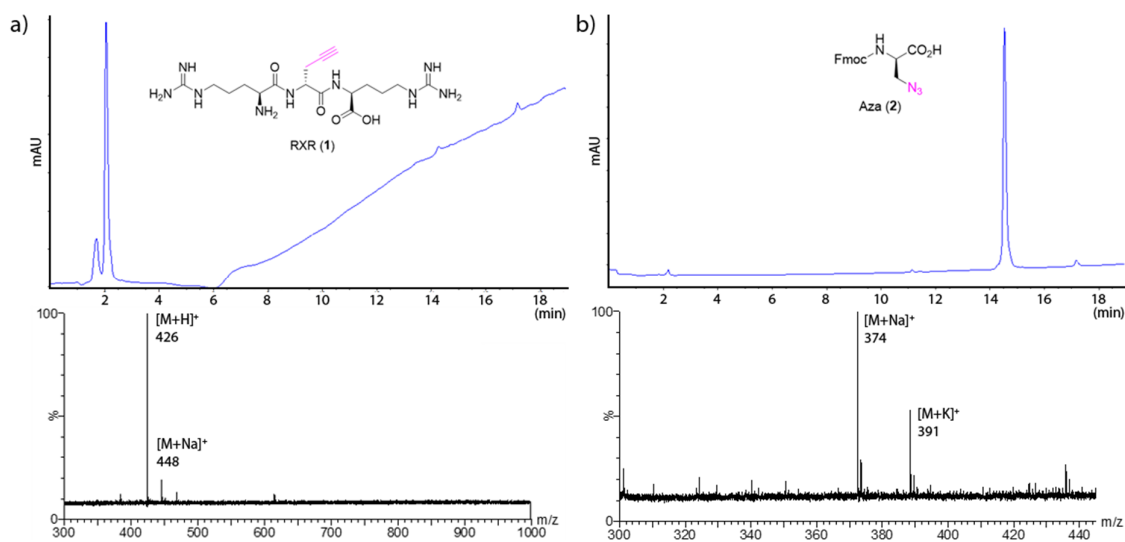


Figure 3.27. Characterization of the reagents used to determine the rate constant for the CuAAC reaction in solution. a) HPLC profile and MALDI mass spectrum of RXR ($[M+H]^+ = 426$ (obs), $[M+H]^+ = 426$ (calcd.)). The purity of RXR was analyzed by HPLC monitored at 230 nm, which yielded a major peak at 2.2 min. The baseline drift is due to the increase of acetonitrile in solvent gradient, which absorbs the light at 230 nm. b) HPLC profile and MALDI mass spectrum of Aza ($[M+Na]^+ = 374$ (obs), $[M+Na]^+ = 374$ (calcd.)). The purity of Aza was analyzed by HPLC monitored at 260 nm, which yielded a single peak at 14.4 min.

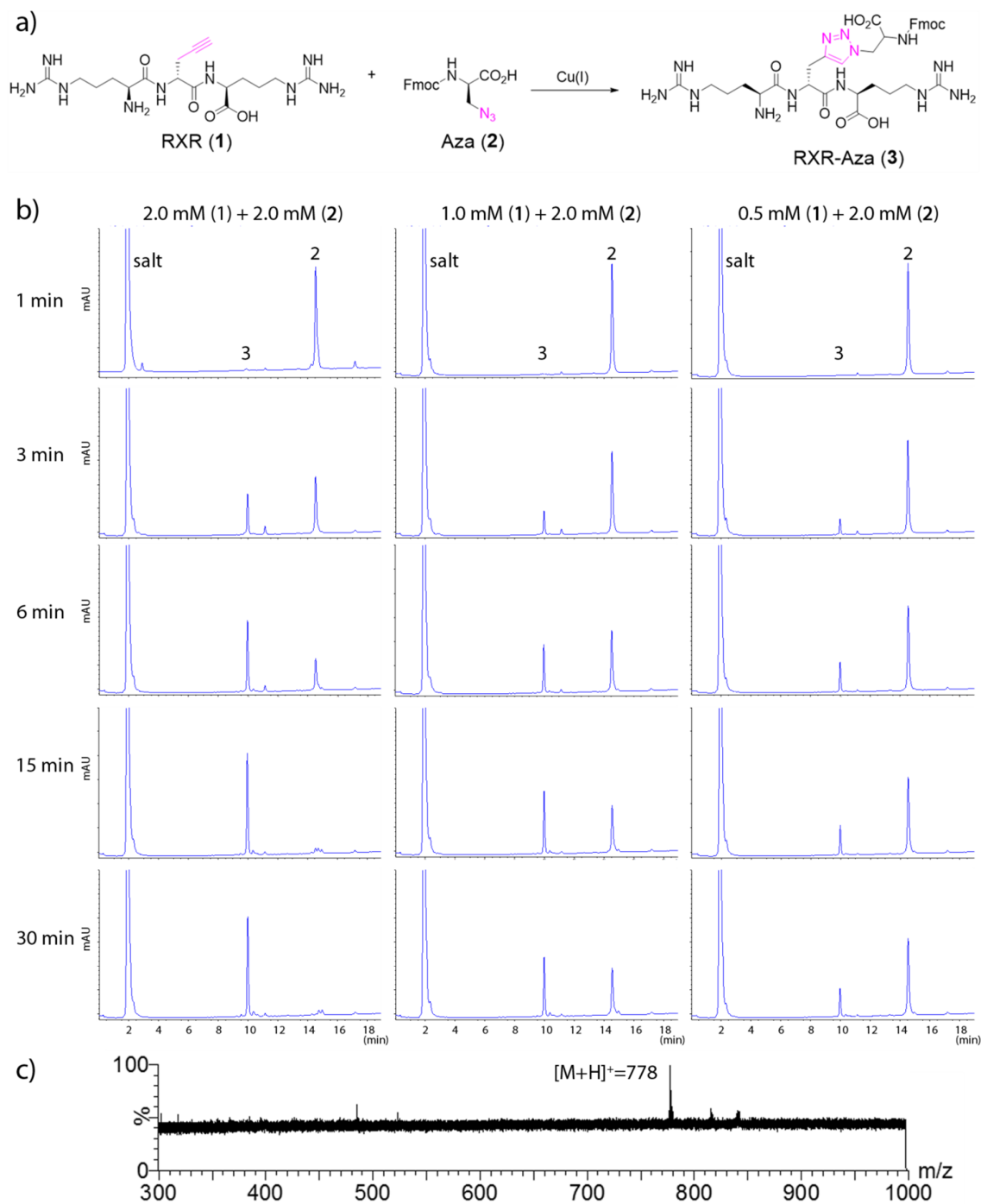


Figure 3.28. Monitoring the CuAAC reaction in solution. (a) The CuAAC reaction of a water-soluble peptide (RXR, **1**) and Fmoc-L- β -azidoalanine (Aza, **2**). (b) Analysis of the CuAAC kinetics by HPLC. The formation of the product (**3**) was monitored at 260 nm with three different concentrations of RXR (2, 1, and 0.5 mM). (c) MALDI mass spectrum of RXR-Aza (**3**) produced from the CuAAC reaction. ([M+H]⁺ = 778 (obs), [M+H]⁺ = 778 (calcd.))

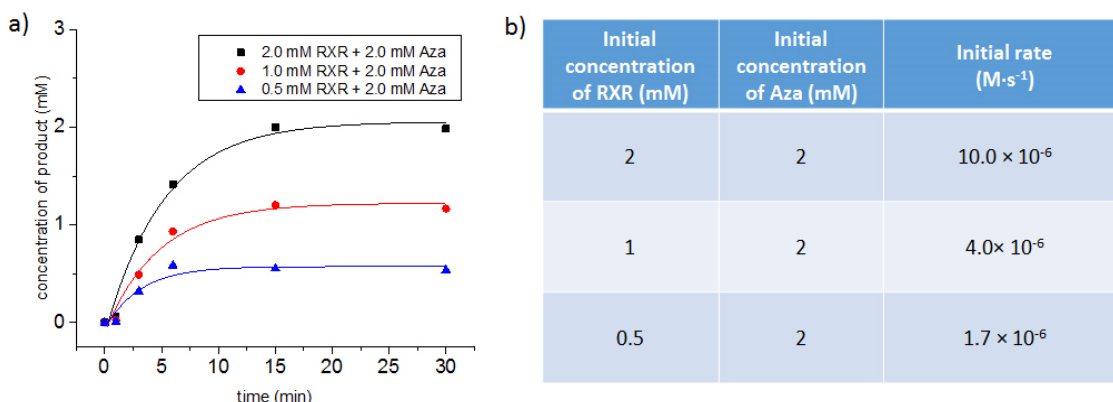


Figure 3.29. Analysis of the CuAAC kinetics. To determine the overall reaction constant, concentrations of the product were plotted and fitted to an exponential function. The initial rates were determined from the slopes at $t=0$. From the three separate experiments, the reaction constant of the CuAAC in bulk solution (1 M KCl, 20 mM Bis-Tris propane, 25 μM TCEP, pH 8.5, 20 mM Cu(I)) was determined to be $2.1 \pm 0.4 \text{ M}^{-1}\cdot\text{s}^{-1}$ ($k = 0.24 \pm 0.01 \text{ M}^{-1}\cdot\text{s}^{-1}$ for single-molecule). The correlation coefficients are $r = 0.86$, 0.93 , and 0.80 for the black, red, and blue curves, respectively.

In previous work, we have found that reactions of small molecules inside the αHL pore exhibit rates within a factor of ten of the rates in solution⁶. In the present case, the low rate constant must reflect the ability of the bulky reagents to partition into the pore⁵⁴ and their ability to become correctly oriented close to the internal wall. The ability of copper to form a complex with Bis-Tris propane⁵⁵ in the reaction buffer might also lead to the slower reaction rate.

3.3. Conclusion

We suggest that the nanoreactor system, which we have extended here, will be applicable to many reactions that occur on a millisecond time scale in aqueous conditions over a broad range of pH values^{44, 56, 57} and at temperatures (close to

100 °C)⁵⁸. The central peptide in our semisynthetic approach might contain single or multiple functional groups from a variety of possibilities. Reversible chemistry is of particular interest, because many events can be recorded in each experiment, allowing the ready determination of rate constants with statistical significance. For example, the incorporation of an aldehyde functionality would enable us to study a variety of reversible reactions such as the formation of hydrates, hemiacetals and acetals, and imines and enamines.

3.4. Materials and methods

3.4.1. Preparation of plasmids and polypeptides

The same methods described in section 2.4.11 to 2.4.15 were used to produce the plasmids and polypeptide fragments (NTF and CTF).

3.4.2. Peptide synthesis

i) SP-Nbz

2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and N^α-Fmoc protected amino acids were from Novabiochem. Hydroxybenzotriazole (HOBt), N-methylpyrrolidone (NMP), N,N-diisopropylamine (DIEA), N,N-dimethylformamide (DMF), dichloromethane (DCM, anhydrous), diethyl ether, acetonitrile (HPLC-grade), guanidine hydrochloride (Gu.HCl), trifluoroacetic acid (TFA), and triisopropylsilane (TIS) were from Sigma-Aldrich. The amino acids for Fmoc-SPPS were from Novabiochem, except for Boc-Thz-OH (Bachem) and Fmoc-L-propargylglycine-OH (Alfa Aesar).

Automated synthesis was performed on a 0.25 mmol scale by using previously reported protocols ³⁴ with a peptide synthesizer (Liberty, CEM). After Fmoc removal from Dawson Dbz resin (Novabiochem), standard chain elongation was performed. For each coupling reaction, 0.5 M (1.2 mmol) HBTU, 0.5 M (1.2 mmol) HOBt and 0.2 M Fmoc-Xaa (1.2 mmol) in DMF were added to the resin. Subsequently, 2.0 M DIEA in NMP was added, and the coupling reaction was carried out in a microwave chamber. The Fmoc protecting group was then removed by treatment with 20% (v/v) piperidine in DMF. Side-chain protection was as follows: Fmoc-Tyr(tBu)-OH, Fmoc-Thr(tBu)-OH and Fmoc-Asn(Trt)-OH. The N-terminal residue was incorporated as Boc-Thz-OH.

After the completion of the chain assembly, the resin-bound protected peptide was washed with anhydrous DCM, after which 50 mM p-nitrophenyl chloroformate ³⁴ in DCM (20 mL) was added. After 40 min with N₂ purging, the resin was washed with DCM and treated with two portions of 0.5 M DIEA in DMF (2 X 10 mL) for 15 min each. The resin was then washed with DMF, followed by DCM, and dried under an N₂ stream. Peptides were cleaved from the resin and side-chain and N-terminal protecting groups were removed by treatment with TFA containing 5% (v/v) DCM, 2.5% (v/v) TIS and 2.5% (v/v) water. After 2 h, the mixture was concentrated by evaporating the solvent. The SP[Thz¹¹⁴-Gly¹²⁶]-Nbz was precipitated with cold diethylether and isolated by centrifugation with several diethylether washes.

ii) RXR

On a Arg-preloaded (H-Arg(Pbf)-2-chlorotrityl) resin (Sigma), standard chain elongation was performed by using Fmoc-L-propargylglycine-OH and Fmoc-Arg(Pbf)-OH as described in the synthetic procedure for SP-Nbz. The resin was washed and dried as described and peptide was cleaved from the resin in 95% TFA solution for 2 h.

3.4.3. NCL

The two sequential NCL reactions were carried out under previously established conditions⁵⁹. CTF (0.5 mM), from which pyruvate had been removed, was mixed with SP-Nbz (5 mM) under Ar in 0.5 mL of degassed NCL buffer (200 mM NaH₂PO₄ buffer (pH 6.8) containing 6 M guanidine hydrochloride (Gu.HCl), 40 mM tris(2-carboxyethyl)phosphine (TCEP), and 20 mM 4-mercaptophenylacetic acid (MPAA)). After overnight reaction at room temperature, most of the unreacted peptide was removed by passing the mixture through a centrifugal filter (Amicon, MWCO 3k, 500 µL). Then, the mixture was desalted on a P6 column (Bio-Rad) and analyzed by LC-MS. The N-terminal Thz group was subsequently cleaved by treatment with 0.2 M

MeONH₂.HCl in 200 mM NaH₂PO₄ buffer (adjusted to pH 4.0) containing 6 M Gu.HCl, 40 mM TCEP for 4 h at room temperature. For the next round of ligation, the buffer was then replaced with NCL buffer containing 200 mM MPAA by repeated dilution and concentration with a centrifugal filter (Amicon, MWCO 3k). NTF- α thioester (0.1 mM) was mixed with the first ligation product (0.05 mM) and the reaction was allowed to proceed overnight. The final ligation product was purified by gel filtration followed by ion-exchange chromatography. The purified product was desalted with a P6 column and characterized by SDS-PAGE and LC-MS: ([M+H]⁺ = 34994 (obs), 34994 (calcd.)).

3.4.4. Ionic Current Recordings and Single-Molecule Chemistry

Ionic current measurements were performed with a planar bilayer apparatus at room temperature. Data were analysed with pClamp (version 10.1, Molecular Devices). OriginPro8.5 was used for plotting IV curves and statistical analyses. For single-molecule click chemistry, substrate APB400 or APB700 (100 mM in 1 M KCl, 20 mM Bis-Tris propane (pH 8.5), 25 μ M TCEP) was added in 2.5 μ L portions to the *trans* compartment to give final concentrations of 0.5 mM to 2.0 mM. CuBr₂ was dissolved in the same buffer and mixed with 2 equivalents of Na ascorbate to reduce Cu(II) to Cu(I) prior to addition to both the *cis* and *trans* compartments at 20 mM final concentration.

3.4.5. Analytical HPLC

The purified synthetic peptide (SP-Nbz) was further examined on an analytical reverse-phase column (Polaris C18 5 μ , 100 mm \times 4.6 mm) on an Agilent 1260 Infinity HPLC at a flow rate of 1 mL min⁻¹ with a gradient from 5 to 95% of eluent B in eluent A, over 13 min (eluant A: 0.1% TFA in water; eluant B: 0.1% TFA in acetonitrile). The separation was monitored at 220 nm.

The synthetic monomer, purified by ion-exchange chromatography, was further analyzed for purity with a Polaris C18 column by using a ternary mobile phase: eluant

A: 0.1% formic acid in water; eluant B: 0.1% formic acid in acetonitrile; eluant C: 10 mM triethylammonium bicarbonate, pH 8.5. A linear gradient of 2-35 % eluant B in eluant C was applied over 10 min, followed by a gradient of 35-98% eluant B in eluant A over 6 min, at a flow rate of 1 mL min⁻¹. The analysis was monitored at 280 nm.

3.4.6. Peptide purification

Preparative peptide purification was carried out with a Dionex UltiMate 3000 HPLC. Peptides were dissolved at 10 mg mL⁻¹ in DMSO prior to injection and separation was performed by using a Vydac C18 column (250 × 22 mm, 10-15 μm) at a flow rate of 15 mL min⁻¹ with a gradient from 30 to 75% of eluant B in eluant A, over 30 min (eluant A: 0.1% TFA in water; eluant B: 0.1% TFA in acetonitrile). The separation was monitored at 220 nm. The purified peptide was lyophilized and characterized by LC-MS ([M+H]⁺=1511 (obs), 1511 (calcd.)).

3.4.7. Analytical LC-MS

The N-terminal fragment (NTF), C-terminal fragment (CTF), and chemically synthesized peptides (SP-Nbz) were analyzed on a Waters LCT accurate-mass time-of-flight instrument (ESI-TOF MS) by using the positive ion mode. LC-MS employed a Chromolith RP-18e 50 mm × 2 mm column with a linear gradient of 5-100% eluant B in eluant A, over 8 min, with a flow rate of 1 mL min⁻¹ (eluant A: 0.1% formic acid in water; eluant B: 0.1% formic acid in acetonitrile). The mass spectra were generated with MassLynx software (version 4.1, Waters) by combining mass spectra obtained over the major peak in the total ion chromatogram.

3.4.8. Protein Purification

3.4.8.1. Gel filtration

The final ligation product was exchanged into denaturing purification buffer (8 M urea, 50 mM NaH₂PO₄ (pH 7.0), 250 mM NaCl, 5 mM TCEP) by successive concentration, dilution and re-concentration with a centrifugal filter (MWCO, 3k). The concentrated material was injected onto a gel filtration column (Superdex™ 200 10/300 GL), which was equilibrated with the same buffer and eluted at 0.7 mL min⁻¹.

3.4.8.2. Ion-exchange chromatography

The fraction containing the desired protein was collected and concentrated for further purification. The fraction was desalted with a PD-10 column (GE Healthcare) and then subjected to ion-exchange chromatography (HiTrap Q FF, GE Healthcare) by using a gradient of 0-1M KCl in 10 mM Tris.HCl, pH 8.0. Fractions containing the semisynthetic α HL monomer were identified by SDS-PAGE and LC-MS ([M+H]⁺= 34994 (obs), 34994(calcd.)). For storage of the protein as a monomer, the Tris buffer was replaced by 8 M urea, 50 mM Na phosphate (pH 7.0), 250 mM NaCl, 5 mM TCEP by using a centrifugal filter (MWCO, 3k) and the resulting solution stored at -80°C.

3.4.9. Renaturation of the synthetic monomer

The denaturing buffer used for the purification of SM was exchanged for the renaturation buffer (10 mM Tris.HCl, pH 8.0, containing 1 mM EDTA) with a centrifugal filter (Amicon, MWCO 3k) as follows. SM in 8 M urea (500 μ L) was centrifuged at 13,000 g for 20 min (4°C) to a volume of 100 μ L. Renaturation buffer (400 μ L) was added and mixed with the concentrated SM to dilute the urea 5-fold. The urea concentration was further reduced to ~60 mM by repeating the dilution and concentration three more times.

3.4.10. Heptameric α HL pores

3.4.10.1. Heteroheptameric pores (WT_{7-n}SM_n, n = 0-7)

WT α HL monomer was prepared by IVTT (*E. coli* S30 extract Promega, Madison, WI) in the presence of [³⁵S]methionine (1500 Ci mmol⁻¹, 10 mCi mL⁻¹; Amersham Pharmacia Biotech, Piscataway, NJ). To form heteroheptameric pores, the denaturing buffer used for the purification of SM was first exchanged by the renaturation buffer (see above). The concentration of SM was determined from the absorbance at 280 nm measured with a NanoDrop Spectrophotometer (ND-1000, Thermo Scientific) and the calculated extinction coefficient. The IVTT product (~5-10 ng μ L⁻¹) and SM in the renaturation buffer (60 mM urea) were mixed in various concentration ratios (7:0, 6:1, 5:2, 3.5:3.5, 2:5, and 1:6) in the presence of rabbit erythrocyte membranes (rRBCm) according to a previously described procedure ⁶⁰. After 1 h at 37°C, the mixture was centrifuged for 1 min at 21,000 g. The membrane pellet containing homo- and heteroheptamers was washed in MBSA buffer (10 mM 3-[N-morpholino]propane sulphonic acid (MOPS), 150 mM NaCl, pH 7.4, containing 1 mg mL⁻¹ bovine serum albumin) without heating, prior to electrophoresis in a 5% SDS-polyacrylamide gel. The labelled bands corresponding to heptamers with one or more copies of the SM subunit (WT_nSM_{7-n}, where n = 0 to 6) were resolved through the D₈ tails on the SM subunits ⁶¹. WT₆SM₁ was eluted from the gel as described previously ⁶².

3.4.10.2. Homoheptameric pore (SM₇)

SM (50 μ L, 0.60 mg mL⁻¹) in the renaturation buffer was mixed with DPhPC liposomes (10 μ L, 1 mg mL⁻¹)⁶³ and incubated at 37°C for 1 h. The reaction mixture was separated in a 7.5% SDS-polyacrylamide gel (Bio-Rad).

3.4.11. Hemolysis Assays

Rabbit red blood cells were washed ⁶⁴ repeatedly in MBSA and made up to 1% in MBSA. The hemolytic activities of α HL proteins were assayed in a 96-well plate. The

α HL-H₆ monomer (1.08 mg mL⁻¹) had been expressed in *E. coli* by use of the pT7-WT-H₆ plasmid and purified on a Ni-NTA column by the method described for CTF, except that 8 M urea was used instead of 6 M Gu.HCl (**Fig. S11c**). SM (4.02 mg mL⁻¹) was in the renaturation buffer (10 mM Tris-HCl, pH 8.0, containing 1 mM EDTA). Both proteins were serially two-fold diluted with MBSA such that each well contained a volume of 50 μ L. For example, WT α HL or SM (1 μ L) was mixed with MBSA (99 μ L) in the first well. Half of the first well (50 μ L) was transferred to the next well containing MBSA (50 μ L) and mixed. In this way, serial dilution was continued to the 12th well. Rabbit red blood cells (50 μ L, 1% in MBSA) were then added to each well, starting with the well that contained the most dilute protein. Lysis was recorded by monitoring the decrease in light scattered at 595 nm with a microplate reader⁶⁵ (Bio-Rad Laboratories Ltd, UK). The specific hemolytic activity (HC₅₀) was the concentration of protein showing 50% lysis in 100 min at room temperature³⁷.

3.4.12. Limited proteolysis

The synthetic monomer and SM₇ (see above) in the renaturation buffer (20 μ L) were digested separately by the addition of proteinase K in water (2 μ L, 0.1 mg mL⁻¹). After 10 min at room temperature, the proteinase K was inactivated with 40 mM phenylmethanesulfonyl fluoride (PMSF, 1 μ L in ethanol) followed after 30 min by the addition of gel loading buffer (2X). The proteinase K-treated SM₇ was heated for 15 min at 95°C⁶⁶.

3.4.13. Single channel recording and data analysis

Recordings were performed at 21.0 \pm 2.0°C. A bilayer of 1,2-diphytanoyl-sn-glycero-3-phosphatidylcholine (DPhPC, Avanti Polar Lipids) was formed across an aperture of 100 μ m diameter in a Teflon film (Goodfellow) separating the *cis* and *trans*

compartments of the recording apparatus (0.5 mL each). The electrolyte in both the *cis* and *trans* compartments was 20 mM Bis-Tris propane (pH 8.5), 1 M KCl, 25 μ M TCEP. TCEP was added to reduce disulfide bonds that might form between the pores. (Cys residues at the ligation sites were not blocked). Gel-purified α HL heptamers were added to the grounded *cis* compartment. After the insertion of a single pore into a bilayer, the *cis* compartment was manually perfused to prevent further insertions. Currents were measured by using Ag/AgCl electrodes connected to a patch-clamp amplifier (Axopatch 200B, Axon Instruments). Signals were low-pass-filtered at 5 kHz and sampled at 25 kHz with a Digidata 1440A digitizer (Axon Instruments).

3.4.14. β CD kinetics

Aliquots of β CD were added to the *trans* compartment to produce concentrations of 25, 50, 75 and 100 μ M. For the determination of each set of kinetic constants with WT₆SM₁ pores, three separate experiments were performed and data acquired for 5 min were analysed. Events were collected by Single-Channel searches with the Clampfit software (version 10.1). Very short events (<0.1 ms) were excluded. For kinetic analysis, dwell times were collected for each pore, plotted as unbinned cumulative histograms and fitted to single exponentials.

3.4.15. Molecular dynamics simulations

LEaP was used to build topology and coordinate files for the L-propargylglycine amino acid. L-Propargylglycine (Prp) was substituted for Thr-117 and the full-length α HL monomer taken from the crystal structure of the heptameric α HL pore (PDB ID: 7AHL) containing Prp was subjected to energy minimization. The geometric optimization of positions of hydrogen atoms in the monomeric α HL was implemented by a restrained energy minimization method. The energy minimization processes were carried out with AMBER (version 12) as previously reported⁶⁷. The solvent effect in

water was calculated with the Generalized Born method (GBSA)⁶⁸ implemented in AMBER. The convergence procedure was terminated after 1000 iteration cycles to obtain a 0.05 kcal/mol Å gradient threshold.

3.4.16. MALDI analysis

MALDI-TOF mass spectra of the azide substrates (APB400 (Jena Bioscience, Germany) and APB700 (AGTC Bioproduct, UK)) were obtained⁶⁹ with a Waters MALDI Micro MX by using the reflectron mode. Desorption and ionization were produced by irradiation with pulsed UV (337.1 nm) from a nitrogen laser. The instrument was operated at 12 kV in the positive ion mode with an extraction delay time of 500 ns. The samples were solutions (1 mg mL⁻¹) in water/MeCN (6/4) containing 0.1% TFA. The matrix solution was dithranol (Sigma, 10 mg mL⁻¹) in THF with 0.1% TFA. The sample and matrix were mixed in a 1:1 ratio before drying.

3.4.17. CuAAC kinetics in solution

The rate constant of the CuAAC reaction was determined in 1 M KCl, 20 mM Bis-Tris propane, 25 µM TCEP, pH 8.5, the same buffer that was used for the CuAAC reaction in the semisynthetic pore. A water-soluble peptide (ArgPrpArg, RXR) was synthesized and purified by the methods described in the previous sections. Fmoc-*L*-β-azidoalanine (Aza) was obtained from Novabiochem.

The reactions were set up in the first column of a 96-well plate (0.5 mL well volume). 15 µL of 30% H₂O₂ and 15 µL of ethanol were dispensed into all other wells for quenching the reaction at specified time points. The click reactions were started by the addition of Cu(I) solution to the mixture of RXR and Aza (Fig. S22). Aliquots of 10 µL were taken from the reaction wells with a multichannel pipette at 1, 3, 6, 15, and 30 min and transferred to columns of quenching solution. After the aliquots had been quenched, 3 µL was taken from each of wells and injected onto an analytical C18

reverse-phase HPLC column (Polaris C18 A 5u, 100 mm × 4.6 mm) at a flow rate of 1 mL min⁻¹ with a gradient from 5 to 95 % of eluant B in eluant A, over 13 min (eluant A: 0.1 % formic acid in water; eluant B: 0.1 % formic acid in acetonitrile). The concentrations of the analytes were obtained by integrating the peak areas measured at 260 nm.

Reference

1. Bayley, H. & Cremer, P.S. Stochastic sensors inspired by biology. *Nature* **413**, 226-230 (2001).
2. Boersma, A.J. & Bayley, H. Continuous stochastic detection of amino acid enantiomers with a protein nanopore. *Angew Chem Int Ed* **51**, 9606-9609 (2012).
3. Harrington, L., Cheley, S., Alexander, L.T., Knapp, S. & Bayley, H. Stochastic detection of Pim protein kinases reveals electrostatically enhanced association of a peptide substrate. *Proc Natl Acad Sci U S A* **110**, E4417-4426 (2013).
4. Branton, D. et al. The potential and challenges of nanopore sequencing. *Nat Biotechnol* **26**, 1146-1153 (2008).
5. Bayley, H. Nanopore sequencing: from imagination to reality. *Clin Chem* **61**, 25-31 (2015).
6. Bayley, H., Luchian, T., Shin, S.-H. & Steffensen, M. in *Single Molecules and Nanotechnology*, Vol. 12. (eds. R. Rigler & H. Vogel) 251-277 (Springer Berlin Heidelberg, 2008).
7. Steffensen, M.B., Rotem, D. & Bayley, H. Single-molecule analysis of chirality in a multicomponent reaction network. *Nat Chem* **6**, 603-607 (2014).
8. Pulcu, G.S., Mikhailova, E., Choi, L.-S. & Bayley, H. Continuous observation of the stochastic motion of an individual small-molecule walker. *Nature Nanotechnology* **10**, 76-83 (2015).
9. Shin, S.H., Luchian, T., Cheley, S., Braha, O. & Bayley, H. Kinetics of a reversible covalent-bond-forming reaction observed at the single-molecule level. *Angew Chem Int Ed* **41**, 3707-3709 (2002).
10. Luchian, T., Shin, S.H. & Bayley, H. Single-molecule covalent chemistry with spatially separated reactants. *Angew Chem Int Ed* **42**, 3766-3771 (2003).
11. Lu, S., Li, W.W., Rotem, D., Mikhailova, E. & Bayley, H. A primary hydrogen-deuterium isotope effect observed at the single-molecule level. *Nat Chem* **2**, 921-928 (2010).
12. Choi, L.S. & Bayley, H. S-nitrosothiol chemistry at the single-molecule level. *Angew Chem Int Ed* **51**, 7972-7976 (2012).
13. Boersma, A.J., Brain, K.L. & Bayley, H. Real-time stochastic detection of multiple neurotransmitters with a protein nanopore. *ACS Nano* **6**, 5304-5308 (2012).

14. Hammerstein, A.F., Shin, S.H. & Bayley, H. Single-molecule kinetics of two-step divalent cation chelation. *Angew Chem Int Ed* **49**, 5085-5090 (2010).
15. Banerjee, A. et al. Molecular bases of cyclodextrin adapter interactions with engineered protein nanopores. *Proc Natl Acad Sci U S A* **107**, 8165-8170 (2010).
16. Chin, J.W. Expanding and reprogramming the genetic code of cells and animals. *Annu Rev Biochem* **83**, 379-408 (2014).
17. Wang, K. et al. Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat Chem* **6**, 393-403 (2014).
18. Sachdeva, A., Wang, K., Elliott, T. & Chin, J.W. Concerted, rapid, quantitative, and site-specific dual labeling of proteins. *J Am Chem Soc* **136**, 7785-7788 (2014).
19. Dawson, P.E., Muir, T.W., Clarklewis, I. & Kent, S.B.H. Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779 (1994).
20. Kent, S.B.H. Total chemical synthesis of proteins. *Chemical Society Reviews* **38**, 338-351 (2009).
21. Kent, S.B.H. Bringing the science of proteins into the realm of organic chemistry: total chemical synthesis of SEP (synthetic erythropoiesis protein). *Angew Chem Int Ed* **52**, 11988-11996 (2013).
22. Hackeng, T.M., Griffin, J.H. & Dawson, P.E. Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology. *Proc Natl Acad Sci U S A* **96**, 10068-10073 (1999).
23. Bang, D., Makhatadze, G.I., Tereshko, V., Kossiakoff, A.A. & Kent, S.B. Total chemical synthesis and X-ray crystal structure of a protein diastereomer: [D-Gln 35]ubiquitin. *Angew Chem Int Ed* **44**, 3852-3856 (2005).
24. Muir, T.W. Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* **72**, 249-289 (2003).
25. Muralidharan, V. & Muir, T.W. Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat Methods* **3**, 429-438 (2006).
26. Shah, N.H. & Muir, T.W. Inteins: nature's gift to protein chemists. *Chem Sci* **5**, 446-461 (2014).
27. Olschewski, D. & Becker, C.F. Chemical synthesis and semisynthesis of membrane proteins. *Mol Biosyst* **4**, 733-740 (2008).

28. Komarov, A.G., Linn, K.M., Devereaux, J.J. & Valiyaveetil, F.I. Modular strategy for the semisynthesis of a K⁺ channel: investigating interactions of the pore helix. *ACS Chem Biol* **4**, 1029-1038 (2009).
29. Grosse, W., Essen, L.O. & Koert, U. Strategies and perspectives in ion-channel engineering. *Chembiochem* **12**, 830-839 (2011).
30. Devaraneni, P.K. et al. Semisynthetic K⁺ channels show that the constricted conformation of the selectivity filter is not the C-type inactivated state. *Proc Natl Acad Sci U S A* **110**, 15698-15703 (2013).
31. Bayley, H., Cheley, S., Harrington, L. & Syeda, R. Wrestling with native chemical ligation. *ACS Chem Biol* **4**, 983-985 (2009).
32. Kent, S.B. Total chemical synthesis of proteins. *Chem Soc Rev* **38**, 338-351 (2009).
33. Gentle, I.E., De Souza, D.P. & Baca, M. Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjug Chem* **15**, 658-663 (2004).
34. Blanco-Canosa, J.B. & Dawson, P.E. An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew Chem Int Ed* **47**, 6851-6855 (2008).
35. Song, L. et al. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859-1866 (1996).
36. Henricus, M. & Bayley, H., Vol. Thesis (Master of Science) (University of Oxford, 2010).
37. Walker, B., Krishnasastri, M., Zorn, L., Kasianowicz, J. & Bayley, H. Functional expression of the alpha-hemolysin of *Staphylococcus aureus* in intact *Escherichia coli* and in cell lysates. Deletion of five C-terminal amino acids selectively impairs hemolytic activity. *J Biol Chem* **267**, 10902-10909 (1992).
38. Mantri, S., Sapra, K.T., Cheley, S., Sharp, T.H. & Bayley, H. An engineered dimeric protein pore that spans adjacent lipid bilayers. *Nat Commun* **4**, 1725 (2013).
39. Huang, K.S., Bayley, H., Liao, M.J., London, E. & Khorana, H.G. Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments. *J Biol Chem* **256**, 3802-3809 (1981).

40. Walker, B., Braha, O., Cheley, S. & Bayley, H. An intermediate in the assembly of a pore-forming protein trapped with a genetically-engineered switch. *Chem Biol* **2**, 99-105 (1995).
41. Cheley, S. et al. Spontaneous oligomerization of a staphylococcal alpha-hemolysin conformationally constrained by removal of residues that form the transmembrane beta-barrel. *Protein Eng* **10**, 1433-1443 (1997).
42. Olson, R., Nariya, H., Yokota, K., Kamio, Y. & Gouaux, E. Crystal structure of staphylococcal LukF delineates conformational changes accompanying formation of a transmembrane channel. *Nat Struct Biol* **6**, 134-140 (1999).
43. Gu, L.Q., Braha, O., Conlan, S., Cheley, S. & Bayley, H. Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. *Nature* **398**, 686-690 (1999).
44. Gu, L.Q. & Bayley, H. Interaction of the noncovalent molecular adapter, beta-cyclodextrin, with the staphylococcal alpha-hemolysin pore. *Biophys J* **79**, 1967-1975 (2000).
45. Kolb, H.C., Finn, M.G. & Sharpless, K.B. Click chemistry: diverse chemical function from a few good reactions. *Angew Chem Int Ed* **40**, 2004-2021 (2001).
46. Iha, R.K. et al. Applications of orthogonal "click" chemistries in the synthesis of functional soft materials. *Chem Rev* **109**, 5620-5686 (2009).
47. Debets, M.F. et al. Bioconjugation with strained alkenes and alkynes. *Accounts Chem Res* **44**, 805-815 (2011).
48. Kolb, H.C. & Sharpless, K.B. The growing impact of click chemistry on drug discovery. *Drug Discov Today* **8**, 1128-1137 (2003).
49. Movileanu, L., Cheley, S. & Bayley, H. Partitioning of individual flexible polymers into a nanoscopic protein pore. *Biophys J* **85**, 897-910 (2003).
50. Himo, F. et al. Copper(I)-catalyzed synthesis of azoles. DFT study predicts unprecedented reactivity and intermediates. *J Am Chem Soc* **127**, 210-216 (2005).
51. Moczydlowski, E. in *Ion Channel Reconstitution*. (ed. C. Miller) 75-113 (Springer US, 1986).
52. Worrell, B.T., Malik, J.A. & Fokin, V.V. Direct evidence of a dinuclear copper intermediate in Cu(I)-catalyzed azide-alkyne cycloadditions. *Science* **340**, 457-460 (2013).
53. Lang, K. & Chin, J.W. Bioorthogonal reactions for labeling proteins. *ACS Chem Biol* **9**, 16-20 (2014).

54. Movileanu, L. & Bayley, H. Partitioning of a polymer into a nanoscopic protein pore obeys a simple scaling law. *Proc Natl Acad Sci U S A* **98**, 10137-10141 (2001).
55. Sadeghi, S.M., Ferreira, C.M.H., Vandenberghe, S. & Soares, H.M.V.M. Graphic data analysis and complex formation curves as modeling and optimization tools for characterization of Cu–(buffer)_x–(OH)_y systems involving BTP or BES in aqueous solution. *Journal of Coordination Chemistry* **68**, 777-793 (2014).
56. Maglia, G. et al. DNA strands from denatured duplexes are translocated through engineered protein nanopores at alkaline pH. *Nano Letters* **9**, 3831-3836 (2009).
57. Franceschini, L., Mikhailova, E., Bayley, H. & Maglia, G. Nucleobase recognition at alkaline pH and apparent pK(a) of single DNA bases immobilised within a biological nanopore. *Chem Commun* **48**, 1520-1522 (2012).
58. Kang, X.F., Gu, L.Q., Cheley, S. & Bayley, H. Single protein pores containing molecular adapters at high temperatures. *Angew Chem Int Ed* **44**, 1495-1499 (2005).
59. Bang, D., Chopra, N. & Kent, S.B. Total chemical synthesis of crambin. *J Am Chem Soc* **126**, 1377-1383 (2004).
60. Cheley, S., Braha, G., Lu, X.F., Conlan, S. & Bayley, H. A functional protein pore with a "retro" transmembrane domain. *Protein Sci* **8**, 1257-1267 (1999).
61. Howorka, S., Cheley, S. & Bayley, H. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nat Biotechnol* **19**, 636-639 (2001).
62. Miles, G., Bayley, H. & Cheley, S. Properties of *Bacillus cereus* hemolysin II: A heptameric transmembrane pore. *Protein Science* **11**, 1813-1824 (2002).
63. Mantri, S., Sapra, K.T., Cheley, S., Sharp, T.H. & Bayley, H. An engineered dimeric protein pore that spans adjacent lipid bilayers. *Nat Commun* **4** (2013).
64. Hammerstein, A.F., Jayasinghe, L. & Bayley, H. Subunit dimers of alpha-hemolysin expand the engineering toolbox for protein nanopores. *J Biol Chem* **286**, 14324-14334 (2011).
65. Cheley, S. et al. Spontaneous oligomerization of a staphylococcal alpha-hemolysin conformationally constrained by removal of residues that form the transmembrane beta-barrel. *Protein Eng* **10**, 1433-1443 (1997).

66. Walker, B., Krishnasastri, M., Zorn, L. & Bayley, H. Assembly of the oligomeric membrane pore formed by Staphylococcal alpha-hemolysin examined by truncation mutagenesis. *J Biol Chem* **267**, 21782-21786 (1992).
67. Cornell, W.D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J Am Chem Soc* **118**, 2309-2309 (1996).
68. Spassov, V.Z., Yan, L. & Szalma, S. Introducing an implicit membrane in generalized Born/solvent accessibility continuum solvent models. *J Phys Chem B* **106**, 8726-8738 (2002).
69. Robertson, J.W.F. et al. Single-molecule mass spectrometry in solution using a solitary nanopore. *Proc Natl Acad Sci U S A* **104**, 8207-8211 (2007).

Appendix

DNA and amino acid sequence

atg: start codon

T7

Ccatgg - *NdeI*

ACTAGT - *SpeI*

NTF_{FL}-thioester, **Intein**, **CBD**

TAATACGACTCACTATAGGG T7F

CCGCTGAGCAATAACTAGC T7R

- pTXB3-NTF_{FL}

TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGT
TAACTTTAAGAAGGAGATATAAC**catgg**CAGATTCTGATATTAATATTAAAACCGGTACTACAG
ATATTGGAAGCAATACTACAGTAAAAACAGGTGATTTAGTCACTTATGATAAAGAAAATGGCAT
GCACAAAAAGTATTTTATAGTTTATCGATGATAAAAATCACAATAAAAAACTGCTAGTTATT
AGAACAAAAGGTACCATTGCTGGTCAATATAGAGTTTATAGCGAAGAAGGTGCTAACAAAAGTG
GTTTAGCCTGGCCTTCAGCCTTTAAGGTACAGTTGCAACTACCTGATAATGAAGTAGCTCAAAT
ATCTGATTACTATCCAAGAAATTCGATTGATACAAAAGAGTATATGAGTACTTTAACTTATGGA
TTCAACGGTAATGTTACTGGT**TGCATCACGGGAGATGCACTAGTTGCCCTACCCGAGGGCGAGT**
CGGTACGCATCGCCGACATCGTGCCGGGTGCGCGGCCCAACAGTGACAACGCCATCGACCTGAA
AGTCCTTGACCGGCATGGCAATCCCGTGCTCGCCGACCGGCTGTTCCACTCCGGCGAGCATCCG
GTGTACACGGTGCGTACGGTCGAAGGTCTGCGTGTGACGGGCACCGCGAACCACCCGTTGTTGT
GTTTGGTTCGACGTCGCCGGGGTGCCGACCCTGCTGTGGAAGCTGATCGACGAAATCAAGCCGGG
CGATTACGCGGTGATTCAACGCAGCGCATTACGCGTCGACTGTGCAGGTTTGGCCGCGGgAAA
CCCGAATTTGCGCCCAACCTACACAGTCGGCGTCCCTGGACTGGTGCGTTTCTTGGAAGCAC
ACCACCGAGACCCGGACGCCCAAGCTATCGCCGACGAGCTGACCGACGGGCGGTTCTACTACGC
GAAAGTCGCCAGTGTACCGACGCCGGCGTGCAGCCGGTGTATAGCCTTCGTGTCGACACGGCA
GACCACGCGTTTATCACGAACGGGTTCGTCAGCCACGCTACTGGCCTCACCGGTCTGAACCTAG
GCCTCACGACAAATCCTGGTGTATCCGCTTGGCAGGTCAACACAGCTTATACTGCGGGACAATT
GGTCACATATAACGGCAAGACGTATAAATGTTTGCAGCCCCACACCTCCTTGGCAGGATGGGAA
CCATCCAACGTTCTGCCTTGTGGCAGCTTCAATGACTGCAGGAAGGGGATCCGGCTGCTAACA
AAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCA**CCGCTGAGCAATAACTAGC**ATAACCCCTTGG
GGCCTCTAAACGGGTCTTGAGGGGTTTTTGTCTGAAAGGAGGAACCTATATCCGGAT

Protein

Sequence (377 residues):

MADSDINIKTGTTDIGSNNTVKTGDLVTYDKENGMHKKVFYSFIDDKNHNKLLLVIRTKG
TIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPNRSIDTKEYMSTLTYG
FNGNVTG**CITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHS**

GEHPVYTVRTVEGLRVGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVD
 CAGFARGKPEFAPTTYTVGVPGLVRFLEAHRDPDAQIADELTDGRFYYAKVASVTDAG
 VQPVYSLRVDTADHAFITNGFVSHA TGLTGLNSGL TTNPGVSAWQVNTAYTAGQLVTYNG
 KTYKCLQPHTSLAGWEPSNVPALWQLQ

NTF_{FL}-MESNa :

ADSDINIKTGTTDIGSNTTVKTGDLVITYDKENGMHKKVFYSFIDDKNHNKLLVIRTKGTIAGQ
 YRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFGNVTGQ
 (where, Q is a thioester formed by MESNa)

- pTXB3-NTF_{Δ6-3M}

TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGT
 TAACTTTAAGAAGGAGATATAACcatggCAGATTCTGATATTAATATTAAAACCGGTACTACAG
 ATATTGGAAGCAATACTACAGTAAAAACAGGTGATTTAGTCACTTATGATAAAGAAAAATGGCAT
 GCACAAAAAAGTATTTTATAGTTTTATCGATGATAAAAAATCACAATAAAAAAATGCTAGTTATT
 AGAACAAAAGGTACCATTGCTGGTCAATATAGAGTTTATAGCGAAGAAGGTGCTAACAAAAGTG
 GTTTAGCCTGGCCTTCAGCCTTTAAGGTACAGTTGCAACTACCTGATAATGAAGTAGCTCAAAT
 ATCTGATTACTATCCAAGAAATTTCGATTGATACAAAAGAGTATATGTGCATCACGGGAGATGCA
 CTAGTTGCCCTACCCGAGGGCGAGTCGGTACGCATCGCCGACATCGTGCCGGGTGCGCGGCCCA
 ACAGTGACAACGCCATCGACCTGAAAAGTCCTTGACCGGCATGGCAATCCCGTGCTCGCCGACCG
 GCTGTTCCACTCCGGCGAGCATCCGGTGTACACGGTGCGTACGGTCAAGGTCTGCGTGTGACG
 GGCACCGCGAACCACCCGTTGTTGTGTTTGGTCGACGTCGCCGGGGTGCCGACCCTGCTGTGGA
 AGCTGATCGACGAAATCAAGCCGGGCGATTACGCGGTGATTCAACGCAGCGCATTCAGCGTCGA
 CTGTGCAGGTTTTGCCCGCGGgAAACCCGAATTTGCGCCCAACCTACACAGTCGGCGTCCCT
 GGACTGGTGCGTTTTCTTGGAAGCACACCACCGAGACCCGGACGCCCAAGCTATCGCCGACGAGC
 TGACCGACGGGCGGTCTACTACGCGAAAGTCGCCAGTGTCACCGACGCCGGCGTGCAGCCGGT
 GTATAGCCTTCGTGTGACACGGCAGACCACGCGTTTATCACGAACGGGTTCGTGAGCCACGCT
 ACTGGCCTCACCGGTCTGAACTCAGGCCTCACGACAAATCCTGGTGTATCCGCTTGGCAGGTCA
 ACACAGCTTATACTGCGGGACAATTGGTCACATATAACGGCAAGACGTATAAATGTTTGCAGCC
 CCACACCTCCTTGGCAGGATGGGAACCATCCAACGTTCTGCCTTGTGGCAGCTTCAATGA

NTF_{Δ6-3M}-MESNa :

ADSDINIKTGTDDIGSNTTVKTGDLVTYDKENGMHKKVFYSFIDDKNHNKLLVIRTKGTIAGQ
 YRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPNSIDTKEYMO

- pTXB3-NTF_{Δ6-3F}

TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGT
 TAACTTTAAGAAGGAGATATAACcatggCAGATTCTGATATTAATATTAAAACCGGTACTACAG
 ATATTGGAAGCAATACTACAGTAAAAACAGGTGATTTAGTCACTTATGATAAAGAAAAATGGCAT
 GCACAAAAAAGTATTTTATAGTTTTATCGATGATAAAAAATCACAATAAAAAAATGCTAGTTATT
 AGAACAAAAGGTACCATTGCTGGTCAATATAGAGTTTATAGCGAAGAAGGTGCTAACAAAAGTG
 GTTTAGCCTGGCCTTCAGCCTTTAAGGTACAGTTGCAACTACCTGATAATGAAGTAGCTCAAAT
 ATCTGATTACTATCCAAGAAATTTCGATTGATACAAAAGAGTATTTCGTCATCACGGGAGATGCA
 CTAGTTGCCCTACCCGAGGGCGAGTCGGTACGCATCGCCGACATCGTGCCGGGTGCGCGGCCCA
 ACAGTGACAACGCCATCGACCTGAAAAGTCCTTGACCGGCATGGCAATCCCGTGCTCGCCGACCG
 GCTGTTCCACTCCGGCGAGCATCCGGTGTACACGGTTCGTACGGTTCGAAGGTCTGCGTGTGACG
 GGCACCGCGAACCACCCGTTGTTGTGTTTGGTTCGACGTCGCCGGGGTGCCGACCCTGCTGTGGA
 AGCTGATCGACGAAATCAAGCCGGGCGATTACGCGGTGATTCAACGCAGCGCATTCAGCGTCGA
 CTGTGCAGGTTTTGCCCGCGGgAAACCCGAATTTGCGCCCAACCTACACAGTCGGCGTCCCT
 GGACTGGTGCCTTTCTTGGAAGCACACCACCGAGACCCGGACGCCCAAGCTATCGCCGACGAGC
 TGACCGACGGGCGGTCTACTACGCGAAAGTCGCCAGTGTACCGACGCCGGCGTGCAGCCGGT
 GTATAGCCTTCGTGTGACACGGCAGACCACGCGTTTATCACGAACGGGTTCGTGAGCCACGCT
 ACTGGCCTCACCGGTCTGAACTCAGGCCTCACGACAAATCCTGGTGTATCCGCTTGCGCAGGTCA
 ACACAGCTTATACTGCGGGACAATTGGTCACATATAACGGCAAGACGTATAAATGTTTGCAGCC
 CCACACCTCCTTGCGAGGATGGGAACCATCCAACGTTCTGCCTTGTGGCAGCTTCAATGA

NTF_{Δ6-3F}-MESNa :

ADSDINIKTGTDDIGSNTTVKTGDLVTYDKENGMHKKVFYSFIDDKNHNKLLVIRTKGTIAGQ
 YRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPNSIDTKEYFO

- pT7-CTF-CDH

TAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTA
AGAAGGAGATATACAT**atg**TGCGATACCGGCAAAATTGGAGGCCTTATTGGTGCAAATGTTTCG
ATTGGTCATACACTTAAGTATGTTCAACCTGATTTCAAACAATTCTCGAGAGCCCAACTGATA
AAAAAGTAGGCTGGAAAGTGATATTTAACAATATGGTGAATCAAAATTGGGGACCATACGATCG
AGATTCTTGGAACCCGGTATATGGCAATCAACTTTTCATGAAGACTAGAAATGGTTCTATGAAA
GCAGCAGATAACTTCCTTGATCCTAACAAAGCAAGTTCCTATTATCTTCAGGGTTTTCACCAG
ACTTCGCTACAGTTATTACTATGGATAGAAAAGCATCCAAACAACAAACAAATATAGATGTAAT
ATACGAACGAGTTCGTGATGATTACCAATTGCATTGGACTTCAACAAATTGGAAAAGTACCAAT
ACTAAAGATAAATGGACAGATCGTTCTTCAGAAAGATATAAAATCGATTGGGAAAAAGAAGAA
TGACAAATGATGACGATGATGACGACGATGATCACCACCATCACCATCATTGATAA

CTF

CDTGKIGGLIGANVSIGHTLK YVQPDFK TILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPV
YGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRD
DYQLHWTSTNWKGTNTKDKWTD RSSERYKIDWEKEEMTNDDDDDDDDHHHHHH

- pT7-CTF_{Δ6}

TAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTA
AGAAGGAGATATACAT**atg**TGTACTTTAACTTATGGA **Δ6** GGTGATGAT
ACCGGCAAA **Δ6** ATTAATGTTTCGATTGGTCATACACTTAAGTATGTTT
AACCTGATTTCAAACAATTCTCGAGAGCCCAACTGATAAAAAAGTAGGCTGGAAAGTGATATT
TAACAATATGGTGAATCAAATTTGGGGACCATACGATCGAGATTCTTGGAACCCGGTATATGGC
AATCAACTTTTCATGAAGACTAGAAATGGTTCCTATGAAAGCAGCAGATAACTTCCTTGATCCTA
ACAAAGCAAGTTCCCTATTATCTTCAGGGTTTTTCACCAGACTTCGCTACAGTTATTACTATGGA
TAGAAAAGCATCCAAACAACAAACAAATATAGATGTAATATACGAACGAGTTCGTGATGATTAC
CAATTGCATTGGACTTCAACAAATTTGGAAAGGTACCAATACTAAAGATAAAATGGACAGATCGTT
CTTCAGAAAGATATAAAATCGATTGGGAAAAAGAAGAAATGACAAATGATGACGATGATGACGA
CGATGATCACCACCATCACCATCATTGATAA

CTF_{Δ6}

CTLTYG **Δ6** GDDTGKI **Δ6** NVSIGHTLKYVQPDFKTILESPDCKKVGWKVIFNNMVN
QNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVITMDRKAS
KQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWTDRSSERYKIDWEKEEMTNDDDDDDDH
HHHHH

Amino acid sequence of the semisynthetic monomers

– SM_{FL}:

ADSDINIKTGTTDIGSNTTVKTGDLVITYDKENGMHKKVFYSFIDDKNHNKKLLVIRTKGTIAGQ
YRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLT**YGFNGNVT****G**CD
TGKIGGLIGANVSIGHTLKYVQPDFK**T**ILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYG
NQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVITMDRKASKQQT**NIDVIYERVRDDY**
QLHWTSTNWKGTNTKDKWTD**RSSERYKIDWEKEEMT**NDDDDDDDDHHHHHH

– TSM_{Δ6-3M}

ADSDINIKTGTTDIGSNTTVKTGDLVITYDKENGMHKKVFYSFIDDKNHNKKLLVIRTKGTIAGQ
YRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEY**M**CTLT**YGGDDTGKINV**
SIGHTLKYVQPDFK**T**ILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSM
KAADNFLDPNKASSLLSSGFSPDFATVITMDRKASKQQT**NIDVIYERVRDDYQLHWTSTNWKGT**
NTKDKWTD**RSSERYKIDWEKEEMT**NDDDDDDDDHHHHHH

– TSM_{Δ6-3F}

ADSDINIKTGTTDIGSNTTVKTGDLVITYDKENGMHKKVFYSFIDDKNHNKKLLVIRTKGTIAGQ
YRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEY**F**CTLT**YGGDDTGKINV**
SIGHTLKYVQPDFK**T**ILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSM
KAADNFLDPNKASSLLSSGFSPDFATVITMDRKASKQQT**NIDVIYERVRDDYQLHWTSTNWKGT**
NTKDKWTD**RSSERYKIDWEKEEMT**NDDDDDDDDHHHHHH

The terminal sequences of NTFs are underlined and represented as bold.

Modeller script

```
# Homology modeling by the automodel class
from modeller import *
from modeller.automodel import * # Load the automodel class

log.verbose()
env = environ()

# directories for input atom files
env.io.atom_files_directory = './'

# Building multi-chain models with symmetry
class mymodel(automodel):
    def special_restraints(self, aln):
        rsr = self.restraints
        at = self.atoms
        sA = selection(self.chains['A']).only_atom_types('CA')
        sB = selection(self.chains['B']).only_atom_types('CA')
        sC = selection(self.chains['C']).only_atom_types('CA')
        sD = selection(self.chains['D']).only_atom_types('CA')
        sE = selection(self.chains['E']).only_atom_types('CA')
        sF = selection(self.chains['F']).only_atom_types('CA')
        sG = selection(self.chains['G']).only_atom_types('CA')
        self.restraints.symmetry.append(symmetry(sA, sB, 1.0))
        self.restraints.symmetry.append(symmetry(sB, sC, 1.0))
        self.restraints.symmetry.append(symmetry(sC, sD, 1.0))
        self.restraints.symmetry.append(symmetry(sD, sE, 1.0))
        self.restraints.symmetry.append(symmetry(sE, sF, 1.0))
        self.restraints.symmetry.append(symmetry(sF, sG, 1.0))
        self.restraints.symmetry.append(symmetry(sG, sA, 1.0))
    def user_after_single_model(self):
        self.restraints.symmetry.report(1.0)

a = mymodel(env,
             alnfile = 'TBMd6.ali', # alignment filename
             knowns   = '7AHL',     # codes of the templates
             sequence  = 'TBMd6')   # code of the target
a.starting_model= 1                # index of the first model
a.ending_model   = 5                # index of the last model
                                   # (determines how many models to calculate)
a.md_level = refine.slow           # refinement level

a.make()                           # do homology modelling
```

The file name in green and the number of models in pink can alter according to the structure than can be modelled. In this case the structure having the lowest energy are selected among five predicted models.

Alignment file (e.g. TBMd6.ali)

```
>P1;7AHL
structureX:7AHL:1:A:293:G::::
ADSDINIKTGTDDIGSNTTVKTGDLVITYDKENGMHKKVFYSFIDDKNHNKKLLVIRTK
GTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPNRSIDTKEYMSTL
TYGFNGNVTGDDTGKIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNM
VNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVI
TMDRKASKQQTNIIDVIYERVRDDYQLHWTSTNWKGTNTKDKWTDSSERYKIDWEKEE
MTN/
ADSDINIKTGTDDIGSNTTVKTGDLVITYDKENGMHKKVFYSFIDDKNHNKKLLVIRTK
GTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPNRSIDTKEYMSTL
TYGFNGNVTGDDTGKIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNM
VNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVI
TMDRKASKQQTNIIDVIYERVRDDYQLHWTSTNWKGTNTKDKWTDSSERYKIDWEKEE
MTN/
ADSDINIKTGTDDIGSNTTVKTGDLVITYDKENGMHKKVFYSFIDDKNHNKKLLVIRTK
GTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPNRSIDTKEYMSTL
TYGFNGNVTGDDTGKIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNM
VNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVI
TMDRKASKQQTNIIDVIYERVRDDYQLHWTSTNWKGTNTKDKWTDSSERYKIDWEKEE
MTN/
ADSDINIKTGTDDIGSNTTVKTGDLVITYDKENGMHKKVFYSFIDDKNHNKKLLVIRTK
GTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPNRSIDTKEYMSTL
TYGFNGNVTGDDTGKIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNM
VNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVI
TMDRKASKQQTNIIDVIYERVRDDYQLHWTSTNWKGTNTKDKWTDSSERYKIDWEKEE
MTN/
ADSDINIKTGTDDIGSNTTVKTGDLVITYDKENGMHKKVFYSFIDDKNHNKKLLVIRTK
GTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPNRSIDTKEYMSTL
TYGFNGNVTGDDTGKIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNM
VNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVI
TMDRKASKQQTNIIDVIYERVRDDYQLHWTSTNWKGTNTKDKWTDSSERYKIDWEKEE
MTN/
ADSDINIKTGTDDIGSNTTVKTGDLVITYDKENGMHKKVFYSFIDDKNHNKKLLVIRTK
GTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPNRSIDTKEYMSTL
TYGFNGNVTGDDTGKIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNM
VNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVI
TMDRKASKQQTNIIDVIYERVRDDYQLHWTSTNWKGTNTKDKWTDSSERYKIDWEKEE
MTN/
ADSDINIKTGTDDIGSNTTVKTGDLVITYDKENGMHKKVFYSFIDDKNHNKKLLVIRTK
GTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPNRSIDTKEYMSTL
TYGFNGNVTGDDTGKIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNM
VNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVI
TMDRKASKQQTNIIDVIYERVRDDYQLHWTSTNWKGTNTKDKWTDSSERYKIDWEKEE
MTN/
*
```