# Engineered α-hemolysin pores with chemically and genetically-fused functional proteins

*A dissertation by*

SHIKSHA MANTRI


Submitted to the University of Oxford for the degree of

Doctor of Philosophy

**Major Subject: Chemical Biology**

St. Cross College, Oxford

# DECLARATION

The work described in this thesis was carried out in Prof. Hagan Bayley's laboratory, Chemistry Research Laboratory at the University of Oxford from Oct'09 to Sept'12. Transmission electron microscopy (TEM) (Chapter 2, section 2.16) was done by Dr. Thomas Sharp in the Division of Structural Biology (STRUBI), Nuffield Department of Clinical Medicine (NDM) at the University of Oxford. The work on αHL- λ-exo genetic fusion (Chapter 3, sections 3.8-3.18) and the trimer fusion construct of λ-exo (Chapter 4) was done in collaboration with Dr. K Tanuj Sapra.

# ACKNOWLEDGEMENTS

Just saying “thanks” would belittle the indebt I feel for the following people for their contributions in my life in Oxford and beyond. Put simply, I owe this to:

- Prof. Hagan Bayley for being a very supportive supervisor throughout.

- Dr. K Tanuj Sapra, my mentor, critic, teacher, and closest friend in Oxford. Apart from (heroically) correcting and reshaping much of this thesis, he has taught me how to think science. I hope that we collaborate on scientific ventures in the future as well and that I am able to beat him at chess one day. I owe him many ‘parties’ and fond memories.

- Dr. Stephen Cheley (Steve) for spotting the dimer on SDS-PAGE and insisting that I characterize it (read Chapter 2 for the exciting results there on). I owe an iPad and another award to his hunch. He taught me all the tricks and trade of molecular biology. If one day I become famous for my work on genetic engineering, I would still owe it to him and Leon.

- Dr. Leon Harrington for setting me up in the lab and being an extremely patient teacher.

- Pinky Raychaudhuri for helping me get settled outside the lab.

- Dr. Anne Hammerstein for being a great housemate and, Dr. Manuela Zanetti and Dr. Ofer Wilner for reading sections of this thesis.

- Dr. Manuela Zanetti, Dr. Lingbing Kong for their creative inputs in my projects from time to time. Dr. Mariam Ayub for the car lifts, Dr. Lajos Holfer for his help in the MD simulations and Dr. Gokce Su Pulcu for her help with the FPLC. Joongoo Lee, Dr. Shuo Huang and all the aforementioned people for being such good company.

- Oxford University Press for the Clarendon Scholarship and St. Cross College for the amazing lunches.

- Prof. Colin Kleanthous and Prof. Stefan Howorka, my viva examiners.

- And finally the most important people of my life, my best friends- Saurabh Shintre, Gargi Golwalkar, Pooja Rastogi, Mudrika Khandelwal (Murgi), and Dr. Abhilasha Goyal, my parents- Mr. Johns Mantri and Mrs. Sunita Mantri and my sister, Ms. Khushboo Mantri for always being there, just outside my little bubble.

- And God… that thesis writing is over!

# LIST OF ABBREVIATIONS

| | |
|---|---|
| α1 | αHL monomer |
| $(α1)_2$ | αHL monomer dimer |
| α7 | αHL heptamer |
| $(α7)_2$ | αHL heptamer dimer |
| αHL | α-hemolysin |
| BSA | bovine serum albumin |
| CD | cyclodextrin |
| γCD | γ cyclodextrin |
| DOC | sodium deoxycholate |
| DNA | deoxyribonucleic acid |
| dNTP | deoxynucleoside triphosphate |
| DPhPC | 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine |
| dsDNA | double-stranded DNA |
| EDTA | ethylenediaminetetraacetic acid |
| egg-PC | L-α-phosphatidylcholine (95%) |
| λ-exo | λ-exonuclease |
| exo3 | fusion trimer of λ-exo |
| f-PCR | forward PCR |
| GABA | gamma-aminobutyric acid |
| h | hour(s) |
| HCl | hydrochloric acid |
| $I_o$ | open pore current value |
| $I_{res}$ | residual current value |
| IMAC | immobilized metal affinity chromatography |
| IVTT | *in vitro* transcription and translation |
| KCl | potassium chloride |
| $M_{app}$ | apparent molecular weight |
| MBP | maltose binding protein |
| MBSA | buffer containing 10 mM MOPS, 150 mM NaCl, 0.1% (w/v) BSA, pH 7.4 |

| | |
|---|---|
| min | minute(s) |
| MOPS | 3-[*N*-morpholino]propane sulphonic acid |
| MspA | *Mycobacterium smegmatis* porin A |
| $M_w$ | molecular weight |
| nACh | nicotinic acetylcholine |
| NaCl | sodium chloride |
| NN | E111N/K147N |
| nt | nucleotide |
| PAGE | polyacrylamide gel electrophoresis |
| PCR | polymerase chain reaction |
| $P_i^-$ | phosphate ions at pH 8 |
| PMSF | phenylmethylsulfonyl fluoride |
| POPC | 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine |
| POPG | 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'–*rac*-glycerol) |
| PTFE | polytetrafluoroethylene |
| RNA | ribonucleic acid |
| r-PCR | reverse PCR |
| RR | rectification ratio ($I_{+v}$/ $I_{-v}$) |
| rRBC | rabbit red blood cells |
| rRBCm | rabbit red blood cell membrane |
| s | seconds |
| S.D. | standard deviation |
| SDS | sodium dodecyl sulfate |
| SEC | size exclusion chromatography |
| SG linker | linker encoding for serines and glycines |
| ssDNA | single stranded DNA |
| TEM | transmission electron microscopy |
| Tris | 2-amino-2-hydroxymethyl-propane-1,3-diol |
| WT | wild type |

*For my country, India*

# TABLE OF CONTENTS

DECLARATION.....i
ACKNOWLEDGEMENTS.....ii
LIST OF ABBREVIATIONS.....iii
ABSTRACT.....1

## CHAPTER 1

### SYNTHESIZING PROTEIN COMPLEXES-AN INTRODUCTION

Introduction.....3
1.1. Natural protein complexes.....3
1.2. Engineering new protein complexes.....7
1.3. Minimalist design of protein surfaces.....8
1.4. Use of bulky interface residues for homo-multimer formation.....11
1.5. Covalently linking proteins with disulfide bonds.....12
1.6. Driving protein complex formation using electrostatic interactions.....14
1.7. Non-covalent protein-protein association using ligands.....16
1.8. Using the biotin-streptavidin linkage.....18
1.9. Metal mediated protein complexation.....20
1.10. Metal mediated protein assembly using non-natural ligands.....22
1.11. Genetic fusion of proteins.....24
1.11.1. Genetically fused proteins for imaging.....25
1.11.2. Genetically fused proteins for building new materials.....27
1.11.3. Genetically fused proteins for sensing.....29
1.11.4. Genetic fusion to make new classes of proteins.....30
1.12. Self-assembling peptides.....31
Outlook.....35
References.....37

## CHAPTER 2

### ENGINEERING A DIMERIC α-HEMOLYSIN PORE THAT SPANS PROXIMAL BILAYERS

Introduction.....50
2.1. Constructing minimal cells/tissues.....50
2.2. Engineering a protein that can insert in two adjacent bilayers.....53
2.2.1. Structure of αHL.....54
2.2.2. Single αHL molecules can be electrically characterised: applications of αHL.....55
2.2.3. αHL forms an ideal scaffold for protein engineering.....56

2.2.4. Designing a double bilayer inserting pore using αHL ..... 58

Results ..... 58
2.3. Design rationale of the $(\alpha7)_2$ pore ..... 58
2.4. Molecular dynamics (MD) simulations of cap-to-cap $(\alpha7)_2$ showed seven disulfide bridges without steric clashes in the interface ..... 60
2.5. $(\alpha7)_2$ forms spontaneously during K237C purification ..... 61
2.6. Structural characterization of $(\alpha7)_2$ ..... 62
2.7. $(\alpha7)_2$ is a fully folded structure ..... 63
2.8. $(\alpha7)_2$ is robust to mutagenesis ..... 64
2.9. Possible $(\alpha7)_2$ structures ..... 65
2.10. Electrical characteristics of $(\alpha7)_2$ suggested a longer and symmetric pore structure ..... 67
2.11. γCD blocked $(\alpha7)_2$ from both the *cis* and *trans* sides confirming the presence of two β-barrels ..... 72
2.12. Kinetics of γCD binding ..... 75
2.13. Phosphate binding to an $(\alpha7)_2$ mutant confirms the precise cap-to-cap structure ..... 79
2.14. Effect of freeze-thawing on $I_o$ and $P_i^-$ blocking characteristics ..... 81
2.15. Simultaneous insertion of $(\alpha7)_2$ in two bilayers ..... 83
2.15.1. Single channel current measurements of $(\alpha7)_2$ inserted in a planar lipid bilayer and a liposome ..... 83
2.15.2. Cis γCD blocking events stopped upon insertion in liposome ..... 84
2.15.3. $(\alpha7)_2$ is conducting even after insertion in a liposome: possible explanations ..... 86
2.16. Electron microscopy of $(\alpha7)_2$ ..... 87
2.17. Imipramine transfer across two proximal bilayers connected by $(\alpha7)_2$ 91
2.17.1. Variation in imipramine blocking prior to liposome insertion over different experiments ..... 95
2.17.2. Detecting γCD entrapped in liposomes from single channel electrical measurements of $(\alpha7)_2$ inserted in two bilayers ..... 96
2.17.3. Alternative experiments to demonstrate small molecule transfer across two bilayers using $(\alpha7)_2$ ..... 97
2.18. Theoretical model of $(\alpha7)_2$ explained its electrical properties ..... 98
2.18.1. Electrical model for γCD blocking ..... 100
2.18.2. Electrical model explaining current block due to liposome insertion ..... 103
2.18.3. Electrical model at other potentials ..... 108

Conclusions and outlook ..... 110

Methods ..... 115
2.19. Computational modelling of $(\alpha7)_2$ ..... 115
2.20. Expression and purification of $(\alpha7)_2$ and barrel mutant M113R ..... 115
2.21. Proteinase K assay ..... 118
2.22. Molecular modeling and graphics ..... 118
2.23. Single channel recording ..... 119
2.24. $P_i^-$ binding experiments ..... 121
2.25. Preparation of liposomes ..... 121
2.26. Electron Microscopy and Image processing ..... 122

References ..... 123

# CHAPTER 3

## COUPLING λ-EXONUCLEASE TO THE α-HEMOLYSIN NANOPORE


Introduction ..... 130
3.1. DNA processing enzymes ..... 130
3.2. Recombination in bacteriophage λ: role of λ-exonuclease *in vivo* ..... 130
3.3. λ-exonuclease (λ-exo) ..... 132
3.3.1. Structure of λ-exo ..... 132
3.3.2. λ-exo is a highly processive enzyme ..... 133
3.3.3. Choice of co-factors and DNA substrates ..... 133
3.3.4. Bulk activity assays ..... 134
3.3.5. Single molecule activity studies ..... 136
3.3.6. Mechanism of DNA cleavage ..... 140
3.4. Coupling λ-exo to the αHL nanopore ..... 148
3.4.1. Nanopores for studying single molecule enzymatic activity ..... 148
3.4.2. Nanopores for DNA sequencing ..... 150
3.4.3. Applications of the λ-exo-αHL pore complex ..... 152

Results …. ..... 154
3.5. Covalent attachment of λ-exo and αHL using disulfide bonds ..... 154
3.5.1. Cysteine mutants of αHL ..... 154
3.5.2. Cysteine mutants of λ-exo ..... 155
3.6. Strategy to latch λ-exo on αHL by non-covalent interactions ..... 160
3.7. Aligning λ-exo and αHL using electrostatic interactions ..... 163
3.8. Genetic fusion of λ-exo and αHL ..... 167
3.8.1. Genetic fusion constructs of λ-exo and αHL monomers ..... 167
3.8.2. Fusion construct of αHL dimer and λ-exo (EAA10) ..... 169
3.8.3. Expression, purification and characterization of the fusion constructs ..... 171
3.8.4. Hemolytic activity assays of the αHL-λ-exo fusion constructs ..... 174
3.8.5. Exonuclease activity assays of the αHL-λ-exo fusion constructs ..... 175
3.8.6. Full length AE10 protein shows hemolytic and exonuclease activities ..... 179
3.8.7. AE10 gets truncated into its constituent αHL and λ-exo parts ..... 181
3.8.8. Spontaneously formed oligomers of αHL-λ-exo fusion constructs 183
3.8.9. Single channel electrical recordings ..... 185
3.8.10. Summary of the αHL- λ-exo fusion constructs' activities ..... 198
3.8.11. Determining the oligomeric state of AE10 using SEC ..... 199
3.8.12. Oligomerization of AE constructs ..... 203
3.8.13. Purification strategies for AE10 and AE13 hetero-heptamers ..... 215
3.9. Construction of a fusion trimer of λ-exo with concatenated monomers ..... 219

Conclusions and outlook ..... 219

Methods ..... 223

3.10. Mutagenesis of αHL and λ-exo ..... 223
3.10.1. Point mutagenesis of αHL and λ-exo ..... 223
3.10.2. Construction of αHL-λ-exo fusion constructs. ..... 225
3.11. Expression and purification of αHL–λ-exo fusion constructs ..... 228
3.11.1. *In vitro* expression ..... 228
3.11.2. *In vivo* expression ..... 229
3.12. Estimation of $M_{app}$ ..... 232
3.13. Determination of number of active cysteines using PEG-OPSS ..... 232
3.14. Activity assays ..... 233
3.14.1. Hemolytic activity assay ..... 233
3.14.2. Exonuclease activity assay ..... 233
3.14.3. Exonuclease activity measurements in presence of SDS ..... 235
3.15. Electrical measurements in planar lipid bilayers ..... 235
3.16. Western blot ..... 236
3.17. Size exclusion chromatography (SEC) ..... 237
3.17.1. Calibration of the size exclusion column ..... 237
3.17.2. SEC of the αHL-λ-exo fusion constructs ..... 238
3.18. Oligomerization of αHL-λ-exo fusion constructs. ..... 238
3.18.1. Oligomerization using rabbit red blood cell membranes ..... 238
3.18.2. Oligomerization using detergent ..... 239

References ..... 240

# CHAPTER 4

## DESIGNING A CONCATENATED FUSION CONSTRUCT OF λ-EXONUCLEASE

Introduction ..... 252
4.1. Constructing a concatemer of λ-exo (exo3) to make the λ-exo-αHL pore complex ..... 252
4.2. Fusion constructs of other multimeric proteins ..... 253
4.3. Exo3 variants for mechanistic studies of λ-exo ..... 254
4.4. Synthesis of the exo3 construct ..... 254
4.5. Expression of the exo3 protein ..... 258
4.6. Exonuclease activity of exo3 ..... 259

Future work and outlook ..... 260

Methods ..... 263
4.7. Construction of the concatenated λ-exo construct ..... 263
4.8. Protein expression of exo3 ..... 266
4.9. Exonuclease activity assay ..... 266

References ..... 267

## CHAPTER 5

CONCLUSIONS AND OUTLOOK……..……..269

APPENDIX……..……..……...276

# ABSTRACT

Protein engineering could be used to bring two proteins together, which don't normally interact, in an oriented configuration. Using computer modelling and experimental work involving mutagenesis, a new dimer complex, $(\alpha7)_2$, was engineered with two α-hemolysin (αHL) heptamers (α7) units linked *via* disulfide bridges in a cap-to-cap orientation. The structure of $(\alpha7)_2$ was confirmed by biochemical analysis, transmission electron microscopy (TEM) and single-channel electrical recording. Importantly, it was shown that the one of two transmembrane β barrels of $(\alpha7)_2$ can insert into an attoliter liposome, while the other spans a planar lipid bilayer. $(\alpha7)_2$ pores spanning two bilayers were also observed by TEM. In potential, $(\alpha7)_2$ could be used for small molecule transfer between micron-sized vesicles (minimal cells) and would have applications in forming proto-tissues from minimal cells.

Another target has been to couple a highly processive exonuclease, λ-exonuclease (λ-exo), which functions as a trimer, with the α7 pore for DNA sequencing and single molecule studies of λ-exo. Several genetic fusion constructs of λ-exo and αHL were screened and optimized for activity. By linking the N-terminus of λ-exo monomer to the C-terminus of the αHL monomer (α1), a new kind of processive exonuclease (AE) was synthesized that can form pores in bilayers. AE and wild-type α1 could be integrated into hetero-heptamers with different number of AE subunits. To achieve a hetero-heptamer with only one λ-exo trimer molecule mounted on the αHL cap, a concatemer of

λ-exo (exo3) was made by genetically linking the monomers of λ-exo with 15 and 17 amino acid linkers. The immediate next step is to link exo3 to α1 and then to co-assemble the exo3-α1 fusion construct with α1 to make the λ-exo-αHL pore complex. Using similar strategies as described in this thesis, other proteins could be linked to αHL increasing the scope of the nanopore technology.

# CHAPTER 1

## SYNTHESIZING PROTEIN COMPLEXES - AN INTRODUCTION

## Introduction

### 1.1. Natural protein complexes

Nature has taken the course of evolution to fabricate multicellular organisms starting from simple chemicals in a bottom-up fashion. Proteins are composed of twenty amino acids, whose unique combinations in a polypeptide chain determine the overall fold and function of the protein (Voet and Voet, 2011). The chemical diversity of protein surfaces allow proteins to bind to a variety of small molecules such as ligands (De Meyts and Whittaker, 2002; Gao and Skolnick, 2012) and even macromolecules like RNA (Glisovic et al., 2008; Klinge et al., 2011), DNA (Pabo and Sauer, 1984; Zhang et al., 2011) and other proteins (Jones and Thornton, 1996). Most proteins form transient or permanent associations with other proteins to perform their physiological functions (Jones and Thornton, 1996; Nooren and Thornton, 2003b). Bacterial pore toxins such as the Staphylococcal α-hemolysin and some DNA processing proteins such as λ-exonuclease of bacteriophage λ are composed of identical subunits (homo-oligomers) (Kovall and Matthews, 1997; Song *et al.*, 1996) (Fig 1.1a, b). Hetero-protein complexes are formed of non-identical subunits, which often have different functions to perform. For example, DNA polymerase III holoenzyme is composed of two multi-subunit DNA polymerases associated with nine other

protein subunits (Glover and McHenry, 2001; Kelman and O'Donnell, 1995). The α subunit has polymerase activity, the ε subunit shows 3'→5' exonuclease activity, whereas the θ subunit stimulates the ε subunit's proofreading activity. Other subunits are involved in other tasks such as binding single strand binding proteins and ATP.

Large protein complexes composed of smaller protein subunits have evolved to function as nanomachines within the cell and are vital for transcription (Kadonaga, 1998), protein quality control (Finley, 2009), inter and intra cellular signaling, cell metabolism (Nooren and Thornton, 2003a) and other tasks (Alberts, 1998). The ribosome is a prime example of a cellular nanomachine, which is employed for translation of proteins from mRNA in all the three domains of life (bacteria, archae and eukaryotes). Ribosomes are composed of two asymmetric subunits formed of a macromolecular complex of proteins and RNA. The eukaryotic ribosome (80S) is formed of RNA associated to a large subunit (60S) composed of 49 proteins (Klinge et al., 2011) and to a smaller subunit (20S), which contains 33 proteins (Rabl et al., 2011).

The proteasome, a large multi-protein complex found ubiquitously in all eukaryotes and archaea, performs the important task of regulating the concentration of proteins within the cell and degrading misfolded proteins (Bochtler *et al.*, 1999; Smalle and Vierstra, 2004). The 26S proteosome is a 1.5 MDa protein complex consisting of a barrel shaped core composed of at least 32 different subunits and a capping regulatory particle of 19 subunits (Lander *et al.*, 2012; Murata *et al.*, 2009). Another multi-protein complex, which is also a

crucial cellular machine, is the ATP synthase motor. ATP synthases are complexes (~500 kDa) formed of proteins with distinct functions; the ATP-driven $F_1$ motor (~380 kDa) and the transmembrane proton-driven $F_o$ motor (Yoshida *et al.*, 2001). The $F_1$ and $F_o$ motors are in-turn hetero-oligomers. In bacterial ATP synthases, the $F_1$ motor consists of five different types of subunits, $\alpha_3\beta_3\gamma\delta\varepsilon$, with three copies of *α* and *β* subunits and single copies of the *γ*, *δ* and *ε* subunits. The subunit architecture of the $F_o$ motor is $ab_2c_n$, with $n$ =10 – 14, depending on the species (Mitome *et al.*, 2004; Stock *et al.*, 1999a; Tomashek and Brusilow, 2000) (Fig. 1.1c).

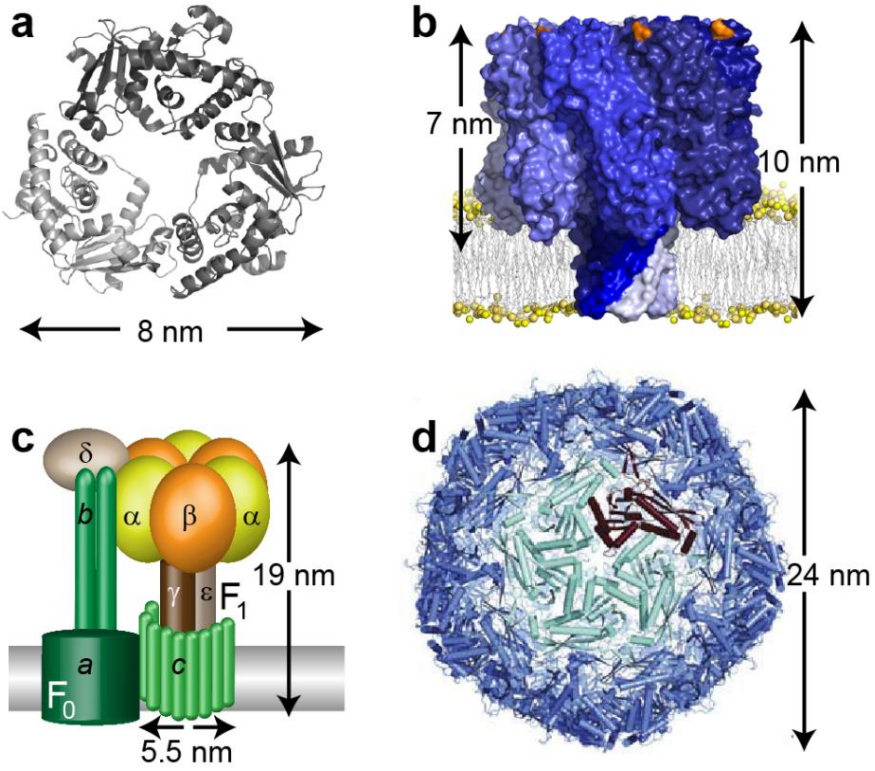



**Figure 1.1. Examples of natural homo- and hetero-oligomer protein complexes.** Some proteins are functional as homo-oligomers formed of identical subunits. Figures show (**a**) the homo-trimer of the DNA processing enzyme, λ-exonuclease (PDB 1AVQ) and (**b**) the homo-heptamer of the pore forming toxin of *Staphylococcus aureus* α-hemolysin (PDB 7AHL). Hetero-protein complexes are also prevalent in nature where two or more different proteins associate. (**c**) The ATP synthase consists of two motor proteins, $F_1$ and $F_o$. $F_1$ motor consists of five different types of subunits, $\alpha_3\beta_3\gamma\delta\varepsilon$, and the $F_o$ motor consists of three subunits, $ab_2c_n$, with n = 10 – 14. Dimensions shown correspond to the yeast F1 c10 complex (Stock *et al.*, 1999b). (**d**) Some viral

and bacterial proteins self-assemble to form nanometer sized hollow compartments to encapsulate genetic material and proteins. Cartoon representation of a bacterial encapsulin (PDB 3DKT). Monomers (one of them is highlighted in maroon) assemble into pentamers (one of them is shown in teal) with a five-fold symmetry axis. Twelve pentamers further assemble to form the 60-mer encapsulin shells. Reprinted by permission from Macmillan Publishers Ltd: [Nature Structural and Molecular Biology] (Sutter *et al.*, 2008), copyright (2008).

In prokaryotes and viruses, homogeneous capsules of unique architectures form by self-assembly of identical protein subunits. In bacteria and archaea, S-layer proteins self-assemble into monolayers forming an envelope around the cell and function as a shield against foreign attack and local changes in environment (Sara and Sleytr, 2000). Other bacterial proteins such as carboxysomes and encapsulins form nano-compartments that encapsulate enzymes (Kerfeld *et al.*, 2005; Sutter *et al.*, 2008) (Fig. 1.1d). These nano-compartments could function as metabolic reaction chambers and/ or protect the enzymes from proteolytic degradation.

In viruses, capsid proteins form high-order supramolecular structures of various shapes and sizes to encapsulate genetic material (Suttle, 2005). Icosahedral structures range from 18 - 200 nm in size and rod shaped structures could be > 2 µm (Callaway *et al.*, 2001; Mannige and Brooks, 2010; Shepherd *et al.*, 2006). The assembly or disassembly of the capsid proteins can be triggered by external factors (Bothner *et al.*, 2005). The dynamic metastable structure of the capsids allows stimulus dependent intake and release of encapsulated cargo (Konecny *et al.*, 2006; Liepold *et al.*, 2005). For example, the cowpea chlorotic mottle virion (CCMV) contains 180 metal binding sites that can in turn control the reversible opening and closing of 60 pores of ~2 nm each (Speir *et al.*,

1995). The CCMV capsid can also be disintegrated into protein dimers and reassembled by controlling the pH (Johnson and Speir, 1997).

### 1.2. Engineering new protein complexes

The underlying principles of protein-protein association have been elucidated by the wealth of available crystal structure data of homo and hetero-protein complexes (Jones and Thornton, 1996; Keskin *et al.*, 2008). Favourable electrostatic interactions, shape complementarity, burial of hydrophobic patches and formation of new hydrogen bonds are the key elements involved in protein-protein interactions (Chothia, 1975; Nooren and Thornton, 2003a; Reichmann *et al.*, 2007). The binding interactions between protein partners could be tuned, for example, by introducing point mutations that increase the affinity between proteins (Lee and Vasmatzis, 1997; Pokkuluri *et al.*, 2002), or entirely disrupt it (Lim *et al.*, 2011; Wu and Wong, 2005). For example, streptavidin, a tetrameric protein, was shown to function at the level of monomer by disrupting the inter-subunit interactions (Wu and Wong, 2005) and increasing the stability of the monomer (Lim *et al.*, 2011). Two residues (T76 and V125) in the monomer-monomer interface of streptavidin were changed to arginine to create electrostatic repulsions. Two more hydrophobic residues were changed to threonine to increase the hydrophilicity and solubility of the monomer. Introducing charged residues such as glutamates, arginine and lysines further increased the stability of the monomer by forming intra-molecular salt bridges (Lim *et al.*, 2011). These charged residues also have a higher β-sheet forming propensity (Minor and Kim, 1994) than the smaller wild-type (WT) residues of streptavidin.

Alternatively, two proteins could be brought together, which don't normally interact in an oriented configuration using the tools of protein engineering. Electrostatics (Seebeck *et al.*, 2006), bridging between engineered cysteine residues (Ballister et al., 2008; Miranda et al., 2009), redesign of protein surfaces (Binz *et al.*, 2004; Fellouse *et al.*, 2005; Fellouse *et al.*, 2004; Ye *et al.*, 2008), biotin-streptavidin anchorage (Hess *et al.*, 2005; Ringler and Schulz, 2003), genetic fusion (Mao *et al.*, 2004; Moll *et al.*, 2002; Moreau *et al.*, 2008; Sinclair *et al.*, 2011) use of natural ligands or their analogues (Kitov *et al.*, 2008; Kitov *et al.*, 2000; Solomon *et al.*, 2005) and metal ligand binding (Salgado *et al.*, 2008) have been used to make complexes of two or more non-interacting proteins.

### 1.3. Minimalist design of protein surfaces

The best example of protein engineering in nature is the generation of antibodies against any antigen by the immune system (Kuby, J, Immunology 2nd edition, W.H. Freeman and Company, 1994). Natural antibodies are generated starting from combinatorial arrays of template structures that evolve into antigen-binding structures with dissociation constants ($K_d$) lying in the nanomolar (nM) range (Tonegawa, 1983).

Sidhu and co-workers engineered antibodies starting from an antigen binding fragment (Fabs) as the template against specific antigen targets by redesigning the Fabs surface (Fellouse, Wiesmann *et al.* 2004; Fellouse, Li *et al.* 2005; (Ye *et al.*, 2008). Combinatorial libraries of the complementarity-determining region

(CDR) loops of a humanized Fab were made consisting of just two amino acids, tyrosine and serine (Fellouse *et al.*, 2005). Tyrosine was chosen because its side chain can interact with a wide range of functional groups and was introduced to make the key binding contacts (Fellouse *et al.*, 2005; Koide and Sidhu, 2009; Lo Conte *et al.*, 1999). Like tyrosine, serine is also a prevalent residue in the binding site of natural antibodies. Serine was used in conjunction with tyrosine to provide flexibility to the loops that formed unique binding surfaces for different antigens. Using *in vitro* phage display techniques (Bradbury *et al.*, 2011), out of a library of ~$10^{10}$ mutants, specific Fabs antibodies against six target antigens were found with $K_d$ between 50 nM and 5 μM. The minimalist binding surfaces of the engineered antibodies consisted of Ser and Tyr at just 28 to 36 positions and were specific for the targets. The crystal structure of the engineered Fabs domain bound to the extracellular domain of the human death receptor (hDR5-ECD) showed that 99% of the solvent accessible part of binding interface consisted of only tyrosine residues (Fig. 1.2a). The bulky side chain of tyrosine was contributing to a large surface area of interaction while the small hydrophilic side chain of serine acted as an auxiliary.

Extending this approach of minimalist design of the protein binding interface, Koide and co-workers designed ‘monobodies’ starting from a smaller single domain protein scaffold against three protein targets (Koide *et al.*, 2007). For the binding surface scaffold, the BC, DE and FG loops of the $10^{th}$ fibronectin type III domain of human fibronectin (FNfn10) were chosen and combinatorial libraries of all the three loops were made (Fig. 1.2b). The designed monobodies

showed very high affinities (nanomolar $K_d$) for their cognate targets and very few members of the library showed cross-reactivity or affinity for the other targets. The crystal structure of the FNfn10 monobody, MBP-74, for maltose binding protein (MBP) revealed that the monobody mimicked the natural binding partner of MBP, β-cyclodextrin (βCD). The aromatic part of the tyrosine side chains of MBP-74 mimicked the sugar backbone of βCD and the tyrosine hydroxyl groups made the same polar contacts as the hydroxyl groups of βCD with MBP.

Other small proteins made of amino acid repeats have also been engineered to bind specific protein targets with pico-molar affinities (Binz *et al.*, 2004; Boersma and Pluckthun, 2011; Gebauer and Skerra, 2009). Plückthun and co-workers introduced the concept of designed ankyrin repeat proteins (DARPins), which have 4-5 repeat motifs of engineered ankyrins (Binz *et al.*, 2003). Ankyrin repeat motifs were constructed with six positions randomized to any amino acid except cysteine, proline and glycine, and one residue was kept variable between asparagine, histidine or tyrosine in each monomer (Binz *et al.*, 2004). Out of the large library thus created, DARPins could be isolated against a variety of targets. DARPins have been engineered to act as antibody mimetic molecules and receptor antagonists, and its therapeutic applications are being probed (Binz *et al.*, 2004; Parizek *et al.*, 2012; Pecqueur *et al.*, 2012).

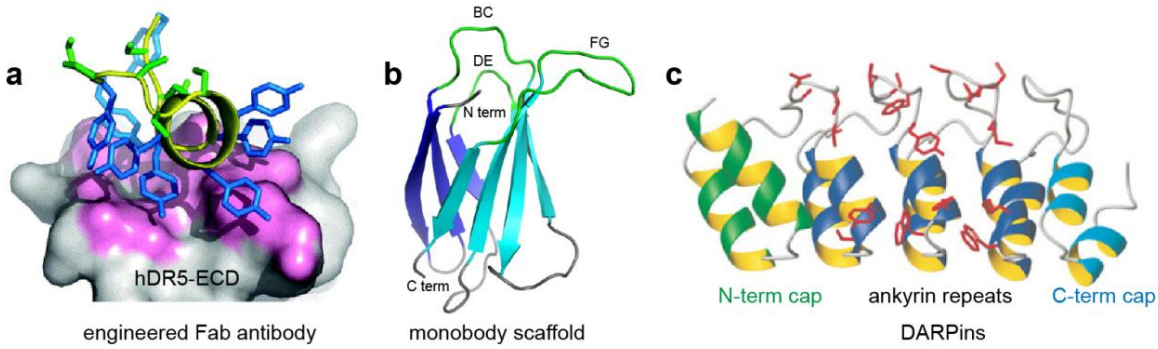



**Figure 1.2. Design of monobodies against protein targets and key binding contacts.** (**a**) Crystal structure of the designed antigen binding fragment (Fabs) antibody (yellow cartoon) and the extracellular domain of the antigen hDR5-ECD (surface). Tyrosine side chains (blue sticks) of CDR-H3 made majority of the contacts with the residues of hDR5-ECD (pink surface), while the serine residues (green sticks) acted as auxiliaries providing loop flexibility (Fellouse *et al.*, 2005). (**b**) Cartoon representation of the small protein domain of the human fibronectin (FNfn10) used by Koide and co-workers as a framework for building 'monobodies' against specific cognate targets. The monobody scaffold consisted of seven β-strands and the three loops, BC, FG and DE that were randomized (Koide *et al.*, 2007). (**c**) Ribbon structure of a designed ankyrin repeat protein (DARPin) against MBP. It consisted of three ankyrin repeat (AR) domains (blue) and N-terminal (green) and C-terminal (cyan) AR cap proteins. Red sticks show locations of residues that were randomized (Binz *et al.*, 2004). Figure adapted (**a**) from (Fellouse *et al.*, 2005) and (**b**) from (Koide *et al.*, 2007) Figure (**c**) reprinted by permission from Macmillan Publishers Ltd: [Nature Biotechnology] (Binz *et al.*, 2004), copyright (2004).

## 1.4. Use of bulky interface residues for homo-multimer formation

Instead of remodeling entire loops of proteins, Grueninger *et al* made homo-multimers of several proteins by adding just point mutations on the desired protein-protein interfaces (Grueninger *et al.*, 2008). The tetrameric protein L-rhamnulose-1-phosphate aldolase (Rua) was designed to self-assemble as nanofibres. Two mutants of Rua, termed RuaA and RuaD, were designed to yield tail-tail (RuaA) and head-head (RuaD) dimers by introducing bulky non-polar side chains of phenylalanine and tryptophan at the tail and head surfaces, respectively (Fig. 1.3a, b). In RuaD, the head also contained serine mutations. A double mutant of Rua (RuaE) with the RuaD head and RuaA tail mutations

assembled head-to-tail forming long fibrous multimers (Fig. 1.3c).

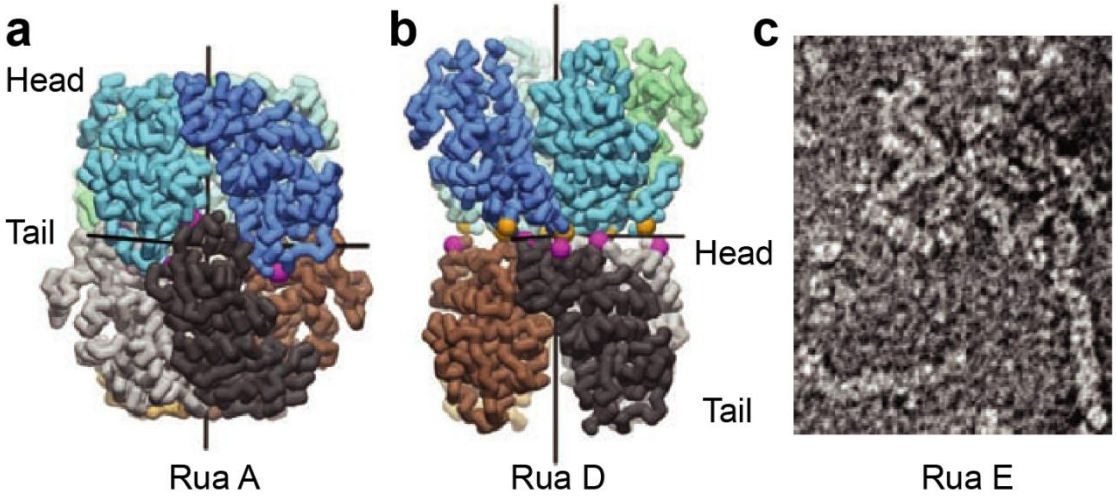



**Figure 1.3. Use of bulky amino acids to make protein homo-multimers.** Engineered multimers of the tetrameric L-rhamnulose-1-phosphate aldolase (Rua). (**a**) Phenylalanine was added to the Rua tail face (RuaA) to yield D4-symmetric octamers with the tetramer subunits associating *via* their tails. (**b**) Tryptophan and serine mutations in the head face of Rua (RuaD) led to the formation of head-head octamers. (**c**) The double mutant of Rua, with the afore-mentioned head and tail face mutations (RuaE), led to the formation of head-to-tail associated multimer fibres. Figures from (Grueninger *et al.*, 2008). Reprinted with permission from AAAS.

## 1.5. Covalently linking proteins with disulfide bonds

Protein surfaces can be modified with residues such as cysteines for site-specific covalent attachment of small chemical groups and other proteins (Ballister et al., 2008; Lim et al., 2011; Ringler and Schulz, 2003; Swain et al., 2010). The hexameric pores of Hcp1 (~17 kDa; *Pseudomonas aeruginosa*) of unknown *in vivo* function were designed to form protein nanotubes by genetically introducing cysteine mutations at the top and bottom surfaces of Hcp1 (Ballister *et al.*, 2008) (Fig. 1.4a). The residues for cysteine mutations were chosen based on the high-resolution structure of a P6 crystal of Hcp1 rings stacked head to tail. Under non-reducing conditions, Hcp1 rings with cysteine mutations, R157C and G90C, on the top and bottom faces,

respectively self-assembled head-to-tail into long tubular structures up to ~100 nm in length. Small molecules were attached with maleimide linkages to engineered cysteine residues in the Hcp1 lumen. As these nanotubes could be disassembled in reducing conditions, its therapeutic applications in drug delivery were proposed. Drugs could be encapsulated in the nanotubes and subsequently released in the reducing environment of the cytoplasm. Using a similar strategy, with cysteine mutations on the top and bottom faces and dithiothreitol (DTT) as the linking molecule, homo-multimers of the 11-mer trp RNA-binding attenuation protein (TRAP) formed nanofibres with very high aspect ratios (diameter to length ratio) ranging up to 12 (Miranda *et al.*, 2009). Engineered TRAP nanofibres with conducting or semi-conducting materials filled in the lumen could be potentially used as nano-wires.

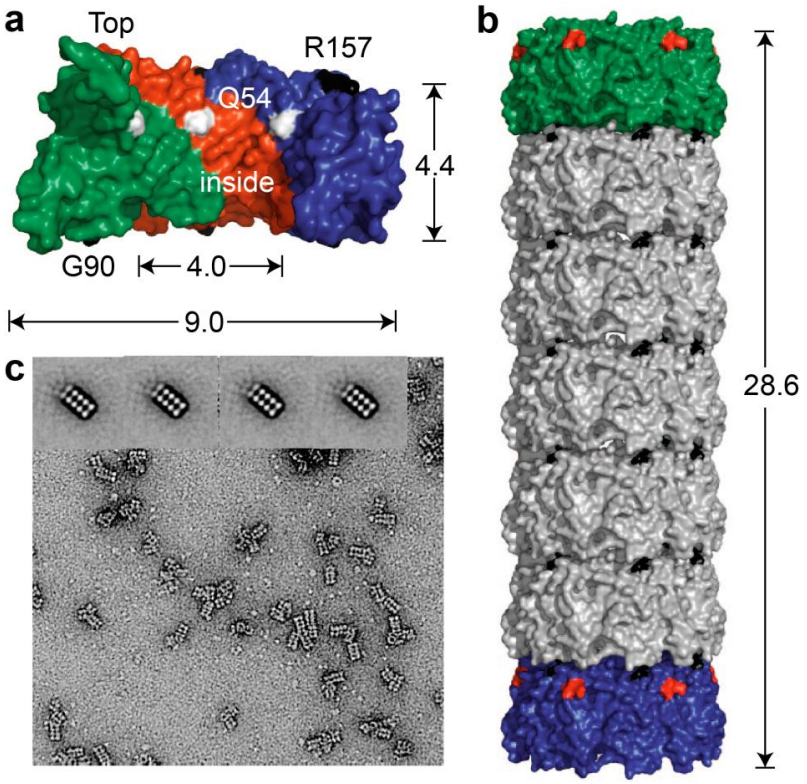



**Figure 1.4. Using disulfide bridging to form protein nanotubes.** To make nanotubes of the hexameric protein pore Hcp1, a double cysteine mutant, R157C (top face) and G90C (bottom face) was made. (**a**) Cross section of the Hcp1 pore showing the location of the residues chosen for mutation into

cysteine. Q54 (white), a residue in the Hcp1 lumen, was mutated to cysteine to encapsulate cargo molecules. (**b**) A schematic model of an Hcp1 nanotube showing five Hcp1 ring subunits (gray) with double cysteine mutations (black) attached head to tail, with two capping subunits with single cysteine mutations, G90C (green) and R157C (blue) and C terminus (red). (**c**) EM images show the different lengths of nanotubes that were formed when capping subunits were added. Figures from (Ballister *et al.*, 2008).

### 1.6. Driving protein complex formation using electrostatic interactions

The electrostatic interaction energy is proportional to the inverse of the distance between charges (Griffiths, 1999). Electrostatics thus play a crucial role in protein-protein interfaces where charged side chains are in close proximity (Dong and Zhou, 2006; Kukic and Nielsen, 2010; Sheinerman and Honig, 2002). Formation of many natural protein complexes is driven by electrostatic interactions (Buckle *et al.*, 1994; Maeda *et al.*, 2009). For example, the bacterial ribonuclease, barnase, binds to its polypeptide blocker, barstar, steered by electrostatic attraction (Buckle *et al.*, 1994). Two connexon proteins associate *via* hydrogen bonds and electrostatic interactions between their extracellular regions to form gap junctions (Maeda *et al.*, 2009). Synthetic protein complexes have also been made by placing opposite charges on the binding partners. Positively charged GFP (GFP-R10) and an HIV protease bearing a deca-arginine tag (R10) were encapsulated in negatively charged capsids (Fig. 1.5) (Seebeck *et al.*, 2006; Worsdorfer *et al.*, 2011). Four lumen residues in each monomer of the *Aquifex aeolicus* lumazine synthase (AaLS) capsid were changed into glutamates. The capsid formed of 60 (T=1) or 180 (T=3) subunits was thus endowed with 240 or 720 extra negative charges in its interior (AaLS-neg) (Seebeck *et al.*, 2006). GFP encapsulation had a strong dependence on the electrostatic interactions between GFP and AaLS. Only when a combination of the GFP-R10 and the AaLS-neg were co-expressed,

GFP could be substantially encapsulated. 3-4 GFP-R10 molecules were encapsulated per T=3 AaLS-neg capsid.

The AaLS-neg capsid was optimized for binding a positively charged HIV protease with much higher capacity by directed evolution (Worsdorfer *et al.*, 2011). When tagged with a R10 tag (HIV-R10), HIV protease could also be encapsulated in co-expressed AaLS-neg capsids and was less toxic to cell growth. A library of AaLS-neg mutants were made by repeated rounds of error-prone polymerase chain reactions (PCR) for enhanced encapsulation of HIV-R10. An AaLS-neg mutant with seven mutations in total, with three new negatively charged residues and three cationic and one negative residues removed was the best HIV-R10 binder and bound to average of seven molecules per capsid.

Using this AaLS capsid variant and a super-charged GFP with 36 extra positive charges, the average number of encapsulated GFP molecules was increased to 74±10 per T=3 capsid (Worsdorfer *et al.*, 2012). The entrapped GFP molecules had higher thermal stability than free protein and were protected from protease action.

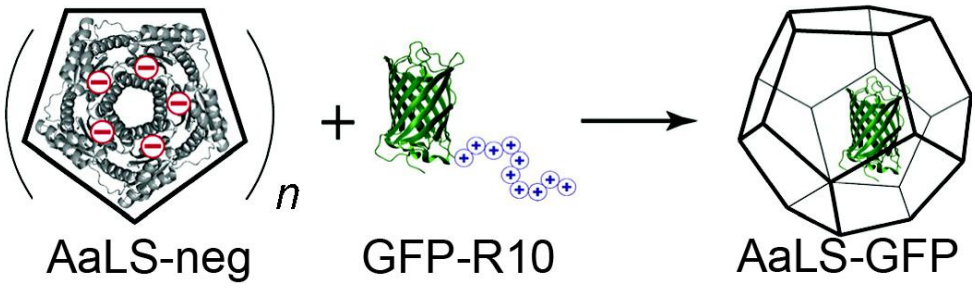



**Figure 1.5. Protein association *via* electrostatic interactions.** Schematic illustration of the concept of GFP encapsulation inside viral capsids employing electrostatics interactions. GFP (green ribbon structure) was rendered positive

using a R10 tail (GFP-R10). Four negatively charged residues were engineered in each monomer of the lumazine synthase capsid protein (AaLS) making it highly negatively charged (AaLS-neg). Co-purification of GFP-R10 and AaLS-neg led to encapsulation of ~4 GFP molecules on average in the AaLS capsid. Reprinted with permission from (Seebeck *et al.*, 2006). Copyright (2006) American Chemical Society.

### 1.7. Non-covalent protein-protein association using ligands

Protein complex formation can be mediated using ligands that are known to bind the proteins of interest. Encapsulation of proteins within a viral capsid was directed and enhanced using RNA aptamers that could bind to both the cargo and the capsid proteins (Fiedler *et al.*, 2010; Rhee *et al.*, 2011). Fluorescent proteins were entrapped in the virus-like particles of the bacteriophage Qβ (Qβ VLP). The gene of Qβ VLP monomers (QP) was tagged such that the transcribed mRNA contained two binding domains; a RNA sequence to bind to the interior of assembled Qβ VLPs and an aptamer to bind to an arginine rich peptide (Rev). Various GFP spectral variants were engineered with Rev tags (GFP-Rev) and co-expressed with the QP protein. The mRNA with dual adapters tethered the GFP-Rev proteins within assembled Qβ VLPs.

Like aptamers, use of bivalent ligands that bind to both protein partners of interest has been another approach to make protein complexes. Two pentameric Shiga toxin (Stx1) protein units were brought face-to-face using a ligand with ten $P^k$-trisaccharide groups, each of which bound in the carbohydrate binding site of a Stx1 monomer (Fig. 1.6a, b) (Kitov *et al.*, 2000).

Another pentameric protein, serum amyloid P component (SAP) was attached face-to-face to Stx1 using heterobivalent ligands with binding moieties for each

protein (Kitov *et al.*, 2008; Solomon *et al.*, 2005). The bifunctional ligand consisted of $P^k$-trisaccharide and cyclic glycerol pyruvate analogues to bind Stx1 and SAP, respectively (Fig. 1.6c, d). The 1, 2-*O*-carboxyethylidene derivative of glucose was used to mimic the cyclic glycerol pyruvate ligand of SAP. Linkers were not used to connect the functional groups in order to minimize entropy losses. The molecular modeling of the ligand bound to Stx1 and SAP showed that the binding moieties fit in the respective binding sites and did not lead to unfavourable interactions between the proteins. In *in vitro* assays, pentamers of Stx1 and SAP were observed in a 1:1 complex.

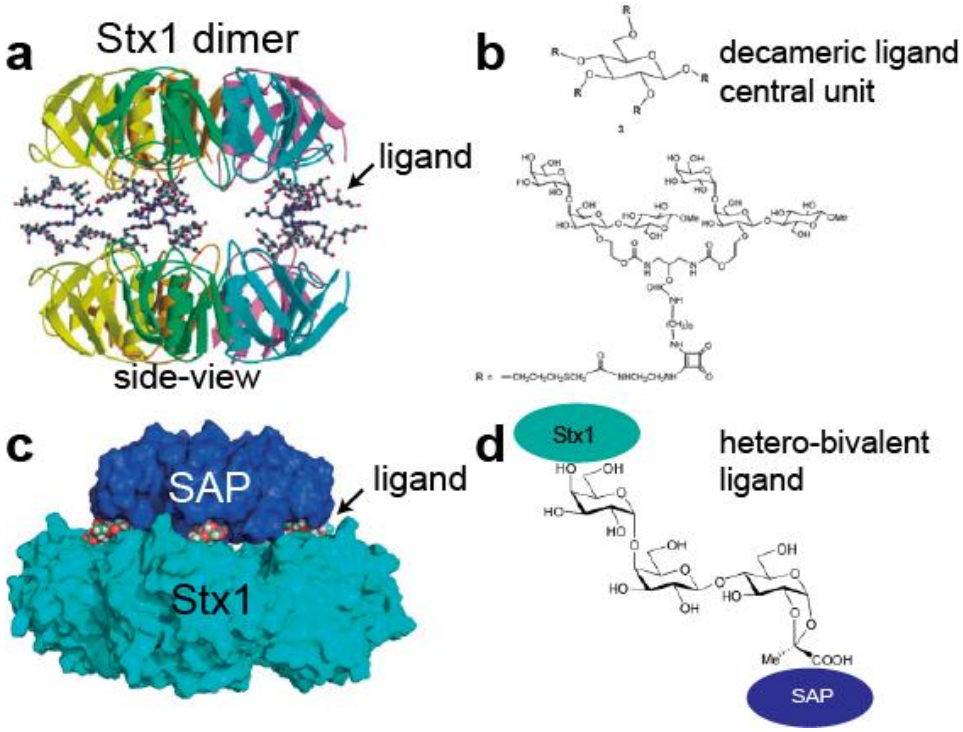



**Figure 1.6. Protein complexes mediated by natural ligand analogues.** Models showing (**a**) side view of the crystallographic face-to-face complex of the pentameric Shiga toxin (Stx1) protein (ribbon structure) connected by a decameric $P^k$-trisaccharide ligand (balls and sticks). (**b**) Structure of the ligand consisting of 10 $P^k$-trisaccharide groups used to bring two Stx1 molecules face-to-face. Figure reprinted by permission from Macmillan Publishers Ltd: [Nature] (Kitov *et al.*, 2000), copyright (2000). (**c**) Space filling model showing the hetero-complex of pentameric proteins serum amyloid P (SAP) (blue) with Stx1 (teal) using a hetero-bifunctional ligand (spheres). (**d**) Structure of the hetero-bifunctional ligand containing $P^k$-trisaccharide and cyclic glycerol pyruvate analogues to bind Stx1 and SAP, respectively. The absence of a long linker

connecting the two functional groups helped prevent any conformational entropy losses. Figure reprinted with permission from John Wiley and Sons (Kitov *et al.*, 2008), Angewandte Chemie International Edition, copyright (2008).

### 1.8. Using the biotin-streptavidin linkage

The strong affinity between biotin and streptavidin ($K_d \approx 10^{-14}$ M) has also been used to cross-link proteins. Biotin and streptavidin functionalized microtubules were self-assembled on surfaces coated with kinesin motor proteins (Hess *et al.*, 2005) (Fig. 1.7). Linear and circular microtubules were biotinylated and partially coated with streptavidin. Random movement of these microtubule filaments on the kinesin covered surfaces led to frequent collisions and cross-linking between microtubules *via* biotin-streptavidin-biotin linkages. The biotin-streptavidin anchorages between the microtubules were stable in thermal equilibrium but movement of the kinesin motors could shear associated microtubules apart. The tip-to-tip cross-linked microtubules dynamically assembled as nano-wires and nano-spools over time.

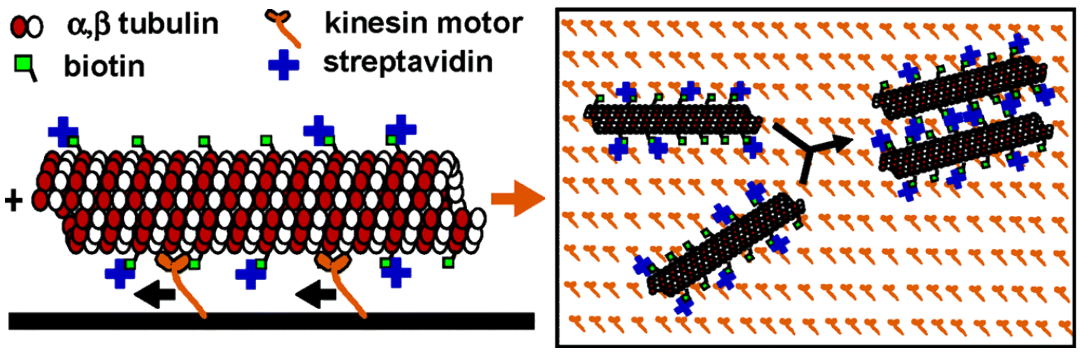



**Figure 1.7. Dynamic self-assembly using biotin-streptavidin anchorages.** Biotinylated microtubules were partially coated with streptavidin. Collisions between the microtubules while diffusing randomly on a surface coated with kinesin motor proteins led to inter-molecular cross-linking *via* biotin-streptavidin−biotin anchorages. Figure adapted with permission from (Hess *et al.*, 2005). Copyright (2000) American Chemical Society.

Ringler *et al.* used streptavidin-biotin linked 2-D networks of the tetrameric aldolase, L-rhamnulose-1-phosphate aldolase (Rua) (Ringler and Schulz, 2003). Cysteines were placed at the solvent exposed surfaces of each Rua monomer and functionalized with biotin. Each molecule of the cubic Rua tetramer, with eight exposed biotins could thus bind four molecules of streptavidin, one on each face (Fig. 1.8). Additional streptavidin molecules were then added to act as four-way connectors between these Rua-streptavidin building blocks to generate large 2-D networks. A His-6 tag on each Rua monomer allowed assembly of the streptavidin-linked Rua networks on lipid monolayers containing Ni-NTA head groups. It has been proposed that lattices of other proteins such as membrane proteins and enzymes could be created by genetically fusing them to Rua.

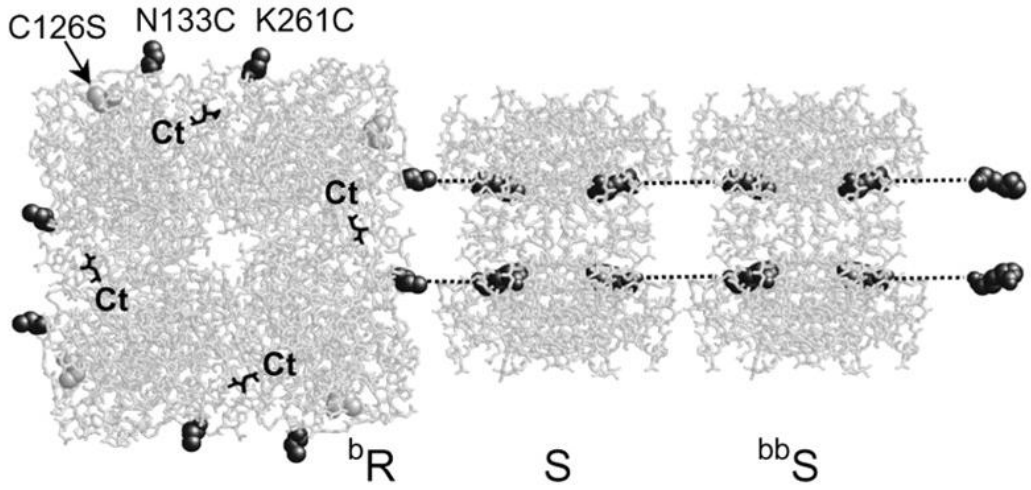



**Figure 1.8. Protein networks with streptavidin-biotin linkages.** Model of the subunit of a 2-D network formed by the C4-symmetric enzyme L-rhamnulose-1-phosphate aldolase (Rua) and D2 symmetric streptavidin. Each subunit consisted of three components, the Rua tetramer ($^{b}$R) and two streptavidin molecules (S and $^{bb}$S). Each monomer of $^{b}$R tetramer was engineered to contain two cysteines (N133C and K261C), each of which was labeled with a biotin molecule (black spheres). The two biotins on each side of $^{b}$R (only one side shown in figure) bound to a streptavidin molecule (S), which further bound to another streptavidin molecule ($^{bb}$S). With S acting as a four-way connector, a 2-D network of this system was created. The C terminal (Ct) of each monomer contained a His-6 tag, which was later used to tether the

network on lipid bilayers containing Ni-NTA head groups. Figure from (Ringler and Schulz, 2003). Reprinted with permission from AAAS.

### 1.9. Metal mediated protein complexation

More than a third of proteins bind metal ions *via* their side chains (Barondeau and Getzoff, 2004). Examples of natural protein complexes bridged *via* metal ions are also known (Emsley *et al.*, 2000; Zhou *et al.*, 2005). Zhou *et al.* found that interfacial metal ions could also mediate antibody-antigen recognition (Zhou *et al.*, 2005). The presence of $Ca^{2+}$ enhanced the binding affinity between a CD4-reactive antibody and CD4 by 55,000 times. In another study, the crystal structure of the I domain of integrin α2β1 in complex with a collagen peptide revealed a bridging metal ion in the protein-protein interface (Emsley *et al.*, 2000). Similar crystals formed with different divalent metal ions, viz. $Mg^{2+}$, $Mn^{2+}$, $Co^{2+}$, $Ni^{2+}$, $Cd^{2+}$, $Ni^{2+}$ and $Zn^{2+}$. The glutamate side chain of collagen coordinated with the metal ion situated on the integrin surface. This drove other protein residues in proximity as well, which then formed hydrogen bonds, salt bridges and Van der Waals contacts.

The selectivity and directionality of multi-dentate metal coordination can be used to link protein surfaces (Salgado *et al.*, 2010). Attaching proteins *via* metal coordination has the added advantage of reversibility and thermodynamic control by external factors such as pH, temperature and the presence of external ligands or other metal ions (Radford *et al.*, 2011; Salgado *et al.*, 2010). In metalloproteins, a common motif for bidentate coordination to metal ions is a i, i+4 bis-His clamp on α-helices (Arnold and Haymore, 1991). A hemeprotein

cytochrome cb562 was engineered with four histidine residues such that they formed two bis-His clamps. The cb562 variant, MBPC-1, with two bis-His motifs could bind to different metal(II) ions and assemble into oligomers with different stoichiometries depending on the metal ion present (Fig. 1.9). The geometry of the metal ion coordination determined the symmetry of the final supramolecular complex formed. Addition of $Cu^{2+}$ led to anti-parallel C2 symmetric dimers, and coordination with $Ni^{2+}$ resulted in homo-trimers (Salgado *et al.*, 2009). Depending on the $Zn^{2+}$ concentration, MBPC-1 formed dimeric or tetrameric oligomers (Salgado *et al.*, 2007; Salgado *et al.*, 2009).

The binding interface of the MBPC-1 protein was further modified with ten amino acids chosen computationally (Brodin *et al.*, 2012). Apart from the two metal binding motifs already present, a third metal binding site was engineered such that after C2 dimer formation, one $Zn^{2+}$ binding site was left open to bind to another C2 dimer. This promoted self-assembly of the C2 dimers in 1-D arrays, 2-D sheets, or linear unbranched hollow tubes depending on the pH and, the protein and $Zn^{2+}$concentrations.

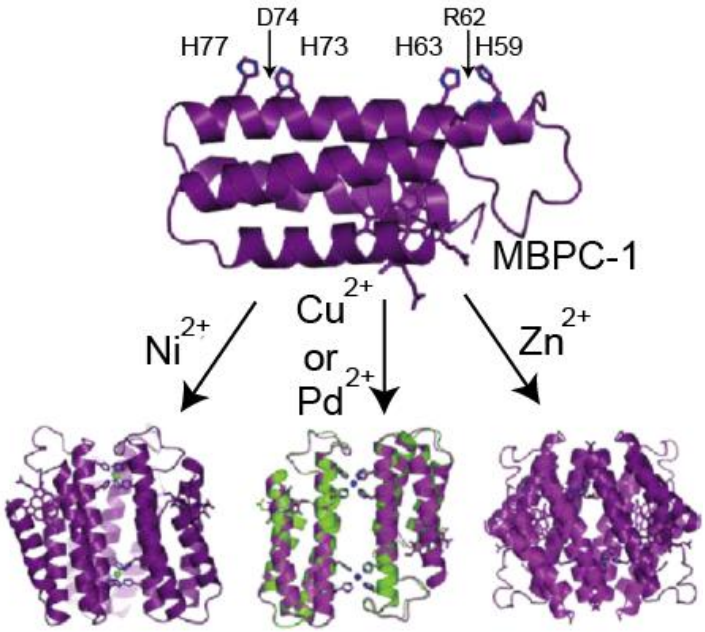



**Figure 1.9. Protein complexes bridged by metal ions.** Cartoon representation of the four-helix-bundle hemeprotein cytochrome cb562 variant (MBPC-1) with two bis-His motifs (H73, H77 and H59, H63), a commonly found structural motif in metal binding proteins. Metal binding to the bis-His motifs led to the self-assembly of MBPC-1 into oligomers with different stoichiometries depending on the stereochemical preference of the metal ion used. A trimer with C3 symmetry was yielded with $Ni^{2+}$, a C2 symmetric dimer with $Cu^{2+}$ and $Pd^{2+}$ and a D2 symmetric tetramer with $Zn^{2+}$. Figure from (Radford *et al.*, 2011). Copyright (2011), with permission from Elsevier.

## 1.10. Metal mediated protein assembly using non-natural ligands

Metal-mediated protein cross-linking could be combined with the use of non-natural ligands. Streptavidin molecules were linked in tandem using a bi-functional linker (ligand) and $Fe^{2+}$ ions to form a 1-D protein network (Burazerovic *et al.*, 2007). The ligand contained a bis-biotin ($Biot_2$) moiety at one end to bind streptavidin and a ter-pyridine (terpy) group at the other end to chelate $Fe^{2+}$ ions. Each $Fe^{2+}$ ion was bound by two ligand molecules to form a $[Fe(Biot_2\text{-terpy})_2]^{2+}$ complex which then acted as a connector between adjacent streptavidin molecules and led to the formation of long linear complexes of streptavidin (Fig. 1.10). The streptavidin network was used for mineralization of calcite into millimeter-sized wires.

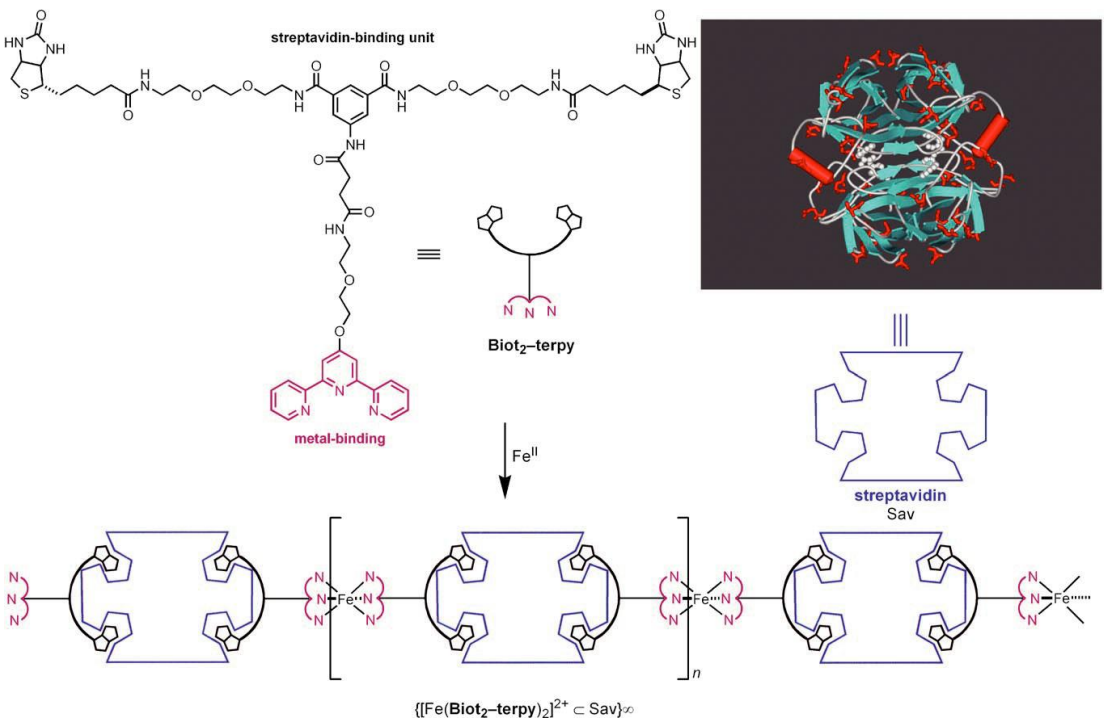



**Figure 1.10. Use of metal coordination and ligand binding to form protein networks.** Structure of the bis-biotinylated (black) terpyridine (violet) ($Biot_2$-terpy) ligand used in conjunction with Fe(II) ions to form linear polymers of streptavidin (Sav) (cartoon, right inset). Each of the bis-biotin groups of $Biot_2$-terpy bound to a different streptavidin molecule. The ter-pyridine groups of two $Biot_2$-terpy molecules bound to Fe(II) ions, which then acted as a connector between the streptavidins and leading to the formation of very long 1-D chains. Figure reprinted with permission from John Wiley and Sons (Burazerovic *et al.*, 2007), Angewandte Chemie International Edition, copyright (2007).

In another example, an iodoacetamide derivative of phenanthroline (Phen) was linked to an engineered cysteine residue in a hydrophobic crevice of the MBPC-1 protein described in section 1.9 (MBP-Phen1) (Radford and Tezcan, 2009). Upon adding $Ni^{2+}$ ions, MBP-Phen1 rearranged as homotrimers where $Ni^{2+}$ ions were coordinated by the Phen groups and histidine (H77) side chains (Fig. 1.11). Similarly other synthetic ligands could be used to generate protein complexes with varying stoichiometries and binding strengths. Although this approach is very powerful, it cannot be generalized to all proteins and used to make hetero-protein complexes selectively.

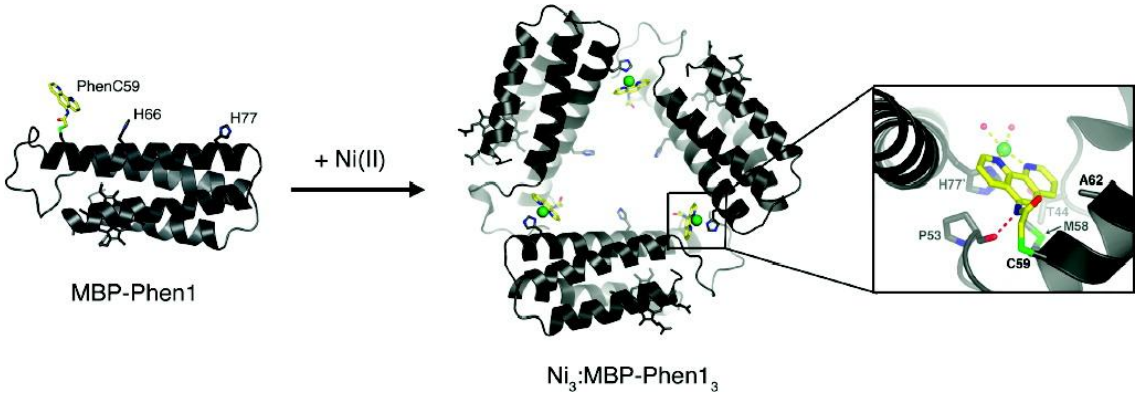



**Figure 1.11. Linking non-natural ligands to proteins for metal mediated protein complexes.** To achieve metal-mediated protein complex formation, metal binding ligands could be chemically attached to proteins at specific positions. The MBPC-1 protein was engineered with a cysteine residue (C59), which was then modified with a phenanthroline ligand (PhenC59, yellow sticks) (MBP-Phen1). Addition of $Ni^{2+}$ ions (green spheres) led to the formation of a C3 symmetric homo-trimer of MBP-Phen1. Figure reprinted with permission from (Radford and Tezcan, 2009). Copyright (2009) American Chemical Society.

## 1.11. Genetic fusion of proteins

Two or more different proteins can be physically constrained to be at predetermined distances by genetically fusing them together. In *in vitro* techniques, the genes of the proteins of interest are placed in tandem spaced by linkers. The resultant gene upon expression yields the proteins linked to each other by the linker sequences.

In nature, fusion of genes can occur due to chromosomal translocation, interstitial deletion or during transcription (Akiva et al., 2006; Davis and Barr, 1997; Meyer et al., 2009; Santo et al., 2012). The rearrangement of genes into fusion constructs could be lethal and in some cases could act as oncogenes (Mitelman et al., 2007; Tomlins et al., 2005; Villanueva, 2012). Interestingly, the use of genetically fused proteins has also been implicated for cancer therapy (Orentas *et al.*, 2012; Staerz *et al.*, 1985). Bevan and colleagues designed

bispecific antibodies by genetically fusing two antibodies that targeted CD3 on T cells and a surface target antigen on cancer cells (Staerz et al., 1985). Such antibodies could thus be used to link T cells specifically with cancer cells allowing the T cells to exert their cytokine release activity locally.

Genetically fused proteins have also been used for vivo imaging (Lorenz *et al.*, 2006), bio-sensing (Moreau *et al.*, 2008; Völlenkle *et al.*, 2004), elucidating intra-cellular signaling pathways (Dueber et al., 2003; Miyawaki et al., 1997), and for the construction of new proteinaceous and inorganic materials (Mao *et al.*, 2004; Padilla *et al.*, 2001; Sinclair *et al.*, 2011).

### 1.11.1. Genetically fused proteins for imaging

Proteins have been genetically fused with photoactivable fluorescent proteins for imaging purposes like photo-activated localization microscopy (PALM) (Betzig *et al.*, 2006). Red fluorescent protein (RFP), GFP and its cyan (CFP) and yellow (YFP) spectral variants were genetically fused to different cellular proteins (Lorenz *et al.*, 2006). The precise localization of the cellular proteins in the Golgi apparatus, the endoplasmic reticulum, peroxisomes, mitochondria and autophagosomes could be tracked by fluorescence-microscopy. Fusion constructs with fluorescent proteins has also been used to elucidate in vivo protein-protein interactions (Lorenz *et al.*, 2006; Yan and Marriott, 2003).

GFP mutants with shifted excitation wavelengths have also been fused together FRET measurements in single cells (Janetopoulos *et al.*, 2001; Mahajan *et al.*, 1998; Miyawaki *et al.*, 1997; Nguyen and Daugherty, 2005; Rizzo *et al.*, 2004).

In the seminal work by Tsien and co-workers, a tandem gene of CFP, calmodulin, calmodulin binding peptide M13 and YFP with enhanced fluorescence output was prepared (Miyawaki *et al.*, 1997). Calmodulin responded to changes in $Ca^{2+}$ concentrations by binding to the M13 peptide. This conformational change in turn led to a decrease in the distance between the fluorescing proteins and hence, a change in the FRET output (Fig. 1.12). Complementary DNA of the fusion construct was localized and expressed in various cell organelles and the FRET output was used to track $Ca^{2+}$ dynamics in single cells.

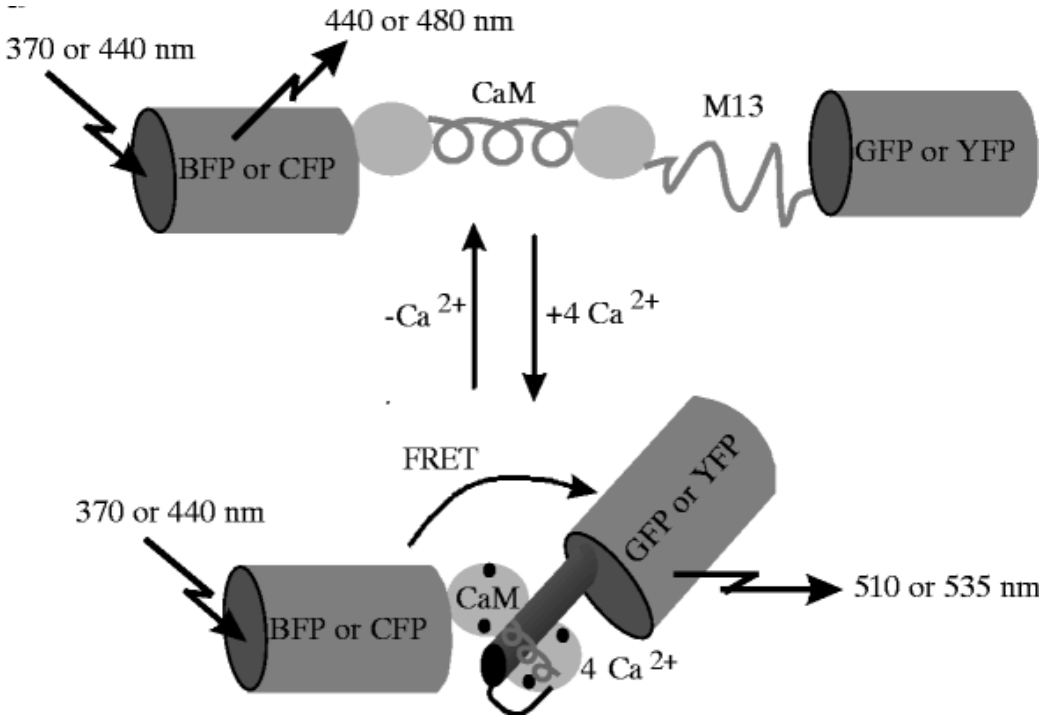



**Figure 1.12. Use of GFP fusion constructs for FRET measurements.** A tandem gene containing the genes of two GFPs (BFP or CFP and GFP or YFP) interspaced with genes for calmodulin (CaM) and the M13 peptide was constructed. The spectral characteristics of the fusion construct changed depending on the concentration of $Ca^{2+}$. Addition of $Ca^{2+}$ led to CaM and M13 binding. The resulting conformational change in CaM brought the GFPs closer, which led to an increase in the FRET output. Figure reprinted by permission from Macmillan Publishers Ltd: [Nature] (Miyawaki et al., 1997), copyright (1997).

### 1.11.2. Genetically fused proteins for building new materials

The genetic fusion approach has also been used to create proteinaceous materials with unique architectures. Proteins with oligomerization stoichiometries as low as two have been combined to create multimeric structures of distinct architectures (Padilla *et al.*, 2001). The design of pre-defined structures requires the geometry of the symmetry axes of the individual oligomerizing domains to be fixed (Fig. 1.13). The rigidity between the oligomerizing units was imposed by genetic fusion of the protein monomers with short α-helical linkers. A protein cage was designed with subunits consisting of a trimer forming protein (bromoperoxidase) connected *via* a linker of 9 residues to a protein that assembled as dimers (influenza virus matrix protein M1). Another fusion construct of two dimeric proteins M1 and carboxylesterase yielded networks and bundles of 4 nm wide filaments. Larger protein networks of other geometries, formed of proteins with matching rotational symmetries have also been demonstrated (Sinclair *et al.*, 2011). The ability to design unique architectural shapes of protein assemblies is crucial for its use as a scaffold for positioning other proteins.

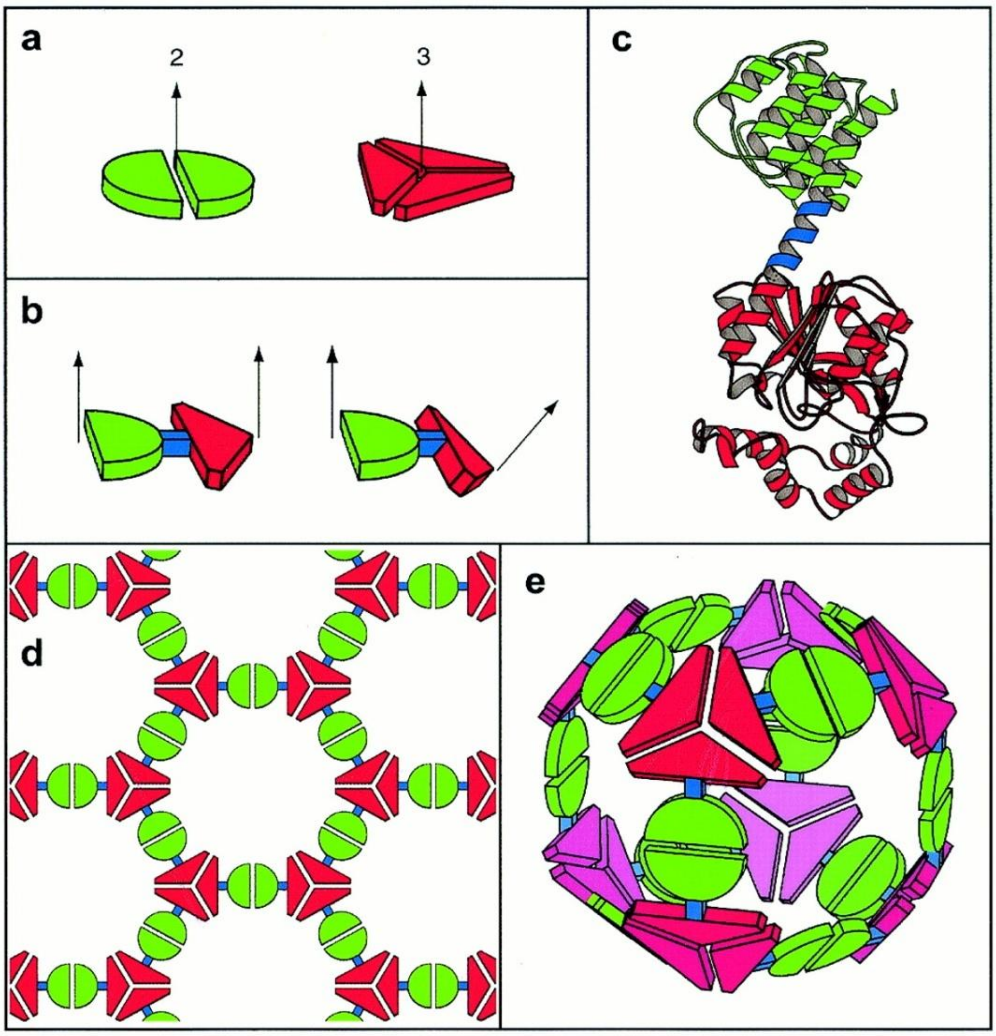



**Figure 1.13. Design principle of self-assembling protein nanostructures.** Schematic showing a strategy of building self-assembling nanostructures from two smaller protein units. (**a**) Protein units used to create nanostructures consisted of a dimeric protein (influenza virus matrix protein M1) represented as two green semicircles, and a trimeric protein (bromoperoxidase) three red triangles, with the axes of symmetry shown as black arrows. (**b**) The monomer subunits of the dimer and trimer proteins were linked genetically into a new protein, where each of the monomers acted as 'oligomerization domains'. Length of the rigid linker (blue) decided the geometry between the two oligomerization domains. (**c**) Ribbon structure of a fusion construct of the oligomerization domains (red and green) linked to each other by a short and rigid α-helical linker (blue). The red monomers trimerized and the green monomers dimerized forming supramolecular complexes whose structure depended on the relative geometry of the symmetry axes of the two oligomerization domains. (**d**) With symmetry axes pointing in the same direction (b, left), a molecular layer was formed by self-assembly. (**e**) With an arrangement of the oligomerization domains as shown in (**b**) (right), a cubic cage was formed. Figure from (Padilla *et al.*, 2001).

The architecture of viral capsids and self-assembling bacterial proteins has also been used to attain highly ordered and symmetric arrangements of peptides and proteins (Chatterji *et al.*, 2004; Mao *et al.*, 2004; Moll *et al.*, 2002). Four peptides that could control the nucleation of ZnS, CdS, CoPt and FePt nanoparticles were genetically fused to the N-terminus of the M13 bacteriophage capsid protein gP8 (Mao *et al.*, 2004). The gP8 self-assembly into high order symmetric structures worked as a rigid scaffold for specific orientation of the fusion peptides and hence, the nucleation of the mineral particles added later. The organic scaffold of the virus could then be removed by heating up to 500°C where upon the mineral nanocrystals re-arranged into 1-D nanowires. Depending on the symmetry of the capsid protein employed and the peptide used, a variety of single-crystal nanostructures could be constructed.

### 1.11.3. Genetically fused proteins for sensing

A nanopatterned molecular affinity matrix was made of a chimeric bacterial S-layer protein with streptavidin (Moll *et al.*, 2002). Streptavidin monomers were added to the S-layer streptavidin fusion construct monomers in a ratio of 3:1 in order to form streptavidin tetramers. The self-assembly of the S-layer proteins placed the covalently linked streptavidin molecules in a regular patterned array. In another fusion construct, immunoglobulin G (IgG) binding Z-domain was linked with S-layer proteins (Völlenkle *et al.*, 2004). Self-assembly of the S-layer led to the formation of a densely packed IgG binding affinity matrix, which could be coated on microbeads and used for extracorporeal blood purification.

### 1.11.4. Genetic fusion to make new classes of proteins

Inspired by ATP dependent $K^+$ channels in nature, Vivaudou and co-workers built ion channel coupled receptors by genetically fusing a GPCR with a $K^+$ channel (Caro *et al.*, 2012; Moreau *et al.*, 2008). Upon ligand binding, the GPCR underwent a conformational change, which was transmitted to the $K^+$ channel causing a change in ionic current through it. Based on the GPCR used, muscarinic M2 receptor or the dopaminergic D2 receptor, corresponding ligands and antagonists were shown to induce gating responses in the $K^+$ channel, Kir 6.2. Various deletion mutants of the C and N termini of both the proteins were made to optimize the signal response of the $K^+$ channel upon ligand binding to the GPCR (Fig. 1.14).

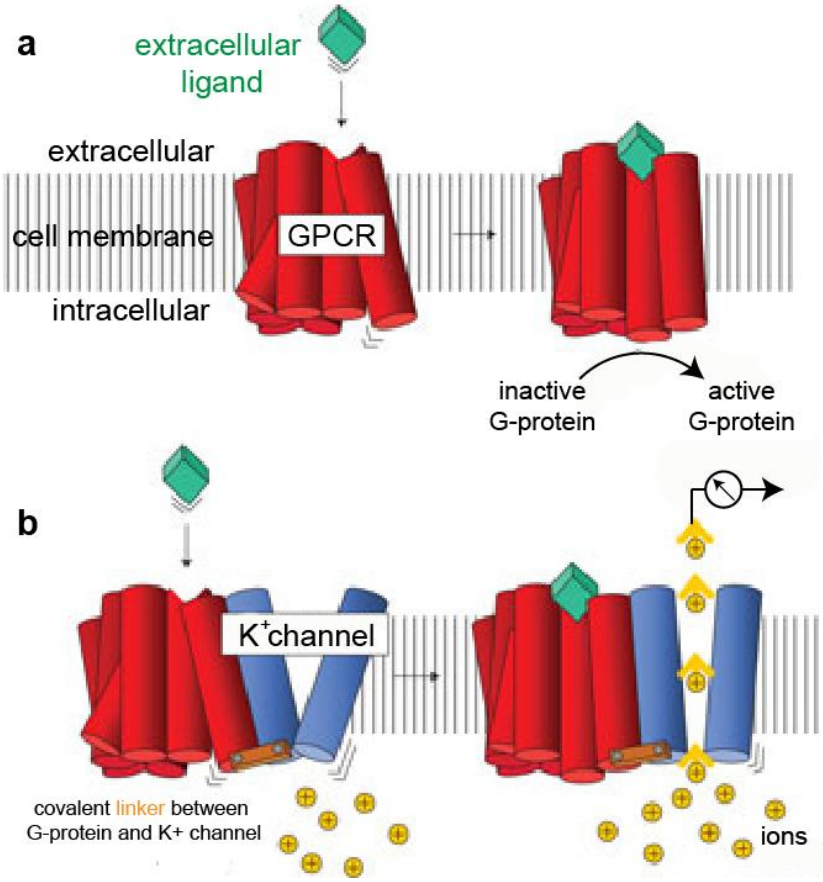



**Figure 1.14. Principle of ion channel coupled receptors.** (**a**) Transmembrane G protein-coupled receptors (GPCRs) (red) change conformation on binding to specific ligands (green) added to the extracellular

side. (**b**) When fused to a $K^+$ channel (blue), conformational change due to ligand binding in the GPCR could be mechanically translated to the $K^+$ channel changing its electric conductance. Moreau *et al* genetically fused the M2 GPCR with the Kir 6.2 $K^+$ channel making a new class of ions channels (ion channel coupled receptors) that responded to the presence of the M2 ligand acetylcholine and antagonist atropine. Figure adapted by permission from Macmillan Publishers Ltd: [Nature Nanotechnology] (Moreau *et al.*, 2008), copyright (2008).

In the seminal work of Chandrasegaran and co-workers on Zn finger nucleases (ZFNs), the cleavage domain of an endonuclease, FokI, was genetically fused to Zn finger repeats yielding new restriction enzymes that could cleave DNA at predetermined sequences (Kim *et al.*, 1996). The Zn finger repeats, which were made with different combinations of Zn fingers recognized and bound to specific DNA sequences whereupon the FokI enzyme exerted its nuclease action. Since then, much progress has been made on the design of ZFNs and restriction enzymes could be programmed to cut DNA at any desired sequence (Carroll, 2011). Besides genetic engineering, ZFNs would have direct applications in gene therapy for treating certain genetic disorders (Li *et al.*, 2011) and HIV-AIDS (Gaj *et al.*, 2012).

### 1.12. Self-assembling peptides

As model systems for protein engineering, peptides have been designed for self-assembly into supramolecular nano-structures like fibers (Zhang *et al.*, 2002), nanospheres (Matsuura *et al.*, 2005), nanotubes (Reches and Gazit, 2007) and dendrimer networks (Ryadnov, 2007). Circular peptides with alternating D and L-amino acids were designed by Ghadiri *et al.* that self-assembled to form nanotubes (Ghadiri *et al.*, 1993). On changing the hydrophilic residues of the cyclic peptides to hydrophobic residues, these

nanotubes could insert in bilayers and mimic ion channel behaviour (Ghadiri *et al.*, 1994). Peptides formed of natural L-amino acids only could also be assembled into high order structures (Matsuura *et al.*, 2005; Ryadnov, 2007).

Peptide nanospheres were inspired from the design of self assembling spherical DNA structures whose basic component consisted of a three-way junction with complementary sticky ends (Matsuura *et al.*, 2003). The building block of the peptide nanosphere was a self-dimerizing peptide (sequence CFKFEFKE) (Matsuura *et al.*, 2005). A three-way junction of the peptide was made by linking the terminal cysteines of three peptides with a C3-symmetric iodoacetamidated core molecule (Fig. 1.15). Each arm of the C3-symmetric peptides formed antiparallel β-sheets with arms of other C3-symmetric peptides at acidic pH (3.3). This led to the formation of dodecahedron type nanospheres of average diameter 19.1 ± 4.1 nm.

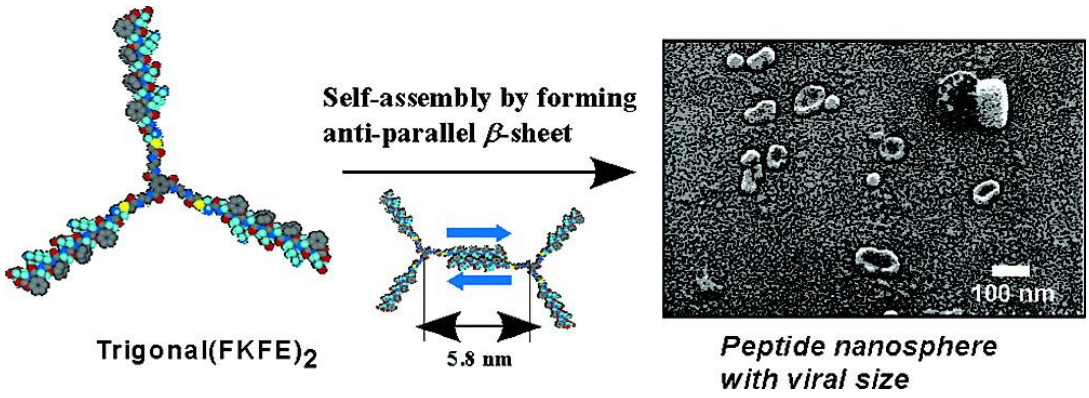



**Figure 1.15. Design of self-assembling peptide nanosphere.** A trigonal peptide (Triagonal(FKFE)$_2$) was made by reacting the terminal cysteines of three peptides with a C3-symmetric iodoacetamidated core molecule. Each arm of Triagonal(FKFE)$_2$ made anti-parallel sheets with an arm of another Triagonal(FKFE)$_2$ molecule, self-assembling into viral sized nanospheres. Figure reprinted with permission from (Matsuura *et al.*, 2005). Copyright (2005) American Chemical Society.

A common motif for self-assembly in natural bio-systems consists of repeats of hydrophobic and polar amino acid sequences, such as HPPHPPP (H= hydrophobic, P=polar), also termed as the leucine zipper (Busch and Sassone-Corsi, 1990). A leucine zipper peptide (SD1), which can spontaneously form dimers was engineered with arginine residues; with a final sequence of QEIA**R**LEQEIA**R**LEYEIA**R**LE-am (Ryadnov, 2007). In the engineered peptide, the guanidinium group of arginine formed salt bridges with two glutamates, one of the same helix and the other belonging to a helix of another dimer. Increased dimer-dimer association led to the formation of large supramolecular networks. When used in conjunction with a branching trimeric peptide (SD2), SD1 formed a network of salt bridges with SD2 and the peptides arranged synergistically into ring shaped networks. Self-assembling peptides such as leucine zippers could be attached on proteins to achieve directed assembly of homo- and hetero-protein complexes.

**Table 1.1** A summary of possible methods to make synthetic complexes of non-interacting proteins are described. Advantages and disadvantages of each method are also listed.

| **Methods to make protein complexes** | **Advantages** | **Disadvantages** |
|---|---|---|
| Surface redesign with randomized amino acids | 1. Protein complexes with very low dissociation constants ($K_d$ ~ nM) could be made<br>2. Could be designed for any protein target with less / no cross-reactivity | 1. Requires screening of the binding partners amongst a large library (~$10^{10}$)<br>2. Requires burial of large surface area of interacting proteins |
| Non-polar surface burial | Requires little modification of protein surfaces | 1. Adding non polar residues to water- |

| | | |
|---|---|---|
| | | exposed protein surfaces might compromise protein stability<br>2. Difficult to make specific hetero-oligomers |
| Disulfide cross-linking | 1. Easy to introduce cysteines on protein surfaces<br>2. Site specific covalent attachment is possible<br>3. Protein complex formation is reversible by using reducing/ oxidizing agents | 1. Precise orientation of protein molecules could be difficult to achieve in each case<br>2. Homo-oligomers will always be a side-product unless the cysteine thiols are modified with linkers for hetero-oligomer formation only |
| Electrostatics | 1. Protein complexes could be formed with high affinity constants<br>2. Affinity can be modulated by external elements such as ionic strength of buffer, pH, etc. making protein complex formation a reversible process<br>3. Contaminating homo-oligomers are not formed | 1. Extensive redesign of protein surfaces with charged groups can affect protein function and stability<br>2. Cannot be generalized to all protein partners |
| Ligand mediated | 1. Requires no or little modification of protein surfaces<br>2. Depending on ligands used, affinity constants of protein complexes can be high | Cannot be generalized to all protein partners as not all proteins consist of ligand binding sites on their surface |
| Metal ion mediated | 1. Requires less modification of protein surfaces<br>2. Affinity can be modulated by external elements | 1. Difficult to make hetero-oligomer selectively<br>2. Even homo-oligomer formation cannot be generalized to all proteins |
| Streptavidin-biotin anchorage | Protein complexes form with the ~$10^{-14}$ M affinity | 1. Irreversible complex formation<br>2. Distance between proteins of interest becomes more than the |

| | | |
|---|---|---|
| | | width of streptavidin used as the connector |
| Genetic fusion | 1. In principle, any two or more proteins can be brought at predetermined distances<br>2. Proteins of interest are expressed in equimolar or desired molar ratios | 1. Not all proteins fold well when attached to another protein. Insoluble fusion protein expression is a huge drawback<br>2. Irreversible complex formation |

## Outlook

The principles of protein-protein interactions derived from natural protein complexes have been applied on combining two or more proteins that normally do not interact, into new proteins. The chemical diversity of protein surfaces offers limitless possibilities of site-specific functionalization. Protein surfaces have been chemically modified by mutagenesis of existing residues into other natural and unnatural amino acids (Fellouse *et al.*, 2005; Fellouse *et al.*, 2004; Liu and Schultz, 2010), or by placing small molecules such as ligands to bind other proteins (Radford and Tezcan, 2009). All protein engineering methods discussed in this chapter (Table 1.1) have their advantages and disadvantages, which weigh differently depending on the nature of the protein being used.

Applications of engineered protein complexes have been shown for therapeutics or drug delivery (Ballister *et al.*, 2008; Boersma and Pluckthun, 2011), sensing (Moll *et al.*, 2002; Moreau *et al.*, 2008), single molecule studies (Lorenz *et al.*, 2006; Miyawaki *et al.*, 1997) and synthesis of nanostructures

(Mao *et al.*, 2004; Padilla *et al.*, 2001). The use of protein components in building nanodevices has also been implicated because designed protein assemblies surpass the range and throughput of current fabrication techniques based on electron beam lithography and atomic force microscopy (Astier *et al.*, 2005; Moll *et al.*, 2002). Designed protein arrays have a direct use in catalysis, coatings, and biosensing (Yeates, 2011).

The proof of principle studies mentioned in the sections above demonstrated how proteins could be brought in proximity by different approaches, laying a foundation for construction of new materials, which exploit the properties of all the proteins involved.

# References


Akiva, P., Toporik, A., Edelheit, S., Peretz, Y., Diber, A., Shemesh, R., Novik, A., and Sorek, R. (2006). Transcription-mediated gene fusion in the human genome. Genome research *16*, 30-36.

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell *92*, 291-294.

Arnold, F.H., and Haymore, B.L. (1991). Engineered metal-binding proteins: purification to protein folding. Science *252*, 1796-1797.

Astier, Y., Bayley, H., and Howorka, S. (2005). Protein components for nanodevices. Current opinion in chemical biology *9*, 576-584.

Ballister, E.R., Lai, A.H., Zuckermann, R.N., Cheng, Y., and Mougous, J.D. (2008). In vitro self-assembly of tailorable nanotubes from a simple protein building block. Proceedings of the National Academy of Sciences of the United States of America *105*, 3733-3738.

Barondeau, D.P., and Getzoff, E.D. (2004). Structural insights into protein-metal ion partnerships. Current opinion in structural biology *14*, 765-774.

Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. Science *313*, 1642-1645.

Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grutter, M.G., and Pluckthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. Nature biotechnology *22*, 575-582.

Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P., and Pluckthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. Journal of molecular biology *332*, 489-503.

Bochtler, M., Ditzel, L., Groll, M., Hartmann, C., and Huber, R. (1999). The proteasome. Annual review of biophysics and biomolecular structure *28*, 295-317.

Boersma, Y.L., and Pluckthun, A. (2011). DARPins and other repeat protein scaffolds: advances in engineering and applications. Current opinion in biotechnology *22*, 849-857.

Bothner, B., Taylor, D., Jun, B., Lee, K.K., Siuzdak, G., Schultz, C.P., and Johnson, J.E. (2005). Maturation of a tetravirus capsid alters the dynamic properties and creates a metastable complex. Virology *334*, 17-27.

Bradbury, A.R., Sidhu, S., Dubel, S., and McCafferty, J. (2011). Beyond natural antibodies: the power of in vitro display technologies. Nature biotechnology *29*, 245-254.

Brodin, J.D., Ambroggio, X.I., Tang, C., Parent, K.N., Baker, T.S., and Tezcan, F.A. (2012). Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. Nature chemistry *4*, 375-382.

Buckle, A.M., Schreiber, G., and Fersht, A.R. (1994). Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. Biochemistry *33*, 8878-8889.

Burazerovic, S., Gradinaru, J., Pierron, J., and Ward, T.R. (2007). Hierarchical self-assembly of one-dimensional streptavidin bundles as a collagen mimetic for the biomineralization of calcite. Angew Chem Int Ed Engl *46*, 5510-5514.

Busch, S.J., and Sassone-Corsi, P. (1990). Dimers, leucine zippers and DNA-binding domains. Trends in genetics : TIG *6*, 36-40.

Callaway, A., Giesman-Cookmeyer, D., Gillock, E.T., Sit, T.L., and Lommel, S.A. (2001). The multifunctional capsid proteins of plant RNA viruses. Annual review of phytopathology *39*, 419-460.

Caro, L.N., Moreau, C.J., Estrada-Mondragon, A., Ernst, O.P., and Vivaudou, M. (2012). Engineering of an artificial light-modulated potassium channel. PloS one *7*, e43766.

Carroll, D. (2011). Genome engineering with zinc-finger nucleases. Genetics *188*, 773-782.

Chatterji, A., Ochoa, W., Shamieh, L., Salakian, S.P., Wong, S.M., Clinton, G., Ghosh, P., Lin, T., and Johnson, J.E. (2004). Chemical conjugation of heterologous proteins on the surface of Cowpea mosaic virus. Bioconjugate chemistry *15*, 807-813.

Chothia, C. (1975). Structural invariants in protein folding. Nature *254*, 304-308.

Davis, R.J., and Barr, F.G. (1997). Fusion genes resulting from alternative chromosomal translocations are overexpressed by gene-specific mechanisms in alveolar rhabdomyosarcoma. Proceedings of the National Academy of Sciences of the United States of America *94*, 8047-8051.

De Meyts, P., and Whittaker, J. (2002). Structural biology of insulin and IGF1 receptors: implications for drug design. Nature reviews Drug discovery *1*, 769-783.

Dong, F., and Zhou, H.X. (2006). Electrostatic contribution to the binding stability of protein-protein complexes. Proteins *65*, 87-102.

Dueber, J.E., Yeh, B.J., Chak, K., and Lim, W.A. (2003). Reprogramming control of an allosteric signaling switch through modular recombination. Science *301*, 1904-1908.

Emsley, J., Knight, C.G., Farndale, R.W., Barnes, M.J., and Liddington, R.C. (2000). Structural basis of collagen recognition by integrin alpha2beta1. Cell *101*, 47-56.

Fellouse, F.A., Li, B., Compaan, D.M., Peden, A.A., Hymowitz, S.G., and Sidhu, S.S. (2005). Molecular recognition by a binary code. Journal of molecular biology *348*, 1153-1162.

Fellouse, F.A., Wiesmann, C., and Sidhu, S.S. (2004). Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. Proceedings of the National Academy of Sciences of the United States of America *101*, 12467-12472.

Fiedler, J.D., Brown, S.D., Lau, J.L., and Finn, M.G. (2010). RNA-directed packaging of enzymes within virus-like particles. Angew Chem Int Ed Engl *49*, 9648-9651.

Finley, D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. Annual review of biochemistry *78*, 477-513.

Gaj, T., Guo, J., Kato, Y., Sirk, S.J., and Barbas, C.F., 3rd (2012). Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. Nature methods *9*, 805-807.

Gao, M., and Skolnick, J. (2012). The distribution of ligand-binding pockets around protein-protein interfaces suggests a general mechanism for pocket formation. Proceedings of the National Academy of Sciences of the United States of America *109*, 3784-3789.

Gebauer, M., and Skerra, A. (2009). Engineered protein scaffolds as next-generation antibody therapeutics. Current opinion in chemical biology *13*, 245-255.

Ghadiri, M.R., Granja, J.R., and Buehler, L.K. (1994). Artificial transmembrane ion channels from self-assembling peptide nanotubes. Nature *369*, 301-304.

Ghadiri, M.R., Granja, J.R., Milligan, R.A., McRee, D.E., and Khazanovich, N. (1993). Self-assembling organic nanotubes based on a cyclic peptide architecture. Nature *366*, 324-327.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS letters *582*, 1977-1986.

Glover, B.P., and McHenry, C.S. (2001). The DNA polymerase III holoenzyme: an asymmetric dimeric replicative complex with leading and lagging strand polymerases. Cell *105*, 925-934.

Griffiths, D.J. (1999). Introduction to electrodynamics, 3rd edn (Upper Saddle River, NJ ; London: Prentice Hall).

Grueninger, D., Treiber, N., Ziegler, M.O., Koetter, J.W., Schulze, M.S., and Schulz, G.E. (2008). Designed protein-protein association. Science *319*, 206-209.

Hess, H., Clemmens, J., Brunner, C., Doot, R., Luna, S., Ernst, K.-H., and Vogel, V. (2005). Molecular self-assembly of "nanowires" and "nanospools" using active transport. Nano Lett *5*, 629-633.

Janetopoulos, C., Jin, T., and Devreotes, P. (2001). Receptor-mediated activation of heterotrimeric G-proteins in living cells. Science *291*, 2408-2411.

Johnson, J.E., and Speir, J.A. (1997). Quasi-equivalent viruses: a paradigm for protein assemblies. Journal of molecular biology *269*, 665-675.

Jones, S., and Thornton, J.M. (1996). Principles of protein-protein interactions. Proceedings of the National Academy of Sciences of the United States of America *93*, 13-20.

Kadonaga, J.T. (1998). Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. Cell *92*, 307-313.

Kelman, Z., and O'Donnell, M. (1995). DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. Annual review of biochemistry *64*, 171-200.

Kerfeld, C.A., Sawaya, M.R., Tanaka, S., Nguyen, C.V., Phillips, M., Beeby, M., and Yeates, T.O. (2005). Protein structures forming the shell of primitive bacterial organelles. Science *309*, 936-938.

Keskin, O., Gursoy, A., Ma, B., and Nussinov, R. (2008). Principles of protein-protein interactions: what are the preferred ways for proteins to interact? Chemical reviews *108*, 1225-1244.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences of the United States of America *93*, 1156-1160.

Kitov, P.I., Lipinski, T., Paszkiewicz, E., Solomon, D., Sadowska, J.M., Grant, G.A., Mulvey, G.L., Kitova, E.N., Klassen, J.S., Armstrong, G.D*., et al.* (2008). An entropically efficient supramolecular inhibition strategy for Shiga toxins. Angew Chem Int Ed Engl *47*, 672-676.

Kitov, P.I., Sadowska, J.M., Mulvey, G., Armstrong, G.D., Ling, H., Pannu, N.S., Read, R.J., and Bundle, D.R. (2000). Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. Nature *403*, 669-672.

Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., and Ban, N. (2011). Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. Science *334*, 941-948.

Koide, A., Gilbreth, R.N., Esaki, K., Tereshko, V., and Koide, S. (2007). High-affinity single-domain binding proteins with a binary-code interface. Proceedings of the National Academy of Sciences of the United States of America *104*, 6632-6637.

Koide, S., and Sidhu, S.S. (2009). The importance of being tyrosine: lessons in molecular recognition from minimalist synthetic binding proteins. ACS chemical biology *4*, 325-334.

Konecny, R., Trylska, J., Tama, F., Zhang, D., Baker, N.A., Brooks, C.L., 3rd, and McCammon, J.A. (2006). Electrostatic properties of cowpea chlorotic mottle virus and cucumber mosaic virus capsids. Biopolymers *82*, 106-120.

Kovall, R., and Matthews, B.W. (1997). Toroidal structure of l-exonuclease. Science *277*, 1824-1827.

Kukic, P., and Nielsen, J.E. (2010). Electrostatics in proteins and protein-ligand complexes. Future medicinal chemistry *2*, 647-666.

Lander, G.C., Estrin, E., Matyskiela, M.E., Bashore, C., Nogales, E., and Martin, A. (2012). Complete subunit architecture of the proteasome regulatory particle. Nature *482*, 186-191.

Lee, B., and Vasmatzis, G. (1997). Stabilization of protein structures. Current opinion in biotechnology *8*, 423-428.

Li, H., Haurigot, V., Doyon, Y., Li, T., Wong, S.Y., Bhagwat, A.S., Malani, N., Anguela, X.M., Sharma, R., Ivanciu, L.*, et al.* (2011). In vivo genome editing restores haemostasis in a mouse model of haemophilia. Nature *475*, 217-221.

Liepold, L.O., Revis, J., Allen, M., Oltrogge, L., Young, M., and Douglas, T. (2005). Structural transitions in Cowpea chlorotic mottle virus (CCMV). Physical biology *2*, S166-172.

Lim, K.H., Huang, H., Pralle, A., and Park, S. (2011). Engineered streptavidin monomer and dimer with improved stability and function. Biochemistry *50*, 8682-8691.

Liu, C.C., and Schultz, P.G. (2010). Adding new chemistries to the genetic code. Annual review of biochemistry *79*, 413-444.

Lo Conte, L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. Journal of molecular biology *285*, 2177-2198.

Lorenz, H., Hailey, D.W., and Lippincott-Schwartz, J. (2006). Fluorescence protease protection of GFP chimeras to reveal protein topology and subcellular localization. Nature methods *3*, 205-210.

Maeda, S., Nakagawa, S., Suga, M., Yamashita, E., Oshima, A., Fujiyoshi, Y., and Tsukihara, T. (2009). Structure of the connexin 26 gap junction channel at 3.5 A resolution. Nature *458*, 597-602.

Mahajan, N.P., Linder, K., Berry, G., Gordon, G.W., Heim, R., and Herman, B. (1998). Bcl-2 and Bax interactions in mitochondria probed with green fluorescent protein and fluorescence resonance energy transfer. Nature biotechnology *16*, 547-552.

Mannige, R.V., and Brooks, C.L., 3rd (2010). Periodic table of virus capsids: implications for natural selection and design. PloS one *5*, e9423.

Mao, C., Solis, D.J., Reiss, B.D., Kottmann, S.T., Sweeney, R.Y., Hayhurst, A., Georgiou, G., Iverson, B., and Belcher, A.M. (2004). Virus-based toolkit for the directed synthesis of magnetic and semiconducting nanowires. Science *303*, 213-217.

Matsuura, K., Murasato, K., and Kimizuka, N. (2005). Artificial peptide-nanospheres self-assembled from three-way junctions of beta-sheet-forming peptides. Journal of the American Chemical Society *127*, 10148-10149.

Matsuura, K., Yamashita, T., Igami, Y., and Kimizuka, N. (2003). 'Nucleo-nanocages': designed ternary oligodeoxyribonucleotides spontaneously form nanosized DNA cages. Chem Commun (Camb), 376-377.

Meyer, C., Brieger, A., Plotz, G., Weber, N., Passmann, S., Dingermann, T., Zeuzem, S., Trojan, J., and Marschalek, R. (2009). An interstitial deletion at 3p21.3 results in the genetic fusion of MLH1 and ITGA9 in a Lynch syndrome family. Clinical cancer research : an official journal of the American Association for Cancer Research *15*, 762-769.

Minor, D.L., Jr., and Kim, P.S. (1994). Measurement of the beta-sheet-forming propensities of amino acids. Nature *367*, 660-663.

Miranda, F.F., Iwasaki, K., Akashi, S., Sumitomo, K., Kobayashi, M., Yamashita, I., Tame, J.R., and Heddle, J.G. (2009). A self-assembled protein nanotube with high aspect ratio. Small *5*, 2077-2084.

Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. Nature reviews Cancer *7*, 233-245.

Mitome, N., Suzuki, T., Hayashi, S., and Yoshida, M. (2004). Thermophilic ATP synthase has a decamer c-ring: indication of noninteger 10:3 H+/ATP ratio and permissive elastic coupling. Proceedings of the National Academy of Sciences of the United States of America *101*, 12159-12164.

Miyawaki, A., Llopis, J., Heim, R., McCaffery, J.M., Adams, J.A., Ikura, M., and Tsien, R.Y. (1997). Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. Nature *388*, 882-887.

Moll, D., Huber, C., Schlegel, B., Pum, D., Sleytr, U.B., and Sára, M. (2002). S-layer-streptavidin fusion proteins as template for nanopatterned molecular arrays. ProcNatlAcadSciUSA *99*, 14646-14651.

Moreau, C.J., Dupuis, J.P., Revilloud, J., Arumugam, K., and Vivaudou, M. (2008). Coupling ion channels to receptors for biomolecule sensing. Nature nanotechnology *3*, 620-625.

Murata, S., Yashiroda, H., and Tanaka, K. (2009). Molecular mechanisms of proteasome assembly. Nature reviews Molecular cell biology *10*, 104-115.

Nguyen, A.W., and Daugherty, P.S. (2005). Evolutionary optimization of fluorescent proteins for intracellular FRET. Nature biotechnology *23*, 355-360.

Nooren, I.M., and Thornton, J.M. (2003a). Diversity of protein-protein interactions. The EMBO journal *22*, 3486-3492.

Nooren, I.M., and Thornton, J.M. (2003b). Structural characterisation and functional significance of transient protein-protein interactions. Journal of molecular biology *325*, 991-1018.

Orentas, R.J., Lee, D.W., and Mackall, C. (2012). Immunotherapy targets in pediatric cancer. Frontiers in oncology *2*, 3.

Pabo, C.O., and Sauer, R.T. (1984). Protein-DNA recognition. Annual review of biochemistry *53*, 293-321.

Padilla, J.E., Colovos, C., and Yeates, T.O. (2001). Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proceedings of the National Academy of Sciences of the United States of America *98*, 2217-2221.

Parizek, P., Kummer, L., Rube, P., Prinz, A., Herberg, F.W., and Pluckthun, A. (2012). Designed Ankyrin Repeat Proteins (DARPins) as Novel Isoform-Specific Intracellular Inhibitors of c-Jun N-Terminal Kinases. ACS chemical biology *7*, 1356-1366.

Pecqueur, L., Duellberg, C., Dreier, B., Jiang, Q., Wang, C., Pluckthun, A., Surrey, T., Gigant, B., and Knossow, M. (2012). A designed ankyrin repeat protein selected to bind to tubulin caps the microtubule plus end. Proceedings of the National Academy of Sciences of the United States of America *109*, 12011-12016.

Pokkuluri, P.R., Raffen, R., Dieckman, L., Boogaard, C., Stevens, F.J., and Schiffer, M. (2002). Increasing protein stability by polar surface residues: domain-wide consequences of interactions within a loop. Biophysical journal *82*, 391-398.

Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A., and Ban, N. (2011). Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. Science *331*, 730-736.

Radford, R.J., Brodin, J.D., Salgado, E.N., and Tezcan, F.A. (2011). Expanding the utility of proteins as platforms for coordination chemistry. Coordin Chem Rev *255*, 790-803.

Radford, R.J., and Tezcan, F.A. (2009). A superprotein triangle driven by nickel(II) coordination: exploiting non-natural metal ligands in protein self-assembly. Journal of the American Chemical Society *131*, 9136-9137.

Reches, M., and Gazit, E. (2007). Biological and chemical decoration of peptide nanostructures via biotin-avidin interactions. Journal of nanoscience and nanotechnology *7*, 2239-2245.

Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H., and Schreiber, G. (2007). The molecular architecture of protein-protein binding sites. Current opinion in structural biology *17*, 67-76.

Rhee, J.K., Hovlid, M., Fiedler, J.D., Brown, S.D., Manzenrieder, F., Kitagishi, H., Nycholat, C., Paulson, J.C., and Finn, M.G. (2011). Colorful virus-like particles: fluorescent protein packaging by the Qbeta capsid. Biomacromolecules *12*, 3977-3981.

Ringler, P., and Schulz, G.E. (2003). Self-assembly of proteins into designed networks. Science *302*, 106-109.

Rizzo, M.A., Springer, G.H., Granada, B., and Piston, D.W. (2004). An improved cyan fluorescent protein variant useful for FRET. Nature biotechnology *22*, 445-449.

Ryadnov, M.G. (2007). A self-assembling peptide polynanoreactor. Angew Chem Int Ed Engl *46*, 969-972.

Salgado, E.N., Faraone-Mennella, J., and Tezcan, F.A. (2007). Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. Journal of the American Chemical Society *129*, 13374-13375.

Salgado, E.N., Lewis, R.A., Faraone-Mennella, J., and Tezcan, F.A. (2008). Metal-mediated self-assembly of protein superstructures: influence of secondary interactions on protein oligomerization and aggregation. Journal of the American Chemical Society *130*, 6082-6084.

Salgado, E.N., Lewis, R.A., Mossin, S., Rheingold, A.L., and Tezcan, F.A. (2009). Control of protein oligomerization symmetry by metal coordination: C2 and C3 symmetrical assemblies through Cu(II) and Ni(II) coordination. Inorganic chemistry *48*, 2726-2728.

Salgado, E.N., Radford, R.J., and Tezcan, F.A. (2010). Metal-directed protein self-assembly. Accounts of chemical research *43*, 661-672.

Santo, E.E., Ebus, M.E., Koster, J., Schulte, J.H., Lakeman, A., van Sluis, P., Vermeulen, J., Gisselsson, D., Ora, I., Lindner, S.*, et al.* (2012). Oncogenic activation of FOXR1 by 11q23 intrachromosomal deletion-fusions in neuroblastoma. Oncogene *31*, 1571-1581.

Sara, M., and Sleytr, U.B. (2000). S-Layer proteins. Journal of bacteriology *182*, 859-868.

Seebeck, F.P., Woycechowsky, K.J., Zhuang, W., Rabe, J.P., and Hilvert, D. (2006). A simple tagging system for protein encapsulation. Journal of the American Chemical Society *128*, 4516-4517.

Sheinerman, F.B., and Honig, B. (2002). On the role of electrostatic interactions in the design of protein-protein interfaces. Journal of molecular biology *318*, 161-177.

Shepherd, C.M., Borelli, I.A., Lander, G., Natarajan, P., Siddavanahalli, V., Bajaj, C., Johnson, J.E., Brooks, C.L., 3rd, and Reddy, V.S. (2006). VIPERdb: a relational database for structural virology. Nucleic acids research *34*, D386-389.

Sinclair, J.C., Davies, K.M., Venien-Bryan, C., and Noble, M.E. (2011). Generation of protein lattices by fusing proteins with matching rotational symmetry. Nature nanotechnology *6*, 558-562.

Smalle, J., and Vierstra, R.D. (2004). The ubiquitin 26S proteasome proteolytic pathway. Annual review of plant biology *55*, 555-590.

Solomon, D., Kitov, P.I., Paszkiewicz, E., Grant, G.A., Sadowska, J.M., and Bundle, D.R. (2005). Heterobifunctional multivalent inhibitor-adaptor mediates specific aggregation between Shiga toxin and a pentraxin. Organic letters *7*, 4369-4372.

Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., and Gouaux, J.E. (1996). Structure of staphylococcal a-hemolysin, a heptameric transmembrane pore. Science *274*, 1859-1865.

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

Staerz, U.D., Kanagawa, O., and Bevan, M.J. (1985). Hybrid antibodies can target sites for attack by T cells. Nature *314*, 628-631.

Stock, D., Leslie, A.G., and Walker, J.E. (1999a). Molecular architecture of the rotary motor in ATP synthase. Science *286*, 1700-1705.

Stock, D., Leslie, A.G.W., and Walker, J.E. (1999b). Molecular architecture of the rotary motor in ATP synthase. Science *286*, 1700-1705.

Sutter, M., Boehringer, D., Gutmann, S., Gunther, S., Prangishvili, D., Loessner, M.J., Stetter, K.O., Weber-Ban, E., and Ban, N. (2008). Structural basis of enzyme encapsulation into a bacterial nanocompartment. Nature structural & molecular biology *15*, 939-947.

Suttle, C.A. (2005). Viruses in the sea. Nature *437*, 356-361.

Swain, M., Thirupathi, R., Krishnarjuna, B., Eaton, E.M., Kibbey, M.M., Rosenzweig, S.A., and Atreya, H.S. (2010). Spontaneous and reversible self-assembly of a polypeptide fragment of insulin-like growth factor binding protein-2 into fluorescent nanotubular structures. Chem Commun (Camb) *46*, 216-218.

Tomashek, J.J., and Brusilow, W.S. (2000). Stoichiometry of energy coupling by proton-translocating ATPases: a history of variability. Journal of bioenergetics and biomembranes *32*, 493-500.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R.*, et al.* (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science *310*, 644-648.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575-581.

Villanueva, M.T. (2012). Genetics: Gene fusion power. Nature reviews Clinical oncology *9*, 188.

Voet, D., and Voet, J.G. (2011). Biochemistry, 4th edn (Hoboken, NJ: John Wiley & Sons).

Völlenkle, C., Weigert, S., Ilk, N., Egelseer, E., Weber, V., Loth, F., Falkenhagen, D., Sleytr, U.B., and Sára, M. (2004). Construction of a functional S-layer fusion protein comprising an immunoglobulin G-binding domain for development of specific adsorbents for extracorporeal blood purification. ApplEnvironMicrobiol *70*, 1514-1521.

Worsdorfer, B., Pianowski, Z., and Hilvert, D. (2012). Efficient in vitro encapsulation of protein cargo by an engineered protein container. Journal of the American Chemical Society *134*, 909-911.

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

Wu, S.C., and Wong, S.L. (2005). Engineering soluble monomeric streptavidin with reversible biotin binding capability. The Journal of biological chemistry *280*, 23225-23231.

Yan, Y., and Marriott, G. (2003). Analysis of protein interactions using fluorescence technologies. Current opinion in chemical biology *7*, 635-640.

Ye, J.D., Tereshko, V., Frederiksen, J.K., Koide, A., Fellouse, F.A., Sidhu, S.S., Koide, S., Kossiakoff, A.A., and Piccirilli, J.A. (2008). Synthetic antibodies for specific recognition and crystallization of structured RNA. Proceedings of the National Academy of Sciences of the United States of America *105*, 82-87.

Yeates, T.O. (2011). Nanobiotechnology: protein arrays made to order. Nature nanotechnology *6*, 541-542.

Yoshida, M., Muneyuki, E., and Hisabori, T. (2001). ATP synthase--a marvellous rotary engine of the cell. Nature reviews Molecular cell biology *2*, 669-677.

Zhang, J., McCabe, K.A., and Bell, C.E. (2011). Crystal structures of lambda exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. Proceedings of the National Academy of Sciences of the United States of America *108*, 11872-11877.

Zhang, S., Marini, D.M., Hwang, W., and Santoso, S. (2002). Design of nanostructured biological materials through self-assembly of peptides and proteins. Current opinion in chemical biology *6*, 865-871.

Zhou, T., Hamer, D.H., Hendrickson, W.A., Sattentau, Q.J., and Kwong, P.D. (2005). Interfacial metal and antibody recognition. Proceedings of the National Academy of Sciences of the United States of America *102*, 14575-14580.

# CHAPTER 2

## ENGINEERING A DIMERIC α-HEMOLYSIN PORE THAT SPANS PROXIMAL BILAYERS

## Introduction

### 2.1. Constructing minimal cells/tissues

Nature allows cells to take up nutrients from the environment and interact with other cells through an intense network of proteins embedded in their membranes (Voet and Voet, 2011). Nanometer-sized vesicles encased in lipids (water-in-water, w/w system) have been modelled as minimal cells/protocells if they meet functional criteria of a cell, (Kurihara *et al.*, 2011; Oberholzer *et al.*, 1995; Szostak *et al.*, 2001). One of the most important requirements of a cell is its ability to communicate with its external environment. The Bayley laboratory has suggested that aqueous droplets connected through droplet interface bilayers (Holden *et al.*, 2007) as functional compartmented assemblies, and significant progress has been made in this area (Maglia *et al.*, 2009a; Villar *et al.*, 2011). To achieve transfer of small analytes from the bulk exterior to the enclosed interior of a single lipid bilayer demarcated aqueous container (Noireaux and Libchaber, 2004; Villar *et al.*, 2011), stimulus responsive membranes (Cisse *et al.*, 2007b; Lvov *et al.*, 2001) and pores such as α-Hemolysin (αHL) (Ishitsuka *et al.*, 2010; Okumus *et al.*, 2009) (Fig. 2.1) have been used. By modulating the αHL protein density in the aqueous vesicles, controlled translocation of analytes can be achieved (Cisse *et al.*, 2007a). The

αHL pore can also allow compartments in droplet networks to communicate chemically and electrically with each other and the entire assembly to communicate with the external environment (Villar *et al.*, 2011).

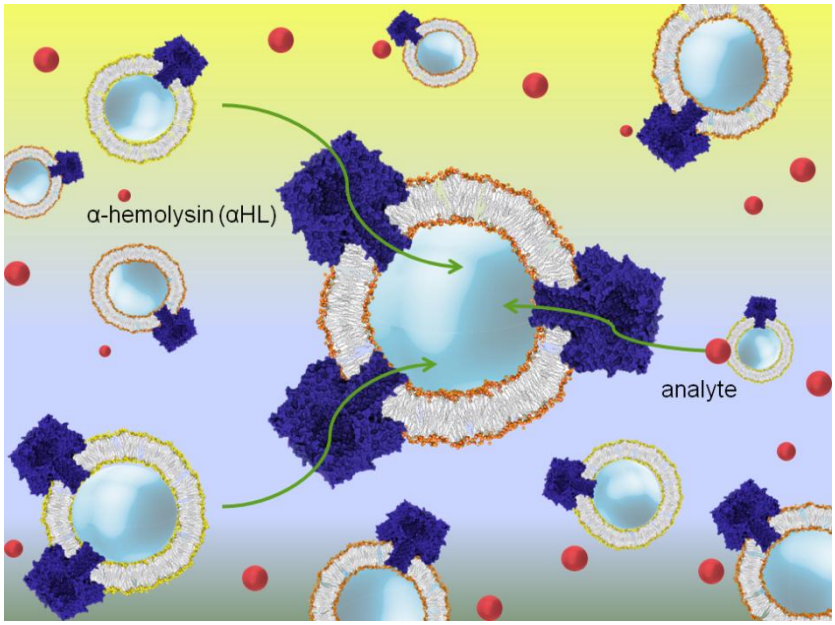



**Figure 2.1. Vesicles as protocells.** Small aqueous vesicles encased in lipids, such as liposomes, have been modelled as minimal cells or protocells. An essential requirement for a protocell is communication with the environment for which pores such as αHL have been used (Noireaux and Libchaber, 2004).

Far less work has been done on constructing minimal tissues (or prototissues), i.e. assemblies of small aqueous compartments with tissue-like properties (Woolfson and Bromley, 2011). A higher-order assembly of protocells could be used to model functional prototissues (Fig. 2.2). A key requirement of functional prototissues is intercellular communication – chemical and/or electrical. It would thus be advantageous to be able to connect vesicles or networks through their encapsulating bilayers. Connecting proximal bilayers would require a protein that spans two bilayers (Figs. 2.2, 2.3c).

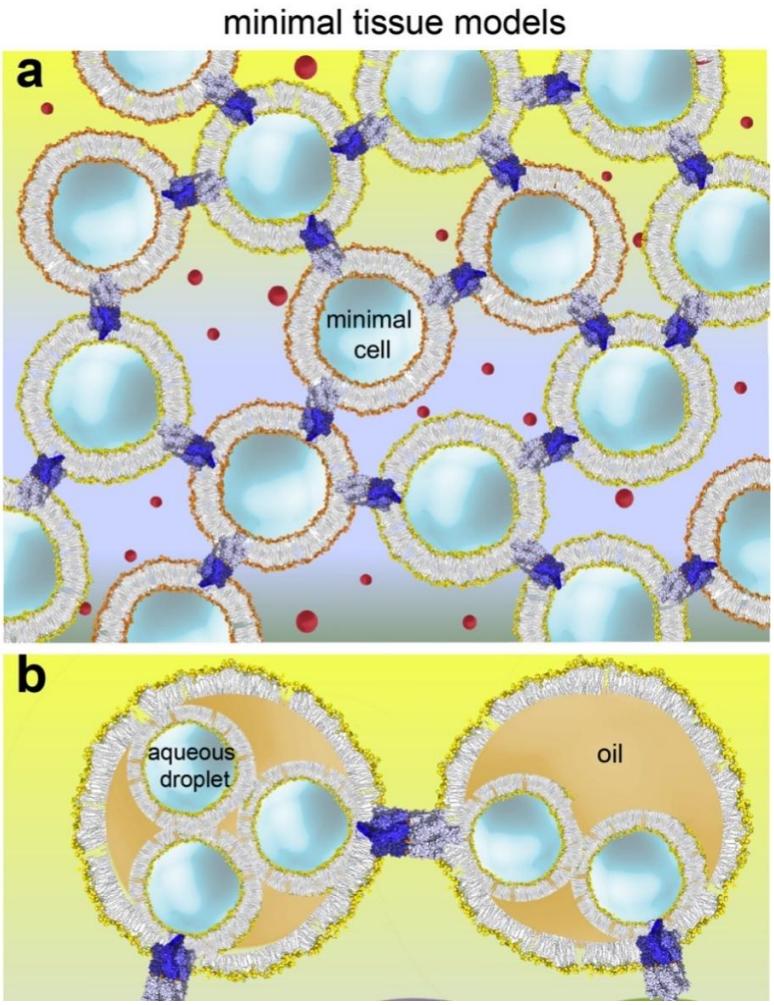



**Figure 2.2. Minimal tissue models.** (**a**) Vesicles made porous to small molecules by the use of transmembrane pores such as αHL have been modelled as protocells. In order to build a functional prototissue, communication between constituent protocells could be achieved with protein pores that insert in two proximal bilayers. (**a**) Minimal cells communicating through protein conduits that span both bilayers. (**b**) Interconnected monolayer-encased droplet networks (Villar *et al.*, 2011).

Double bilayers are a recurrent feature in Nature, and occur in organelles such as mitochondria, chloroplasts and other plastids, autophagosomes and cell nuclei. For example, the outer and inner membranes of the nucleus are separated by 20 – 40 nm (Bupp *et al.*, 2007). Cargo of small molecules and proteins are ferried across the double membranes of organelles through transiently interacting proteins inserted in each of the membranes, e.g., the TIM/TOM complex in the mitochondria (Pfanner and Meijer, 1997), or through

permanently associated protein complexes interconnecting the two bilayers, like the nucleopore complex (D'Angelo and Hetzer, 2008). A notable example of a protein connecting two bilayers is of gap junctions, where the interdigitating extracellular loops of hemi-channels or connexons in individual bilayers are connected by an intense network of hydrogen bonds and salt bridges to form full channels.

A bottom-up engineering approach to synthetic biology demands dissecting naturally occurring processes and mimicking them (Schwille, 2011). Such an approach has been taken to rationally design a protein element to physically connect two proximal bilayers (analogous to active communication in a biological tissue) of the constituent protocells (Fig. 2.2).

## 2.2. Engineering a protein that can insert in two adjacent bilayers

As the building block for engineering a protein that can insert in two bilayers simultaneously, the well studied αHL pore was used (Bhakdi *et al.*, 1996; Song *et al.*, 1996; Walker *et al.*, 1992). The folding and structure of αHL has been studied in atomic detail (Song *et al.*, 1996; Valeva *et al.*, 1996). αHL has been a subject of extensive mutagenesis studies. αHL is also a remarkably stable molecule, easy to express and robust to mutagenesis. It is resistant to the action of proteases (Walker *et al.*, 1992), stable in high concentrations of urea (Japrung *et al.*, 2010), SDS (Bhakdi *et al.*, 1981) and a wide pH range (Maglia *et al.*, 2009b). The αHL pore is functional even at temperatures close to the boiling point of water (Kang *et al.*, 2005). These properties make the αHL pore an ideal scaffold for protein engineering.

### 2.2.1. Structure of αHL

αHL is one of the several β-barrel pore-forming toxins secreted by the *Staphylococcus aureus.* The water soluble monomers of αHL (α1; 33.2 kDa) bind to susceptible membranes and assemble to form membrane permeating heptamers (α7) (Bhakdi and Tranum-Jensen, 1991; Thelestam and Blomqvist, 1988). The crystal structure of the detergent solubilized α7 pore revealed a mushroom shaped molecule, ~10 nm in height with an ~8.5 nm broad globular, extracellular cap with a water filled vestibule and an ~2.5 nm wide membrane permeating β barrel (Song *et al.*, 1996) (Fig. 2.3a, b). The region between the cap domain and the β barrel in α7 (rim region) is composed of mainly aromatic residues and also contains a phosphocholine binding site (Galdiero, 2004). This helps to anchor α7 molecules vertically on lipid bilayers.

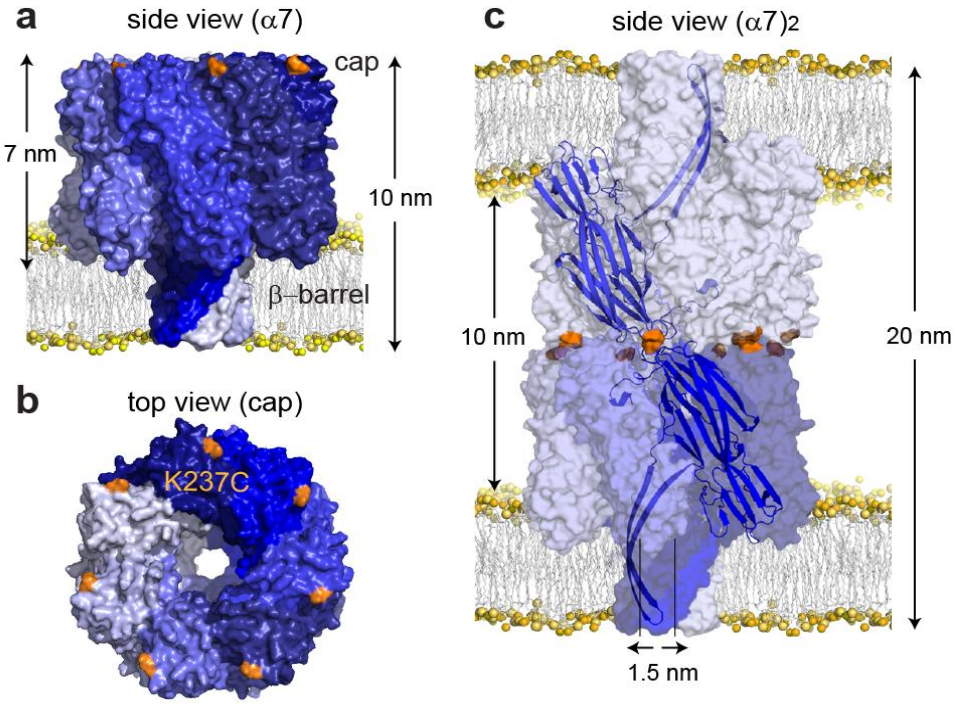



**Figure 2.3. Engineering the αHL heptamer dimer pore.** (**a**) Space filling model of the mushroom shaped α7 pore inserted in a bilayer. Each monomer is shown in a different shade of blue. (**b**) Top-view of α7 shows the central pore and a flat cap surface. To engineer a pore that could span two bilayers

simultaneously, α7 units were linked cap-to-cap *via* disulfide bonds between cysteine residues (K237C, orange) on the cap of each monomer (α1) subunit. (**c**) Space filling model of the double bilayer spanning dimer pore, $(\alpha7)_2$, of α7. Models of α7 (PDB ID – 7AHL) and $(\alpha7)_2$ were created in PyMOL.

### 2.2.2. Single αHL molecules can be electrically characterised: applications of αHL

The vectorial insertion of α7 in *in vitro* formed lipid bilayers (Fang *et al.*, 1997; Krasilnikov *et al.*, 2000) permits straightforward electrical measurements of the nanopore. Translocation of small molecules and large biological macromolecules such as peptides, DNA and RNA through the α7 pore can be detected as a blockade in the electric current passing through it (Akeson *et al.*, 1999; Gu *et al.*, 1999; Meller *et al.*, 2000). This has allowed the study of single molecule kinetics, stochastic sensing, and DNA sequencing using the α7 pore (Bayley and Cremer, 2001; Bayley and Jayasinghe, 2004; Braha *et al.*, 2000; Cherf *et al.*, 2012; Hammerstein *et al.*, 2010; Lu *et al.*, 2010).

αHL adapters such as cyclodextrins (CD) can lodge within the α7 β-barrel for several milliseconds to seconds (Gu *et al.*, 1999; Gu *et al.*, 2001) (Fig. 2.4a). As the CD bound pore displays a lower conductance than the open α7 pore, CD binding and unbinding to single α7 pores can be detected electrically as transient blockades in $I_o$ (Gu *et al.*, 1999) (Fig. 2.4b). Small organic molecules that can be trapped in the internal hydrophobic cavity of CD can also be detected using the α7 pore (Gu *et al.*, 1999). The binding-unbinding kinetics can be used to determine the nature of the analyte (Fig. 2.4c).

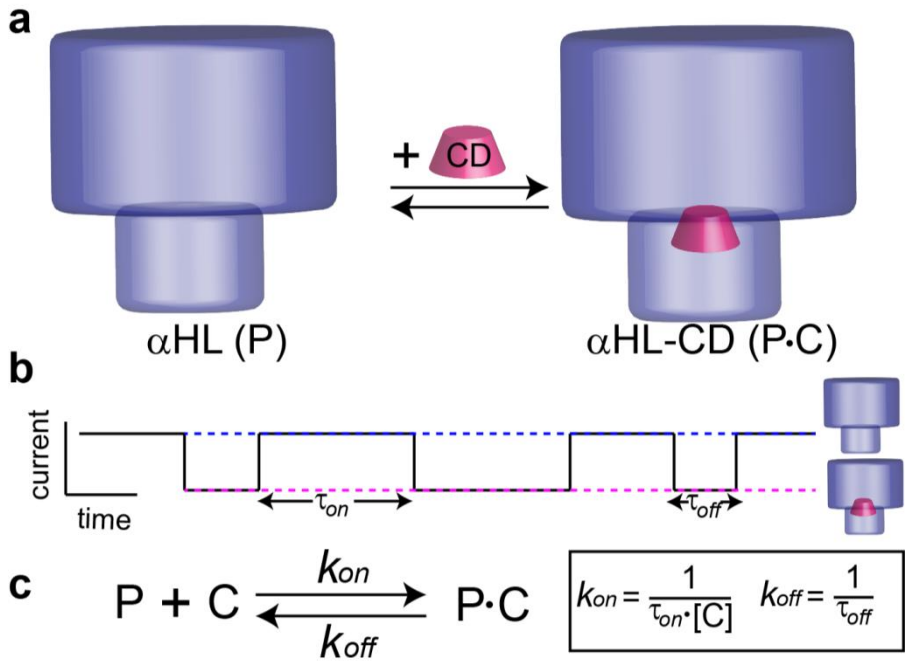



**Figure 2.4. CD binding to α7.** (**a**) Cartoon showing CD (pink) can bind near the constriction of the β-barrel of α7 (blue) (Gu *et al.*, 1999). (**b**) CD binding and unbinding is seen as regular blockades in the open pore current of α7. From single channel current traces, the mean dwell time of CD within the pore ($\tau_{off}$) and the frequency of the binding ($\tau_{on}$) can be deduced. (**c**) The values of $\tau_{off}$ and $\tau_{on}$ can be used to determine the kinetic rate constants, $k_{on}$ and $k_{off}$, of CD (C) binding to the pore (P).

The α7 pore fitted with heptakis-(6-deoxy-6-amino)-β-cyclodextrin ($am_7$βCD) can also differentiate between all the DNA and RNA nucleotides (Astier *et al.*, 2006). The β barrel of α7 contains three recognition sites for nucleobases of ssDNA or RNA strands stretched within the α7 pore (Stoddart *et al.*, 2009). This has led to development of the nanopore technology for DNA sequencing (Cherf *et al.*, 2012; Clarke *et al.*, 2009). Methods to attach an exonuclease to the α7 pore for base-by-base cleavage and recognition are currently being sought (Chapter 3).

### 2.2.3. αHL forms an ideal scaffold for protein engineering

αHL has also been a subject of extensive mutagenesis and protein engineering

studies (Cheley *et al.*, 1997; Jayasinghe *et al.*, 2006; Walker *et al.*, 1995). Engineered variants of α7 have been designed to possess unique properties such as size selection, rectification, triggered activity, reversible gating and blocker binding (Astier *et al.*, 2005; Bayley and Cremer, 2001; Cheley *et al.*, 2002; Gu *et al.*, 1999; Maglia *et al.*, 2009a). A photo-switchable pore was created by mutagenesis of a vestibule residue R104 to a cysteine and functionalizing it with a caging molecule (2-bromo-2-(nitrophenyl)acetic acid) (Chang *et al.*, 1995). The derivatized α1 could assemble into heptameric prepores (inactive state) but not lyse rabbit erythrocytes. Upon irradiation at 300 nm, the caging molecules dissociated restoring the pore forming activity of αHL.

Like the R104 residue, several other charged and neutral residues can also be modified to cysteine without loss in activity (Walker and Bayley, 1995). Cysteine mutants of the β barrel of α7 have been employed for studying single molecule chemical kinetics (Bayley *et al.*, 2008). Derivatized cysteine mutants of the αHL lumen and cap surface have also been made (Howorka and Bayley, 2002; Howorka *et al.*, 2000; Howorka *et al.*, 2004; Rotem *et al.*, 2012). The natural carbohydrate ligand of the lectin from *Bauhinia purpurea* was tethered to the T9C residue, located near the αHL pore opening (Howorka *et al.*, 2004). Using heteromeric pores with only one or two engineered α1 units, the kinetics of monovalent and bivalent binding between lectin molecules in solution and the ligands on αHL could be probed. In another example, ssDNA and aptamers were covalently linked to the N17C residue, which is located at the edge of the αHL cap pore opening (Howorka and Bayley, 2002; Rotem *et al.*, 2012). Thrombin could form a cation-stabilized quadruplex with these aptamers and its

binding could be detected at the single molecule level. These studies have demonstrated that many residues on the αHL cap could be mutated to cysteines and linked to molecules such as ligands, DNA and PEG. Extending this approach, proteins could also be linked to the αHL cap, thereby increasing the nanopore's functionality.

### 2.2.4. Designing a double bilayer inserting pore using αHL

The design blueprint of a pore that could insert in two bilayers was the hexameric connexon protein, which forms gap junctions between two adjacent biological cells (Evans and Martin, 2002; Maeda *et al.*, 2009; Nakagawa *et al.*, 2010). Advantageously, the flat cap surface of αHL offered the possibility of connecting two α7 units cap-to-cap, resulting in a dimer complex, $(α7)_2$, with the lipid bilayer inserting β-barrels facing in opposite directions (Fig. 2.3c). Hexameric Hcp1 pores (*Pseudomonas aeruginosa*), which have a flat cap surface like αHL, could be linked head-to-tail with cysteine point mutations on the top and the bottom surfaces of the Hcp1 monomers (Ballister *et al.*, 2008). As αHL doesn't contain any native cysteines, we surmised that the caps of two α7 pores could also be linked using thiol-cross linking.

# Results and Discussion

## 2.3. Design rationale of the $(α7)_2$ pore

To covalently link two α7 units to form the $(α7)_2$ pore, the surface exposed cap residue K237 was mutated to cysteine (Fig. 2.3a, b). The K237C mutation has been studied previously and has no effect on heptamer formation

characteristics, and hemolytic activity on rabbit erythrocytes even when modified with a bulky chemical group (Walker and Bayley, 1995) (Fig. 2.5). The unitary conductance of the heptameric (α7) K237C pore in planar lipid bilayers (Methods) was 857±6 pS (+50 mV, mean±s.e.m; *n*=169, *N*=9, where *n* is number of single events in *N* independent experiments) in 1 M KCl, 25 mM Tris.HCl, 50 μM EDTA, pH 8.0, was found to be similar to wild-type (WT) α7 conductance (721±6 pS at +40 mV in 1M KCl, 10 mM potassium phosphate, pH 7.5) reported elsewhere (Gu *et al.*, 2001). Hence, K237C mutation seemed a good choice for linking two α7 units cap-to-cap without compromising the stability and properties of the α7 units.

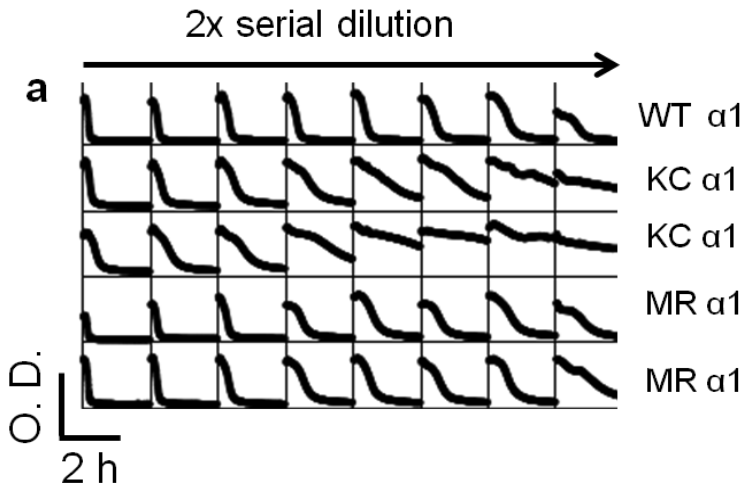
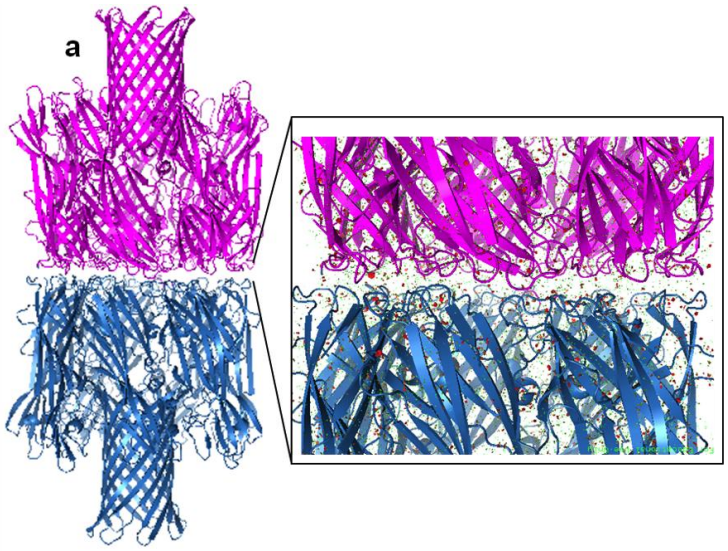



**Figure 2.5. Hemolytic activity of αHL mutants.** The hemolytic activity of pore forming toxins such as αHL can be assayed using rabbit erythrocytes. αHL monomers (α1) can assemble as heptamers on rabbit blood cells and lyse them. The cell lysis can be monitored as a decrease in the optical density (O.D.) of the cells at 595 nm. (**a**) WT, K237C/D8H6 (KC) and M113R/K237C/D8H6 (MR) α1 were prepared by in vitro transcription and translation in the presence of $^{35}$S methionine. SDS-polyacrylamide gel electrophoresis and autoradiography showed that they were at closely similar concentrations. The proteins were serially 2x diluted in each lane of a microtiter plate. All the afore mentioned proteins showed hemolytic activity on rabbit erythrocytes monitored over two hours. Assays were done in duplicates.

## 2.4. Molecular dynamics (MD) simulations of cap-to-cap $(\alpha7)_2$ showed seven disulfide bridges without steric clashes in the interface

The feasibility of a cap-to-cap $(\alpha7)_2$ complex of the K237C mutant bridged through disulfide bonds was evaluated using MD simulations. Two α7 units were docked cap-to-cap in PyMOL, where all the cysteine thiols were within disulfide bridging distance (~0.2 nm) (Methods). $(\alpha7)_2$ was solvated in a buffer containing 500 mM NaCl, and a 1 ns all atom MD simulation was performed using the GROMACS software (Hess *et al.*, 2008). It was found that in the energy minimized output structure of $(\alpha7)_2$ disulfide bridges were possible between all the seven cysteine pairs without causing any steric clashes between the interface residues (Fig. 2.6a). Energy analysis revealed a favourable electrostatic interaction and Lennard Jones potential between the α7 caps (Fig. 2.6b).




b

| | Energy (kJ /mol) |
|---|---|
| Short range electrostatics | -7400 ± 91 |
| Short range Lennard Jones | -4900 ± 10 |
| Average number of hydrogen bonds | 179 |

**Figure 2.6. Energy minimized cap-to-cap $(\alpha7)_2$.** Two α7 molecules were docked cap-to-cap in PyMOL such that the cysteines of each α7 unit were within disulfide bridging distance of cysteines of the adjacent α7 unit. (**a**) Cartoon representation of the cap-to-cap $(\alpha7)_2$ after an all-atom MD simulation of 1 ns in 500 mM NaCl. Zoomed inset of the dimeric interface showed absence of steric clashes (red disks) after 1 ns of simulation. (**b**) Table lists values of short-range electrostatics energy, short-range Lennard Jones energy and the number of hydrogen bonds in the interface of the two α7 units. Energy

calculations were performed using the GROMACS software (v4.0) and models were prepared in PyMOL. For visualization of steric clashes, a Python script 'show_bumps' (Thomas Holder) was used.

### 2.5. $(\alpha7)_2$ forms spontaneously during K237C purification

An 8x-Asp/6x-His (D8H6) tag was genetically fused to the C-terminus of the αHL K237C α1 gene to aid purification using immobilized metal affinity chromatography (IMAC). K237C/D8H6 α1 was expressed in *E. coli* and subsequently purified on Ni-NTA agarose. In the purified fractions of K237C/D8H6, a mixture of α1, monomer dimer $(\alpha1)_2$, α7 and a species of $M_{app}$≈190,000 was obtained (Fig. 2.7a). The $M_{app}$≈190,000 species was attributed to $(\alpha7)_2$.

The D8H6 peptide tag on the C-terminal is known to promote pre-oligomerization of WT α1 into α7 (our unpublished results). Thus, α1 spontaneously oligomerized to α7 which subsequently formed $(\alpha7)_2$ during purification. α7 is stable in SDS contained in the non-reducing sample buffer used for SDS-PAGE (Bhakdi *et al.*, 1981). Because each K237C α1 carried a D8H6 tail, the pre-oligomerized α7 also carried seven D8H6 tags, and hence several extra negative charges as compared to WT α7. The small hydrodynamic radius of the folded α7 structure in SDS and the several extra negative charges per heptamer caused α7 to migrate with $M_{app}$≈120,000, rather than $M_{app}$≈230,000 based on its molecular mass in SDS-PAGE (Kleinschmidt *et al.*, 1999). The dimer, $(\alpha7)_2$, with 14 D8H6 tags, would be also expected to migrate more quickly than anticipated based on its mass. Therefore, the $M_{app}$≈190,000 band, lying above the α7 band, was attributed to $(\alpha7)_2$.

## 2.6. Structural characterization of $(\alpha7)_2$

To confirm that the $M_{app}$≈190,000 species in SDS-PAGE (Fig. 2.7a) was $(\alpha7)_2$ formed of α7 units linked *via* disulfide bonds, I extracted the band from a polyacrylamide gel containing SDS. The gel purified protein band ($M_{app}$≈190,000) in 10 mM Tris.HCl, 1 mM EDTA, pH 8.0, was treated with 3 M βME for 15 min at 25°C, resulting in α7 species only (*lane 4*, Fig. 2.7b). Upon heating $(\alpha7)_2$ at 95°C for 10 min in the absence of a reducing agent, it dissociated to give two distinct species, α1 and $(\alpha1)_2$ (*lane 3*, Fig. 2.7b). The relative band intensities of α1 and $(\alpha1)_2$ were approximately equal, 18 and 17 absorbance units (a. u.) respectively, and added up to the intensity of $(\alpha7)_2$ (36 a. u.), implying that $(\alpha7)_2$ was composed of α1 and $(\alpha1)_2$ in a ratio of 2:1 (mol/mol). On treating $(\alpha7)_2$ with 3 M βME for 15 min at 25°C and on subsequent heating at 95°C for 10 min, $(\alpha7)_2$ reduced to α1 only (*lane 5*, Fig. 2.7b, c). These experiments proved that the introduced cysteines bridged the caps of α7 to form $(\alpha7)_2$.

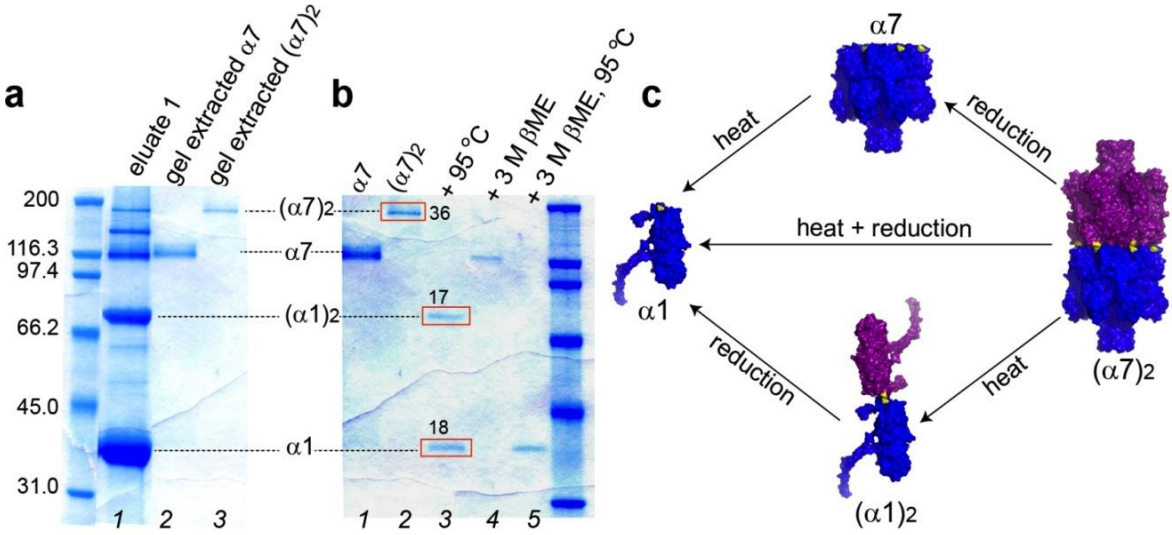



**Figure 2.7. Expression, purification and structural characterization of $(\alpha7)_2$.** (**a**) SDS-PAGE gel showing purification of $(\alpha7)_2$. Eluates of $Ni^{2+}$-NTA purified αHL mutant K237C/D8H6 contained a mixture of the monomer α1, the monomer dimer $(\alpha1)_2$, the heptamer α7, and the heptamer dimer $(\alpha7)_2$.

Extraction of α7 and $(α7)_2$ from a preparative gel (Methods) yielded pure products. (**b**) Heating $(α7)_2$ at 95°C for 10 min yielded α1 and $(α1)_2$ (red boxes, *lane 3*) whose relative band intensities of 18 and 17 absorbance units (a. u.) added up approximately to the intensity of $(α7)_2$ (36 a. u.) (red box, *lane 2*), respectively. To ascertain that $(α7)_2$ was composed of α7 units linked by disulfide bonds, $(α7)_2$ was treated with 3 M βME for 15 min at 25°C, which reduced $(α7)_2$ to α7 (*lane 4*). Upon heating at 95°C followed by reduction with 3 M βME for 15 min, $(α7)_2$ dissociated to α1 (*lane 5* labeled +3 M □ME, 95°C) (*lane 5*). (**c**) Based on these results, a schematic model of the structural composition and dissociation of $(α7)_2$ was proposed.

Hemolytic assays of $(α7)_2$ and α7 were also performed. Preformed α7 molecules could not lyse red blood cells (Fig. 2.8). $(α7)_2$ was also hemolytically inactive suggesting that even trace amounts of free $(α1)_2$ and α1 molecules were not present in the purified fraction of $(α7)_2$.

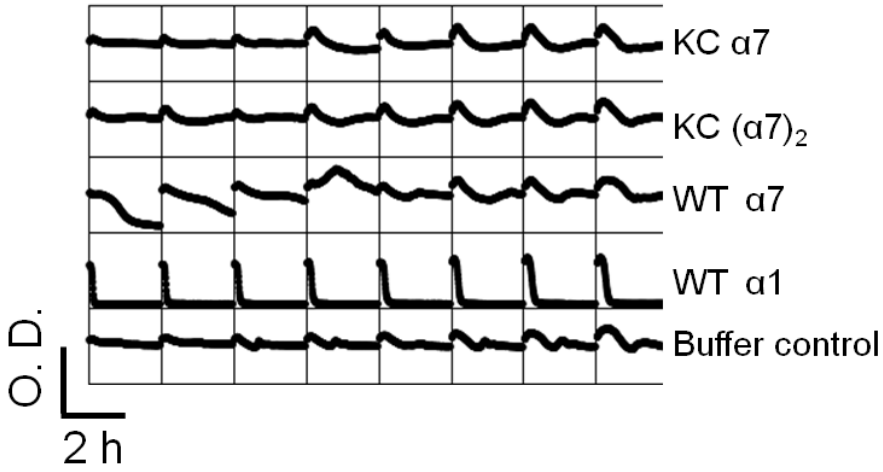



**Figure 2.8. Hemolytic activity of $(α7)_2$.** Hemolytic assays of K237C/D8H6 (KC) α7 and $(α7)_2$ were performed on rabbit erythrocytes. WT α7 was not hemolytically active. KC α7 and KC $(α7)_2$ were also hemolytically inactive. This implied free α1 was not present in purified $(α7)_2$.

### 2.7. $(α7)_2$ is a fully folded structure

Proteinase K cleaves at the glycine rich central domain and near the N-terminus of α1 (Walker *et al.*, 1992), regions inaccessible in the folded α7 structure (Cheley *et al.*, 1997). SDS gel extracted α7 was treated with 0.5 mg $ml^{-1}$ proteinase K and subsequently heated at 95°C for 10 min, yielding untruncated

α1, confirming the reported results. We hypothesized that like SDS stabilized α7 in buffer, a folded two-barrel $(α7)_2$ structure would be resistant to proteinase K. Proteinase K treated $(α7)_2$ upon heating at 95°C for 10 min yielded full-length α1 and $(α1)_2$, and reduced to α7 on treating with 3 M βME for 15 min at 25°C (Fig. 2.9). This showed that gel extracted $(α7)_2$ had a folded structure like α7 and was therefore resistant to proteinase K.

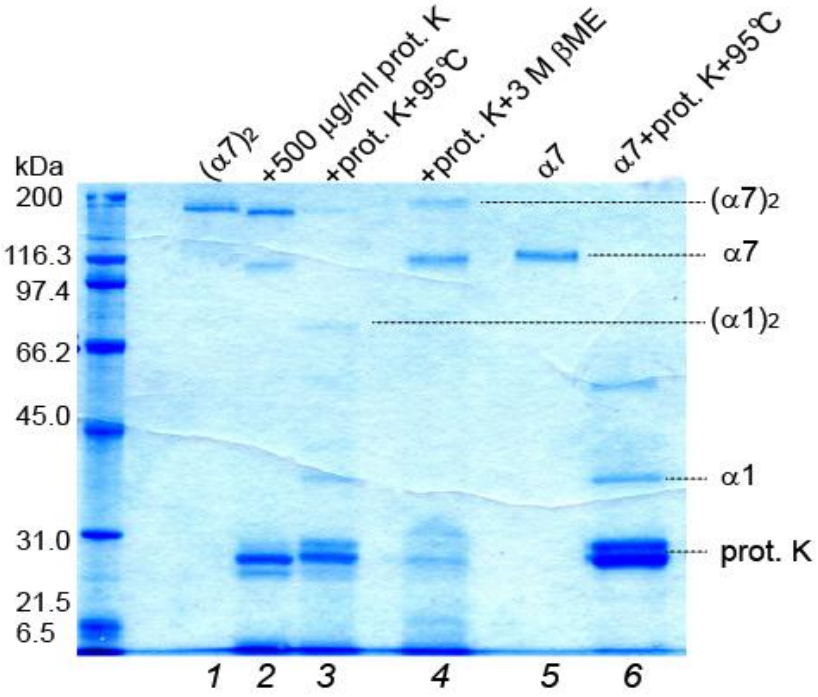



**Figure 2.9. $(α7)_2$ resistance to proteolysis.** SDS-PAGE gel showing that SDS-gel extracted $(α7)_2$ (*lane 1*) was unaffected by proteinase K (prot. K) treatment like α7 (*lane 2*). Proteinase K treated $(α7)_2$ dissociated to full-length α1 and $(α1)_2$ on heating at 95°C for 10 min (*lane 3*). Proteinase K treated $(α7)_2$ gave α7 on reducing with 3 M βME for 15 min (*lane 4*). Proteinase K treated α7 yielded full-length α1 upon heating at 95°C for 10 min (*lanes 5, 6*).

## 2.8. $(α7)_2$ is robust to mutagenesis

The β barrel of α7 is robust to extreme chemical and genetic modifications, which has allowed the development of α7 pores with interesting properties such as enhanced ion selectivity, rectification, size exclusion, etc. (section 2.3.3). The Met 113 residue in the WT barrel of $(α7)_2$ was changed to Arg and Phe. Like

K237C/D8H6 $(\alpha7)_2$, the M113R/K237C/D8H6 (M113R) and M113F/K237C/D8H6 (M113F) mutants of $(\alpha7)_2$ also formed spontaneously in the cell lysate of *E.coli* expressing the respective α1. The $(\alpha7)_2$ barrel mutants showed similar current characteristics as K237C/D8H6 $(\alpha7)_2$ (Fig. 2.14, section 2.10). This result demonstrated that like α7, $(\alpha7)_2$ is also robust to mutagenesis.

### 2.9. Possible $(\alpha7)_2$ structures

Although SDS-PAGE suggested a single species of $(\alpha7)_2$, $(\alpha7)_2$ might be formed from structures other than a perfectly aligned cap-to-cap structure (Fig. 2.3c), because of the different possible cross-linking patterns between the α7 units. Four structures, other than the the precise cap-to-cap $(\alpha7)_2$ (Fig. 2.3c), are possible with a minimum of one and a maximum of two disulfide bonds between the α7 units (Fig. 2.10). The α7 units of $(\alpha7)_2$ were modeled as regular heptagons in which the vertices represented the cysteine residues of each α1 subunit. To form a valid $(\alpha7)_2$ structure, the two heptagons were constrained to have at least one common vertex, i.e., a disulfide bond. According to this model, four structures other than the correctly aligned cap-to-cap $(\alpha7)_2$ (Fig. 2.3b) are possible with a minimum of one and a maximum of two disulfide bonds between the α7 units. As compared to the fixed structures **c** and **d** (Fig. 2.10) different conformations of disulfide bonds in structures **a** and **b** (Fig. 2.10) would lead to other $(\alpha7)_2$ orientations including the exact cap-to-cap structure (Fig. 2.10e). The aligned cap-to-cap $(\alpha7)_2$ structure can form with any number of disulfide bonds, from one to seven (Fig. 2.10e). When α7 units or heptagons are linked at three or more positions, only the aligned cap-to-cap structure is geometrically possible.

$(\alpha7)_2$ structures linked *via* different number of disulfides would yield different ratios of α1 and $(\alpha1)_2$ on heating at 95°C. Assuming that a perfect cap-to-cap arrangement of $(\alpha7)_2$ was formed with three or more disulfide bonds, the molar ratios of $(\alpha1)_2$ and α1 after heating would lie between 3:8 (three disulfide bonds) and 1:0, i.e., $(\alpha1)_2$ as the only product (7 disulfide bonds). Structure **a** (Fig. 2.10), formed with one disulfide linkage, would yield a mixture of α1 and $(\alpha1)_2$ in a molar ratio of 12:1 on heating. Similarly, structures **b-d** (Fig. 2.10) with two disulfide linkages would yield α1 and $(\alpha1)_2$ in a ratio of 5:1. Experimentally, on heating $(\alpha7)_2$ at 95°C, $(\alpha1)_2$ and α1 were obtained in a molar ratio of 2 to 1, from which we deduce that the average number of disulfides in $(\alpha7)_2$ was ~3.3 (section 2.6). Hence, it was concluded that the majority of the $(\alpha7)_2$ molecules were in an aligned cap-to-cap arrangement (Fig. 2.10e).

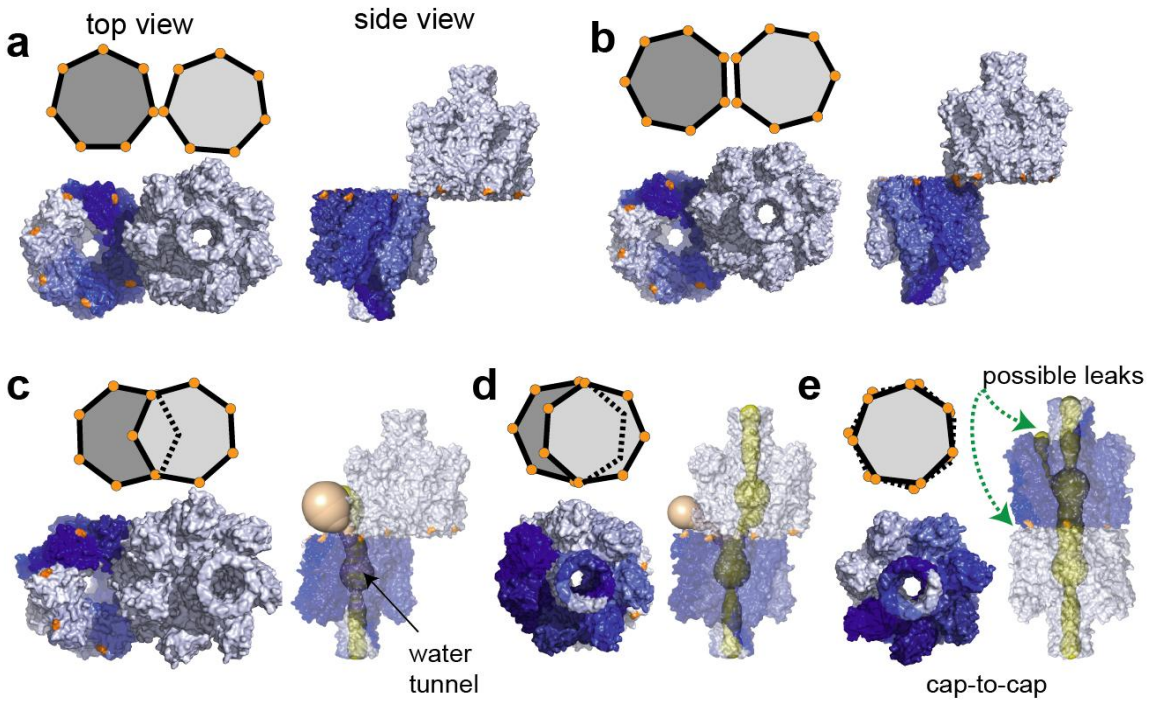



**Figure 2.10. Possible disulfide linked structures of $(\alpha7)_2$.** Schematic diagrams and space-filling models of all the possible $(\alpha7)_2$ structures. Each α7 unit is represented as a regular heptagon (grey) with orange vertices denoting the cysteine locations. (**a**) $(\alpha7)_2$ structure with a single disulfide bridge between

*cis*- and *trans*-α7. (**b – d**) $(\alpha7)_2$ structures with two disulfide bridges between *cis*- and *trans*-α7. The electrical conductance of structures **a – c** would be similar to α7 because the *trans*-α7 was not obstructed by *cis*-α7. In structure **d**, *cis*- and *trans*-α7 are connected *via* a water tunnel (yellow, visualized by MOLE analysis (Petrek *et al.*, 2007)) increasing the effective length of the pore. A water tunnel entering from the unobstructed side (wheat bulb) of *trans*-α7 could also be visualized. Hence, the conductance of **d** might not be similar to α7. (**e**) Different dihedral angles about the disulfide bonds of structures **a** and '**b**' will result in different $(\alpha7)_2$ structures including the axially aligned cap-to-cap orientation. However, the aligned cap-to-cap orientation is the only structure geometrically possible if the *cis*-α7 and the *trans*-α7 units are linked by three or more disulfides. Water-filled channels passing through the central pore and between the α1 units were visualized by using MOLE. Another possible path for ions is at the interface of the *cis*-α7 and the *trans*-α7 units.

## 2.10. Electrical characteristics of $(\alpha7)_2$ suggested a longer and symmetric pore structure

α7 readily inserts vectorially in *in vitro* lipid bilayers (Fang *et al.*, 1997; Krasilnikov *et al.*, 2000). On adding a suitable electrolyte on both sides of the bilayer and on applying a bias electric potential, the electrical activity of single α7 molecules inserted in planar lipid bilayers or droplet interface bilayers can be studied (Fang *et al.*, 1997; Krasilnikov *et al.*, 2000; Villar *et al.*, 2011). Insertion of individual α7 molecules is seen as a step-wise increase in the electrical conductance (Fang *et al.*, 1997; Krasilnikov *et al.*, 2000).

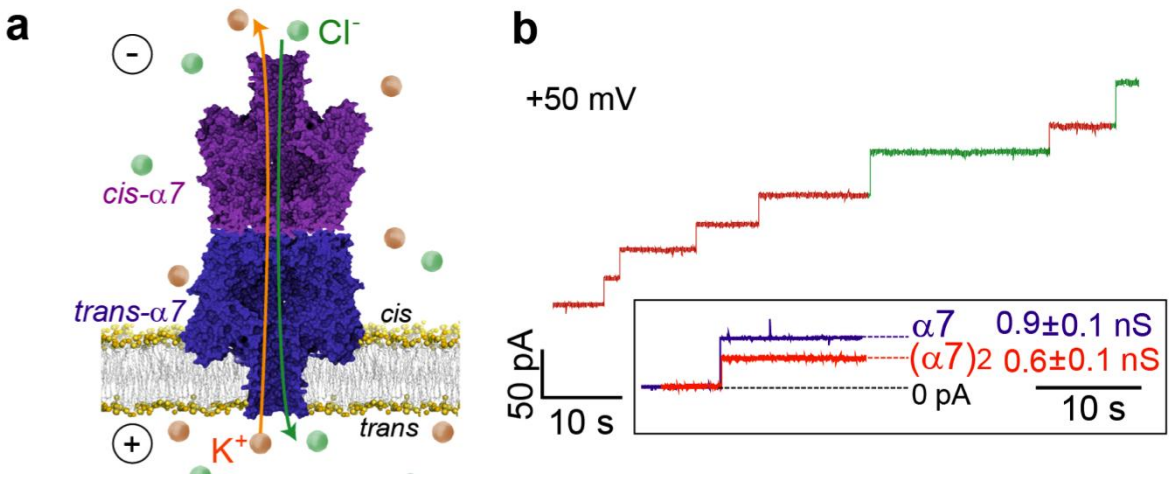



**Figure 2.11. Electrical characteristics of $(\alpha7)_2$.** (**a**) $(\alpha7)_2$ was electrically characterized using a planar lipid bilayer set up. A planar lipid bilayer (≈1.2 x$10^4$

– 3 x$10^4$ μ$m^2$, determined by capacitance measurements) formed on an aperture separated two equimolar buffer reservoirs (1 M KCl, 25 mM Tris.HCl, 50 μM EDTA, pH 8.0) (Maglia *et al.*, 2010). The flow of ions through single protein pores due to an applied potential was amplified and measured as an electric current. (**b**) A representative trace showing rapid insertion of $(\alpha7)_2$ in a planar DPhPC bilayer at +50 mV. The mean conductance (*G*) of majority of the $(\alpha7)_2$ channels (red steps) was 570±4 pS (*n*=284, *N*=18) as compared to that of α7, *G*=857±6 pS (*n*=169, *N*=9). Some $(\alpha7)_2$ channels of unitary conductance like α7 were also observed (green steps).

Like α7, $(\alpha7)_2$ readily inserted in planar lipid bilayers and formed pores through insertion of one of the α7 transmembrane barrels (termed *trans*-α7) (Fig. 2.11a). We surmise that the other $\alpha7$ unit (*cis*-$\alpha7$) is surrounded by buffer and perhaps stabilized by residual traces of SDS (Fig. 2.11a). At +50 mV, 284 of 377, i.e. 75±4% (percentage of pores±95% confidence interval of binomial distribution (Methods, equation (2.22))), of the $(\alpha7)_2$ pores had a mean conductance of 570±4 pS (mean±s.e.m.; *n*=284, *N*=18, 1 M KCl, 25 mM Tris, 50 μM EDTA, pH 8.0) (red), which was ~66% of the value found for α7, 857±6 pS, (mean±s.e.m.; *n*=169, *N*=9) (blue, Fig. 2.11b).

The histograms of the open pore current ($I_o$) of $(\alpha7)_2$ at each of the three potentials, +50 mV, +100 mV and +160 mV showed at least two populations (red and green) (Fig. 2.12). A double Gaussian function was used to fit the $I_o$ histogram of $(\alpha7)_2$. At +50 mV, 75% of the $(\alpha7)_2$ pores had a mean $I_o$ of 27±4 pA (mean±standard deviation (S.D.); *n*=284, *N*=18; 1 M KCl, 25 mM Tris, 50 μM EDTA, pH 8.0) (Fig. 2.12a, red peak) as compared to the $I_o$ value of $\alpha7$ of 42±4 pA (mean±S.D.; *n*=169, *N*=9) (Fig. 2.12a, blue peak). The mean $I_o$ of the remaining 25% of the $(\alpha7)_2$ pores was 42±4 pA, (mean±S.D.; *n*=93, *N*=18) (Fig. 2.12a, green peak), which corresponded exactly to the value for $\alpha7$.

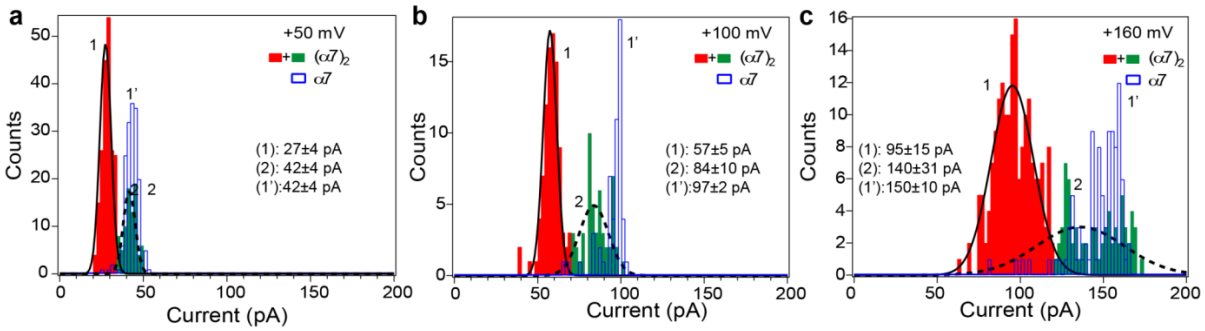



**Figure 2.12. Conductance characteristics of $(\alpha7)_2$ and α7.** Histogram plots of $I_o$ of $(\alpha7)_2$ at (**a**) +50 mV (**b**) +100 mV and (**c**) +160 mV showed at least two populations of $(\alpha7)_2$ channels. At all potentials, majority of $(\alpha7)_2$ channels had lower $I_o$ (red, fit 1) whose mean was ~66% of the $I_o$ of α7 (blue bars, fit 1'). These $(\alpha7)_2$ channels would comprise structures with α7 units arranged exactly cap-to-cap (Fig. 2.3c). The $(\alpha7)_2$ peak with the higher mean $I_o$ (green, fit 2) coincided with the distribution of α7 $I_o$ (blue bars, fit 1') at the same potentials, and should comprise structures **a-c** (Fig. 2.10). Numbers denote the mean $I_o$±S.D. derived from the Gaussian fits (Methods, equation (2.21)).

The two distinct populations of $(\alpha7)_2$ channels suggested more than one $(\alpha7)_2$ structures. If the different structures of $(\alpha7)_2$ as discussed in section 2.9 were considered, they would also have different unitary conductance (G) values. G depends on the length (l) of a pore, its cross-sectional area (A) and the solution ionic conductivity (σ) (equation (2.1)).

$$G = \frac{\sigma A}{l} \qquad (2.1)$$

To estimate the conductance of these protein pores, MOLE (Petrek *et al.*, 2007), an algorithm to visualize water-filled tunnels through protein pores was used. In structures **a** and **b** (Fig. 2.10), there are no obstructions of the α7 pores by each other. Hence, the $I_o$ of these structures would be similar to α7 because just one of the two α7 units of **a** and **b** (Fig. 2.10) span the lipid bilayer. Although the α7 pores partially obstruct each in structure **c** (Fig. 2.10), the *trans*-α7 pore entrance is larger than the constriction in the barrel. Hence, the constriction would limit the conductance of **c** (Fig. 2.10), which would again be similar to α7.

In structure **d** (Fig. 2.10), a continuous water channel was found in between the two α7 units. Another water channel could also be visualized with an outlet at the cap-cap interface of the two α7 units. Unlike the tunnels in structures **a**-**c** (Fig. 2.10), the length of the water tunnel, and hence the effective pore length of structure **d** (Fig. 2.10) is longer than that of α7, implying that its conductance could differ from α7. The aligned cap-to-cap $(α7)_2$, structure **e** (Fig. 2.10), has twice the pore length of α7. However, water tunnels between the α1 subunits of structure **e** were also detected, which could contribute to the current and make its conductance slightly more than half of α7. Another possible inlet for ions is through the cap-cap interface of the two α7 units but it was not detected by the MOLE program.

The 25% of $(α7)_2$ pores that had an open pore conductance similar to α7 were attributed to $(α7)_2$ structures that contain an offset between the α7 units (Fig. 2.10a-c). The lower conductance of the major fraction of the $(α7)_2$ pore (Fig. 2.12, red peak) was consistent with a pore that is longer than α7 (Fig. 2.10d, e) (equation (2.1)). Moreover, as suggested by TEM (section 2.16) and heating analysis, majority of the $(α7)_2$ molecules were precisely cap-to-cap (Fig. 2.3c). Hence, $(α7)_2$ channels corresponding to the red peak (Fig. 2.12) reflect $(α7)_2$ structures with the precise cap-to-cap orientation of its α7 units.

Further experiments were done using $(α7)_2$ channels with mean conductance lying between 570±4 pS only. I-V characteristics of $(α7)_2$ channels and α7 differed in conductance values and rectification properties (Fig. 2.13). The rectification ratio ($-I_{+V}/I_{-V}$) of α7 lied between 1.05 – 1.54 (20 - 200 mV), and

decreased with increasing voltage owing to a difference in charges on the *cis* and *trans* pore entrances (Chapter 3, section 3.8.9.7). Compared to α7, the rectification ratio of $(\alpha7)_2$ did not vary as much and ranged from 1.05 – 1.13 (20 - 200 mV) (Fig. 2.13b). The rectification ratios diverged most strikingly at higher potentials. For example, $(-)I_{+100}/I_{-100}$ was 1.29±0.02 (mean±s.e.m.; *N*=5) for α7 and 1.05±0.05 (mean±s.e.m.; *N*=8) for $(\alpha7)_2$. The I-V characteristics for $(\alpha7)_2$ were roughly ohmic from negative to positive applied potentials, which is consistent with a symmetrical pore structure (Fig. 2.13).

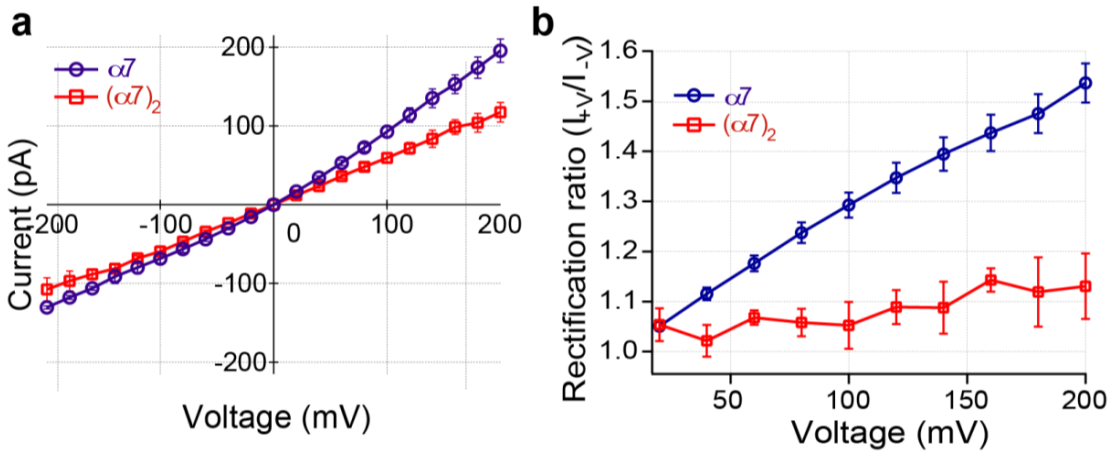



**Figure 2.13. I-V characteristics of $(\alpha7)_2$ as compared to α7.** For plotting the I-V curves and calculating the rectification ratios of cap-to-cap $(\alpha7)_2$, channels with unitary conductance in the range 570±4 pS were used. (**a**) I-V plots of $(\alpha7)_2$ (red) were compared with α7 (blue) ($N \geq 3$). The $I_o$ values of $(\alpha7)_2$ between -200 mV and +200 mV were lower than α7 suggesting that $(\alpha7)_2$ had a higher resistance or a longer pore length than α7 (equation (2.1)). (**b**) Rectification ratios ($-I_{+V}/I_{-V}$) of α7 and $(\alpha7)_2$ were calculated as a function of voltage. Due to a difference in the charges at the *cis* and *trans* pore entrances, the rectification of α7 increased, i.e., the rectification ratio decreased, with increasing applied potential. Experiments were performed in 1 M KCl, 25 mM Tris.HCl, 50 µM EDTA, pH 8.0.

Similarly, M113R and M113F $(\alpha7)_2$ (section 2.8) were also electrically characterized. The pore insertion frequency of M113F $(\alpha7)_2$ was found to be several fold higher at low pH. Hence, planar lipid bilayer measurements of M113F $(\alpha7)_2$ were performed in 1 M KCl, 2 mM succinic acid, pH 4.1. The I-V

characteristics and the rectification ratios of M113R and M113F $(\alpha7)_2$ closely matched with $(\alpha7)_2$ with the WT barrel (K237C/D8H6).

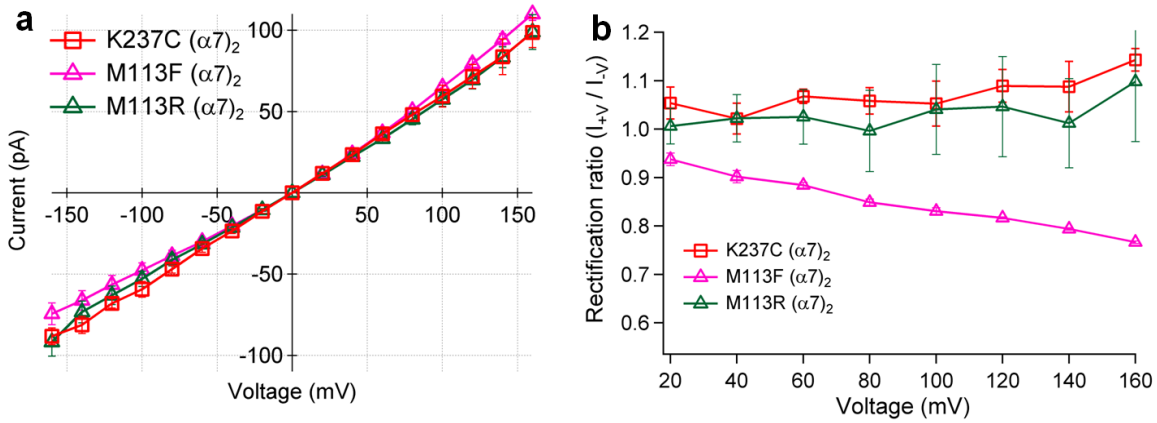



**Figure 2.14. I-V characteristics of barrel mutants of $(\alpha7)_2$.** $(\alpha7)_2$ with the additional barrel mutations of M113R and M113F could be synthesized and purified in a similar way as K237C/D8H6 $(\alpha7)_2$ (Methods). (**a**) The I-V characteristics of M113R/K237C/D8H6 (M113R) and M113F/K237C/D8H6 $(\alpha7)_2$ (M113F) matched closely with K237C/D8H6 $(\alpha7)_2$ (K237C) with the WT barrel. (**b**) The rectification ratios of M113R and M113F were also similar to K237C from 20-160 mV.

## 2.11. γCD blocked $(\alpha7)_2$ from both the *cis* and *trans* sides confirming the presence of two β-barrels

The structural characteristics of $(\alpha7)_2$ were further investigated by determining the binding kinetics of γCD, a α7 adapter (Gu *et al.*, 1999). γCD binds near the constriction of the α7 β-barrel (Fig. 2.4a) and the γCD bound pore shows a lower conductance than the open α7 pore (equation (2.1)) and γCD binding-unbinding can be detected electrically as short blockades in the α7 $I_o$ (Gu *et al.*, 1999) (Fig. 2.4b).

The dimensions of the α7 pore allow γCD to bind only when γCD is added to the *trans* side of the bilayer (Gu *et al.*, 1999). γCD produced a blockade of 66±1% of

$I_o$ of α7 (percent block±s.e.m.; *n*=17354, *N*=6) at +50 mV from the *trans* side of the bilayer (Fig. 2.15a-d). Under the same conditions, γCD blocked $(α7)_2$ from both the *trans* and the *cis* sides (Fig. 2.15e-j). The blocking from the *trans* side was observed with 100% success (*n*=33, *N*=15; where *n* is number of pores in *N* independent experiments), and from the *cis* side in 50 out of 56 channels (89±8%; percentage of pores blocked±95% confidence interval; *n*=56, *N*=19) (Fig. 2.15e-h). This implied that $(α7)_2$ contained two γCD binding sites, one facing each side of the bilayer, which is consistent with the presence of two symmetrically placed β barrels.

γCD, when added to the *cis* side of the bilayer (*cis* γCD binding), produced a smaller block in the $I_o$ of $(α7)_2$ than when added to the *trans* side (*trans* γCD binding) (Table 2.1). For example, at +50 mV, *cis* γCD binding led to a 15±1% (blockade in $I_o$±s.e.m., *n*=22,344, *N*=12) as compared to 55±2% (blockade in $I_o$±s.e.m., *n*=11,068, *N*=7) for *trans* γCD binding. When γCD was added to both the *cis* and *trans* sides of the bilayer, it bound to either the *cis*- or *trans*-α7 barrels or to both the barrels (*cis*-*trans* γCD block). Three kinds of events were observed that could be attributed to binding to either *cis* γCD binding or *trans* γCD binding, or to both sites at the same time. For example, at +50 mV, simultaneous *cis* and *trans* γCD binding blockades of 61±1% (*n*=1121), *cis* γCD binding blockades of 15±2% (*n*=4183) and *trans* γCD binding blockades of 57±1% (*n*=3737) (blockade in $I_o$±s.e.m.; *N*=6) were observed.

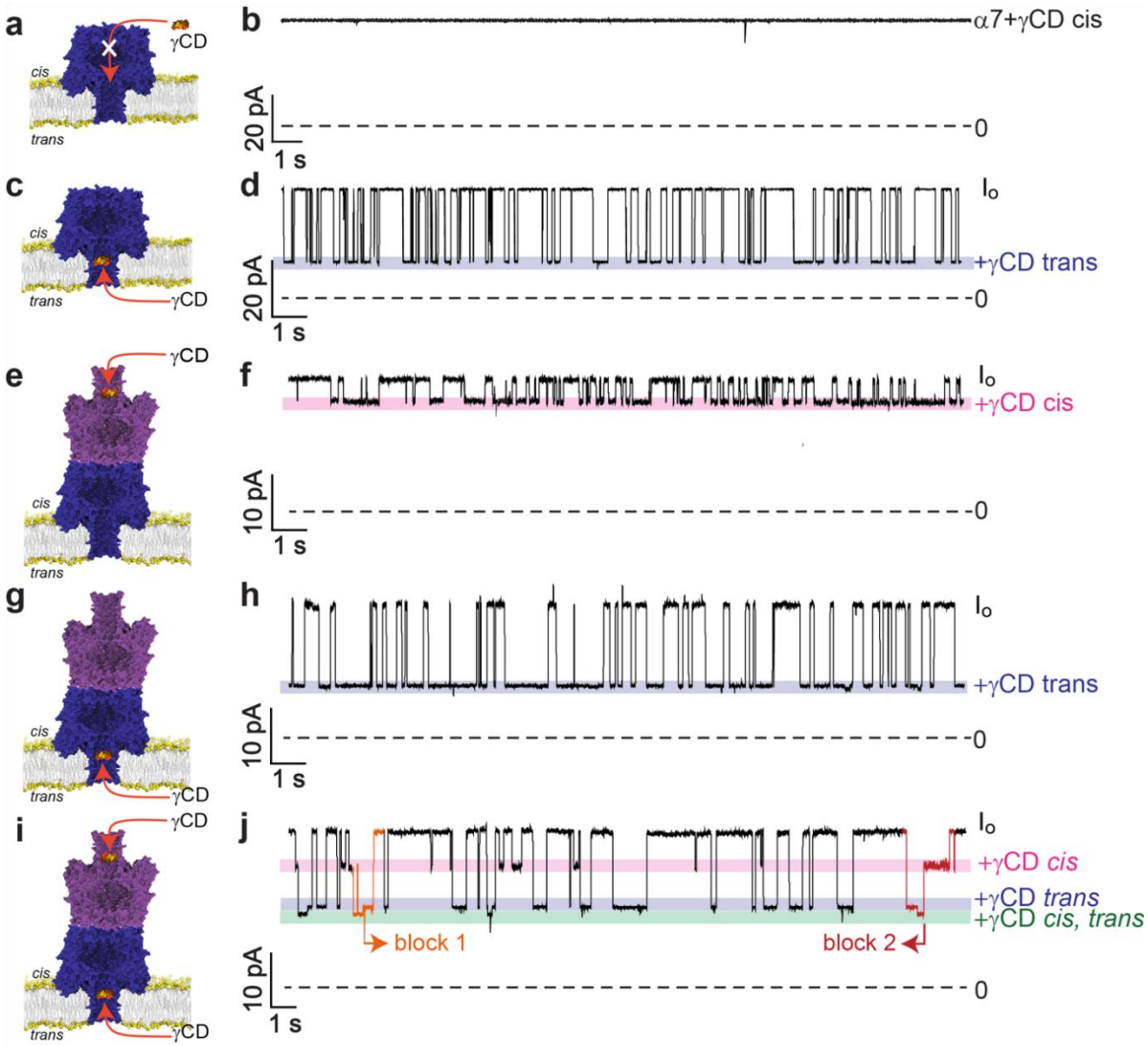



**Figure 2.15. γCD binding to α7 and $(α7)_2$.** (**a, b**) When γCD was added to the *cis* side of α7, no current blocks were observed. (**c**) γCD bound in the α7 barrel and produced transient blocks in $I_o$ when added to the *trans* side. (**d**) Representative current trace showing a 66±1% blockade in $I_o$ (+50 mV) upon *trans* γCD binding to α7. (**e, g**) In comparison to α7, γCD blocked $(α7)_2$ from both the *cis* and *trans* sides of the bilayer. (**f**) Representative current trace of *cis* γCD binding to $(α7)_2$ leading to a 15% blockade in $I_o$ compared to a 55% reduction produced by *trans* γCD binding (**h**). (**i, j**) When γCD was added to both the *cis* and *trans* sides simultaneously, γCD could lodge in both the *cis*- and *trans*-α7 barrels of $(α7)_2$ leading to 3 levels of current blocks: (i) only *cis* (15% blockade), (ii) only *trans* (55% blockade), and (iii) simultaneous *cis* and *trans* γCD bound states (61% blockade) (*n*=1121, *N*=6). The double γCD block (*cis*-*trans* γCD was preceded by an only *cis* (block 1, orange) or after an only *trans* γCD block (block 2, red). In all traces, the blue and pink bars show the *trans* and *cis* γCD blocked pore currents, respectively. For γCD binding studies to $(α7)_2$, only the channels with unitary conductance in the range 570±4 pS were used. All current traces shown were recorded at +50 mV in 1 M KCl, 25 mM Tris.HCl, 50 µM EDTA, pH 8.0 with 40 µM γCD on the *cis* and/or *trans* sides.

Double γCD binding was observed as two types of blocks (block 1 and block 2) depending on the order of γCD binding to *cis*- and *trans*-α7 barrels (Fig. 2.15j). In block 1, the order of γCD binding was (i) first γCD molecule bound to the *cis*-α7 barrel. (ii) Then, another γCD molecule entered the *trans*-α7 barrel leading to double γCD bound state. (iii) The first γCD exited from either the *cis*-α7 barrel leaving a *trans* γCD only blocked state. (iv) The second γCD molecule also exited leading to the $I_o$ level. Similarly, block 2 was when (i) the first γCD molecule bound to the *trans*-α7 barrel. (ii) Then another γCD molecule blocked the *cis*-α7 barrel leading to the double block. The block cleared away when (iii) the first γCD exited from the *trans*-α7 barrel and the current became equivalent to the $(α7)_2$ *cis* γCD bound state. (iv) The second γCD molecule also exited from the *cis*-α7 barrel leading to the $I_o$ level. In summary, the *cis-trans* γCD blockade was preceded by either a *cis* γCD blockade level or a *trans* γCD blockade (Fig. 2.15i, j). This showed that γCD could bind to both the β-barrels of $(α7)_2$ independently.

## 2.12. Kinetics of γCD binding

The electrical current trace of a single channel showing γCD blockades can be used to deduce the mean dwell time of γCD within the pore ($\tau_{off}$) and the mean inter-event time ($\tau_{on}$). The association ($k_{on}$) and dissociation rate constants ($k_{off}$) of γCD binding (Fig. 2.4c) can be deduced from the mean dwell times using equations (2.2.1-2.2.3) (Fig. 2.16).

$$k_{\text{off}} = \frac{1}{\tau_{\text{off}}} \qquad (2.2.1)$$

$$C \times k_{on} = \frac{1}{\tau_{off}} \quad (2.2.2)$$

$$K_d = \frac{k_{off}}{k_{on}} \quad (2.2.3)$$

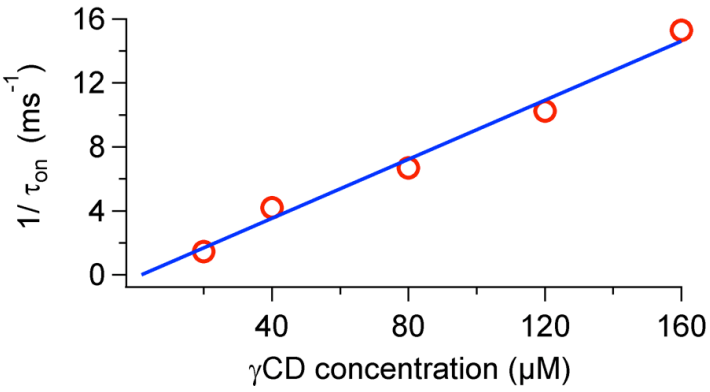



**Figure 2.16. Calculation of association rate constant ($k_{on}$) of γCD binding.** A representative plot showing the inverse of mean inter-event time ($\tau_{on}$) (y) of *trans* γCD binding to $(\alpha7)_2$ plotted *vs.* five different γCD concentrations (x) at +40 mV. The data points were fit with a straight line (y=ax) whose slope gave the value of $k_{on}$.

The association rate constants ($k_{on}$) were similar for *cis* and *trans* γCD binding to $(\alpha7)_2$ and *trans* γCD binding to α7 (Table 2.1). However, the dissociation rate constants ($k_{off}$) for *cis* γCD binding were approximately twice the values determined for *trans* γCD binding (Table 2.1). The $K_d$ values of *cis* γCD binding were approximately twice of *trans* γCD binding to $(\alpha7)_2$ because of a corresponding difference in the $k_{off}$ values (Table 2.1; Fig. 2.17).

The frequency of CD binding is also a function of the electroosmotic flow through the pores, i.e., on the ionic strength and applied potential (Gu *et al.*, 2003). The $K_d$ values of *trans* γCD binding to α7 increased with increasing potentials from -100 mV to +100 mV. In stark contrast, the $K_d$ values of *cis* and *trans* γCD binding to $(\alpha7)_2$ did not vary with the applied potential, from -100 mV to +100 mV. This result suggested that the flow of ions was similar at both

negative and positive potentials, which was also an indication of a symmetric structure of the $(\alpha7)_2$ pore. The physical origin of a smaller block in $I_o$ and weaker *cis* γCD binding compared to *trans* γCD binding in $(\alpha7)_2$ could be attributed to a leak in the structure (Fig. 2.10e) (also see section 2.18.1).

**Table 2.1.** Table lists the kinetic and equilibrium dissociation constants for the interaction of γCD with $(\alpha7)_2$ and α7 in 1 M KCl, 25 mM Tris.HCl, 50 μM EDTA, pH 8.0 at 25°C. The kinetic constant values were obtained from at least three different experiments, in each of which at least three γCD concentrations were used. At each voltage, the $k_{off}$ values were derived for each γCD concentration (20-160 μM) as the inverse of $\tau_{off}$ (equation (2.2.1)) and averaged. The average $k_{off}$ values over 3-8 experiments have been reported in the table. The $k_{on}$ values were obtained in each experiment from the slope of a plot of the inverse of the inter-event time ($\tau_{on}$) *vs.* γCD concentrations (Fig. 2.16) (equation (2.2.2)). The dissociation constant ($K_d$) was calculated for each experiment as $k_{off}/k_{on}$ and the average value over three or more experiments is reported (equation (2.2.3)).

| **Voltage (mV)** | $k_{on}$ **x** $10^4$ ($M^{-1}$ $s^{-1}$) | $k_{off}$ ($s^{-1}$) | $K_d$ (μM) | **blockade of** $I_o$ |
|---|---|---|---|---|
| $(\alpha7)_2$ ***trans*** | (mean±s.e.m) | (mean±s.e.m) | (mean±s.e.m) | (mean±s.e.m) |
| +40 | 7.1±0.7 | 3.7±0.4 | 49±3 | 55±2 |
| -40 | 8.2±1.1 | 3.2±0.2 | 36±2 | 51±2 |
| +50 | 9.0±1.1 | 3.4±0.4 | 42±3 | 55±2 |
| -50 | 6.1±0.6 | 3.3±0.3 | 54±8 | 48±3 |
| +100 | 8.9±1.0 | 4.2±0.8 | 46±7 | 54±5 |
| -100 | 4.5±0.3 | 2.9±0.5 | 64±11 | 45±6 |
| $(\alpha7)_2$ ***cis*** | | | | |
| +40 | 5.8±0.8 | 6.0±0.4 | 85±12 | 16±2 |

| | | | | |
|---|---|---|---|---|
| -40 | 5.6±0.5 | 7.2±1.1 | 110±18 | 17±1 |
| +50 | 6.0±0.5 | 6.9±0.7 | 110±12 | 15±1 |
| -50 | 6.1±1.0 | 6.2±1.0 | 96±9 | 19±1 |
| +100 | 6.3±1.2 | 6.4±2.0 | 110±27 | 13±0.3 |
| -100 | 6.4±1.3 | 10±1.6 | 130±24 | 21±2 |
| **α7 *trans*** | | | | |
| +40 | 9.5±1.2 | 6.1±0.7 | 76±12 | 65±1 |
| -40 | 5.0±0.1 | 3.1±0.4 | 61±6 | 58±1 |
| +50 | 7.8±0.9 | 6.9±0.9 | 110±37 | 66±1 |
| -50 | 7.2±2.3 | 3.0±0.3 | 47±9 | 58±0.4 |
| +100 | 4.3±1.2 | 12.3±2.0 | 342±63 | 68±0.3 |
| -100 | 6.9±2.8 | 1.7±0.2 | 37±10 | 52±0.1 |

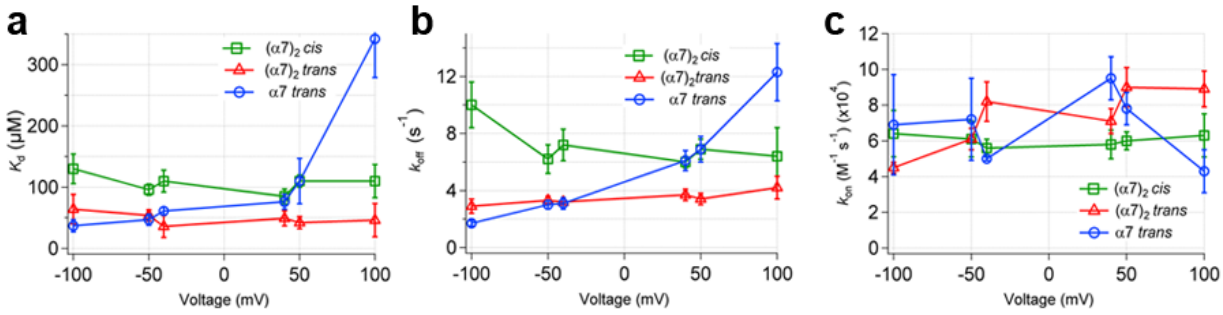



**Figure 2.17. Kinetic constants of γCD binding to α7 and $(α7)_2$.** Figure shows the dependence of (**a**) $K_d$, (**b**) $k_{off}$ and (**c**) $k_{on}$ of *cis*- and *trans*-γCD binding to $(α7)_2$ and *trans*-γCD binding to α7 as a function of voltage. WT α7 is slightly cation selective and due to the net electro-osmotic flow, the $k_{off}$ value for γCD binding increases with increasing positive applied potentials as shown in (**b**) (Gu *et al.*, 2003). Due to the symmetric structure of $(α7)_2$, and hence, no net electro-osmotic flow, the $k_{off}$ values for do not vary appreciably with the applied potential.

## 2.13. Phosphate binding to an $(\alpha7)_2$ mutant confirms the precise cap-to-cap structure

γCD binding confirmed the presence of two barrels in $(\alpha7)_2$ facing each side of the planar lipid bilayer (sections 2.11, 2.12). To further validate the symmetric structure of $(\alpha7)_2$, phosphate anion ($P_i^-$), a *cis* (cap side) blocker of certain α7 mutants (Cheley *et al.*, 2002) was used. Replacing M113 by arginine in α7 causes $P_i^-$ ions to block the pore at low potentials, when added to the *cis* side only. At +20 mV, $P_i^-$ blocked the α7 pore formed from M113R, when added to the *cis* compartment (Fig. 2.18a-c). It was thus hypothesized that an exactly placed cap-to-cap structure of the M113R $(\alpha7)_2$ mutant (section 2.8), would not show $P_i^-$ blockades as there would be no *cis* entrance for $P_i^-$ ions (Fig. 2.18d, e).

When $P_i^-$ was added to the *cis* side of M113R $(\alpha7)_2$, 48 of 65 M113R $(\alpha7)_2$ channels, i. e., 74±11% pores (percentage of pores examined±95% confidence interval), were not blocked. The population (74%) of cap-to-cap M113R $(\alpha7)_2$ as deduced from $P_i^-$ blocking experiments correlated well with the value of 75±4% determined from conductance measurements (*n*=377, *N*=18) (Fig. 2.12). These results in conjunction with γCD binding and TEM experiments (section 2.16) suggest that $(\alpha7)_2$ channels had a perfectly aligned cap-to-cap structure.

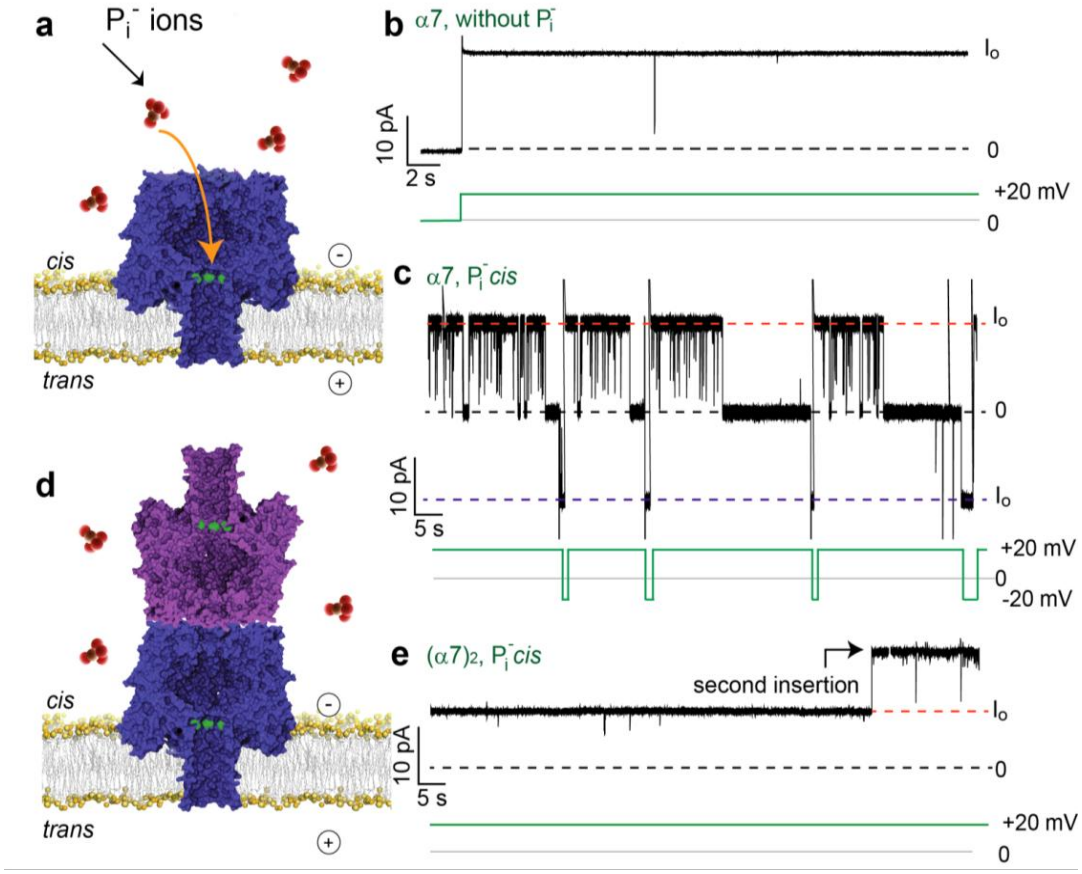



**Figure 2.18. $P_i^-$ binding to M113R $(\alpha7)_2$ mutant.** (**a**) $P_i^-$ ions (red spheres, not to scale) could bind and block the α7 mutant M113R/K237C/D8H6 (M113R) from the *cis* side only. (**b**) M113R α7 did not show any blocking events in the absence of $P_i^-$ ions. (**c**) Representative current trace of a M113R α7 channel when $P_i^-$ ions were added to the *cis* side causing complete blocks in $I_o$ (+20 mV). The blocks could be removed by reversing the potential (-20 mV) indicating the contribution of negative ions to the blocks. (**d**) Cartoon of the cap-to-cap structure of $(\alpha7)_2$ showing that there is no entrance to the *cis* blocking site of $P_i^-$ ions as in α7. Hence $(\alpha7)_2$ should not be blocked by $P_i^-$ ions. (**e**) A representative trace of two $(\alpha7)_2$ channels in the presence of $P_i^-$ ions added to the *cis* side. Neither of the channels showed $P_i^-$ binding at +20 mV. (**a**, **d**) Charges written on the *cis* and the *trans* sides of the bilayer indicate the polarity of the respective electrodes at +20 mV. (**b**, **c**, **e**) The dotted lines represent the $I_o$ levels at +20 mV (red) and -20 mV (blue) and, the blocked level (black) that coincides with the 0 pA level. Green traces below the current traces (black) show the applied potential profile. The experiments were performed in 1 M KCl, 25 mM Tris.HCl, 50 µM EDTA, pH 8.0, with 10 mM sodium phosphate as the source of $P_i^-$ ions.

14 out of the 17 M113R $(\alpha7)_2$ channels, i. e., 82±18% (proportion of channels±95% confidence level), that showed $P_i^-$ blocking had $I_o$ similar to α7 (Fig. 2.19). These results suggest that the $(\alpha7)_2$ channels, which were blocked

by $P_i^-$, had misaligned or misfolded *cis*-α7 units such that $P_i^-$ ions could enter into *trans*-α7 from the *cis* side (Fig. 2.10a-d).

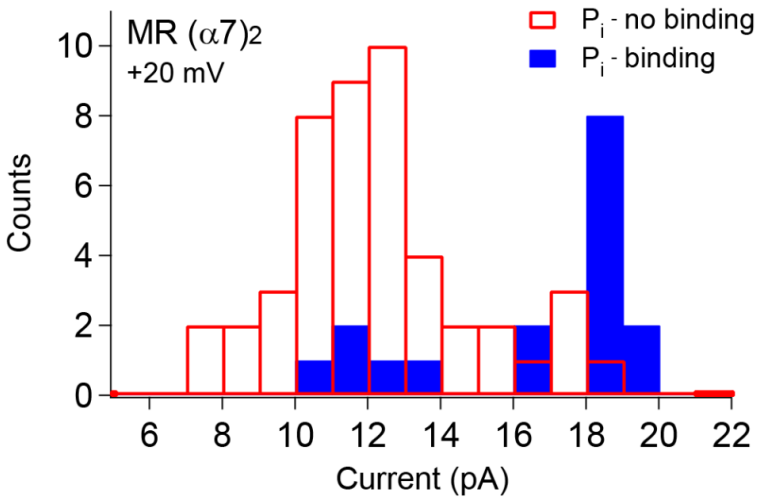



**Figure 2.19. $P_i^-$ blocking of the M113R $(\alpha7)_2$ mutant.** In a perfect cap to-cap arrangement of the *cis*- and *trans*-α7 in the M113R $(\alpha7)_2$ mutant, $(\alpha7)_2$ should not be blocked by $P_i^-$ ions as the cap entrance of *trans*-α7 would be blocked by *cis*-α7. $I_o$ histograms of M113R $(\alpha7)_2$ channels which were blocked (filled blue bars) and which were not blocked (red bars) by $P_i^-$ ions when added to the *cis* side. 17 of 65 channels (*n*=65) were blocked by $P_i^-$ ions. 14 of these 17 channels had $I_o$ similar to that of α7. The pores, which showed $P_i^-$ blocking, could be attributed to structures **a-d** (Fig. 2.10) where the cap entrance of the *trans*-α7 was not blocked by *cis*-α7.

## 2.14. Effect of freeze-thawing on $I_o$ and $P_i^-$ blocking characteristics

The afore mentioned $P_i^-$ blocking experiments were done with an aliquot of M113R $(\alpha7)_2$, which had not been freeze-thawed more than once, termed as 'no freeze-thawed'. To check whether multiple cycles of freeze-thaw led to misfolded $(\alpha7)_2$ pores, the $I_o$ values and $P_i^-$ blocking experiments were done using M113R $(\alpha7)_2$ aliquots which had been freeze-thawed three times (3x freeze-thawed). It was found that the effect of freeze-thawing M113R $(\alpha7)_2$ was drastic. In $P_i^-$ blocking experiments with a 3x freeze-thawed M113R $(\alpha7)_2$ sample, more than 50% of the pores (14 of 25) showed $P_i^-$ binding events as compared to 26% of the pores from a sample that had not been freeze-thawed more than once. The $I_o$ histogram of pores that were blocked by $P_i^-$ was also

much broader in the case of 3x freeze-thawed M113R $(\alpha7)_2$ samples than the protein aliquot which had not been freeze-thawed multiple times (Fig. 2.20a). Further, the open pore current characteristics also differed after freeze-thawing multiple times. At +20 mV, the mean $I_o$ of a 3x freeze-thawed sample was shifted by +2 pA as compared to the 'no freeze-thawed' aliquots (Fig. 2.20b).

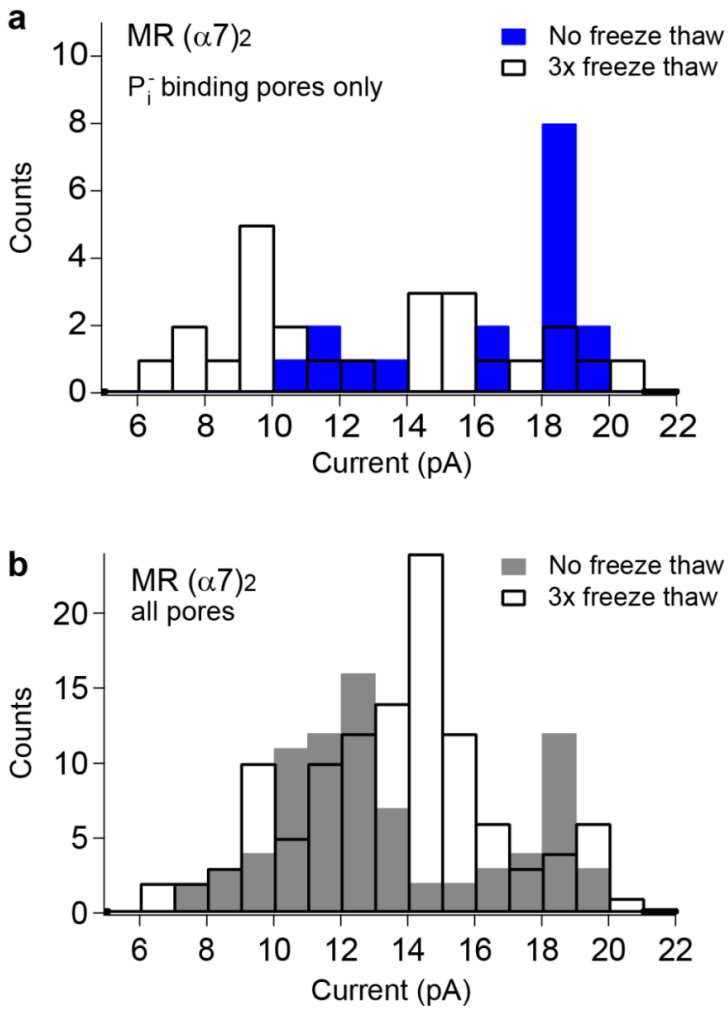



**Figure 2.20. Effect of freeze-thawing on $P_i^-$ binding characteristics and $I_o$ of M113R $(\alpha7)_2$.** (**a**) $I_o$ histograms of channels of 3 times freeze-thawed M113R (MR) $(\alpha7)_2$ aliquots (3x freeze-thaw) and aliquots which had not been freeze-thawed more than once (labelled 'No freeze-thaw') that showed *cis* $P_i^-$ blocking, at +20 mV. $P_i^-$ ion binding was observed in 56±19% (proportion of channels±95% confidence level, *n*=25) of 3x freeze-thaw aliquot channels compared to 26±11% (proportion of channels±95% confidence level, *n*=65) of No freeze-thaw aliquot channels (filled blue bars). (**b**) $I_o$ histograms of No freeze-thaw and 3x freeze-thaw aliquots of M113R $(\alpha7)_2$ channels. The distribution in the $I_o$ value of the 3x freeze-thaw $(\alpha7)_2$ channels was wider than the No freeze-thaw aliquot channels. The mean $I_o$ increased by ~2 pA upon multiple freeze-thawing cycles, at +20 mV.

## 2.15. Simultaneous insertion of $(\alpha7)_2$ in two bilayers

### 2.15.1. Single channel current measurements of $(\alpha7)_2$ inserted in a planar lipid bilayer and a liposome

To validate insertion in two bilayers electrically, we used a planar lipid bilayer set-up and liposomes. Pre-incubation with unilamellar liposomes (25 mg $ml^{-1}$ DPhPC, 1 M KCl, 25 mM Tris.HCl, 50 µM EDTA, pH 8.0; diameter cut off 100 nm) decreased the insertion efficiency of $(\alpha7)_2$ in planar lipid bilayers by >100 fold presumably because both the β-barrels of $(\alpha7)_2$ were already inserted in the liposomes. The insertion efficiency was estimated by comparing the average number of insertions of ~0.2 µl dimer protein (50 mg $L^{-1}$) in 20 min, pre-incubated with and without liposomes.

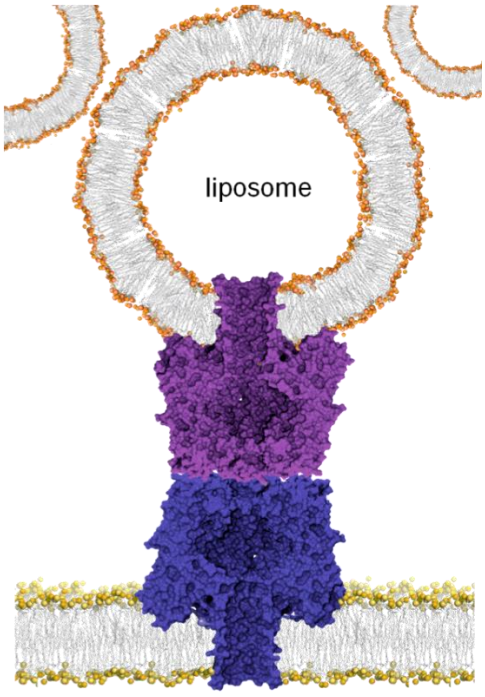



**Figure 2.21. Simultaneous insertion of $(\alpha7)_2$ in a planar lipid bilayer and a liposome.** Cartoon (not to scale) showing $(\alpha7)_2$ with its *cis*-α7 barrel inserted into a unilamellar liposome and the *trans*-α7 barrel inserted into a planar lipid bilayer. Insertion into two bilayers was observed in single-channel electrical recordings (Fig. 2.23).

On adding liposomes to the *cis* side, when $(\alpha7)_2$ was already inserted in a planar lipid bilayer, a permanent blockade of 23±1% (percentage of $I_o$±s.e.m.; in 12 out of *N*=14, where *N* is number of independent experiments) in $I_o$ was observed at +50 mV (Figs. 2.21, 2.23a). Blocks associated with liposome insertion were also observed at other potentials (-50 mV, +40 mV and +200 mV). $(\alpha7)_2$ inserted in two bilayers had lower conductance values than $(\alpha7)_2$ inserted in a single bilayer at all potentials (-200 to +200 mV) (Fig. 2.22a). However, the rectification properties in both the cases remained similar (Fig. 2.22b) (also see section 2.15.3).

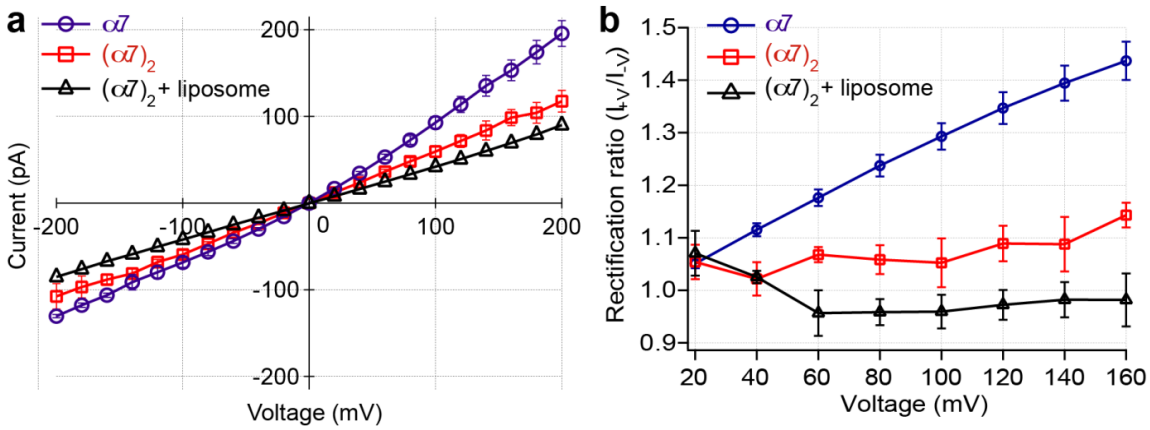



**Figure 2.22. Comparison of I-V characteristics of α7, $(\alpha7)_2$, and liposome/planar bilayer inserted $(\alpha7)_2$.** (**a**) I-V plot of $(\alpha7)_2$ whose two β-barrels were inserted in a planar lipid bilayer and a liposome, each (open black triangles). The current through $(\alpha7)_2$ inserted in two bilayers was less than the current of α7 (open blue circles) and $(\alpha7)_2$ (open red squares) in a planar bilayer, at both positive and negative potentials. (**b**) Plots of rectification ratios ($-I_{+V}/I_{-V}$) of α7, $(\alpha7)_2$, and $(\alpha7)_2$ in two bilayers as a function of voltage. The rectification ratio of α7 increased with increasing potential owing to an asymmetric charge distribution at its *cis* and *trans* pore openings (Chapter 3, section 3.8.9.7). The rectification was decreased by introducing symmetry in $(\alpha7)_2$. The rectification of $(\alpha7)_2$ inserted simultaneously in a planar bilayer and a liposome was similar to $(\alpha7)_2$ inserted in a planar bilayer only.

### 2.15.2. Cis γCD blocking events stopped upon insertion in liposome

To confirm that the observed block in $I_o$ of $(\alpha7)_2$ was because of the insertion of the *cis*-α7 barrel in a liposome, γCD was added to the *cis* side of the planar lipid

bilayer before and after $(\alpha7)_2$ had inserted in a liposome. Upon addition of γCD to the *cis* side after insertion of $(\alpha7)_2$ in a liposome, γCD binding events were not observed (all 6 of *N*=6), confirming that the *cis*-α7 barrel entrance was blocked by the liposome (Fig. 2.23b). If γCD was already present on the *cis* side before the addition of liposomes, *cis* γCD blocking events stopped upon insertion of the *cis*-α7 in a liposome (all 4 of *N*=4) (Fig. 2.23c). On the other hand, if γCD was present on the *trans* side before adding liposomes, *trans* γCD blocking continued even after the insertion of *cis*-α7 in a liposome. However, the magnitude of blocks in $I_o$ due to *trans* γCD binding changed from 60±1% before insertion to 56±1% after insertion (mean±s.e.m.; *n*=4144, *N*=5) (Fig. 2.23d, e). This showed that the liposome occluded the β barrel entrance of *cis*-α7 but the *trans*-α7 unit was conducting.

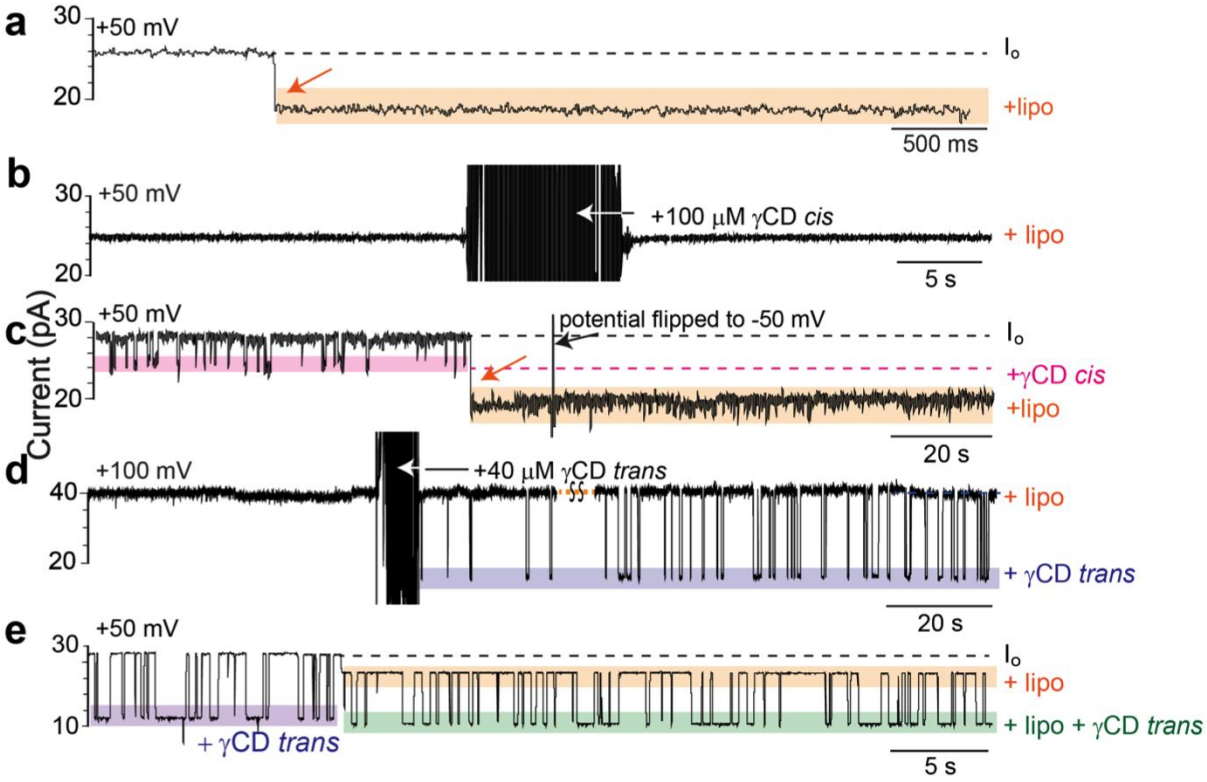



**Figure 2.23. Single channel electrical traces of $(\alpha7)_2$ pores inserted simultaneously in a liposome and a planar lipid bilayer.** Representative electrical traces of liposome insertion or liposome-inserted $(\alpha7)_2$ pores. (**a**)

Insertion of the *cis*-α7 barrel of $(α7)_2$ into a unilamellar liposome is observed as an irreversible block of $I_o$ of 23±1% (percentage block of $I_o$±s.e.m.; *N*=12) at +50 mV. (**b**) An $(α7)_2$ pore that had inserted into a liposome did not display any *cis* γCD (100 μM) binding events at +50 mV. (**c**) C*is* γCD blockades cease after insertion into a liposome at +50 mV. The liposome block could not be reversed by flipping the potential to -50 mV. (**d**) Upon the addition of γCD (40 μM) to the *trans* side of an $(α7)_2$ pore that had inserted into a liposome, transient binding events (purple box) were detected (+100 mV). (**e**) If γCD was present on the *trans* side of $(α7)_2$ before liposome insertion, *trans* γCD binding events (purple box) continued (green box) after liposome insertion at +50 mV. The experiments were performed in 1 M KCl, 25 mM Tris.HCl, 50 μM EDTA, pH 8.0, with 80 μM γCD (*cis* or *trans*) and/or liposomes (2.5 mg $ml^{-1}$, *cis*), as noted.

### 2.15.3. $(α7)_2$ is conducting even after insertion in a liposome: possible explanations

γCD experiments showed that the *cis* entrance of $(α7)_2$ was blocked by the liposome. However, only a 23±1% (percentage block of $I_o$±s.e.m.; *N*=12) blockade was seen in the $I_o$ of $(α7)_2$ upon inserting in a liposome. The residual current ($I_r$) after insertion of the *cis*-α7 barrel in a liposome could be explained by assuming a porous liposome with other $(α7)_2$ pores (Fig. 2.24). $(α7)_2$ pores puncturing the liposome would not allow large molecules like γCD to pass through and hence, *cis*-γCD binding events would not be observed after the liposome insertion event. Alternatively, the $I_r$ could be due to a leak current arising from a flow of ions in between the cap-domains of the two α7 units or between the α1 subunits of *cis*-α7.

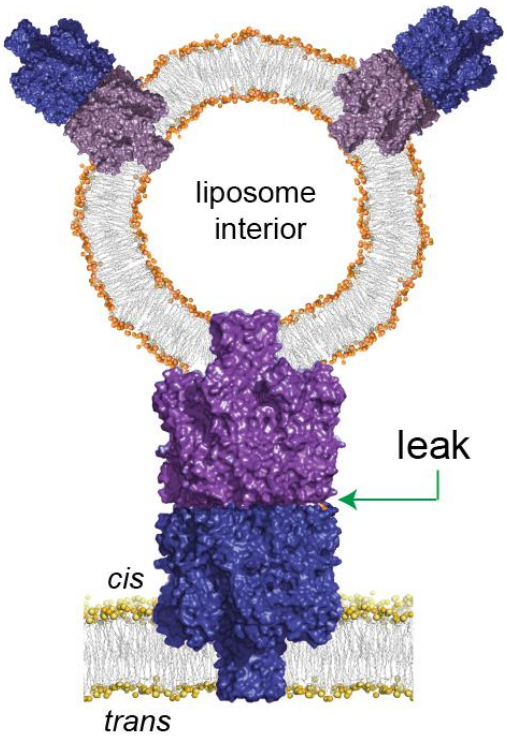



**Figure 2.24. Molecular model of $(\alpha7)_2$ inserted in a planar bilayer and a porous liposome.** The residual conductance of an $(\alpha7)_2$ pore inserted into both a planar bilayer and a liposome could be due to a leak current or additional $(\alpha7)_2$ pores in the liposome, which make it porous (cartoon not to scale).

To explain the empirical current values of $(\alpha7)_2$, a theoretical model was developed (section 2.18). The theoretical model was also used to rigorously examine both the aforementioned possible causes of the $I_r$ of $(\alpha7)_2$ after inserting in a liposome (section 2.18.2). The theoretical model suggested that the $I_r$ of $(\alpha7)_2$ was due a leak current bypassing the liposome blocked *cis*-α7 into the *trans*-α7 unit.

## 2.16. Electron microscopy of $(\alpha7)_2$

The cap-to-cap structure of $(\alpha7)_2$ was further ascertained by performing TEM of $(\alpha7)_2$ and α7. α7 was observed as ring shaped particles with an average diameter of 8.1±0.1 nm (mean±s.e.m., *n*=52) (data not shown) that correlates well with some previous reports (Füssle *et al.*, 1981; Song *et al.*, 1996). In

contrast, both the side and ring views of $(\alpha7)_2$ were visible (*n*=399, *N*=4) (Fig. 2.25a, b). As expected, the ring views of $(\alpha7)_2$ were similar to α7 with a heptamer arrangement and an average diameter of 7.8±0.1 nm (*n*=131) (Fig. 2.25a). Elongated particles were also observed in the $(\alpha7)_2$ protein sample with an average length of 19.6±0.2 nm (mean±s.e.m., *n*=268) and width 8.4±0.1 nm (mean±s.e.m.; *n*=96) (Fig. 2.25b). These values are in good agreement with the expected side view dimensions of a cap-to-cap $(\alpha7)_2$ with length (≈ 20 nm) and width (≈8.5 nm) (Fig. 2.3c).

It should be noted that due to particle clustering (Fig. 2.25a) in TEM, we were not able to distinguish and evaluate the population distribution between the various $(\alpha7)_2$ structures (Fig. 2.10). The side and top view dimensions of discreet particles only (Fig. 2.25a, b *black markers*) were measured (Fig. 2.25c).

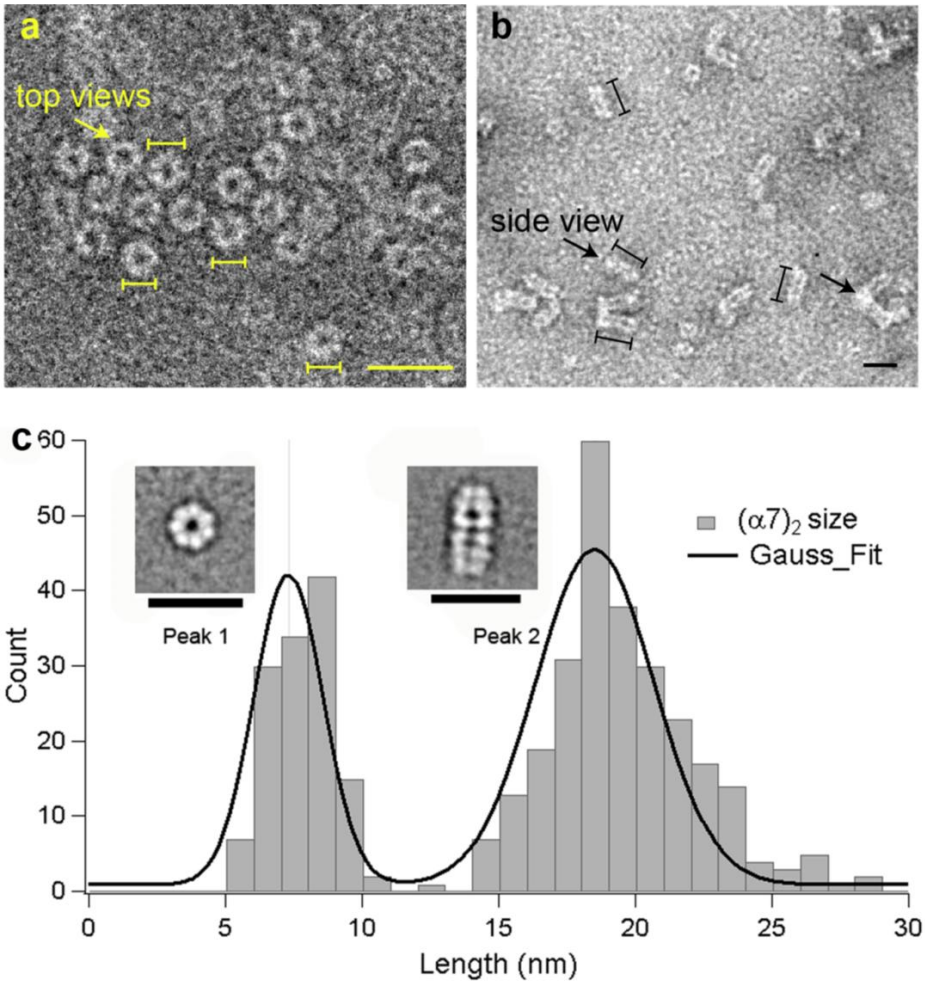



**Figure 2.25. TEM of $(\alpha7)_2$ particles.** (**a**) Uranyl acetate stained TEM images showing the ring views of $(\alpha7)_2$. The average diameter of the particles was 7.8±0.1 nm (mean±s.e.m.; *n*=131). (**b**) Electron micrographs showing side views (*left* arrow) of $(\alpha7)_2$ as elongated particles of 19.6±0.2 nm (mean±s.e.m., *n*=268) and width 8.4±0.1 nm (mean±s.e.m.; *n*=96). $(\alpha7)_2$ molecules were also observed in clusters of three to four molecules associated at their ends, presumably at their transmembrane regions (*right* arrow). Lines drawn next to single particles (**a**: black and **b**: yellow) denote the boundaries used to measure lengths. (**c**) Histogram showing distributions of single particles of $(\alpha7)_2$. Insets show the class average of 309 single particles showing the side view and 224 single particles showing the ring view. Scale bars are 20 nm. The dimensions of single particles observed in EM correlated well with the length (≈20 nm) and width (≈8.5 nm) of $(\alpha7)_2$ in a molecular model (Fig. 2.3c).

TEM showed that $(\alpha7)_2$ could insert in single bilayers of POPC liposomes (Fig. 2.26a). Liposome inserted $(\alpha7)_2$ particles were observed as 17.0±0.4 nm (mean±s.e.m.; *n*=131) protrusions from the liposome surface (Fig. 2.26a). To show that $(\alpha7)_2$ could insert in two bilayers, $(\alpha7)_2$ was mixed with 10 mol%

POPG: POPC liposomes (Fig. 2.26b) and egg-PC liposomes (Fig. 2.26c) in lipid to protein ratios of 20:1 (w/w) and 100:1 (w/w), respectively. Insertion in single liposomes and simultaneous insertion in two liposomes were observed (Fig. 2.26). The average length of $(\alpha 7)_2$ particles inserted in two liposomes was 12.3±0.7 nm (mean±s.e.m.; *n*=15).

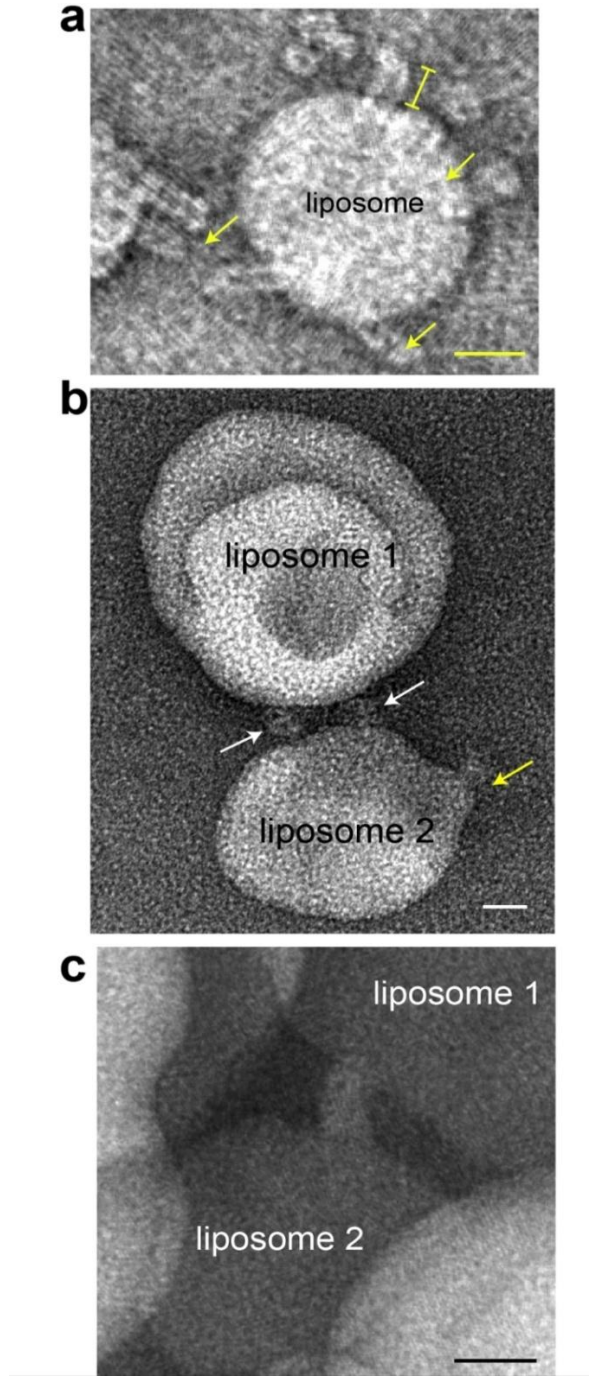



**Figure 2.26. TEM of $(\alpha 7)_2$ particles inserted in a single bilayer and two bilayers simultaneously.** (**a**) TEM view showing $(\alpha 7)_2$ inserted into the bilayer

of a liposome and seen as 17.0±0.4 nm (mean±s.e.m.; *n*=131) protrusions (yellow) from the liposome surface. (**b**, **c**) TEM images of $(\alpha7)_2$ inserted into two liposomes simultaneously. $(\alpha7)_2$ was mixed with (**b**) 10 mol% POPG: POPC liposomes or (**c**) egg-PC liposomes in lipid to protein ratios of 20:1 and 100:1 (w/w) respectively. Scale bars are 20 nm.

### 2.17. Imipramine transfer across two proximal bilayers connected by $(\alpha7)_2$

Single channel recordings suggested that $(\alpha7)_2$ inserted in a planar lipid bilayer could insert in a unilammelar liposome simultaneously. TEM confirmed this and the length and width of protein particles connecting two neighbouring liposomes matched with the dimensions of the molecular model of $(\alpha7)_2$ (Fig. 2.3c). To show that analytes could be transferred across two bilayers connected *via* $(\alpha7)_2$, single channel measurements of $(\alpha7)_2$ pores inserted in a planar lipid bilayer and a liposome containing analytes were performed (Fig. 2.27a).

Imipramine, a drug molecule, was chosen as the analyte (Fig. 2.27b). Imipramine is a small molecule whose translocation through α7 cannot be detected electrically unless a host molecule such as γCD is present in the pore (Gu *et al.*, 1999). Imipramine can bind to the hydrophobic core of γCD, which can be seen as an extra current block arising from the γCD blocked current level (Fig. 2.28a).

Imipramine (1 mM) was encapsulated in DPhPC liposomes (25 mg $ml^{-1}$; 1 M KCl, 25 mM Tris.HCl, 50 µM EDTA, pH 8.0). Imipramine containing liposomes were diluted 10 fold in the *cis* side of $(\alpha7)_2$ inserted in a planar lipid bilayer. γCD (80 µM) was added to the *trans* side of the planar bilayer to detect imipramine translocation through $(\alpha7)_2$ once it inserted in a liposome (Fig. 2.27a).

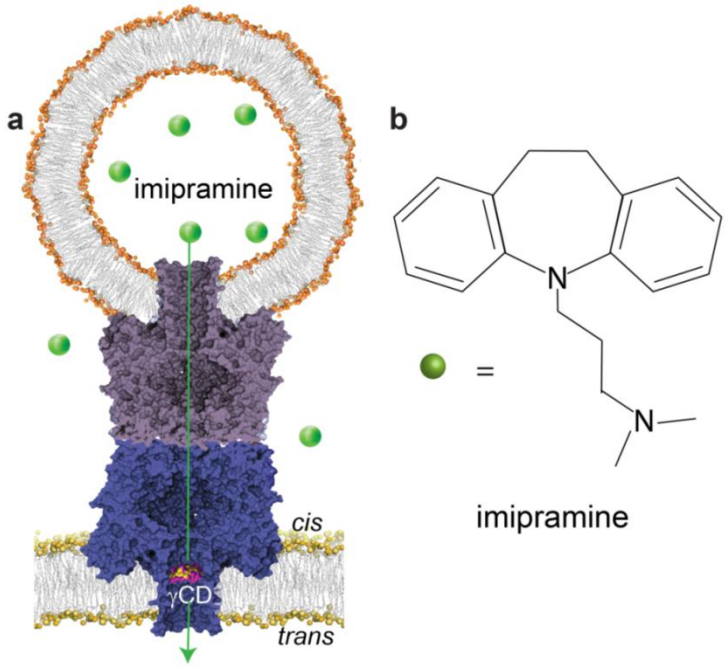



**Figure 2.27. Detection of small molecule transfer between two bilayers *via* $(\alpha7)_2$.** (**a**) Cartoon shows $(\alpha7)_2$ inserted in a planar lipid bilayer and a liposome containing the small drug molecule imipramine. Imipramine translocation could be detected using γCD added to the *trans* side of the planar bilayer. (**b**) Structure of imipramine suggests its amphiphilic nature. Some imipramine could leak out of liposomes as it can intercalate with bilayers (Ahyayauch *et al.*, 2004).

The non-entrapped imipramine in the liposomes buffer was removed using a Sephadex G25 size exclusion column prior to each experiment. However, imipramine translocation events were observed as short spikes in the *trans* γCD bound state as soon as liposomes were added into the recording chamber (Fig. 2.28a, b). Like some other amphiphiles, imipramine can intercalate with bilayers and has been shown to make liposomes porous to small molecules (Ahyayauch *et al.*, 2004). Hence, imipramine events observed before insertion of $(\alpha7)_2$ in liposome could be attributed to the imipramine that had leaked out of liposomes ($Im_{leak}$) gradually with time (Fig. 2.28a, b) (also see section 2.17.2).

After an imipramine loaded liposome inserted in the barrel of *cis*-α7, imipramine diffusion from the liposome into *trans*-α7 led to a sudden increase in the number of imipramine translocation events (3 out of *N*=6). This was deduced by kinetic analysis of the mean inter-event time ($\tau_{on}$), which decreased ~1.3 – 2.2 fold upon insertion of the imipramine containing liposome in the three successful cases (Figs. 2.28, 2.29, Table 2.2).

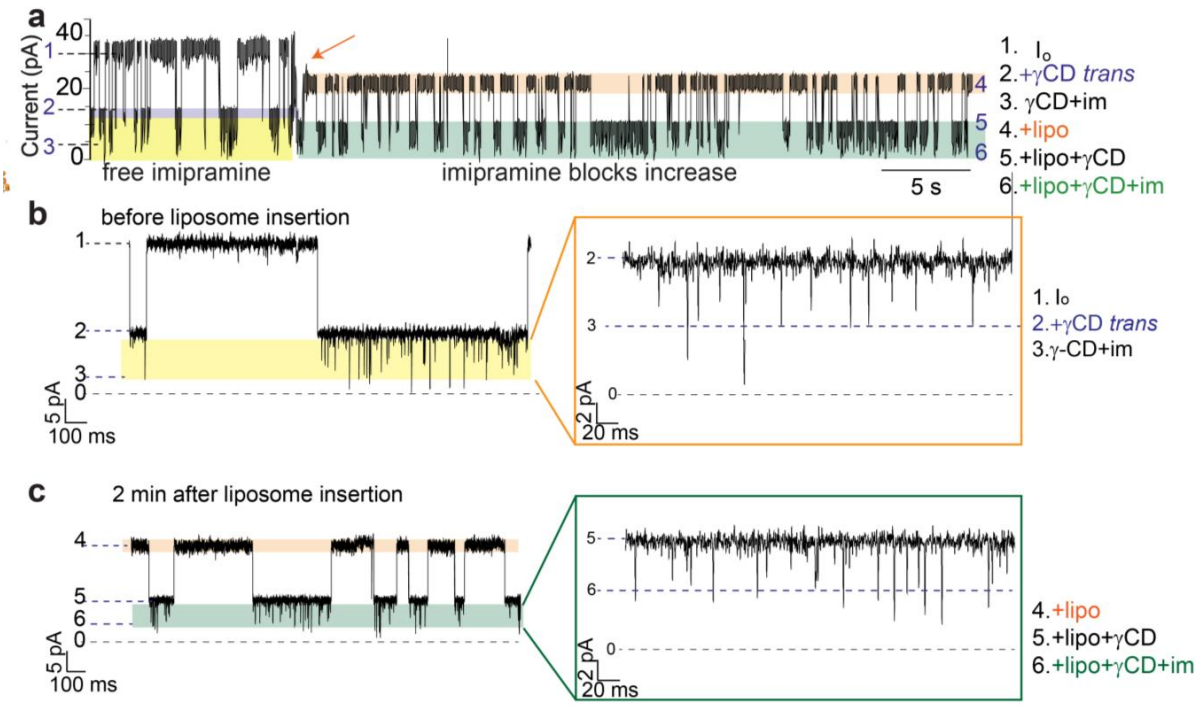



**Figure 2.28. Single channel measurements of imipramine translocation across two bilayers *via* $(\alpha7)_2$.** Liposomes containing imipramine were added to the *cis* side and γCD was added to the *trans* side of a planar lipid bilayer. (**a-c**) Representative electrical traces of $(\alpha7)_2$ showing imipramine translocation as short spikes within the γCD blocked level. (**a**) Imipramine translocation events, observed because of $Im_{leak}$ increased after insertion of $(\alpha7)_2$ in the liposome, implying transfer of imipramine from the liposome to the *trans* side of the planar bilayer. (**b**) and (**c**) show ~10x and ~50x zoomed electrical traces of $(\alpha7)_2$ before and after insertion in the liposome containing imipramine respectively. The increase in imipramine translocation events was quantified from the mean inter-event time ($\tau_{on}$) (Fig. 2.29).

In each of the six experiments, the $\tau_{on}$ of imipramine translocation just before insertion of the $(\alpha7)_2$ in the liposome (termed t=0) was measured from 1 min of

the electrical recording trace and found to be different in each case (section 2.17.1) (Table 2.2).

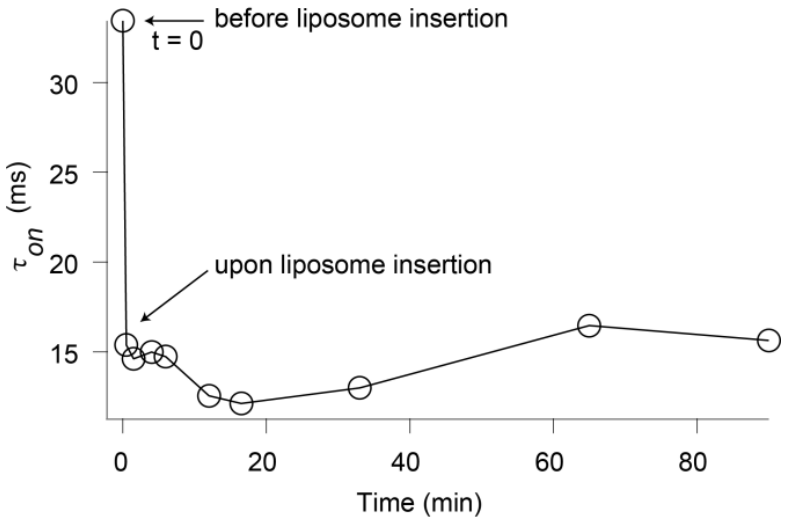



**Figure 2.29. Imipramine binding before and after liposome insertion.** Plot shows $\tau_{on}$ just before (determined from 1 min of recorded data; marked t=0) and after different times of insertion of $(\alpha 7)_2$ in the liposome.

**Table 2.2** Table lists the $\tau_{on}$ values of imipramine blocking calculated from electrical traces 1 min before and after the liposome block event.

| | $\tau_{on}$ (ms) before insertion in liposome | $\tau_{on}$ (ms) after insertion in liposome |
|---|---|---|
| 1 | 33 | 15 |
| 2 | 50 | 31 |
| 3 | 15 | 10 |
| 4 | 26 | 21 |
| 5 | 35 | 30 |
| 6 | 62 | 62 |

### 2.17.1. Variation in imipramine blocking prior to liposome insertion over different experiments

The variation in imipramine translocation events before liposome insertion could be because of different amounts of imipramine encapsulated in the liposomes and hence, the $Im_{leak}$ in the outer buffer in different experiments. The concentration of $Im_{leak}$ ($[Im]_{leak}$) would also depend on the time imipramine can diffuse from the liposomes into the outer solution, i.e., the time between purification of liposomes ($[Im]_{leak}=0$) and insertion of $(\alpha7)_2$ in a liposome.

Being a stochastic event, $(\alpha7)_2$ insertion in a liposome was observed after different times of liposomes addition (16 s – 2.5 h) and hence, the $[Im]_{leak}$ just after the addition of liposomes ($\tau_{on}$ ranging from 33 - 62 ms) and just before the insertion of $(\alpha7)_2$ in a liposome ($\tau_{on}$ decreased to 15 – 62 ms) was different in each experiment ($N$=6) (Table 2.2). In 3 out of 6 experiments, a sudden increase in imipramine binding could be associated with liposome insertion, which occurred after 16 s – 30 min of liposome addition. In the three other cases, where there was no effect of the liposome insertion on imipramine binding, the $(\alpha7)_2$ had inserted in the liposomes after 3 – 16 h of purification and subsequent addition during which the imipramine concentration could have equilibrated between the liposomes and buffer. Liposomes could also burst or fuse into the bilayer releasing their contents.

### 2.17.2. Detecting γCD entrapped in liposomes from single channel electrical measurements of $(\alpha7)_2$ inserted in two bilayers

Since imipramine is a small amphiphilic molecule, it could leak out of liposomes before $(\alpha7)_2$ inserted in a liposome. Hence, increased translocation of imipramine from within a liposome could be clearly observed in only 50% of the experiments (*N*=6).

Due to the small volume of the liposome (~1 attoliter in a 100 nm diameter liposome), only a limited number of analyte molecules can be trapped and hence, only a limited number of translocation events can be observed. To observe continuous translocation of analyte molecules, a very high concentration of analytes needs to be encapsulated within the liposomes. An analyte such as γCD could be used such that it did not translocate across $(\alpha7)_2$ but its binding could be observed electrically. Another advantage of using γCD was that it couldn't leak out of liposomes even if there were other $(\alpha7)_2$ pores puncturing the liposome (Fig. 2.24) or enter through the leak pathways in $(\alpha7)_2$ (Fig. 2.10e).

γCD (10 mM) was encapsulated in DPhPC liposomes (25 mg $ml^{-1}$; 1 M KCl, 25 mM Tris.HCl, 50 µM EDTA, pH 8.0). As the liposome would stabilize the *cis*-α7 barrel, upon liposome insertion, I expected to observe γCD binding to the *cis*-α7 barrel like *trans* γCD blockade events. However, in two out of four single channel electrical measurement experiments, $(\alpha7)_2$ did not insert in the γCD loaded liposomes. In the other two experiments where insertion in a liposome was observed, γCD binding was not detected suggesting that the liposome was

electrically insulated (Gu *et al.*, 2003). This implied that other $(\alpha7)_2$ pores did not insert in the liposome making it porous (Fig. 2.24) and that the residual conductance of $(\alpha7)_2$ after liposome insertion was due to a leak current (sections 2.15.3, 2.18.2).

### 2.17.3. Alternative experiments to demonstrate small molecule transfer across two bilayers using $(\alpha7)_2$

Previous experiments suggested that the interior of the liposome connected to a planar lipid bilayer by $(\alpha7)_2$ was electrically insulated (section 2.17.2). This implied that the transfer of analytes from the liposome through the $(\alpha7)_2$ pore would only be due to diffusion. Charged analytes could be transferred across two bilayers much faster than free diffusion through $(\alpha7)_2$ if a potential were applied between the two bilayers. For example, an electric potential could be applied between two giant unilamellar vesicles (GUV) or multisomes (Villar *et al.*, 2011) connected *via* $(\alpha7)_2$ (Fig. 2.30a). Apart from electrical detection, analyte transfer could also be detected optically with immobilized GUVs (Fig. 2.30b). These experiments would be a direct demonstration of the proposed minimal tissue model made of vesicles or multisomes using $(\alpha7)_2$ (Fig. 2.2) (section 2.1).

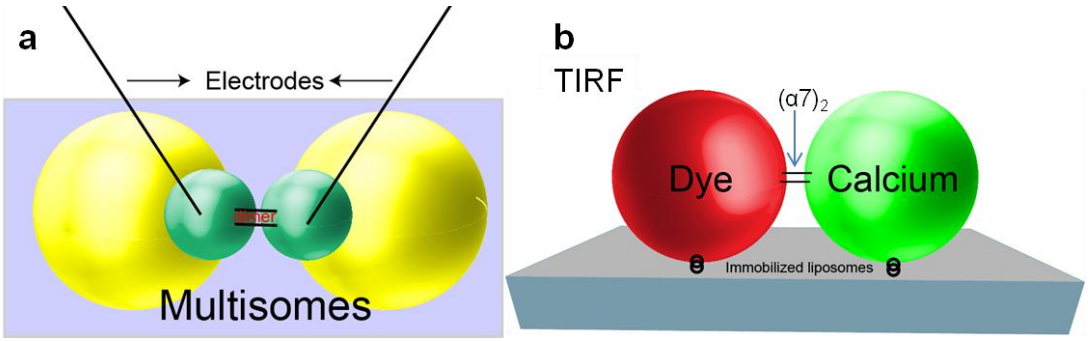



**Figure 2.30. Plausible experiments to demonstrate $(\alpha7)_2$ as a conducting unit between two bilayers.** Figure shows two possible experiments which

could unequivocally prove that small analytes can be transferred across $(\alpha7)_2$ simultaneously inserted in two bilayers. (**a**) Lipid containing aqueous droplets can be encased in an oil droplet to form a bilayer at the interface of the aqueous and the oil droplets (termed multisomes) (Villar *et al.*, 2011). By applying an electric potential between two multisomes connected *via* $(\alpha7)_2$, transfer of small molecules from one multisome to the other through $(\alpha7)_2$ could be detected electrically. (**b**) Similarly, GUVs could be tethered to a surface and inter-connected by $(\alpha7)_2$. Transfer of molecules (like $Ca^{2+}$) from one GUV to the other could be detected optically, for example using total internal reflection fluorescence microscopy (TIRF).

### 2.18. Theoretical model of $(\alpha7)_2$ explained its electrical properties

To explain the electrical properties of $(\alpha7)_2$ and the current blockades observed with $\gamma$CD and liposomes, an electrical model of the pore was built. The model assumes that the *cis*-α7 and *trans*-α7 units are resistors in series, represented by $R_c$ and $R_t$, respectively (Fig. 2.31a, b). The orientation of the *trans*-α7 unit is exactly like α7. Hence, the value of $R_t$ was set to experimentally determined value of the unitary resistance of α7 at +50 mV. The value for $R_c$ was taken as the unitary resistance of α7 at -50 mV, because the orientation of the *cis*-α7 unit is the opposite of that of *trans*-α7, and the opposite of α7 in the usual sign convention. The average resistance values (R) were determined from equation (2.3) and rounded to two significant digits.

$$\mathrm{R} = \frac{1}{n}\sum_{k=1}^{n}\frac{\mathrm{V}}{(\mathrm{I_o})_\mathrm{k}} \qquad (2.3)$$

Where V is the applied potential, $I_o$ is the open pore current and *n* is the total number of single channels used to calculate the average. At +50 mV, *n*=60 for $R_c$ and *n*=165 for $R_t$.

$R_c$ and $R_t$ are both 1.2 GΩ at +50 mV, because the rectification of α7 is weak at low potentials (Fig. 2.13). The resistance of $(α7)_2$ ($R_d$) should be the combined resistance of $R_c$ and $R_t$. The theoretical value of $R_d$ was then calculated by using the equation (2.4) to be equal to 2.4 GΩ.

$$R_d = R_t + R_c \; ; \; R_d = 2.4 \; G\Omega \tag{2.4}$$

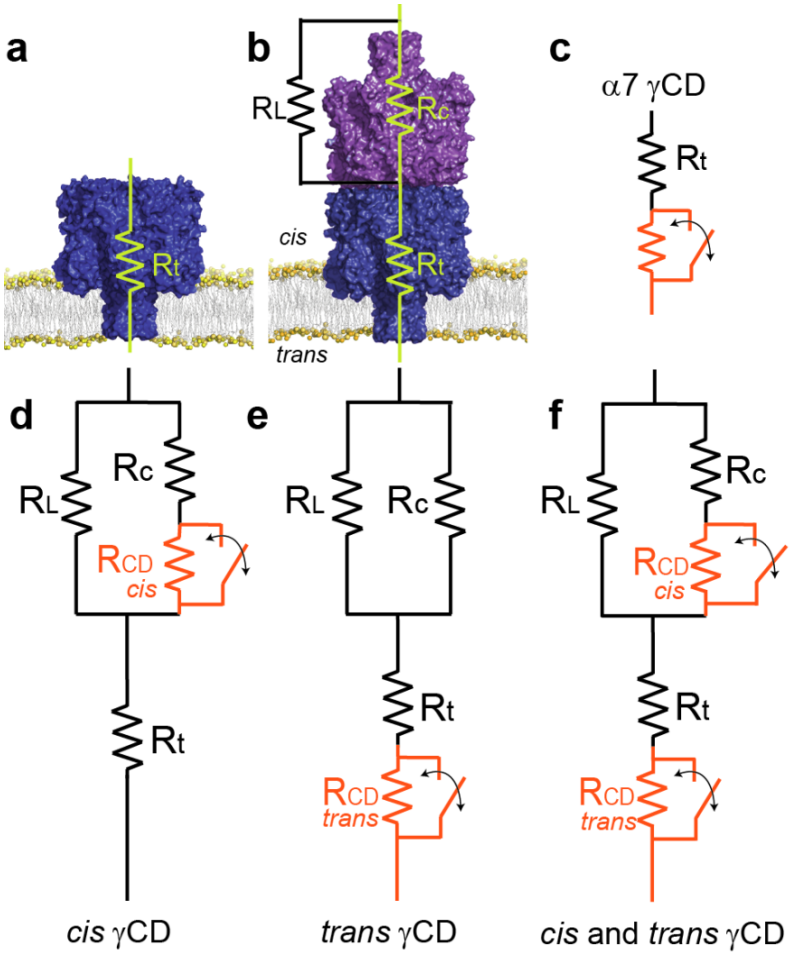



**Figure 2.31. Electrical model of $(α7)_2$.** (**a, b**) Representations of α7 and $(α7)_2$ as electrical resistors. The *cis*- and *trans*-α7 units of $(α7)_2$ were considered as resistors in series, $R_c$ and $R_t$, respectively. An additional 'leak' resistance ($R_L$) was included in parallel with $R_c$ to account for other conductive pathways in the structure. (**c**) α7 blocked with γCD can be represented as a resistor in series with an additional resistor $R_{CD}$ (orange). The switch accounts for the transient nature of the γCD blockade. Electrical models of $(α7)_2$ when γCD is present on (**d**) the *cis* side only, (**e**) the *trans* side only, and (**f**) on both the *cis* and *trans* sides.

The experimentally determined resistance of the cap-to-cap $(\alpha7)_2$ pores ($R_d$) was 1.8 GΩ (*n*=240; equation (2.3)). This implies that a leak in the structure of $(\alpha7)_2$ adds to the measured conduction through the $(\alpha7)_2$ pore (section 2.10). The leak could be due to flow of ions in between the *cis*-α7 and *trans*-α7 units or through small aqueous inlets between the α1 subunits (Fig. 2.10e). To account for the leak current, an additional resistor ($R_L$) was placed in parallel to $R_c$ (Fig. 2.31b). By using the experimentally determined values of $R_d$=1.8 GΩ, $R_c$=1.2 GΩ and $R_t$=1.2 GΩ, the value of $R_L$ was determined to be 1.2 GΩ at +50 mV from equation (2.5).

$$R_d = R_t + \frac{R_L \times R_c}{R_L + R_c} \quad \text{or} \quad R_L = \frac{R_c(R_d - R_t)}{R_d - R_c - R_t} \tag{2.5}$$

### 2.18.1. Electrical model for γCD blocking

Current blockades by γCD can be explained with the same electrical model by including an additional resistance $R_{CD}$ ($R_{CD,\ cis}$ or $R_{CD,\ trans}$) in series (Fig. 2.31c). A switch is included to denote the transient binding of γCD, i.e., the pore with γCD bound has the switch open, while the unoccupied pore has the switch closed (Fig. 2.31c-f). At +50 mV, γCD binding decreases the $I_o$ of α7 by 66% and at -50 mV by 58% (Table 2.1), i. e., the current through the γCD blocked pore was 34% of the α7 $I_o$ ($I_{res,\ \gamma CD\ trans}$) at +50 mV and 42% of $I_o$ of α7 ($I_{res,\ \gamma CD\ cis}$) at -50 mV. Based on these values, the resistance of the γCD blocked α7 pore was calculated at +50 mV ($R\gamma_{CD\ trans}$) and -50 mV ($R\gamma_{CD\ cis}$) to two significant figures. In recordings of $(\alpha7)_2$ at +50 mV, the *trans*-α7 unit is in the orientation of α7 at +50 mV, according to our sign convention, but the *cis*-α7

unit is equivalent to α7 at –50 mV. The values of $R_{CD}$ were determined from experimental data of γCD binding to α7 (Fig. 2.31c) at +50 mV ($R_{CD,\ trans}$) and -50 mV ($R_{CD,\ cis}$) using equations (2.6) and (2.7) respectively.

$$\text{at } +50 \text{ mV}, \qquad R_{\gamma CD\ trans} = \frac{V}{I_{res,\ \gamma CD\ trans}} = 3.5\ G\Omega$$

$$R_{\gamma CD\ trans} = R_{CD,\ trans} + R_t \tag{2.6}$$

$$\text{at } -50 \text{ mV}, \qquad R_{\gamma CD\ cis} = \frac{V}{I_{res,\ \gamma CD\ cis}} = 2.9\ G\Omega$$

$$R_{\gamma CD\ cis} = R_{CD,\ cis} + R_c \tag{2.7}$$

Hence, $R_{CD,\ trans}$=2.3 GΩ and $R_{CD,\ cis}$=1.7 GΩ.

The expected magnitudes of blockades for *cis* and *trans* γCD binding to $(\alpha7)_2$ can be calculated by using values of $R_L$ (equation (2.5)), $R_{CD}$ (equations (2.6) and (2.7)), and equations (2.8)-(2.13) (Fig. 2.31d-f). $R_{CD,\ cis}$ was used for *cis* γCD binding to $(\alpha7)_2$ and $R_{CD,\ trans}$ was used for *trans* γCD binding to $(\alpha7)_2$.

For *cis* γCD binding, the resistance of the *cis* γCD blocked $(\alpha7)_2$ pore ($R_{d,\ \gamma CD\ cis}$) was calculated from equation (2.8) and blockade of $I_o$ ($B_{\gamma CD\ cis}$) could be thus calculated obtained from equation (2.9).

$$R_{d,\ \gamma CD\ cis} = R_t + \frac{R_L \times (R_c + R_{CD,\ cis})}{R_L + R_c + R_{CD,\ cis}} \tag{2.8}$$

$R_{d,\ \gamma CD\ cis}$=2.1 GΩ

$$B_{\gamma CD\ cis} = \frac{\left(I_d - I_{d,\ \gamma CD\ cis}\right)}{I_d} \times 100 = \frac{\left(\frac{V}{R_d} - \frac{V}{R_{d,\ \gamma CD\ cis}}\right)}{\frac{V}{R_d}} \times 100$$

$$B_{\gamma CD\ cis} = \frac{R_{d,\ \gamma CD\ cis} - R_d}{R_{d,\ \gamma CD\ cis}} \times 100 \qquad (2.9)$$

$B_{\gamma CD\ cis}$=14%

For *trans* γCD binding, the resistance of the *trans* γCD blocked $(\alpha 7)_2$ pore ($R_{d,\ \gamma CD\ trans}$) and the blockade of $I_o$ ($B_{\gamma CD\ trans}$) was calculated from equations (2.10) and (2.11).

$$R_{d,\ \gamma CD\ trans} = R_t + R_{CD,\ trans} + \frac{R_L \times R_c}{R_L + R_c} \qquad (2.10)$$

$R_{d,\ \gamma CD\ trans}$=4.1 GΩ

$$B_{\gamma CD\ trans} = \frac{\left(I_d - I_{d,\ \gamma CD\ trans}\right)}{I_d} \times 100 = \frac{\left(\frac{V}{R_d} - \frac{V}{R_{d,\ \gamma CD\ trans}}\right)}{\frac{V}{R_d}} \times 100$$

$$B_{\gamma CD\ trans} = \frac{R_{d,\ \gamma CD\ trans} - R_d}{R_{d,\ \gamma CD\ trans}} \times 100 \qquad (2.11)$$

$B_{\gamma CD\ trans}$=56%

When γCD was present in both the *cis* and *trans* sides, the resistance of the double γCD blocked $(\alpha7)_2$ pore ($R_{d,\ \gamma CD\ cis\text{-}trans}$) and blockade in $I_o$ ($B_{\gamma CD\ cis\text{-}trans}$) was computed using equations (2.12) and (2.13).

$$R_{d,\ \gamma CD\ cis-trans} = R_t + R_{CD\ trans} + \frac{R_L \times (R_c + R_{CD\ cis})}{R_L + R_c + R_{CD\ cis}} \quad (2.12)$$

$R_{d,\ \gamma CD\ cis\text{-}trans}$=4.4 GΩ

$$B_{\gamma CD\ cis-trans} = \frac{(I_d - I_{d,\ \gamma CD\ cis-trans})}{I_d} \times 100 = \frac{\left(\frac{V}{R_d} - \frac{V}{R_{d,\ \gamma CD\ cis-trans}}\right)}{\frac{V}{R_d}} \times 100$$

$$B_{\gamma CD\ trans} = \frac{R_{d,\ \gamma CD\ cis-trans} - R_d}{R_{d,\ \gamma CD\ cis-trans}} \times 100 \quad (2.13)$$

$B_{\gamma CD\ cis\text{-}trans}$=59%

At +50 mV, the percentage reductions in $I_o$ due to *cis* and *trans* γCD binding are 14% and 56%, respectively (Fig. 2.31d, e; Table 2.3), which match the experimental values of 15±1% and 55±2% (Table 2.1). Similarly, the calculated current blockade for simultaneous binding of γCD to both the *cis* and *trans* sites is 59% (Fig. 2.31f), which is similar to the experimental value of 61±1% at +50 mV.

### 2.18.2. Electrical model explaining current block due to liposome insertion

The residual conductance of $(\alpha7)_2$ upon insertion of its *cis*-α7 in a liposome could be explained in two scenarios: (i) the residual conductance was due to

conducting leak pathways in $(\alpha7)_2$ and the *cis*-α7 opening was completely blocked. (ii) Alternatively, the *cis*-α7 barrel could insert into a liposome already hosting $(\alpha7)_2$ pores (Fig. 2.24). In this case, *cis*-α7 would still conduct. Although γCD in liposomes experiments suggested otherwise (section 2.15.2), favouring the model which considers a leak in the structure of $(\alpha7)_2$, both the afore mentioned possibilities were assessed using the electrical model.

**Residual conductance due to a porous liposome.** If we assumed that the leak current was due to a malformed β barrel for the *cis*-α7 unit, which might be the consequence of the absence of a lipid bilayer, upon insertion in liposome the *cis*-α7 unit will be stabilized and the leak will become zero. If *n* $(\alpha7)_2$ pores were inserted in the liposome, the apparent resistance of $(\alpha7)_2$ inserted in the liposome and the planar lipid bilayer, $R_{d,\ lipo}$, will be given by equation (2.15) (Fig. 2.32a).

$$R_{d,lipo} = R_t + R_c + \frac{R_d}{n} \qquad (2.15)$$

Hence, the blockade in current expected upon liposome insertion ($B_{lipo}$) can be estimated from equation (2.16).

$$B_{lipo} = \left(\frac{I_d - I_{lipo}}{I_d}\right) \times 100 = \left(\frac{\frac{V}{R_d} - \frac{V}{R_{lipo}}}{\frac{V}{R_d}}\right) \times 100 = \left(1 - \frac{R_d}{R_{d,\ lipo}}\right) \times 100 \quad (2.16)$$

To estimate the error in the theoretical values of $B_{lipo}$ and $R_{d,\ lipo}$, these values were calculated considering the error in the experimental value of $R_t$. $B_{lipo}$ and $R_{d,\ lipo}$ were calculated with $R_t$=1.2 GΩ (mean value of $R_t$; equation 2.3), $R_t$=1.1 GΩ (mean–S.D.) and $R_t$=1.3 GΩ (mean+S.D.) and plotted as a function of *n* ranging from 1 to 50 (Fig. 2.32b).

The plot of expected current blockade associated with liposome insertion ($B_{lipo}$) *vs. n* suggested that only when a very large number of $(\alpha7)_2$ pores were inserted in the liposome ($n$>50), the calculated values of $B_{lipo}$ were close to the experimental value of the current blockade of 23±1%. Only if the liposome was considered completely porous ($n \rightarrow \infty$), equation (2.15) rearranged to equation (2.17).

$$R_{d,lipo} = R_t + R_c \ = 2.4\ G\Omega \qquad (2.17)$$

Then the value of $B_{lipo}$, calculated using equation (2.16), was equal to 25%, which is close to the experimental value of blockade produced on liposome insertion. As a very large *n* is not practically feasible considering only one $(\alpha7)_2$ pore is inserted in the planar lipid bilayer, it was concluded that the residual conductance upon liposome insertion was due to a leak in the $(\alpha7)_2$ structure.

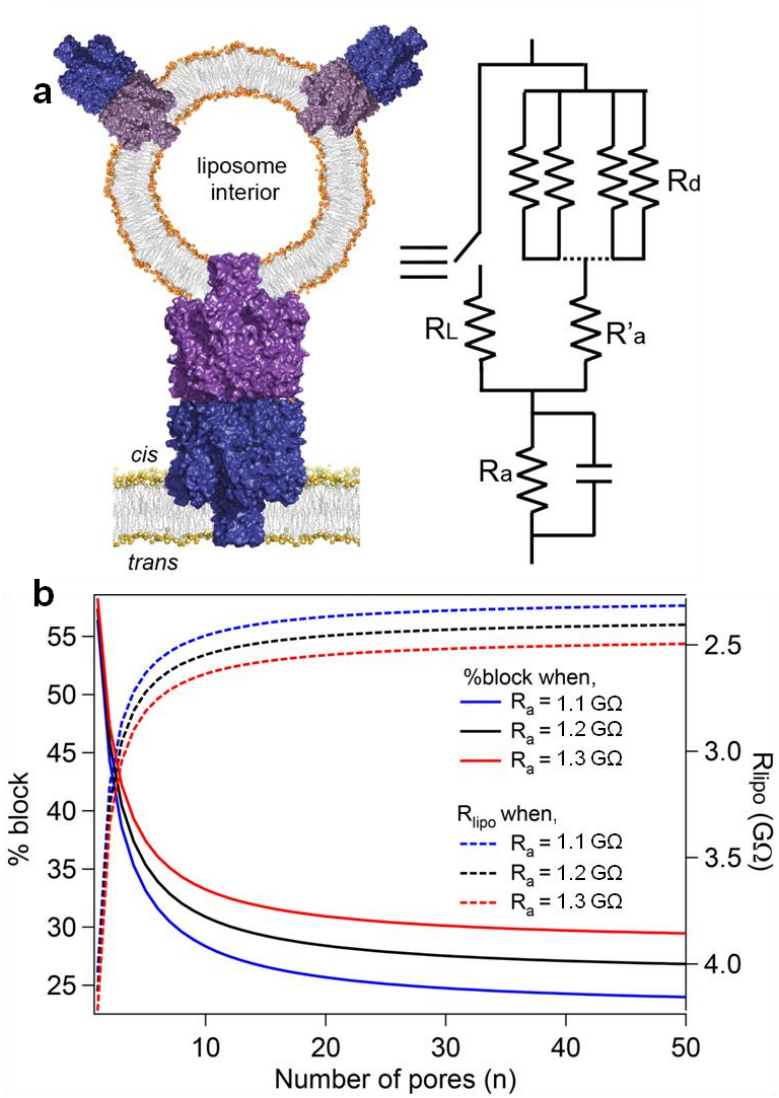



**Figure 2.32. Electrical model of $(\alpha7)_2$ inserted in a porous liposome and a planar lipid bilayer.** (**a**) Cartoon (not to scale) showing $(\alpha7)_2$ inserted in a planar lipid bilayer a liposome punctured with other $(\alpha7)_2$ pores and its equivalent electrical circuit representation. If the leak was due to a malformed *cis*-α7 barrel, upon insertion in liposome the leak should become zero. The resistance of $(\alpha7)_2$ pores inserted in the liposome ($R_d$) should be considered in parallel with each other and in series with $R_c$ and $R_t$. (**b**) The effective resistance of the electrical circuit of (**a**), $R_{d,\ lipo,}$ and the blockade in current due to liposome insert ($B_{lipo}$) were calculated as a function of the number of $(\alpha7)_2$ pores (*n*) inserted in the liposome. To estimate the error in the values of $R_{d,\ lipo}$ (right y-axis) and $B_{lipo}$ (left y-axis), these values were calculated with the experimental values of $R_t$=1.2 GΩ (mean value), $R_t$=1.1 GΩ (mean–S.D.) and $R_t$=1.3 GΩ (mean+S.D.).

**Residual conductance due to a leak in the $(\alpha7)_2$ structure.** The apparent resistance, $R_{d,\ lipo}$, of $(\alpha7)_2$ after insertion of the *cis*-α7 unit into a fully insulating

liposome (i.e. penetrated only by $(\alpha7)_2$ and not by excess pores in the cis compartment) is the sum of the $R_L$ and $R_t$, is given by equation (2.18) (Fig. 2.33a).

$$R_{d,lipo} = R_t + R_L = 2.4\ G\Omega \qquad (2.18)$$

The blockade in current expected upon liposome insertion ($B_{lipo}$) can be estimated from equation (2.16). The value of the current block expected upon liposome insertion at +50 mV is computed to be 25% (Fig. 2.33a), which agrees well with the experimental liposome blockade of 23±1%.

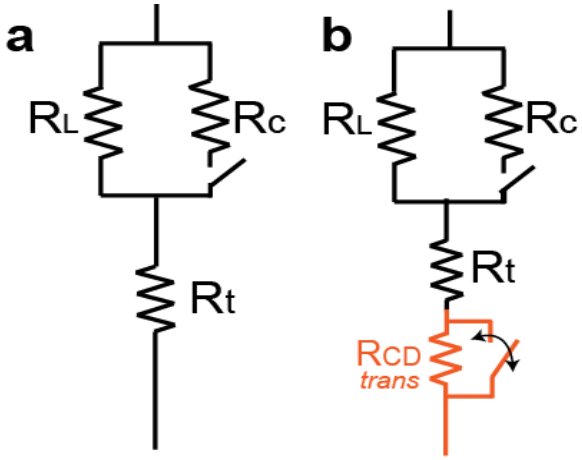



**Figure 2.33. Electrical circuits of $(\alpha7)_2$ inserted into both a planar bilayer and a liposome.** (**a**) The residual current might be attributed to a leak current (Fig. 2.10e). In this case, the current would pass through the leak resistor ($R_L$) and the *trans*-α7 resistor ($R_t$), while the path through $R_c$ would be blocked by the liposome. (**b**) Circuit representing the reversible binding of γCD to the *trans*-α7 unit when $(\alpha7)_2$ is inserted into both a planar bilayer and a liposome. Curved arrow denotes reversible binding of γCD.

When $(\alpha7)_2$ was inserted in liposome and a planar lipid bilayer, the effective resistance ($R_{\gamma CD\ trans,\ lipo}$) and value of *trans* γCD blocks ($B_{\gamma CD\ trans,\ lipo}$) were calculated using equations (2.19) and (2.20) (Fig. 2.33b).

$$R_{\gamma CD\ trans,\ lipo} = R_t + R_{CD\,trans} + R_L \quad (2.19)$$

$R_{\gamma CD\ trans,\ lipo}$=4.9 GΩ

$$B_{\gamma CD\ trans,\ lipo} = \frac{(I_{lipo} - I_{\gamma CD\ trans,\ lipo})}{I_{lipo}} \times 100$$

$$= \frac{\left(\frac{V}{R_{lipo}} - \frac{V}{R_{\gamma CD\ trans,\ lipo}}\right)}{\frac{V}{R_{lipo}}} \times 100 \quad (2.20)$$

$B_{\gamma CD\ trans,\ lipo}$=49%

The *trans* γCD blockade when the *cis* α7 unit was inserted into a liposome was 49% according to the model (Fig. 2.33b), which is reasonably close to the experimental value of 56±1%.

### 2.18.3. Electrical model at other potentials

The electrical model was valid at other potentials as well. The electrical model equations (2.5) – (2.11) were used to calculate the leak conductance and the values of currents blockades by *cis* and *trans* γCD at -50 mV and +100 mV (Table 2.3). At +100 mV, the experimental values of $R_t$, $R_c$ and $R_d$ were 1.1 GΩ, 1.5 GΩ and 1.7 GΩ, respectively. $R_L$ was thus calculated to be 1.0 GΩ (equation (2.5)). From equations (2.6) and (2.7), the values of $R_{CD,\ trans}$ and $R_{CD,\ cis}$ were 2.2 GΩ and 1.6 GΩ, respectively. The values of the *cis* and *trans* γCD current blockades were computed to be 11% and 56%, respectively using equations (2.8) – (2.11). The value of current blockade expected upon liposome insertion considering a leak in the structure was 19% (equations (2.18) and

(2.16)), which precisely matched the experimentally observed value at +100 mV (*N*=3) (section 2.15.1). If the case of a completely porous liposome was assumed, the value of current blockade was calculated to be 34%. The electrical values were computed at -50 mV similarly (Table 2.3). The theoretical values from the electrical model correlated well with the experimentally observed values at +50 mV, -50 mV and +100 mV (Tables 2.1, 2.3).

**Table 2.3.** Values derived from the electrical model of $(\alpha7)_2$ compared to experimental values.

| Applied potential (mV) | Physical parameter/phenomenon | Value from electrical model | Experimental value (mean±s.e.m.) |
|---|---|---|---|
| +50 | $R_{t=}$1.2 GΩ; $R_c$=1.2 GΩ | | |
| | $R_L$ | 1.2 GΩ | |
| | *cis* γCD blockade | 14% of $I_o$ | 15±1% |
| | *trans* γCD blockade | 56% | 55±2% |
| | blockade by γCD both simultaneously *cis* and *trans* | 59% | 61±1% |
| | blockade during liposome insertion | 25% | 23±1% |
| | blockade during liposome insertion with *trans* γCD blockade | 49% | 56±1% |
| +100 | $R_{t=}$1.1 GΩ; $R_c$=1.5 GΩ | | |
| | $R_L$ | 1.0 GΩ | |
| | *cis* γCD blockade | 11% | 13±0.3% |
| | *trans* γCD blockade | 58% | 54±5% |

| | | | |
|---|---|---|---|
| -50 | $R_{t=}$1.2 GΩ; $R_c$=1.2 GΩ | | |
| | $R_L$ | 1.2 GΩ | |
| | *cis* γCD blockade | 14% | 19±1% |
| | *trans* γCD blockade | 51% | 48±3% |

This excellent corroboration between the empirical and theoretical values suggested that the value of the leak conductance ($R_L$) determined from the model was correct. The electrical model also proved that the residual conductance of the $(\alpha7)_2$ pore after liposome insertion was due to a leak in the structure. The model suggests that the leak current accounts for 50% of the current entering the *trans*-α7 unit at +50 mV. However, the physical origin of the leak current is not clear. The single channel electrical measurement studies also imply that the leak is wide enough to allow hydrated ions through, but not wide enough to let large hydrated $P_i^-$ ions on the *cis* side to enter into the *trans*-α7 unit (section 2.13). Computational analysis suggests conductive pathways between the α1 subunits of *cis*-α7. The leak could also arise due to a breathing motion of the *cis*- and *trans*-α7 subunits and enter between their cap domains (Fig. 2.10e).

## Conclusions and outlook

In this chapter, the engineering of a gap-junction mimic using the membrane pore αHL as the starting protein is described. Biochemical, single channel electrical and molecular adapter binding characterization suggested a cap-to-cap arrangement of the α7 units in majority of the disulfide cross-linked $(\alpha7)_2$

molecules. The $(\alpha7)_2$ structure was confirmed by TEM, which showed side and top views of cap-to-cap arrangements with the expected dimensions. An accurate electrical model of $(\alpha7)_2$ was built assuming the two α7 units as series resistors, and an additional resistor parallel to the *cis*-α7 resistor. The calculated values for γCD and liposome blocking using this model correlated well with the experimental values at three different potentials. Simultaneous insertion of $(\alpha7)_2$ in two lipid bilayers has been demonstrated in single channel electrical measurements and TEM. Starting from the αHL monomer gene mutants, $(\alpha7)_2$ - could be purified with the barrel mutations- M113F and M113R (Methods). These initial results suggest that $(\alpha7)_2$ is robust to mutagenesis like α7. More mutants can be made in a similar way to achieve $(\alpha7)_2$ with increased functionalities.

Owing to a single transmembrane β-barrel stem, α7 is able to insert in only one bilayer thereby limiting its functionality. The $(\alpha7)_2$ is however capable of simultaneous insertions and small molecule (imipramine) transfer across two bilayer delimited compartments in a completely aqueous system. Vesicles embedded with α7 pores have been used studying biomolecular reactions at the single molecule level. The α7 pores allow a controlled translocation of small analytes (for example, enzyme cofactor molecules) from the exterior buffer into the vesicles and can be used to maintain a constant concentration of analytes within the vesicles (Ishitsuka *et al.*, 2010; Okumus *et al.*, 2009). Different reactions could be studied simultaneously on a platform containing arrays of vesicles connected by $(\alpha7)_2$ (Fig. 2.34). The throughput of these studies could be increased manifold using engineered $(\alpha7)_2$ pores.

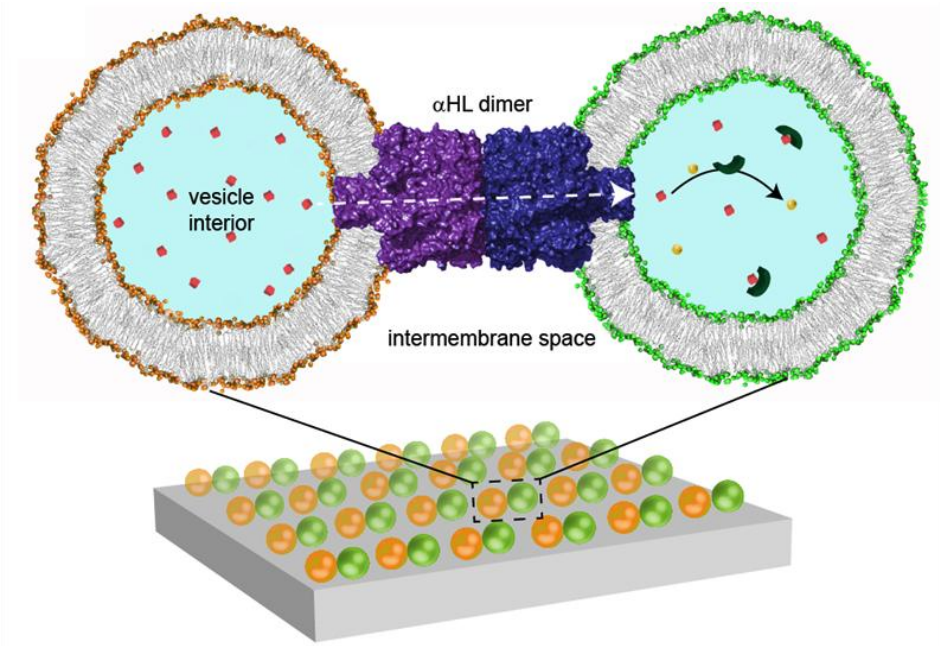



**Figure 2.34. Application of $(\alpha7)_2$ in building high-throughput devices.** Vesicles such as liposomes studded with α7 pores have been used for studying biomolecular reactions at the single molecule level (Okumus *et al.*, 2009). Engineered $(\alpha7)_2$ pores can be used to connect proximal bilayers of vesicles and allow controlled and selective translocation of analytes. A multi-array of such interconnected vesicles could be used to make high-throughput devices for single molecule studies, screening, and other applications.

The construction of a functional $(\alpha7)_2$ pore can also be considered as a first step towards designing a key component in building communicative protocells. We envisage the use of $(\alpha7)_2$ with liposomes to create larger networks as prototissues (Fig. 2.2a).

Recently, it was demonstrated that encapsulating lipid-coated aqueous droplets in an oil droplet, which is further suspended in an aqueous phase (water-oil-water system) creates lipid bilayers at the water-oil-water interface giving the so-called multisomes (Villar *et al.*, 2011). Such a system is an ideal platform to extend the utility of $(\alpha7)_2$. As a simple example, two multisome bilayers could be connected through $(\alpha7)_2$ dissolved in the external aqueous environment (Fig.

2.35). This could form the basic element of a larger interconnected bilayer network.

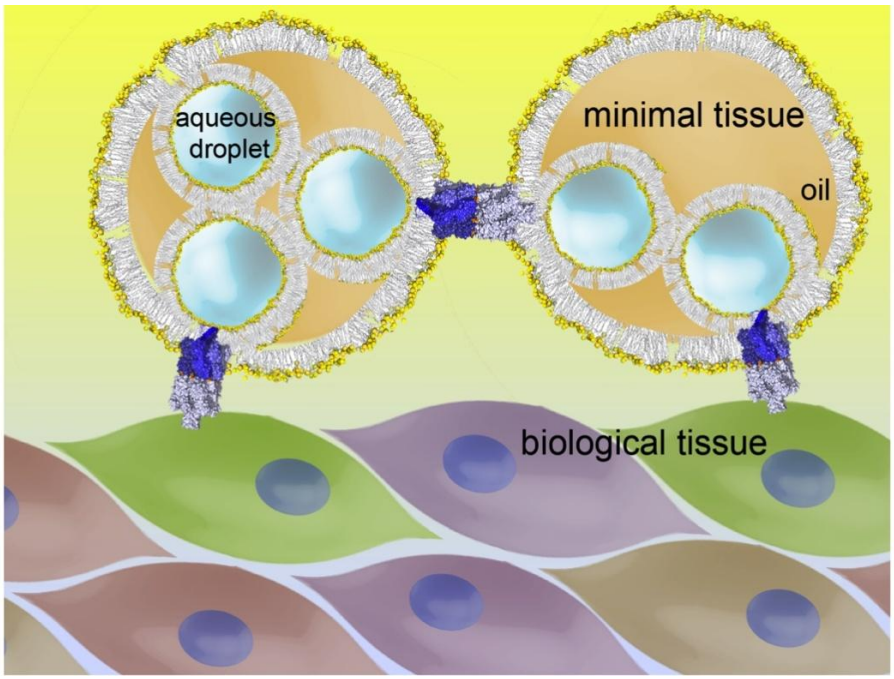



**Figure 2.35. Application of $(\alpha7)_2$ in building prototissues and connecting them with biological tissues.** An interconnected network of aqueous droplets encased in oil and lipid has been modelled as a form of minimal tissue. These minimal tissues or those formed by a network of liposomes (Fig. 2.2) could be connected by $(\alpha7)_2$ pores. Synthetic minimal tissues can also be connected to biological tissues through $(\alpha7)_2$; opening new frontiers in biomedical engineering.

In nature, α7 causes hemostasis and eventually cell death by inserting in the bilayers of single cells (Caiazza and O'Toole, 2003). The double bilayer insertion property of $(\alpha7)_2$ could thus be used to link the interface of a synthetic tissue/cell and a natural one for drug delivery (Eroglu *et al.*, 2000) (Fig. 2.35). An $(\alpha7)_2$ integrated synthetic and natural system would open up new frontiers in biomedical engineering.

The αHL dimer complex is the first example of an engineered membrane pore that can span the length of two artificial lipid bilayers in an aqueous

environment, a prerequisite for constructing complex tissue mimics in a bottom-up fashion. Besides its use in synthetic biology, a stable cap-to-cap αHL dimer also serves as a proof that other proteins, e.g., enzymes, could be attached on the αHL cap for bio-nanotechnology applications.

# Methods

## 2.19. Computational modelling of $(\alpha7)_2$

Two α7 units were docked in a cap-to-cap orientation in PyMOL (Schrodinger). The topology file was generated using GROMACS version 4.0 (www.gromacs.org), and edited to introduce position restraints on the cysteine thiol atoms positioned 0.2 nm apart, which is the S-S bond length. An all atom MD simulation was performed on the docked structure for 1 ns in 500 mM NaCl buffer using the GROMACS OPLS-AA/L all-atom force field. The integration time step was 2 fs with an all bonds constraint. Particle Mesh Ewald (PME) method (Essmann *et al.*, 1995) was used to compute long-range electrostatics interactions and the short-range electrostatic cut-off was kept at 1 nm with a Fourier spacing of 0.32. Steric clashes were visualized using a PyMOL script (show_bumps.py, Thomas Holder) with a Van der Waals radius cut-off of 0.3 Å.

## 2.20. Expression and purification of $(\alpha7)_2$ and barrel mutants: M113R, M113F

The α1 mutation K237C was introduced by site-directed mutagenesis of the α1 wild-type gene inserted into a pT7 vector with a a β-lactamase ampicillin resistance gene. The vector also encoded for a D8H6 tag at the αHL C-terminus to aid in $Ni^{2+}$ affinity purification. The sequence of the primers used for the forward and reverse PCR reactions for the K237C mutation were CTATGGATAGATGCGCATCCAAACAACAAACAAATATAG (forward primer) and CAGAAGTGGTCCTGCAACTTTAT (reverse primer), CTATATTTGTTTGTTGTTTGGATGCGCATCTATCCATAG (reverse primer) and ATAAAGTTGCAGGACCACTTCTG (forward primer), respectively. 150 µl of

Rosetta pLysS DE3 competent cells (Novagen) were transformed with ~2 μg of the mutant DNA by first incubating on ice for 30 min, followed by heat shock at 42°C for 30 s. The transformed cells were plated on ampicillin (200 μg $ml^{-1}$) and chloramphenicol (25 μg $ml^{-1}$) selective LB-agar plates. A single colony was picked and inoculated in 100 ml terrific broth containing ampicillin (200 μg $ml^{-1}$) and chloramphenicol (25 μg $ml^{-1}$), and grown at 37°C for ~16-18 h. 5 ml of this culture was then used to inoculate 100 ml terrific broth containing the antibiotics and grown at 37°C . At O.D. ~0.6 the culture was induced by isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.5 mM) and grown further for 22 h at 20°C. The culture was spun at 3,400 g for 25 min, the supernatant was discarded and the cell pellets were frozen and thawed on ice.

K237C/D8H6 α1 was purified under non-denaturing conditions. The culture was centrifuged at 3,400 xg for 25 min, the cell pellet harvested, and suspended in lysis buffer (50 mM Tris.HCl, 200 mM NaCl, 1% v/v Triton X-100, 10 mM imidazole, 1.2 μl benzonase (Merck), 1 mM $MgCl_2$, 0.1 mM tris (2-carboxyethyl) phosphine (TCEP), pH 8.0). The lysate was sonicated for 5 min in 30 s pulses with intermittent 30 s intervals. The cell pellets could also be lysed by using a commercial lysis reagent such as BugBuster (Novagen) which did not affect the $(α7)_2$ yield (data not shown). The lysate was then centrifuged at 3,400 xg for 15 min and the supernatant was loaded on a pre-equilibrated 2 ml (column volume (c/v)) Ni-NTA Agarose column (Qiagen). The column was washed thrice with 7.5 c/v (15 ml) of wash buffer (50 mM Tris.HCl, 500 mM NaCl, 30 mM imidazole, 0.2% v/v Triton X-100, pH 8.0). The protein was eluted in 6 c/v (12 ml) of elution buffer (50 mM Tris.HCl, 500 mM NaCl, 0.2% v/v Triton X-100, 500

mM imidazole, pH 8.0). All steps of the non-denaturing purification were performed on ice or at 4°C. The $M_{app}$ the purified proteins were estimated from SDS-PAGE after Coomassie blue staining using Quantity One 1-D Analysis Software (Bio-Rad).

The mutants M113R/K237C/D8H6 and M113F/K237C/D8H6 α1 were purified under denaturing conditions. The M113R and M113F mutations were introduced in the pT7 K237C/D8H6 α1 plasmid by another round of site-directed mutagenesis. The transformation protocol and culture growth conditions of the Rosetta pLysS DE3 cells expressing M113R/K237C/D8H6 and M113F/K237C/D8H6 α1 were the same as for K237C/D8H6. The lysate pellet was resuspended in 8 M urea, and centrifuged at 13,000 xg for 30 min at 4°C to remove any insoluble cell debris. All the buffers used in the Ni-NTA purification in denaturing conditions contained 8 M urea. The first three eluate fractions (5 ml each) were combined and dialysed overnight against 50 mM Tris.HCl, 500 mM NaCl, 0.2% v/v Triton X-100, pH 8 using a dialysis membrane (cut-off 3,500 Da, Spectrum) to remove the urea and refold the protein. The dialysed protein was then concentrated ~10x using 10 kDa cut-off centrifugal filter columns (Microcon). SDS-PAGE of the eluates (Laemmli sample buffer) was performed on an 8% (w/v) polyacrylamide gel (length 16 cm, width 1 mm), which was run at +60 V in TGS (25 mM Tris.HCl, 192 mM glycine, 0.1% w/v SDS) running buffer for 16 - 20 h. Gel pieces containing the protein were excised, suspended and crushed in 500 µl 10 mM Tris 1 mM EDTA (TE) (pH 8.0) and left for another 30 - 45 min at 25°C. The crushed gel suspension was loaded onto spin-filtration

columns (Qiagen) and protein fractions eluted by centrifugation at 25,000 xg for 15 min at 4°C.

### 2.21. Proteinase K assay

Proteinase K (50 μg $mL^{-1}$, 500 μg $mL^{-1}$ and 5 mg $mL^{-1}$) was added to in a ratio of 1:9 v/v to purified $(\alpha 7)_2$ (10 mM Tris.HCl, 1 mM EDTA, pH 8.0) and incubated for 5 min at 25°C, when phenylmethylsulfonyl fluoride (PMSF) (2 mM) was added to quench the proteolysis reaction. The samples were analysed using 10% w/v bis-tris polyacrylamide gel electrophoresis (Bio-Rad).

### 2.22. Molecular modeling and graphics

All structures were rendered in PyMOL (Schrodinger) or VMD (Humphrey *et al.*, 1996). The MOLE plug-in (Petrek *et al.*, 2007) for PyMOL was used to visualise water tunnels through the various $(\alpha 7)_2$ pore structures (Fig. 2.10). This algorithm considers the centres of the Van der Waal spheres of the all protein atoms to create a Voronoi mesh. A Voronoi diagram/mesh is a representation of a space in which cells are made around particular points (loci). The cells are made such that every point within each cell is closest to the cell's locus compared to any other loci in that space. In the case of proteins, the Voronoi cells are constructed around the atom centres (loci). If the Voronoi cell edges are far from the locus, this implies empty space around that atom centre and hence possibility of a water channel. The MOLE algorithm then scores the edges of the protein Voronoi mesh depending on the distance of the edges from the central protein atom and its length to find channels. The edges of the Voronoi cells farthest from the central atoms are ranked based on an empirical

cost function and connected to compute optimal egress paths or channels through the protein. For Fig. 2.10b-e, the starting point for tunnel exploration was the centre of the *trans*-α7 unit. The number of tunnels to be searched for in each structure was fixed at a minimum value of 3.

### 2.23. Single channel recording

Biological nanopores such as α-hemolysin (αHL) can be reconstituted in synthetic lipid bilayers (Fang *et al.*, 1997; Krasilnikov and Sabirov, 1989). Electric currents through these nanopores can be measured under a bias potential using suitable electrolytes on both sides of the bilayer (Maglia *et al.*, 2010). Conductance is a function of the cross-sectional area, length and the nature of charges on the mouth of the nanopores (equation (2.1)).

All electrical recordings were done using a chamber with two buffer reservoirs separated by a 30 µm thick Teflon septum with a 125 µm aperture. The aperture was pre-treated with ~2 µl hexadecane in pentane (10% v/v) on each side and the reservoirs filled with 1 ml recording buffer (1 M KCl, 25 mM Tris.HCl, 50 µM EDTA, pH 8.0). A planar bilayer was formed on the aperture by adding ~10 µl 1,2-diphytanoyl phosphatidylcholine (DPhPC) (10 mg $ml^{-1}$) (Avanti Polar Lipids) in pentane on the buffer surface in both the reservoirs. The proteins were added to the electrically grounded side, labelled as '*cis*' (Fig. 2.11a). Ag/AgCl electrodes were used to apply a bias potential, measured as $V_{trans}$ - $V_{cis}$. The electrical current was amplified with a patch-clamp amplifier (Axopatch 200B, Axon Instruments), digitized with a Digidata 1320 A/D convertor (Axon Instruments) and acquired with a 1 *k*Hz filter at a sampling

frequency of 10 *k*Hz. In γCD blocking experiments, 20-160 μM γCD was added to the *cis* or the *trans* side. The electrical recordings were post filtered with low pass 100 Hz Bessel filter. The imipramine blocking experiments were acquired at 50 *k*Hz sampling frequency with a 1 *k*Hz filter. These recordings were not post-filtered for data analysis. All chemicals were purchased from Sigma, unless specified otherwise.

**Statistics analysis.** The $I_o$ histograms of α7 and $(\alpha 7)_2$ were fit with Gaussian functions. The coefficients of the Gaussian functions are reported as mean±S.D. (equation (2.21)).

$$\text{S. D.} = \frac{\text{half width at half maxima}}{\sqrt{2\ln 2}} \tag{2.21}$$

For all other data, the mean±s.e.m. values have been reported. Experiments where only a fraction of $(\alpha 7)_2$ pores displayed a certain behaviour (for example, $P_i^-$ binding) while the other $(\alpha 7)_2$ pores did not, the $(\alpha 7)_2$ pores could be divided into two classes (equivalent to a binomial distribution). The proportion of $(\alpha 7)_2$ pores in each class has been reported with a 95% confidence interval. The confidence interval of a binomial distribution is given by equation (2.22).

$$p \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p\,(1-p)}{n}} \tag{2.22}$$

Where, *p* is the proportion of successes in *n* attempts. For a 95% confidence interval, $z_{1-\alpha/2}$ is equal to 1.96.

### 2.24. $P_i^-$ binding experiments

0.5 M stocks of sodium dihydrogen phosphate ($NaH_2PO_4$) and disodium hydrogen phosphate ($Na_2HPO_4$) were mixed to prepare 0.5 M sodium phosphate, pH 8.0. The 0.5 M sodium phosphate stock was added at a final concentration of 10 mM to the *cis* or *trans* sides of the bilayer. Experiments were always performed with fresh aliquots of $(\alpha7)_2$ stored at -80°C, which had not been freeze-thawed.

### 2.25. Preparation of liposomes

Lipids (25 mg each: 1,2-diphytanoyl phosphatidylcholine (DPhPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'–*rac*-glycerol) (POPG) and L-α-phosphatidylcholine (95%) (egg-PC), Avanti Polar Lipids) were dissolved in 1 ml chloroform, dried under a stream of nitrogen, and kept under high vacuum in a desiccator for 3 to 4 h. The dried lipids were suspended in 1 ml rehydration buffer (1 M KCl, 25 mM Tris.HCl, 50 µM EDTA, pH 8.0), and extruded through 0.1 µm polycarbonate membranes (Avanti) to obtain unilamellar liposomes. For the preparation of liposomes containing analytes, the rehydration buffer also included 1 mM imipramine or 10 mM γCD. To increase the encapsulation efficiency of the analytes, the rehydrated lipids were vortexed, and exposed to five freeze-thaw cycles before extrusion through the polycarbonate membranes. For

electrophysiology experiments, 50 or 100 µl liposomes were added to recording buffer (1 ml final volume) on the *cis* side of the bilayer.

### 2.26. Electron Microscopy and Image processing

To obtain images of $(\alpha7)_2$ inserted into liposomes, $(\alpha7)_2$ was mixed with egg-PC or 10 mol% POPG:POPC liposomes at lipid to protein ratios of 100:1 or 20:1 (w/w), respectively. After incubation for 15 to 60 min, samples were applied to grids and stained with 2% (w/v) uranyl acetate. For single-particle analysis, $(\alpha7)_2$ (50 mg $L^{-1}$) was applied to freshly glow-discharged continuous-carbon grids and left to adsorb for 1 min. After blotting, either 2% (w/v) uranyl acetate or NanoVan® (Nanoprobes, Inc., USA) was used as a negative stain and applied to the grid for ~15 s. Grids were imaged at either 120 kV (NanoVan) or 80 kV (uranyl acetate) with a Tecnai T12 transmission electron microscope fitted with a tungsten filament and a BioTWIN lens. Micrographs were recorded on an Eagle 4k × 4k CCD camera with a final pixel size of 1.675 Å/pixel at a nominal magnification of ×52k.

All image processing was performed using EMAN2. A total of 2,748 particles were selected semi-automatically by using e2boxer and binned by 2. CTF parameters were determined using e2ctf, followed by phase flipping and high-pass filtering. Reference-free class averages were generated assuming no symmetry using e2refine2d.

# References


Ahyayauch, H., Goni, F.M., and Bennouna, M. (2004). Interaction of electrically neutral and cationic forms of imipramine with liposome and erythrocyte membranes. International journal of pharmaceutics *279*, 51-58.

Akeson, M., Branton, D., Kasianowicz, J.J., Brandin, E., and Deamer, D.W. (1999). Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid and polyuridylic acid as homopolymers or as segments within single RNA molecules. BiophysJ *77*, 3227-3233.

Astier, Y., Bayley, H., and Howorka, S. (2005). Protein components for nanodevices. Curr Opin Chem Biol *9*, 576-584.

Astier, Y., Braha, O., and Bayley, H. (2006). Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. Journal of the American Chemical Society *128*, 1705-1710.

Ballister, E.R., Lai, A.H., Zuckermann, R.N., Cheng, Y., and Mougous, J.D. (2008). In vitro self-assembly of tailorable nanotubes from a simple protein building block. Proc Natl Acad Sci *105*, 3733-3738.

Bayley, H., and Cremer, P.S. (2001). Stochastic sensors inspired by biology. Nature *413*, 226-230.

Bayley, H., and Jayasinghe, L. (2004). Functional engineered channels and pores. Mol Membrane Biol *21*, 209-220.

Bayley, H., Luchian, T., Shin, S.-H., and Steffensen, M.B. (2008). Single-molecule covalent chemistry in a protein nanoreactor. In Single Molecules and Nanotechnology, R. Rigler, and H. Vogel, eds. (Heidelberg: Springer), pp. 251-277.

Bhakdi, S., Bayley, H., Valeva, A., Walev, I., Walker, B., Weller, U., Kehoe, M., and Palmer, M. (1996). Staphylococcal alpha-toxin, streptolysin-O, and Escherichia coli hemolysin: prototypes of pore-forming bacterial cytolysins. ArchMicrobiol *165*, 73-79.

Bhakdi, S., Füssle, R., and Tranum-Jensen, J. (1981). Staphylococcal alpha-toxin: oligomerization of hydrophilic monomers to form amphiphilic hexamers

induced through contact with deoxycholate micelles. Proc Natl Acad Sci *78*, 5475-5479.

Bhakdi, S., and Tranum-Jensen, J. (1991). Alpha-toxin of Staphylococcus aureus. MicrobiolRev *55*, 733-751.

Braha, O., Gu, L.-Q., Zhou, L., Lu, X., Cheley, S., and Bayley, H. (2000). Simultaneous stochastic sensing of divalent metal ions. Nature biotechnology *17*, 1005-1007.

Bupp, J.M., Martin, A.E., Stensrud, E.S., and Jaspersen, S.L. (2007). Telomere anchoring at the nuclear periphery requires the budding yeast Sad1-UNC-84 domain protein Mps3. The Journal of cell biology *179*, 845-854.

Caiazza, N.C., and O'Toole, G.A. (2003). Alpha-toxin is required for biofilm formation by Staphylococcus aureus. Journal of bacteriology *185*, 3214-3217.

Chang, C.-Y., Niblack, B., Walker, B., and Bayley, H. (1995). A photogenerated pore-forming protein. ChemBiol *2*, 391-400.

Cheley, S., Gu, L.-Q., and Bayley, H. (2002). Stochastic sensing of nanomolar inositol 1,4,5-trisphosphate with an engineered pore. Chem Biol *9*, 829-838.

Cheley, S., Malghani, M.S., Song, L., Hobaugh, M., Gouaux, J.E., Yang, J., and Bayley, H. (1997). Spontaneous oligomerization of a staphylococcal alpha-hemolysin conformationally constrained by removal of residues that form the transmembrane beta barrel. Protein Eng *10*, 1433-1443.

Cherf, G.M., Lieberman, K.R., Rashid, H., Lam, C.E., Karplus, K., and Akeson, M. (2012). Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. Nature biotechnology *30*, 344-348.

Cisse, I., Okumus, B., Joo, C., and Ha, T. (2007a). Single-Molecule chemistry and biology special feature: fueling protein-DNA interactions inside porous nanocontainers (vol 104, pg 12646, 2007). Proceedings of the National Academy of Sciences of the United States of America *104*, 14878-14878.

Cisse, I., Okumus, B., Joo, C., and Ha, T.J. (2007b). Fueling protein-DNA interactions inside porous nanocontainers. Proceedings of the National Academy of Sciences of the United States of America *104*, 12646-12650.

Clarke, J., Wu, H., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. (2009b). Continuous base identification for single-molecule nanopore DNA sequencing. Nature Nanotechnology *4*, 265-270.

D'Angelo, M.A., and Hetzer, M.W. (2008). Structure, dynamics and function of nuclear pore complexes. Trends Cell Biol *18*, 456-466.

Eroglu, A., Russo, M.J., Bieganski, R., Fowler, A., Cheley, S., Bayley, H., and Toner, M. (2000). Intracellular trehalose improves the survival of cryopreserved mammalian cells. Nat Biotechnol *18*, 163-167.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A Smooth Particle Mesh Ewald Method. J Chem Phys *103*, 8577-8593.

Evans, W.H., and Martin, P.E. (2002). Gap junctions: structure and function. Mol Membr Biol *19*, 121-136.

Fang, Y., Cheley, S., Bayley, H., and Yang, J. (1997). The heptameric prepore of a staphylococcal a-hemolysin mutant in lipid bilayers imaged by atomic force microscopy. Biochemistry *36*, 9518-9522.

Füssle, R., Bhakdi, S., Sziegoleit, A., Tranum-Jensen, J., Kranz, T., and Wellensiek, H.-J. (1981). On the mechanism of membrane damage by Staphylococcus aureus a-toxin. J Cell Biol *91*, 83-94.

Galdiero, S.G.E. (2004). High resolution crystallographic studies of a-hemolysin-phospholipid complexes define heptamer-lipid head group interactions: Implication for understanding protein-lipid interactions. Protein science : a publication of the Protein Society *13*, 1503-1511.

Gu, L.-Q., Braha, O., Conlan, S., Cheley, S., and Bayley, H. (1999). Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. Nature *398*, 686-690.

Gu, L.-Q., Cheley, S., and Bayley, H. (2001). Prolonged residence time of a noncovalent molecular adapter, b-cyclodextrin, within the lumen of mutant a-hemolysin pores. JGenPhysiol *118*, 481-494.

Gu, L.-Q., Cheley, S., and Bayley, H. (2003). Electroosmotic enhancement of the binding of a neutral molecule to a transmembrane pore. ProcNatlAcadSciUSA *100*, 15498-15503.

Hammerstein, A.F., Shin, S.H., and Bayley, H. (2010). Single-molecule kinetics of two-step divalent cation chelation. Angew Chem Int Ed Engl *49*, 5085-5090.

Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput *4*, 435-447.

Holden, M.A., Needham, D., and Bayley, H. (2007). Functional bionetworks from nanoliter water droplets. J Am Chem Soc *129*, 8650-8655.

Howorka, S., and Bayley, H. (2002). Probing distance and electrical potential within a protein pore with tethered DNA. BiophysJ *83*, 3202-3210.

Howorka, S., Movileanu, L., Lu, X., Magnon, M., Cheley, S., Braha, O., and Bayley, H. (2000). A protein pore with a single polymer chain tethered within the lumen. JAmChemSoc *122*, 2411-2416.

Howorka, S., Nam, J., Bayley, H., and Kahne, D. (2004). Stochastic detection of monovalent and bivalent protein-ligand interactions. AngewChemIntEd *43*, 842-846.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J Mol Graph *14*, 33-38, 27-38.

Ishitsuka, Y., Okumus, B., Arslan, S., Chen, K.H., and Ha, T. (2010). Temperature-independent porous nanocontainers for single-molecule fluorescence studies. Analytical chemistry *82*, 9694-9701.

Japrung, D., Henricus, M., Li, Q., Maglia, G., and Bayley, H. (2010). Urea facilitates the translocation of single-stranded DNA and RNA through the α-hemolysin nanopore. Biophysical journal *98*, 1856-1863.

Jayasinghe, L., Miles, G., and Bayley, H. (2006). Role of the amino latch of staphylococcal a-hemolysin in pore formation. The Journal of biological chemistry *281*, 2195-2204.

Kleinschmidt, J.H., Wiener, M.C., and Tamm, L.K. (1999). Outer membrane protein A of E. coli folds into detergent micelles, but not in the presence of monomeric detergent. Protein Sci *8*, 2065-2071.

Krasilnikov, O.V., Merzlyak, P.G., Yuldasheva, L.N., Rodrigues, C.G., Bhakdi, S., and Valeva, A. (2000). Electrophysiological evidence for heptameric

stoichiometry of ion channels formed by Staphylococcus aureus alpha-toxin in planar lipid bilayers. MolMicrobiol *37*, 1372-1378.

Krasilnikov, O.V., and Sabirov, R.Z. (1989). Ion transport through channels formed in lipid bilayers by Staphylococcus aureus alpha-toxin. GenPhysiolBiophys *8*, 213-222.

Kurihara, K., Tamura, M., Shohda, K., Toyota, T., Suzuki, K., and Sugawara, T. (2011). Self-reproduction of supramolecular giant vesicles combined with the amplification of encapsulated DNA. Nat Chem *3*, 775-781.

Lu, S., Li, W.-W., Rotem, D., Mikhailova, E., and Bayley, H. (2010). A primary hydrogen-deuterium isotope effect observed at the single molecule level. Nature chemistry *2*, 921-928.

Lvov, Y., Antipov, A.A., Mamedov, A., Mohwald, H., and Sukhorukov, G.B. (2001). Urease encapsulation in nanoorganized microshells. Nano Lett *1*, 125-128.

Maeda, S., Nakagawa, S., Suga, M., Yamashita, E., Oshima, A., Fujiyoshi, Y., and Tsukihara, T. (2009). Structure of the connexin 26 gap junction channel at 3.5 A resolution. Nature *458*, 597-602.

Maglia, G., Heron, A.J., Hwang, W.L., Holden, M.A., Mikhailova, E., Li, Q., Cheley, S., and Bayley, H. (2009a). Droplet networks with incorporated protein diodes show collective properties. Nat Nanotechnol *4*, 437-440.

Maglia, G., Heron, A.J., Stoddart, D., Japrung, D., and Bayley, H. (2010). Analysis of single nucleic acid molecules with protein nanopores. Methods Enzymol *475*, 591-623.

Maglia, M., Henricus, M., Wyss, R., Li, Q., Cheley, S., and Bayley, H. (2009b). DNA strands from denatured duplexes are translocated through engineered protein nanopores at alkaline pH. Nano Lett *9*, 3831-3836.

Meller, A., Nivon, L., Brandin, E., Golovchenko, J., and Branton, D. (2000). Rapid nanopore discrimination between single polynucleotide molecules. ProcNatlAcadSciUSA *97*, 1079-1084.

Nakagawa, S., Maeda, S., and Tsukihara, T. (2010). Structural and functional studies of gap junction channels. Curr Opin Struct Biol *20*, 423-430.

Noireaux, V., and Libchaber, A. (2004). A vesicle bioreactor as a step toward an artificial cell assembly. Proc Natl Acad Sci *101*, 17669-17674.

Oberholzer, T., Wick, R., Luisi, P.L., and Biebricher, C.K. (1995). Enzymatic RNA replication in self-reproducing vesicles: an approach to a minimal cell. Biochem Biophys Res Commun *207*, 250-257.

Okumus, B., Arslan, S., Fengler, S.M., Myong, S., and Ha, T. (2009). Single molecule nanocontainers made porous using a bacterial toxin. Journal of the American Chemical Society *131*, 14844-14849.

Petrek, M., Kosinova, P., Koca, J., and Otyepka, M. (2007). MOLE: A Voronoi diagram-based explorer of molecular channels, pores, and tunnels. Structure *15*, 1357-1363.

Pfanner, N., and Meijer, M. (1997). The Tom and Tim machine. Current biology : CB *7*, R100-103.

Rotem, D., Jayasinghe, L., Salichou, M., and Bayley, H. (2012). Protein detection by nanopores equipped with aptamers. Journal of the American Chemical Society, published ASAP.

Schwille, P. (2011). Bottom-up synthetic biology: engineering in a tinkerer's world. Science *333*, 1252-1254.

Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., and Gouaux, J.E. (1996). Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. Science *274*, 1859-1865.

Stoddart, D., Heron, A., Mikhailova, E., Maglia, G., and Bayley, H. (2009). Single nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc Natl Acad Sci USA *106*, 7702-7707.

Szostak, J.W., Bartel, D.P., and Luisi, P.L. (2001). Synthesizing life. Nature *409*, 387-390.

Thelestam, M., and Blomqvist, L. (1988). Staphylococcal alpha toxin- recent advances. Toxicon *26*, 51-65.

Valeva, A., Weisser, A., Walker, B., Kehoe, M., Bayley, H., Bhakdi, S., and Palmer, M. (1996). Molecular architecture of a toxin pore: a 15-residue

sequence lines the transmembrane channel of staphylococcal alpha-toxin. EMBO J *15*, 1857-1864.

Villar, G., Heron, A., and Bayley, H. (2011). Formation of droplet networks that function in aqueous environments. Nat Nanotechnol *6*, 803-808.

Voet, D., and Voet, J.G. (2011). Biochemistry, 4th edn (Hoboken, NJ: John Wiley & Sons).

Walker, B., and Bayley, H. (1995). Key residues for membrane binding, oligomerization, and pore-forming activity of staphylococcal a-hemolysin identified by cysteine scanning mutagenesis and targeted chemical modification. J Biol Chem *270*, 23065-23071.

Walker, B., Braha, O., Cheley, S., and Bayley, H. (1995). An intermediate in the assembly of a pore-forming protein trapped with a genetically-engineered switch. ChemBiol *2*, 99-105.

Walker, B.J., Krishnasastry, M., Zorn, L., and Bayley, H. (1992). Assembly of the oligomeric membrane pore formed by staphylococcal alpha-hemolysin examined by truncation mutagenesis. J Biol Chem *267*, 21782-21786.

Woolfson, D.N., and Bromley, E.H.C. (2011). Synthetic biology: a bit of rebranding, or something new and inspiring? The Biochemist *33*, 19-25.

# CHAPTER 3

## COUPLING λ-EXONUCLEASE TO THE α-HEMOLYSIN NANOPORE

## Introduction

### 3.1. DNA processing enzymes

Nature has devised an extremely diverse range of DNA processing enzymes to cleave ssDNA and dsDNA in different ways, with some enzymes requiring specific substrates for reactivity (Marti and Fleck, 2004; Yang, 2011). DNA cleaving enzymes are involved in DNA replication, repair, and recombination, and in the defense of the cell against foreign DNA. Host bacterial cells attack invading bacteriophage DNA with their endogenous restriction systems (Boyer, 1971; Yuan, 1981). The dsDNA breaks thus created are repaired by the bacteriophage using its recombinational systems (Mosig, 1998). Efficient recombination is thus a crucial process for phage chromosome propagation (Hendrix, 2002). Gene exchange/ shuffle amongst different phages is a constant dynamic process and also requires the use of recombination (Hendrix, 2002; Juhala *et al.*, 2000)

### 3.2. Recombination in bacteriophage λ: role of λ-exonuclease *in vivo*

In bacteriophage λ, recombination is carried out by the products of two genes, *exo* and *bet* (Shulman *et al.*, 1970). The *exo* gene encodes for a trimeric, 24 kDa, 5'->3' dsDNA cleaving exonuclease, λ-exonuclease (λ-exo), and the *bet* gene encodes a monomeric, 28 kDa, ssDNA binding protein, β protein. Both λ-

exo and β protein are ring shaped enzymes. Together, λ-exo and β protein form the bacteriophage λ 'Red αβ' recombination system. λ-exo carries out phage DNA recombination through the double-strand break repair (DSBR) and single-strand annealing pathways (Cassuto *et al.*, 1971; Cassuto and Radding, 1971; Takahashi and Kobayashi, 1990) (Fig. 3.1). λ-exo degrades one strand of dsDNA into monophosphate nucleotides leaving a ssDNA overhang to which its partner, β protein and other single-strand binding proteins bind (Little, 1967). β protein binds to ssDNA, promotes renaturation of complementary strands and mediates strands annealing even *in vitro* (Li *et al.*, 1998; Poteete, 2001). The bacteriophage λ 'Red αβ' system has been recently used for genetic engineering and sequencing purposes (Copeland *et al.*, 2001; Poteete, 2001).

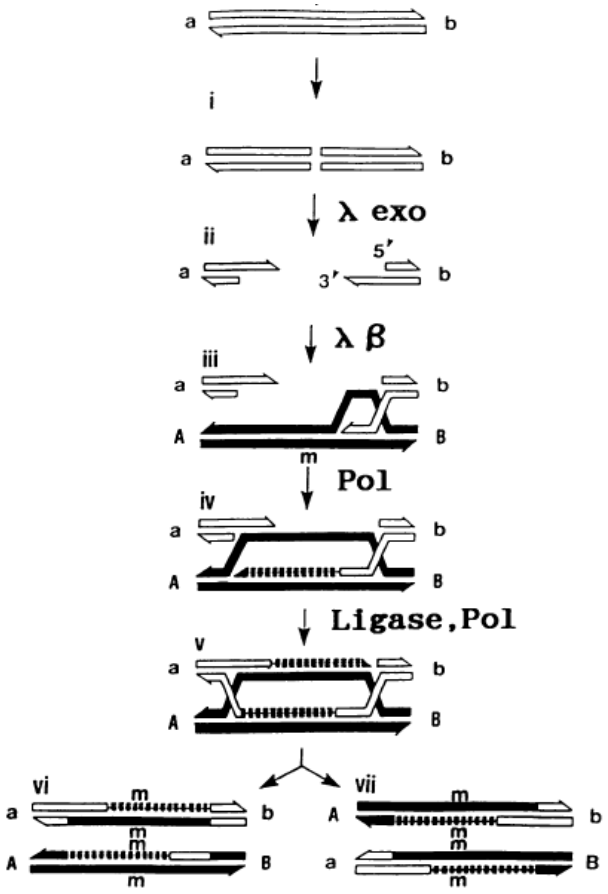



**Figure 3.1 Double strand break repair (DSBR) recombination pathway of λ-exo.** Scheme showing the mechanism of λ exo mediated recombination. λ-exo starts at dsDNA breaks and degrades the 5' strand creating 3' ssDNA.

Other exonucleases can shorten the ssDNA overhang thus produced. β protein (λ β) binds to the ssDNA overhang and promotes annealing with a complementary strand of a homologous duplex. Polymerase (Pol) starts DNA synthesis at the annealed ends and a double-Holliday junction is formed. Ligase fills in the gaps of the breaks and two different products could be formed. Figure from (Takahashi and Kobayashi, 1990).

## 3.3. λ-exonuclease (λ-exo)

### 3.3.1. Structure of λ-exo

The overall fold of λ-exo is similar to type II restriction endonucleases like PvuII and EcoRV (Kovall and Matthews, 1998), and other DNA processing enzymes such as MutH (Lee *et al.*, 2005), RecBCD (Singleton *et al.*, 2004), the β subunit of *E. coli* DNA polymerase III (Kelman and O'Donnell, 1995) and *E. coli* RecE (Zhang *et al.*, 2009). λ-exo forms toroid shaped trimers with a non-crystallographic three-fold symmetry around a central tapering channel (Kovall and Matthews, 1997b) (Fig. 3.2). The wider end of the channel (~3 nm) lets the dsDNA substrate in and the ssDNA product comes out of the other end (~1.5 nm), which is narrow enough to accommodate only ssDNA. Computational studies suggested that 11 nucleotides (nt) of dsDNA can reside within the central channel of λ-exo (Kovall and Matthews, 1998). Nuclease protection assays revealed a similar value of 13–14 nt (Mitsis and Kwagh, 1999).

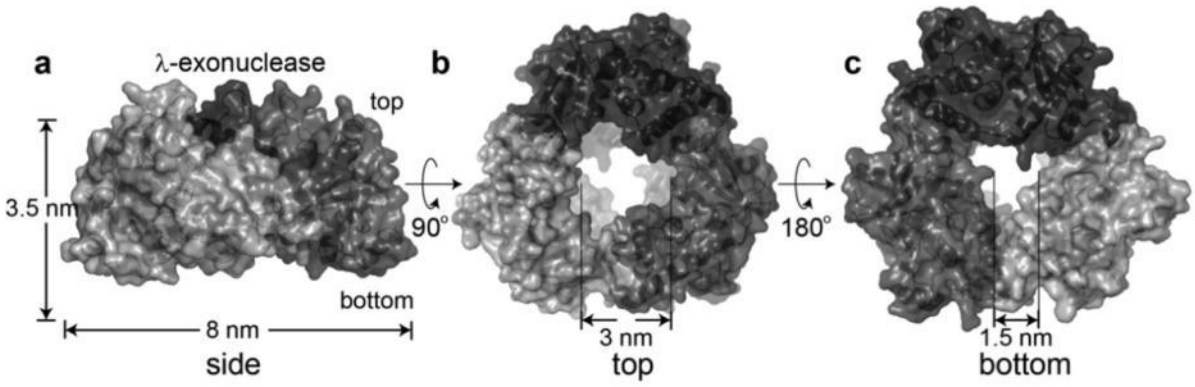



**Figure 3.2 The torroidal structure of λ-exo**. Space filling models of the λ-exo trimer (PDB ID 1AVQ) showing the (**a**) side-view (**b**) top-view and (**c**) the

bottom-view. When viewed from the top or the bottom (**b** and **c**), the three-fold symmetry of λ-exo around a central channel can be seen. The central channel tapers from ~3 nm at the top where dsDNA enters to ~1.5 nm at the bottom through which the ssDNA product exits.

### 3.3.2. λ-exo is a highly processive enzyme

Due to the geometrical constraint of the pore openings with respect to dsDNA and ssDNA, λ-exo should remain linked to the DNA substrate during activity (Kovall and Matthews, 1997a). The topological linkage of λ-exo with DNA, like other ring shaped enzymes, forms the structural basis of its high processivity (Zhang *et al.*, 2011). λ-exo can cleave ~3000 nt of dsDNA at a constant rate before detaching from it (Carter and Radding, 1971; Little, 1967). Xie and co-workers derived that single λ-exo molecules cleaved 18000±8000 nucleotides on an average before dissociating from immobilized DNA substrates in single molecule enzyme activity experiments (van Oijen *et al.*, 2003).

### 3.3.3. Choice of co-factors and DNA substrates

λ-exo is an ATP independent enzyme and $Mg^{2+}$ is the only co-factor required for its exonuclease activity (Lee *et al.*, 2011; Zhang *et al.*, 2011). In the absence of $Mg^{2+}$, λ-exo can also bind other divalent metal ions like $Mn^{2+}$ and $Ca^{2+}$ (Kovall and Matthews, 1997b). However, λ-exo activity is reduced 4 fold with $Mn^{2+}$ as compared to activity in $Mg^{2+}$, and is completely inhibited by $Ca^{2+}$ (Kovall and Matthews, 1997b; Mitsis and Kwagh, 1999).

Like other nucleases, the structure of λ-exo has evolved to specialize in processing only particular DNA substrates (Kovall and Matthews, 1998). λ-exo cannot initiate digestion at nicks and has limited activity on gaps (Carter and

Radding, 1971; Radding and Carter, 1971). The enzyme can act on dsDNA ends only. The rate of digestion depends on the dsDNA ends, 5' recessed ≥ blunt >> 5' overhang (Mitsis and Kwagh, 1999). However, λ-exo cannot initiate digestion on dsDNA when its 3' overhang is larger than > 100 bp due to steric hindrance posed by the secondary structure of the ssDNA (Sriprakash *et al.*, 1975). Even though λ-exo can bind to ssDNA, it can digest ssDNA only distributively. The rate of ssDNA digestion is inversely proportional to the length of ssDNA, and 10–100 times slower than the rate of digestion of dsDNA of the same length (Sriprakash *et al.*, 1975).

### 3.3.4. Bulk activity assays

λ-exo activity has been studied *in vivo* by complementation studies using λ strains with *exo* gene knockouts (λ *exo*$^{-}$ *gam*$^{-}$) (Subramanian *et al.*, 2003). As *exo* function is required for phage λ propagation, λ *exo*- strains cannot grow on *E. coli* (su° recA1) strains. However, if the *E. coli* RecA hosts carry plasmids containing the gene of an active λ-exo mutant, λ *exo*- can engage in homologous recombination *via* the RecBCD pathway of *E.coli* and form plaques.

For studying λ-exo activity *in vitro*, gel-shift assays (Subramanian *et al.*, 2003) and methods employing fluorescent dyes (Tolun and Myers, 2003) or bases (Mitsis and Kwagh, 1999) have been used. Myers and co-workers developed a real-time fluorescence assay using PicoGreen, a dsDNA binding fluorescent dye (Tolun and Myers, 2003). PicoGreen was chosen amongst two other DNA binding dyes, ethidium bromide and SYBR Gold because of its least inhibitory

effect on λ-exo activity and high differential fluorescence output when bound to dsDNA *vs.* ssDNA. As λ-exo processed each end of a linear dsDNA substrate into ssDNA and 5'-mononucleotides, the PicoGreen emission fluorescence decreased, which in turn could be correlated with λ-exo activity kinetics (Fig. 3.3).

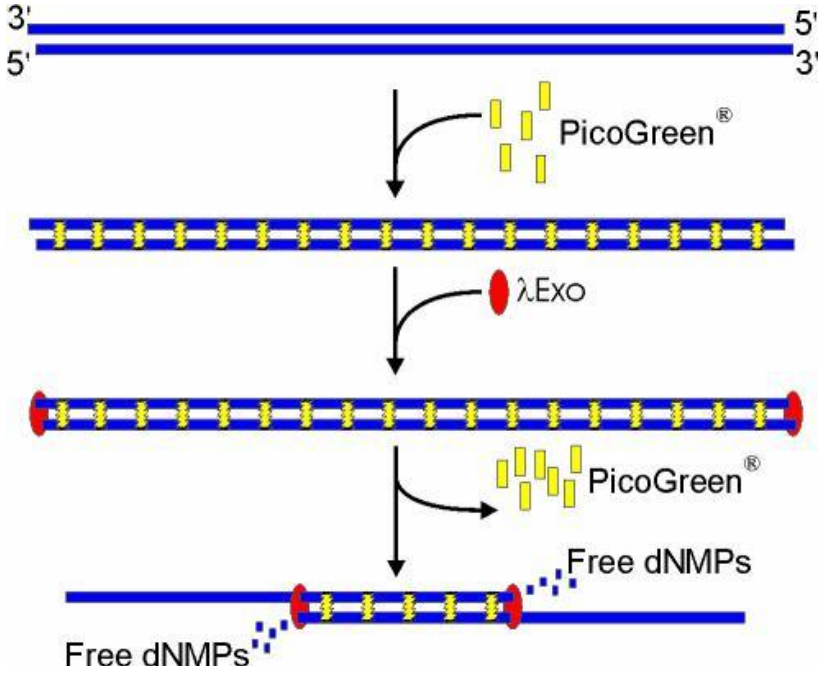



**Figure 3.3 PicoGreen fluorescence assay of exonuclease activity.** Scheme of the activity assay of λ-exo using the dsDNA binding fluorescent dye, PicoGreen. PicoGreen has a high differential fluorescence output when bound to dsDNA than ssDNA. A linear dsDNA template (blue lines) was chosen as the substrate and incubated with PicoGreen (yellow rectangles). Upon adding λ-exo (red ellipses) and subsequently $Mg^{2+}$ to initiate the reaction, λ-exo digested the 5' strands from both ends into mononucleotides (dNMP) displacing the bound PicoGreen dye. λ-exo digestion could thus be observed in real-time as a decrease in the emission fluorescence. Figure reprinted by permission from Oxford University Press, from (Tolun and Myers, 2003).

In the real time PicoGreen fluorescence assay, pUC19 plasmid linearized with KpnI was used as the DNA substrate. The dsDNA and PicoGreen mixture was illuminated at 484 nm (excitation wavelength) and emission was measured at 522 nm. The speed of digestion of dsDNA with phosphorylated ends by λ-exo at pH 7.5 and room temperature was found to be 4-5 nt $s^{-1}$ (Tolun and Myers,

2003). According to some other bulk activity assays, the rate of digestion of dsDNA with similar ends by λ-exo was 12-17 nt $s^{-1}$ (Mitsis and Kwagh, 1999; Subramanian *et al.*, 2003). Similar rates lying between 13 to 32 nt $s^{-1}$ were observed in different single molecule studies (section 3.3.5) (Dapprich, 1999; Perkins *et al.*, 2003; van Oijen *et al.*, 2003).

### 3.3.5. Single molecule activity studies

Single molecule studies of enzymes are pivotal in understanding their dynamics at the molecular level (Min *et al.*, 2005). Single molecule activity of many DNA processing enzymes such as RNA polymerase (Schafer *et al.*, 1991; Wang *et al.*, 1998), helicases (Johnson *et al.*, 2007), RecBCD (Perkins *et al.*, 2004), SSBs (Zhou *et al.*, 2011) and exonucleases (Lee *et al.*, 2011) has been studied using optical techniques (Ha *et al.*, 2012). These studies have thrown light on the enzymes' dynamic features such as phases of enzymatic action (Lee *et al.*, 2011), sequence dependent pausing (Perkins *et al.*, 2003), enzyme backtracking and arrest (Herbert *et al.*, 2006), translocation step sizes and enzyme shuttling on DNA (Ha *et al.*, 2012). Such high-resolution spatial and temporal mechanistic details of enzymes are obfuscated by ensemble averaging in bulk studies (Finkelstein and Greene, 2008; Ha *et al.*, 1996).

Single molecule studies of λ-exo have been very limited. Matsuura *et al.* studied λ-exo digestion of single dsDNA molecules stretched under a DC electric field (Matsuura *et al.*, 2001) (Fig. 3.4). The dsDNA substrate was biotin labeled and anchored on an avidin-coated surface. A dsDNA intercalating dye (YOYO-1) that fluoresced only when bound to dsDNA was used to image the length of

dsDNA in real time. λ-exo digestion of dsDNA was seen as a decrease in the length of dsDNA. The rate of digestion by single λ-exo molecules were thus calculated to be ~1000 nt $s^{-1}$, which is anomalously high as compared to other reported values of λ-exo activity (section 3.3.4). Exonuclease activity of λ-exo was inhibited at high concentrations of YOYO-1 whose intercalation with dsDNA possibly led to a change in the DNA conformation. DNA intercalating dyes have been shown to cause inhibition in exonuclease activity of other exonucleases as well (Tachiiri *et al.*, 2000; Tolun and Myers, 2003). Thus it is possible that YOYO-1 could be affecting λ-exo rate of digestion even at low concentrations.

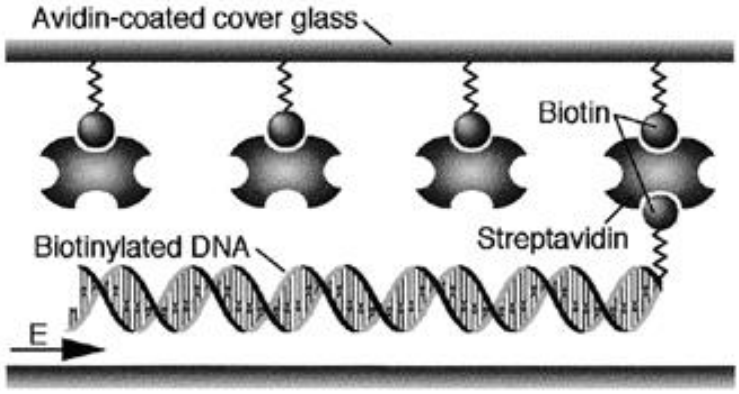



**Figure 3.4 Single molecule DNA digestion by λ-exo.** Experimental set-up to study single DNA molecules digestion by λ-exo. λ DNA molecules were biotinylated at one end and tethered to an avidin-coated cover-glass. The DNA molecules were stretched and straightened by applying a D.C. electric field (E). YOYO-1, a dsDNA intercalating dye was used to image the length of the dsDNA in real time and hence served as a marker to assess the extent of digestion by single λ-exo molecules. Figure reprinted by permission from Oxford University Press, (Matsuura *et al.*, 2001).

Block and co-workers studied the sequence dependent pausing behaviour of single λ-exo molecules using optical tweezers (Perkins *et al.*, 2003). The methodology involved incubating a His-tagged λ-exo with a 7.1 kb dsDNA substrate in presence of $Ca^{2+}$. λ-exo molecules were then anchored stereo-specifically on a glass cover slip using anti His-tag antibodies. The DNA

substrate was attached to a small polystyrene bead. The polystyrene bead with dsDNA was captured using an optical trap and its displacement monitored by back focal plane interferometry (Fig. 3.5). λ-exo digested dsDNA at an average speed of 10±2 nt $s^{-1}$ frequently interspersed with pauses. The longest pause was observed at DNA sequences containing the motif GGCGATTCT. This was confirmed by gel shift assays using short $^{32}P$ labeled oligonucleotides containing the pause sequence. The pausing behaviour was attributed to tight ring stacking interactions between the purine bases (G and A) of the pause sequence and the residue W24 of the central channel.

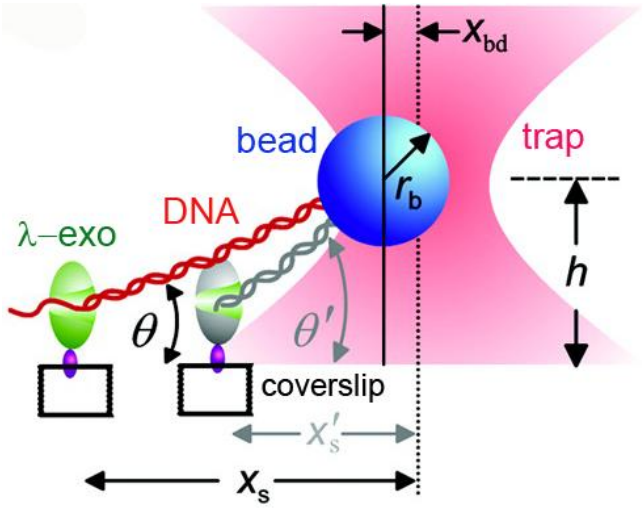



**Figure 3.5 Single molecule studies of λ-exo using an optical trap.** Cartoon showing single molecule optical trap experimental set-up used for λ-exo. Uniform particles such as silica or polystyrene beads can be trapped at the focal point of a microscope objective by a laser beam (Sheetz, 1998). An external force (F) will then displace the trapped bead by a distance (*x*) proportional to the applied force, i. e., F=k*x*, where, k is the proportionality constant (Mehta *et al.*, 1999). λ-exo molecules (green ellipses) were tethered on a cover-glass coated with anti-His antibodies. A 7.1 kb dsDNA substrate, derived from M13 phage DNA, was attached to a small polystyrene bead which was trapped by an optical trap at a height (h) of 200 nm from the cover slip. The stage was moved from its position $x_s$ to $x'_s$ as λ-exo digested the DNA substrate to keep the bead under a constant force of 1.07 pN. This was used to deduce the length of the DNA that remained undigested and the speed of single λ-exo molecules. Figure adapted from (Perkins *et al.*, 2003). Reprinted with permission from AAAS.

Single molecule Förster Resonance Energy Transfer (smFRET) studies by Ha and co-workers revealed three phases of λ-exo activity- (i) an intitation phase, (ii) a distributive phase, and (iii) a processive phase (Lee *et al.*, 2011). The experimental set-up for smFRET experiment of λ-exo consisted of immobilized 5’ phosphorylated dsDNA substrates attached to a polymer coated quartz surface *via* biotin-neutravidin linkage (Fig. 3.6). The FRET donor (Cy5) and acceptor (Cy3) were attached to the strand, which would not be hydrolyzed (3’ strand) spaced by 20 nt. λ-exo was added to the system using a flow-delivery system. On digestion of the unlabeled strand (5’ strand) by λ-exo, the ssDNA generated (3’ strand) coiled leading to a time-averaged decrease in the distance between Cy3 and Cy5, and a corresponding increase in FRET intensity. The three phases of λ-exo were thus deduced by correlating the increase in FRET with the speed of λ-exo activity.

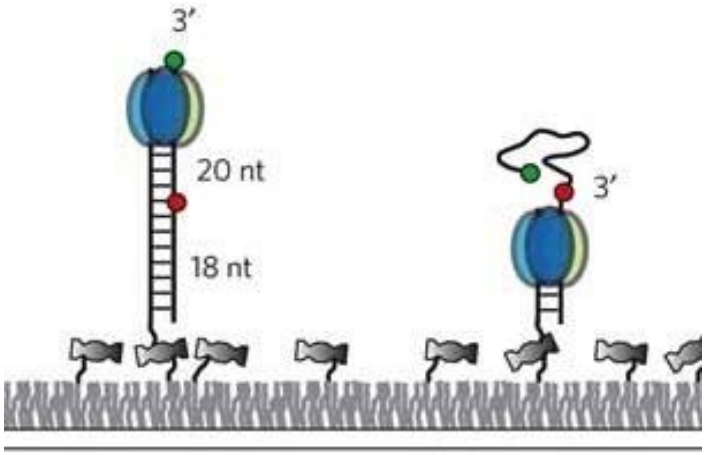



**Figure 3.6 smFRET of λ-exo.** Schematics of the smFRET experiment of λ-exo. FRET requires two chromophores, a donor (D) and an acceptor (A) such that upon excitation, D can transfer its energy to A making it fluoresce (Ha *et al.*, 1996). The efficiency of energy transfer between the chromophores is through non-radiative dipole–dipole coupling and is inversely proportional to the sixth power of the distance between D and A. Thus the emission of A, measured as FRET output, is strongly dependent on the distance between the chromophores and can be used for distances between 1 – 10 nm (Borgia *et al.*, 2008). Two chromophores, D (green) and A (red) were attached on the 3’ strand of dsDNA substrates tethered on a coverslip. Upon λ-exo digestion of the 5’ strand, the 3’ strand turned into ssDNA and coiled. The decrease in distance between the

chromophores led to an increase in FRET. Reprinted by permission from Macmillan Publishers Ltd: [Nature Chemical Biology] (Lee *et al.*, 2011), copyright (2011).

Between the addition of protein and the start of digestion, there was a lag time, which was termed as the initiation phase. This was attributed to the time λ-exo took to bind and form productive complexes with the dsDNA ends. λ-exo degraded the first 20 base pairs (bp) of the duplex DNA substrate distributively with many pauses (distributive phase). These pauses were due to the enzyme falling off from the dsDNA as a whole or as monomers. When the 3' ssDNA overhang was long enough to stably engage the enzyme, λ-exo digested the substrate processively without pausing (processive phase).

The enzymatic phases of λ-exo like RNA polymerase could be assumed to have important biological relevance (Roberts, 2006). The distributive phase with frequent events of enzyme – DNA detachment could allow other viral factors to interact with λ-exo and the DNA, and hence regulate dsDNA degradation *in vivo*. Using the same experimental set-up, Ha and co-workers also found that the rate of λ-exo digestion decreases cooperatively with the number of mismatches in the dsDNA substrate and completely stalls for > 30 min on large dsDNA bubbles.

Thus, single molecule studies of λ-exo have offered insights into its enzymatic action, which could not be derived from *in vivo* or *in vitro* bulk assays.

### 3.3.6. Mechanism of DNA cleavage

The mechanism of λ-exo for dsDNA degradation has been thought to be similar

to other ring-shaped enzymes such as RecE (Zhang *et al.*, 2009). The size and topology of the central pore of λ-exo are very similar to RecE, although the latter forms homo-tetrameric rings for activity. With an analogous mechanism, dsDNA should enter the wider end of the pore, get resected into ssDNA, one of which should be cleaved into mononucleotides (Yang *et al.*, 2011). Zhang *et al.* proposed a similar mechanism of activity with their crystal structure studies of λ-exo in complex with two kinds of DNA, (i) a 5' hydroxylated dsDNA (PDB ID 3SLP), and (ii) an active dsDNA substrate with one of the 5' strands phosphorylated (PDB ID 3SM4) (Zhang *et al.*, 2011). The structures of the λ-exo-dsDNA complex provided an insight into its mechanism of action (Fig. 3.7a).

dsDNA digestion is initiated by the uncoiling of two terminal base pairs from the 5' phosphorylated end, by the apolar residues, V73, A75, A77 and L78, of λ-exo. These apolar residues wedge into the base pairs of two nucleotides downstream from the ssDNA-dsDNA junction forming favourable hydrophobic interactions and split the base pairing as the enzyme moves forward on the DNA. The downstream portion of the dsDNA interacts with the loop residues 42-50 (Fig. 3.7a-c). The positively charged guanidino groups of R45 insert into the sugar phosphate backbone of the downstream minor groove of the DNA and narrow it by 1 – 2 Å. This tilts the DNA relative to the central axis of the protein pore and directs the unwound 5' strand towards the active site. Unwinding of the dsDNA substrate by two base pairs and subsequent alignment of the 5' strand are crucial steps taken by λ-exo before DNA digestion. Thus, removal of the L78 or the R45 residue completely inhibits exonuclease activity.

After unwinding and the correct alignment of 5’ strand with respect to the active site of one of the monomers (subunit *B*), the terminal phosphate group of the 5’ strand is then driven by electrostatic attraction into a positively charged pocket near the active site (Fig. 3.7a, b). The 5’ end makes electrostatic contacts with the side chain of R28 and forms hydrogen bonds with three backbone neighbouring amides, the side chains of three serines and threonines and the indole ring of W24. If the R28 or the W24 residues are removed, λ-exo has reduced activity and processivity (Subramanian *et al.*, 2003; Zhang *et al.*, 2011). The topological interactions between λ-exo and DNA explain its poor activity on dsDNA with mismatches. DNA with mismatches has poor helicity and the 5’ end cannot be aligned with the active site (Lee *et al.*, 2011). If the 5’ end is not phosphorylated, i. e., 5’-OH, it cannot be inserted fully in the active site and forms an inactive complex with λ-exo (Subramanian *et al.*, 2003; Zhang *et al.*, 2011).

The residues of λ-exo required for resecting DNA and making contact with the 5’ phosphate are also found in other proteins are conserved in some other proteins, such as RecE (Zhang *et al.*, 2009), and the alkaline 5’->3’ exonucleases of Kaposi’s sarcoma-associated herpes-virus (KSHV) (Dahlroth *et al.*, 2009) and Epstein-barr virus (EBV) (Buisson *et al.*, 2009).

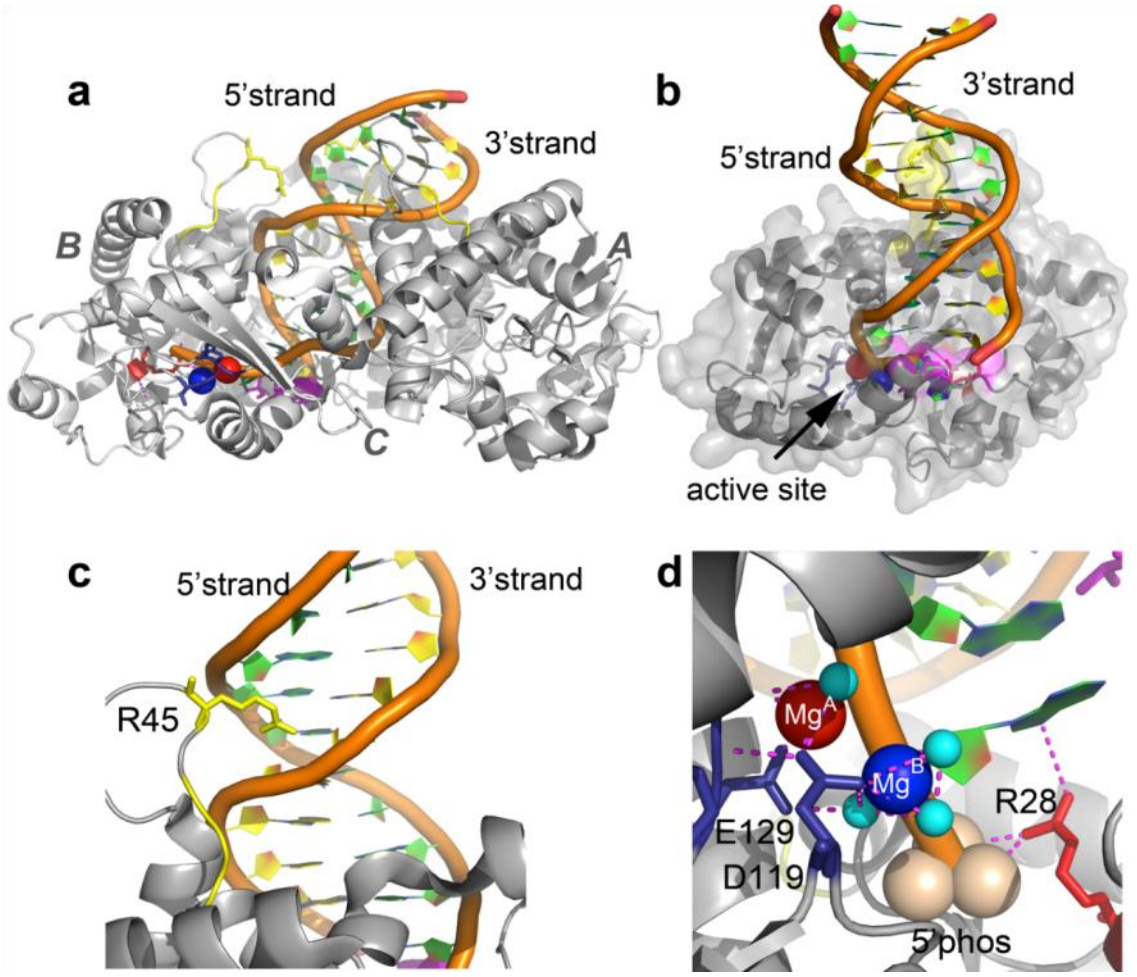



**Figure 3.7 λ-exo interactions with dsDNA substrate.** Cartoons showing (**a**) the crystal structure of λ-exo with a 5' phosphorylated dsDNA substrate. The dsDNA enters *via* the top face of λ-exo and the loop residues 42-50 (yellow, R45 is shown as sticks) tilt it with respect to the central axis of λ-exo to align the 5' end with the active site of one of the monomers (subunit *B*). The dsDNA is first unwound by two base pairs by the hydrophobic residues V73, A75, A77 and L78 (magenta). The 5' phosphate is then pulled into a positive pocket formed by R28 (red sticks) near the active site (blue sticks). Using a two metal (red and blue spheres) DNA cleaving mechanism like some other nucleases (Beese and Steitz, 1991; Nowotny *et al.*, 2005; Pingoud *et al.*, 2005), λ-exo cleaves the last base of the 5' strand while the 3' strand comes out of the central channel. However, as λ-exo unwinds two bases initially, the penultimate base is then pulled into the positive pocket for the next round of cleavage. (**b**) Subunit *B* showing residues as sticks (yellow, magenta and blue) that make contact with the dsDNA. Subunits *A* and *C* are not shown for clarity. (**c**, **d**) Enlarged regions of subunit *B* highlighting some residues that are crucial for the exonuclease activity of λ-exo. (**c**) The side chain of R45 (yellow sticks) inserts into the minor-groove of dsDNA and tilts it. (**d**) The 5'-phosphate group (wheat spheres) is electrostatically pulled by R28 (red sticks) into the active site formed of E85, D119 and K131. The active site also harbours two $Mg^{2+}$ ions, $Mg^{A}$ (red sphere) and $Mg^{B}$ (blue sphere), required for DNA cleavage. $Mg^{A}$ is bound by the side chains of D119, E129, and the backbone carbonyl of L130. $Mg^{B}$ is coordinated by three water molecules (cyan), D119 and the scissile phosphate

of the DNA. Magenta dashed lines represent polar contacts. All figures were rendered in PyMOL using the co-crystal structure of λ-exo with 5'-phosphorylated dsDNA (PDB ID 3SM4).

The active site is formed of the residues E85, D119 and K131. In the subunit harbouring the 5' end of the dsDNA substrate (subunit *B*), a $Mg^{2+}$ ion ($Mg^{A}$) is bound by the side chains of D119, E129 and the backbone carbonyl of L130 (Fig. 3.7d). Another $Mg^{2+}$ ion ($Mg^{B}$) is found 4 Å away from $Mg^{A}$ also exhibiting octahedral coordination. $Mg^{B}$ is coordinated by three water molecules, D119 and the scissile phosphate of the DNA. The two-metal ion structural motif (Yang *et al.*, 2006) is also found in other DNA cleaving enzymes such as the proof reading domain of *E. coli* DNA polymerase I (Beese and Steitz, 1991), RNase H (Nowotny *et al.*, 2005), MutH (Lee *et al.*, 2005) and other type II restriction endonucleases (Pingoud *et al.*, 2005). Some of these proteins also contain residues analogous to the conserved residues, E85, Q157 and Y154 of λ-exo. Hence, by analogy, the mechanism of DNA scission should be similar to other two-metal nucleases where, $Mg^{A}$ activates a water molecule for nucleophilic attack on the phosphate backbone, and $Mg^{B}$ stabilizes the transition state by a bidentate coordination with the leaving phosphate group. It has been postulated that the cleaved nucleotides and possibly $Mg^{B}$ exit through a small rear portal of λ-exo aided by the movement of R28 side chain (Fig. 3.8). After the release of the first base, the next nucleotide to be cleaved is already unwound and its newly exposed phosphate is attracted towards the positive active site. This makes the enzyme move forward on the DNA and the hydrophobic wedge formed by the apolar residues restrict backward movement.

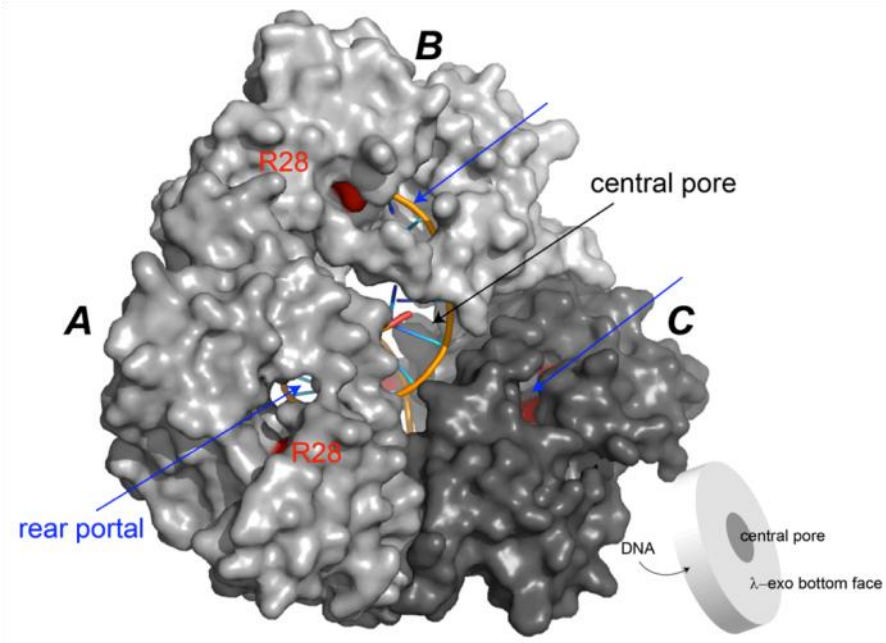



**Figure 3.8 Reorientation of the R28 side chain to eject nucleotides.** The R28 side chain of subunit *A* also bound a phosphate ion like subunits B and C but its R28 side chain was shifted by 8.5 Å. Based on this observation, it was postulated that the cleaved nucleotides are exited through the small portal in the bottom face of the protein and the 3' strand comes out the central channel. A schematic (right bottom) shows the orientation of the λ-exo structure in the figure.

While the 5' strand is degraded into individual nucleotides, the 3' strand comes out of the central channel. The key residues of λ-exo making contact with the dsDNA were illustrated in the co-crystal structure of λ-exo with dsDNA (Fig. 3.9). The interactions of the residues in the central channel with the 3' strand are still not clear as the 3' overhang of the dsDNA substrate used for the co-crystal structure was too short.

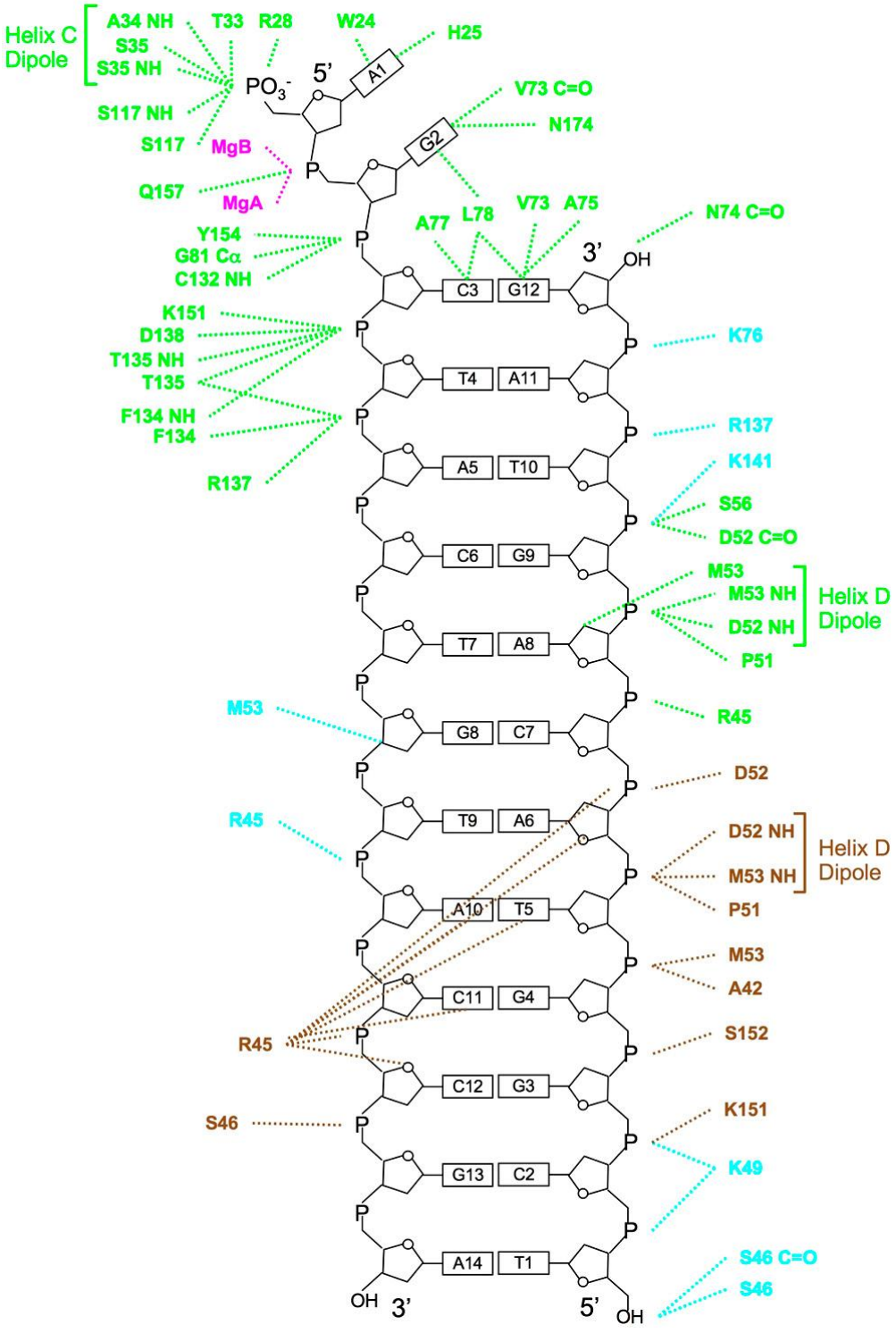



**Figure 3.9 λ-exo residues that interact with dsDNA.** Figure showing the λ-exo residues that make hydrogen bonds (contacts within 3.5 Å) and ion pairs (contacts within 5 Å) with the dsDNA substrate. Residues are colour coded according to the subunits; cyan for subunit *A*, green for subunit *B*, and brown for subunit *C*. The removal of R28, R45, D119, K131 or R137 led to a total loss in exonuclease activity. W24A, L78A or E85A had much diminished levels of exonuclease activity. Mutations of M53, K49, or K76 had only modest effects on the exonuclease activity. Figure from (Zhang *et al.*, 2011).

Unlike other ring shaped DNA processing enzymes such as hexameric helicases, which are driven by hydrolysis of dATP and/ or dTTP (Crampton *et al.*, 2006; Singleton *et al.*, 2000), λ-exo does not work like a motor. Structural alignment of the crystal structures of λ-exo with and without DNA suggested that the subunits do not rearrange themselves around the dsDNA substrate like helicases. The processive mechanism of λ-exo is like an electrostatic ratchet where the negatively charged 5' end of dsDNA is pulled towards the active site, which drives unwinding of the dsDNA by two bases, and backward movement is restricted.

However, many other questions regarding the structure and mechanism of λ-exo still remain unanswered. The trimeric assembly mechanism of λ-exo is not directly evident. Does λ-exo wrap itself as a trimer around a dsDNA or does it thread through the DNA as a preformed trimer? Using size exclusion chromatography (SEC) we were able to show that WT λ-exo exists as trimers in solution (section 3.8.11) in absence of dsDNA (Zhang *et al.*, 2011) and phosphate ions (Kovall and Matthews, 1997b) (section 3.8.11). The inactivity of λ-exo at nicks and gaps and with dsDNA substrates with very long 3' overhangs also suggests that λ-exo threads dsDNA as pre-formed trimers (Carter and Radding, 1971; Mitsis and Kwagh, 1999; Radding and Carter, 1971). In the co-crystal structure of λ-exo with dsDNA, the 5' phosphorylated end of the dsDNA was projected into the active site of one of the monomer subunits. Hence, it is not clear whether all the monomer subunits act in a sequential manner or the active site of only one subunit is used while the other two subunits clamp onto the dsDNA as λ-exo processively degrades 1000s of base-pairs (van Oijen *et*

*al.*, 2003). Studies with hetero-trimers of λ-exo with 1 or 2 inactive subunits will be very informative in this regard. The interactions of the 3' strand with the residues of the central channel also need to be investigated. Further single molecule studies of λ-exo might reveal answers to these outstanding questions.

## 3.4. Coupling λ-exo to the αHL nanopore

### 3.4.1. Nanopores for studying single molecule enzymatic activity

Nanopores such as αHL have been used for stochastic sensing of small molecules and even large macromolecules such as DNA, RNA and proteins (Chapter 2, section 2.3) (Bayley *et al.*, 2000; Cheley *et al.*, 2006; Howorka *et al.*, 2001; Movileanu *et al.*, 2000). The high current density through αHL confers it a high signal to noise ratio and allows resolution of events as small as a few microseconds (Deamer, 2010). Charged molecules of DNA and RNA can be driven into the pore by the electric field of the applied potential (Kasianowicz *et al.*, 1996). Proteins bound to DNA or RNA can also be pulled close to the αHL pore (Astier *et al.*, 2007; Hornblower *et al.*, 2007; Stoddart *et al.*, 2009). The interaction between the protein and the DNA in their complex captured by the αHL pore can then be probed at the single molecule level (Deamer, 2010).

The strength of interaction between a protein and DNA can be quantified using nanopores. This technique has been termed as nanopore force spectroscopy (Dudko *et al.*, 2010). In their pioneering work, Hornblower *et al.* determined the kinetics of ssDNA binding to Exonuclease I (Exo I) of *E.coli* using αHL (Hornblower *et al.*, 2007). Individual Exo I-ssDNA complexes were pulled into the αHL pore under an applied voltage, which was then increased at a constant

ramp. The voltage at which the ssDNA was released from Exo I was used to compute the electrical force on the ssDNA and the kinetic dissociation and association rate constants of the Exo I-ssDNA complex.

In a similar way, it was shown that the complexes of the Klenow fragment with a dsDNA template in the presence and absence of an active NTP substrate could be distinguished using αHL (Benner *et al.*, 2007). The dsDNA template was a ssDNA hairpin whose average lifetime within the pore was ~1 ms. The Klenow fragment could bind to the dsDNA part of the hairpin in solution and be drawn on the pore mouth *via* the ssDNA overhang. In the presence of an active NTP substrate, the average lifetime of the Klenow fragment-DNA complex on the αHL pore was 100 times longer (~200 ms) than in the absence of NTP (2-3 ms).

Akeson and co-workers also showed that the catalytic action of single molecules of T7 DNA polymerase and phi29 DNA polymerase on DNA could be studied in real time using the αHL pore (Lieberman *et al.*, 2010; Olasagasti *et al.*, 2010). The DNA polymerase-DNA complex was captured by the pore under the applied potential. Using the ssDNA as the template, the DNA polymerase synthesized the complementary strand of the ssDNA and pulled it against the applied potential. The movement of the ssDNA within the pore was an indication of the sequential addition of bases by the DNA polymerase. Following this work, it was also shown that phi29 DNA polymerase could be stalled at the αHL pore opening and its pre- and post-translocation states with a DNA substrate could be resolved by the movement of the ssDNA captured within the αHL pore (Dahl *et al.*, 2012).

These studies have shown that single molecules of enzymes could be brought in the proximity of the nanopores using a charged molecule such as DNA and their kinetics studied. However, the duration of the enzymatic studies through nanopores is limited by the frequency and length of enzyme-DNA capture events. Hence, if an enzyme could be linked permanently to the αHL pore, continuous single molecule studies will be possible.

### 3.4.2. Nanopores for DNA sequencing

Apart from its use in studying single molecule chemical and enzyme kinetics and stochastic sensing, the use of the αHL pore and other pores such as the mycobacterium toxin MspA, has been implicated for ultra-fast DNA sequencing (Cherf *et al.*, 2012; Manrao *et al.*, 2012). Engineered αHL pores can distinguish between all the DNA nucleotides individually, termed as base sequencing (Astier *et al.*, 2006; Clarke *et al.*, 2009). The bases of a ssDNA stalled within the αHL pore can also be recognized, termed as strand sequencing (Stoddart *et al.*, 2009). Earlier this year (2012), the Akeson and Gundlach laboratories showed that DNA could be sequenced (strand sequencing) through nanopores using phi29 DNA polymerase (Cherf *et al.*, 2012; Manrao *et al.*, 2012). Cherf *et al.* used a 94 mer ssDNA template to which a 23 nt primer was annealed (Cherf *et al.*, 2012). A 15 or 25 nt blocking oligonucleotide was added to anneal adjacent to the 23 nt primer, with 2 acridine residues attached to its 5’ end to inhibit activity of DNA polymerase in the solution. The inactive DNA-enzyme complex was captured by the nanopore under an applied potential of +180 mV. As the ssDNA overhang was pulled further inside the pore vestibule, the blocking

oligonucleotide unzipped, and the bases of the ssDNA could be read at a median speed of 2.5 nt $s^{-1}$. When the blocking oligonucleotide completely dissociated from the ssDNA, the phi29 polymerase encountered the 3'OH of the 23 nt primer, and started synthesizing the complementary strand to the ssDNA, pulling it through the αHL pore against the applied potential. The sequence of the ssDNA template was read electrically again, but now in the opposite direction at ~40 nt $s^{-1}$ (median rate) (Fig 3.10).

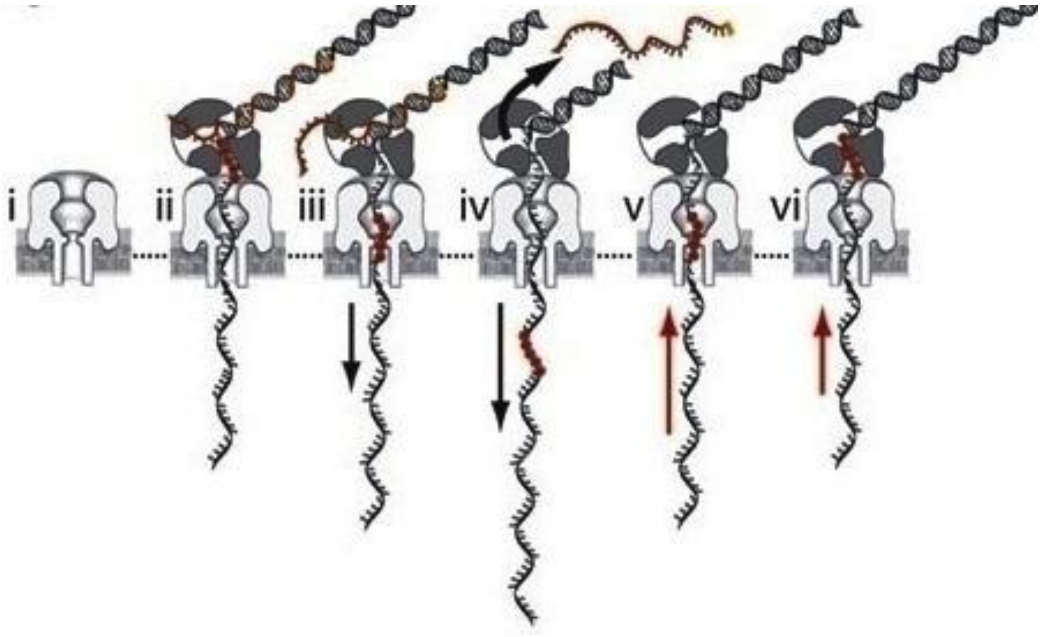



**Figure 3.10 DNA sequencing through α-hemolysin (αHL).** A dsDNA substrate was prepared with a ssDNA overhang (grey). A blocking oligomer (red) with two acridine residues at its 5′ end was used to anneal to the ssDNA overhang of the dsDNA substrate and prevent action of T7 DNA polymerase in solution. (**i-iii**) The complex of phi29 DNA polymerase and dsDNA was pulled into the αHL pore by applying an electric potential. This also caused the blocking oligomer to unzip from the ssDNA overhang. The ssDNA overhang was pulled into the αHL pore and read base by base. (**iv-vi**) After the blocking oligomer was completely unzipped, the phi29 DNA polymerase started synthesizing dsDNA and the same stretch of DNA was ratcheted backwards and proof read. Reprinted by permission from Macmillan Publishers Ltd: [Nature Biotechnology] (Cherf *et al.*, 2012), copyright (2012).

Exonucleases are also potential candidates for use in DNA sequencing as exonucleases cleave DNA in a sequential manner with high processivities and

speeds (Chan, 2005; Foldes-Papp *et al.*, 2001; Werner *et al.*, 2003). In late 1980's, Jett *et al.* proposed the use of processive exonucleases to sequence ~ 40 kb DNA with fluorescently labelled bases optically, at a speed of 100-1000 bp $s^{-1}$ (Davis *et al.*, 1991; Jett *et al.*, 1989). In 2009, the Bayley laboratory demonstrated that the method of exonuclease sequencing could be integrated with the nanopore technology (Astier *et al.*, 2006; Clarke *et al.*, 2009). Exo I was added with ssDNA and $Mg^{2+}$ to a planar lipid bilayer measurement chamber (Chapter 2, section 2.10). The ssDNA was processively hydrolysed by Exo I into nucleotides, which were then recognized by an engineered αHL pore fitted with a cyclodextrin adapter (base sequencing). This method however did not allow identification of bases in the same order in which they were cleaved by the Exo I enzymes. This drawback could be circumvented if the enzyme was placed in the proximity of the nanopore.

### 3.4.3. Applications of the λ-exo-αHL pore complex

λ-exo could be attached to the αHL pore for single molecule studies of the enzyme and also for DNA sequencing, such that the central pores of the two proteins are aligned. λ-exo could cleave very long dsDNA substrates into nucleotides and ssDNA and feed them in to the αHL pore. The ssDNA generated from the dsDNA could be stalled within the αHL pore and sequentially read (Fig. 3.11a). The λ-exo-αHL pore complex could also be used for base sequencing, by detecting bases in the order in which they are generated (Fig. 3.11b). Thereby, the interactions between single λ-exo molecules and dsDNA could also be probed using the λ-exo αHL pore complex.

Nanopore based single molecule studies also have the added advantages of small sample amounts, minimal and cost-effective sample preparation, inexpensive and portable hardware and the capability to be parallelized for high throughput measurements (Branton *et al.*, 2008; Deamer, 2010). Expensive fluorescent labels are also not required which might interfere with enzyme activity (Claessen *et al.*, 2010; Ray *et al.*, 2010; Tolun and Myers, 2003).

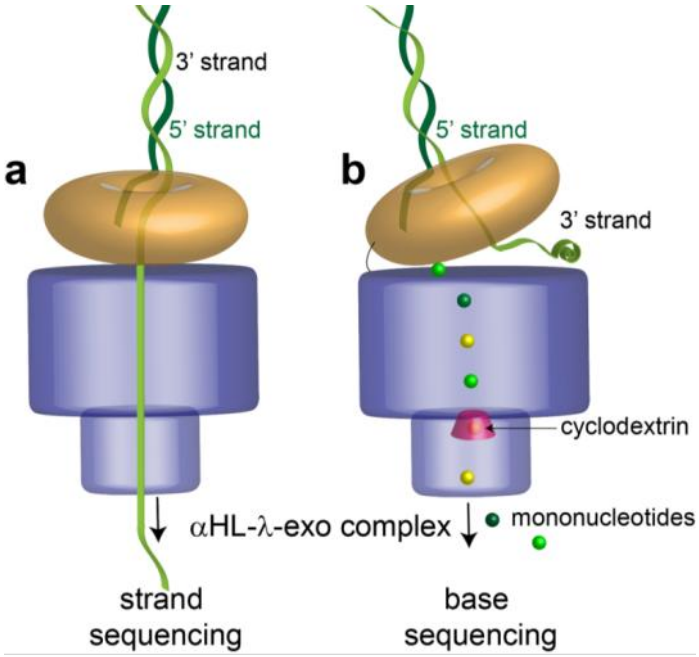



**Figure 3.11 DNA sequencing using the λ-exo-αHL pore complex.** The λ-exo-αHL pore complex will have a direct application in DNA sequencing of long dsDNA substrates. λ-exo would cleave the 5' strand of the dsDNA substrate into mononucleotides leaving the 3' strand (light green) intact and feed these products into αHL. The λ-exo-αHL pore complex could then be used for (**a**) strand sequencing where bases of the unhydrolyzed ssDNA are recognized (Cherf *et al.*, 2012) or (**b**) base sequencing, where an engineered αHL pore with a cyclodextrin adapter (pink) (Clarke *et al.*, 2009) could differentiate between the mononucleotides in the order in which they are cleaved.

Mainly two approaches were followed for attaching λ-exo on the cap of αHL with most emphasis on the second one, (i) covalent attachment using disulfide bonds and (ii) genetic fusion of αHL and λ-exo joint with a connecting flexible linker.

# Results

## 3.5. Covalent attachment of λ-exo and αHL using disulfide bonds

### 3.5.1. Cysteine mutants of αHL

The construction of the cap-to-cap dimer of αHL shows that water soluble proteins could be attached to the cap of αHL using disulfide bridges without loss in αHL activity (Chapter 2). The K237C αHL mutant used for the construction of the αHL heptamer dimer seemed a good anchor point to attach other proteins. As αHL is a heptamer and λ-exo a trimer, to overcome the asymmetry constraint, an additional cysteine mutation was added on αHL cap at position 287 (Fig. 3.12a). The double cysteine mutant of αHL, K237C/E287C/D8H6 (C2), could present a reactive cysteine to all the monomers in the λ-exo trimer. The C2 mutant was expressed using the *E.coli* T7 S30 extract for coupled *in vitro* transcription translation (IVTT) (Promega). The IVTT system contains T7 RNA polymerase and all other components required for translation of RNA into protein such as rNTPs, tRNAs, IPTG and salts. The hemolytic activity of C2 was assessed using rabbit blood cells (rRBC), whose lysis was monitored as a decrease in the light scattering at 595 nm over 2 h (Fig. 3.12b). C2 was also found to have WT like heptamer formation characteristics on rRBCm (Fig. 3.12c).

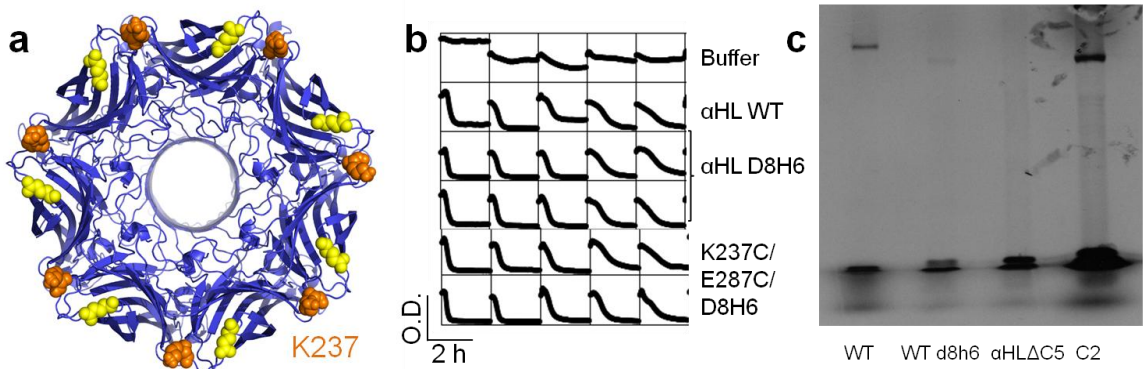



**Figure 3.12. Structure and activity of the double cysteine mutant of αHL.** To maximize the number of links between λ-exo and the cap of αHL, a double cysteine mutant of αHL (K237C/E287C/D8H6; C2) was made. (**a**) Top-view of the αHL showing the cap residues K237 (orange) and E287 (yellow) which were mutated to cysteines. (**b**) Rabbit blood cells were used for assaying hemolytic activity, whose lysis can be seen as a decrease in the optical density (O.D.). Hemolytic activity assay of *in vitro* expressed C2 (2 µl of a 25 µl IVTT reaction) showed that the activity of αHL was preserved even with the cysteine mutations. (**c**) Autoradiograph of $^{35}$S methionine labeled proteins shows that C2 formed heptamers on rRBCm like WT αHL. Heptamers with the seven D8H6 tags migrate faster than WT α7 (Chapter 2, section 2.5).

### 3.5.2. Cysteine mutants of λ-exo

Before designing cysteine mutants of λ-exo, the native cysteines of λ-exo were analyzed. Each monomer of WT λ-exo already contains four reduced cysteines. However, none of them appeared to be suitable for disulfide linkage with αHL to form the λ-exo-αHL pore complex (Fig. 3.13a).

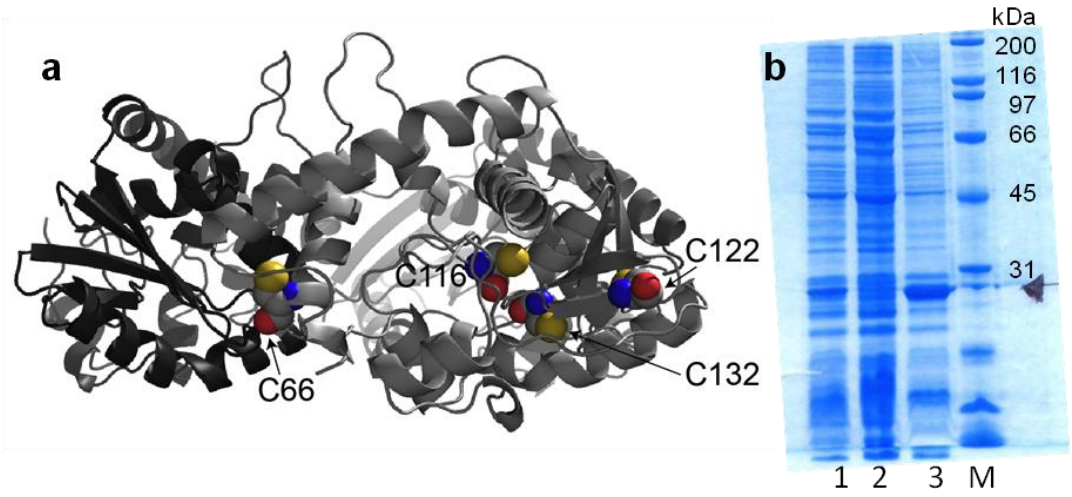



**Figure 3.13. Native cysteines of λ-exo and expression of a cysteine-less mutant of λ-exo.** (**a**) Cartoon structure of λ-exo showing the location of the four

native cysteines, C66, C112, C122 and C132 (shown as spheres), per monomer. The thiol sulfur atoms (yellow) of all the native cysteines of λ-exo seem buried (PDB ID 1AVQ). Cysteines of only one monomer have been shown for clarity. To introduce cysteine mutations in λ-exo for covalent linkage with αHL, the native cysteine residues of λ-exo were removed in several combinations. (**b**) SDS-PAGE showing that the cysteine-less mutant of λ-exo (C4) precipitated during expression in *E.coli. Lane 1* shows the whole lysate of C4, *lane 3* is the lysate pellets which contained most of the protein (marked with a black arrow) and *lane 2* is the lysate supernatant which did not contain much C4 protein. The C4 protein also did not possess any exonuclease activity.

To ensure that the native cysteines of λ-exo did not interfere with the covalent attachment of λ-exo to the αHL pore. Hence, a cysteine-less mutant of λ-exo was prepared and termed C4. All the native cysteines of λ-exo, viz., C66, C112, C122 and C132 (Fig. 3.13a), were mutated to serine. Removal of all native cysteines led to insoluble expression in *E. coli* and total loss of exonuclease activity (Fig. 3.13b).

More cysteine knockout mutants of λ-exo were designed with Dr. Stephen Cheley. Three cysteines, C116, C122 and C132 (Fig. 3.13a) were mutated to serine (C3 mutant). Like C4, soluble expression of C3 could not be achieved either, and exonuclease activity was not observed as well. A point mutant of C66, seemingly the most surface accessible native cysteine (Fig. 3.13a), was also changed to serine (C66S) and over-expressed in *E.coli*. The removal of the C66 residue did not have any effect on the exonuclease activity, and the C66S mutant had activity comparable to WT λ-exo. The C66S mutant could thus be used as a background for additional cysteine mutants.

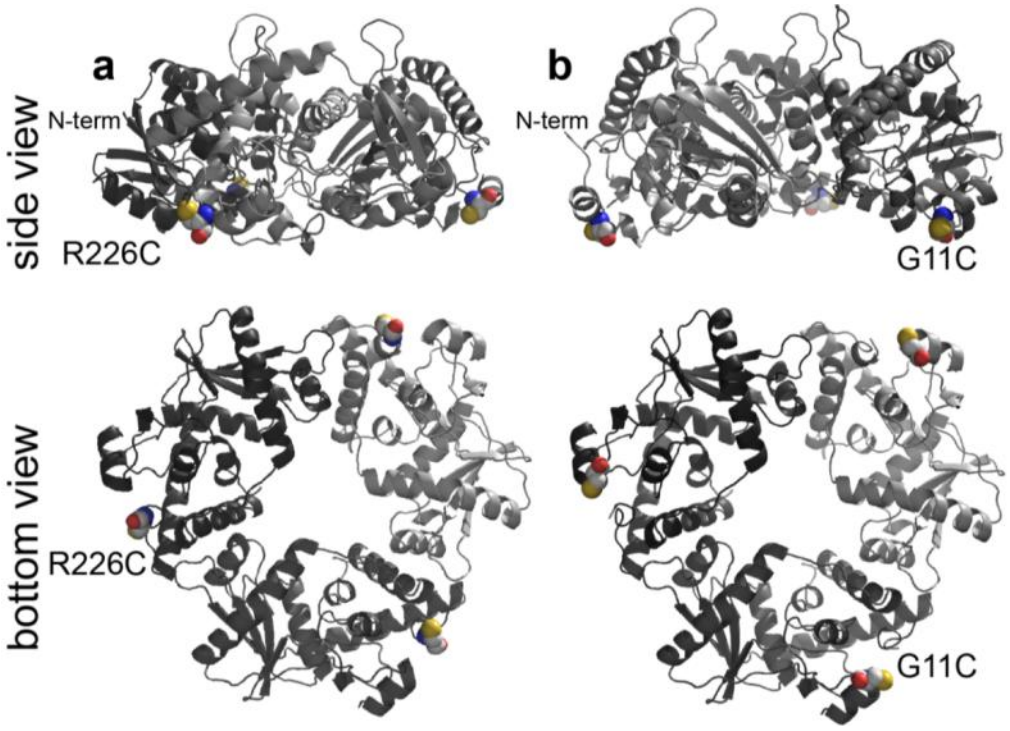



**Figure 3.14. Cysteine mutants of λ-exo.** To link λ-exo and αHL, cysteine point mutants of the C-terminal residue, R226 and a loop residue G11 were prepared. Side and bottom views of λ-exo showing the location of the newly introduced cysteines in place of (**a**) R226 and (**b**) G11. Upper panels show the side views and lower panels show the bottom-views of λ-exo.

Surface accessible residues on the bottom face of λ-exo (Fig. 3.2c) were scanned for genetic mutation to cysteine. A single cysteine mutant of the C-terminal residue, R226 were prepared in the WT λ-exo (R226C) and the C66S λ-exo backgrounds (R226C/C66S) (Fig. 3.14a). A cysteine was introduced at position 11, occupied by glycine in the WT protein, in the C66S λ-exo background (G11C/C66S) (Fig. 3.14b). The cysteine mutants were over-expressed in *E.coli*. G11C/C66S, R226C and R226C/C66S mutant proteins, all showed WT like exonuclease activity in PicoGreen based fluorescence assays (section 3.3.4) (Fig. 3.15).

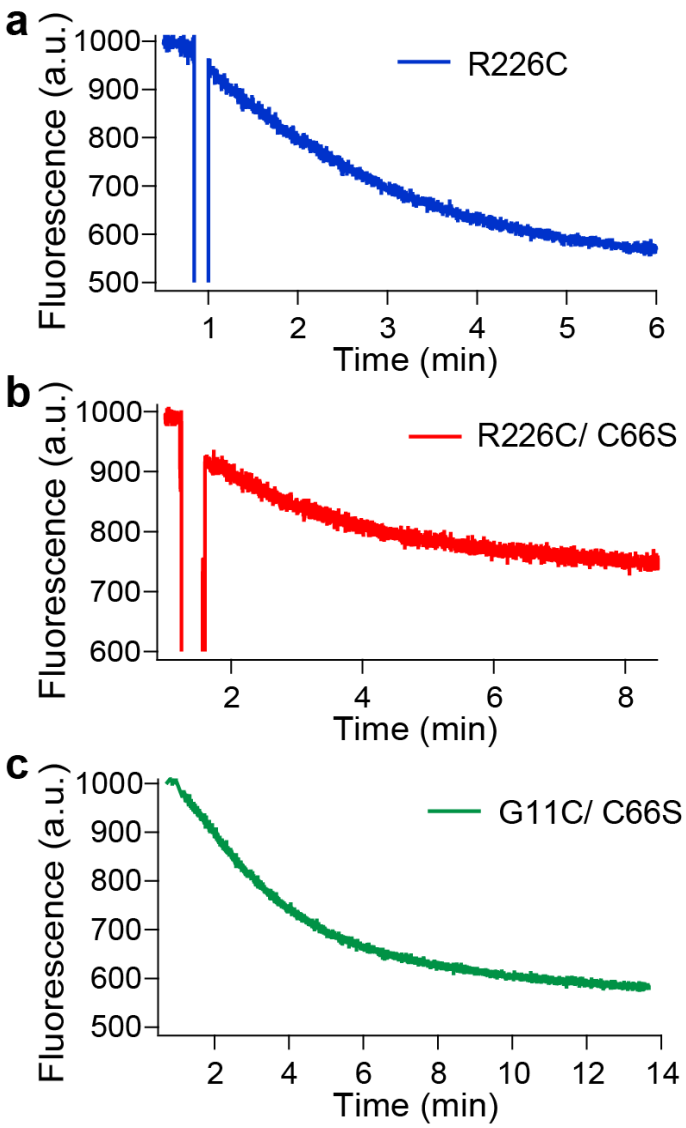



**Figure 3.15. Activity assays of the λ-exo cysteine mutants.** The cysteine mutants of λ-exo were assayed for activity. PicoGreen, a fluorescent dsDNA binding dye, was used for the activity assays. Exonuclease activity causes conversion of dsDNA to ssDNA, which can be observed as a decrease in fluorescence. Exonuclease activity assays of (**a**) R226C (70 nM), (**b**) R226C/C66S (16 nM), and (**c**) G11C/C66S λ-exo (22 nM). Values in paranthesis denote the final concentration of the protein monomers in the assay. All these mutants displayed exonuclease activity. The small gap in the fluorescence traces of (**a**) and (**b**) denote the point of protein addition.

To determine the surface accessibility and reactivity of the newly introduced cysteine residues, a 5 kDa poly-ethylene glycol derivative containing an ortho-pyridyl disulfide group (PEG-OPSS) was used (Movileanu *et al.*, 2003). The reaction of thiol groups with PEG-OPSS is irreversible under non-reducing

conditions. The 5 kDa PEG group attaches to all the reactive cysteines of the protein (Fig. 3.16a). The increase in the protein's $M_{app}$ on SDS-PAGE can thus be used to quantify the number of cysteines in the protein.

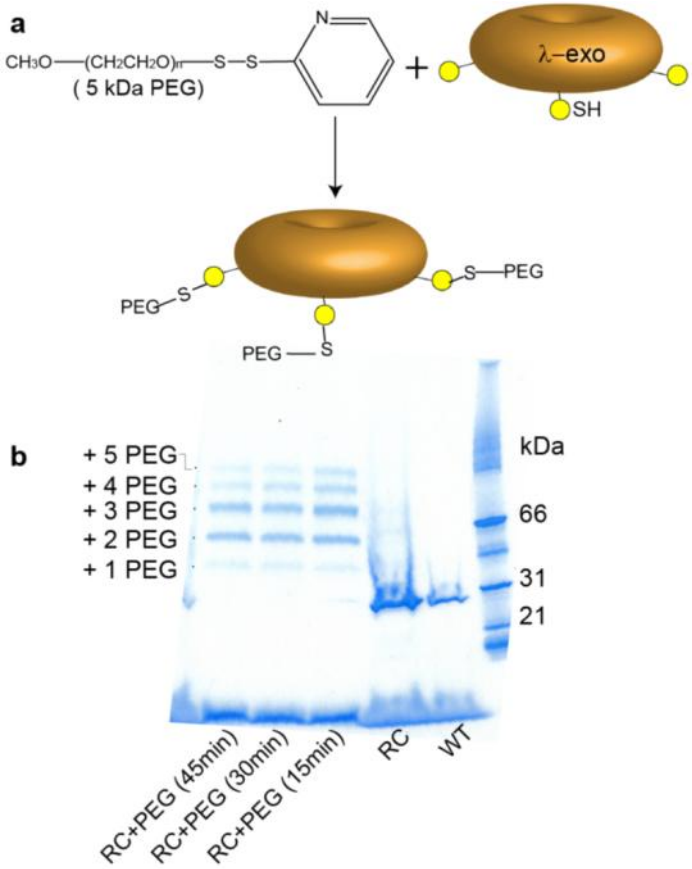



**Figure 3.16. Cysteine quantification using ortho-pyridyl disulfide (OPSS) poly-ethylene glycol derivative (PEG-OPSS)**. (**a**) PEG-OPSS is a highly reactive reagent used for quantification of the number of reduced cysteines in a protein. Upon reaction, the heavy PEG group attaches itself to all the available cysteines of the protein, which can be detected on SDS-PAGE. (**b**) SDS-PAGE of the reaction products of λ-exo R226C (RC) mutant with the 5 kDa PEG-OPSS reagent after different reaction times (15 min, 30 min and 45 min). λ-exo migrates as a monomer in denaturing SDS-PAGE (2% SDS in sample buffer) below the 31 kDa $M_w$ marker. After a 15 min reaction, up to 5 PEG molecules were seen attached to RC suggesting the presence of 5 reactive cysteines.

The λ-exo R226C protein was treated with monomethoxy poly(ethylene glycol)-*o*-pyridyl disulfide ($M_w$≈5 kDa) reagent. Surprisingly, five bands were observed above the λ-exo monomer band in SDS-PAGE suggesting that the R226C protein had five active cysteines per monomer (Fig. 3.16b). This implied that

alongwith the 226C residue, the four native cysteines of λ-exo were accessible to small molecules. Hence, small homo-bifunctional linkers could not be used to link λ-exo and αHL together. If αHL were first derivatized with a linker and λ-exo added subsequently, λ-exo would be linked the αHL cap in different orientations (Fig. 3.17). Possibly, covalent cross-linking could be combined with other methods (sections 3.6-3.8) to drive the λ-exo and αHL proteins in the correct orientation.

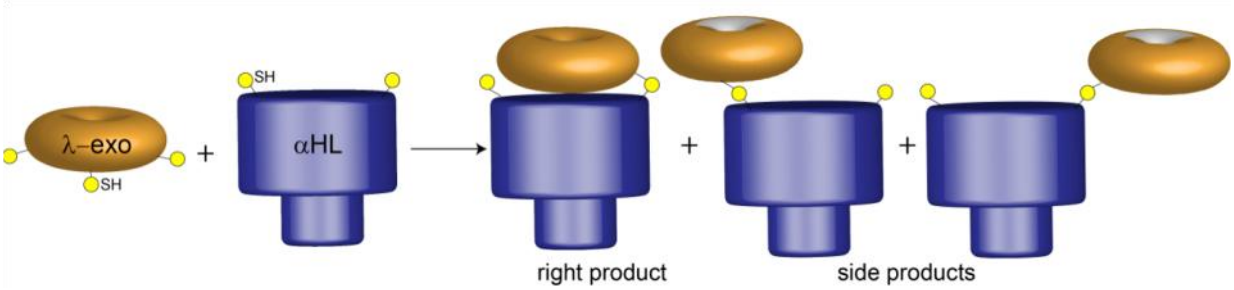



**Figure 3.17. Shortcomings of covalently linking αHL and λ-exo employing their cysteine mutants.** All the native cysteines of λ-exo were found to be reactive towards small molecules. Thus small cross-linker molecules could not be used. An alternate strategy could be first derivatizing αHL with a small linker and adding λ-exo subsequently. However, in this method several side products would be obtained due to cross-linking of λ-exo and αHL in different orientations. It would be very difficult to separate the product with the right orientation of λ-exo on top of αHL from the side products.

### 3.6. Strategy to latch λ-exo on αHL by non-covalent interactions

To place a λ-exo molecule on top of the αHL cap in the correct orientation non-covalent interactions could be employed. A small fragment of the αHL monomer was genetically fused to the λ-exo monomer so that it could bind to the remaining αHL monomer fragment (Fig. 3.18).

A truncated mutant of αHL mutant, termed αHLΔC5, was prepared by deleting the last five C-terminal residues, Glu289, Glu290, Thr291, Met292 and Asn293. The same five residues were genetically linked to the C-terminus of λ-exo,

named exo+C5 mutant. According to the crystal structure of αHL heptamer (α7; PDB ID 7AHL) these five residues form a β sheet structure and hydrogen bond with the rest of the αHL monomer. We surmised that the five-residue extension on the exo+C5 monomer could lodge within a αHLΔC5 monomer (Fig. 3.18). The strategy to link one molecule of λ-exo to αHL was as follows. If a hetero-heptamer were made containing just one αHLΔC5 subunit amongst six WT αHL monomer (α1) subunits, in principle, only one exo+C5 trimer molecule could be linked to the αHL cap. The conformational specificity of hydrogen bonds could ensure correct alignment of the two proteins. The next step of this strategy was to further strengthen the link between αHL and λ-exo by cysteine-coupled crosslinking.

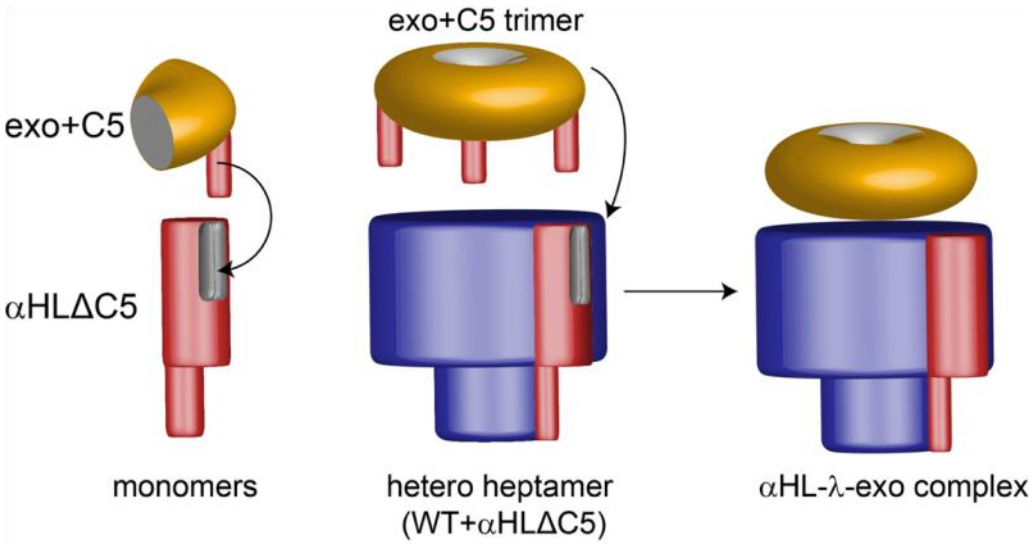



**Figure 3.18. Concept of λ-exo latching to αHL by non-covalent interactions.** A truncated mutant of αHL was prepared by deleting five residues from its C-terminus. The same five residues were introduced in the C-terminus of λ-exo (exo+C5). As the 5 C-terminal residues make favourable hydrogen bonds with the rest of the α1 structure, the monomers of exo+C5 (orange) and αHLΔC5 (red) should latch together when the proteins are in proximity. Hetero-heptamers with one αHLΔC5 subunit and six WT α1 units (blue) could be used in order to link just one exo+C5 trimer on the αHL cap.

Both the genes of exo+C5 and αHLΔC5 were made by homologous

recombination. Exo+C5 was expressed in *E.coli* and was found to retain full exonuclease activity (Fig. 3.19b). αHLΔC5 was expressed *in vitro* using the *E.coli* T7 S30 extract IVTT system (Promega). However, the αHLΔC5 monomer precipitated and did not show hemolytic activity (Fig. 3.19a). This is in agreement with previous studies of αHL C-terminal truncation mutagenesis (Sangha *et al.*, 1999; Walker *et al.*, 1992a). The instability of αHLΔC5 and other mutants of the aforementioned previous studies suggested that the C-terminal region is required for the correct folding of αHL. Hence, more deletion mutants of αHL were not made and an entirely different method employing electrostatics was adopted.

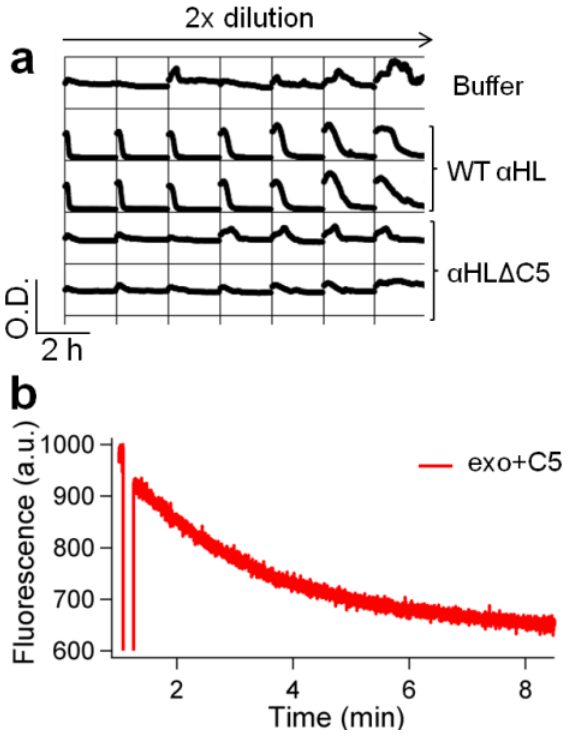



**Figure 3.19. Activity assays of the αHL mutant αHLΔC5 and λ-exo mutant exo+C5.** (**a**) Hemolytic activity assay showing that the inactivity of αHLΔC5 on rRBC. As judged by SDS-PAGE of WT αHL and αHLΔC5 (data not shown), the proteins were expressed in similar quantities *in vitro* using the S30 IVTT system. (**b**) Exo+C5, expressed in *E.coli* showed full exonuclease activity, as determined by the PicoGreen fluorescence assay (section 3.3.4). The final concentration of the protein in the assay was 15 nM.

### 3.7. Aligning λ-exo and αHL using electrostatic interactions

Introducing electrostatic interactions in the αHL-λ-exo interface could also be a plausible method to bring the proteins in the correct orientation, i. e., with aligned central pores. To make charged mutants of αHL and λ-exo, as a first step, their vacuum electrostatic maps were analyzed (Fig. 3.20a, d).

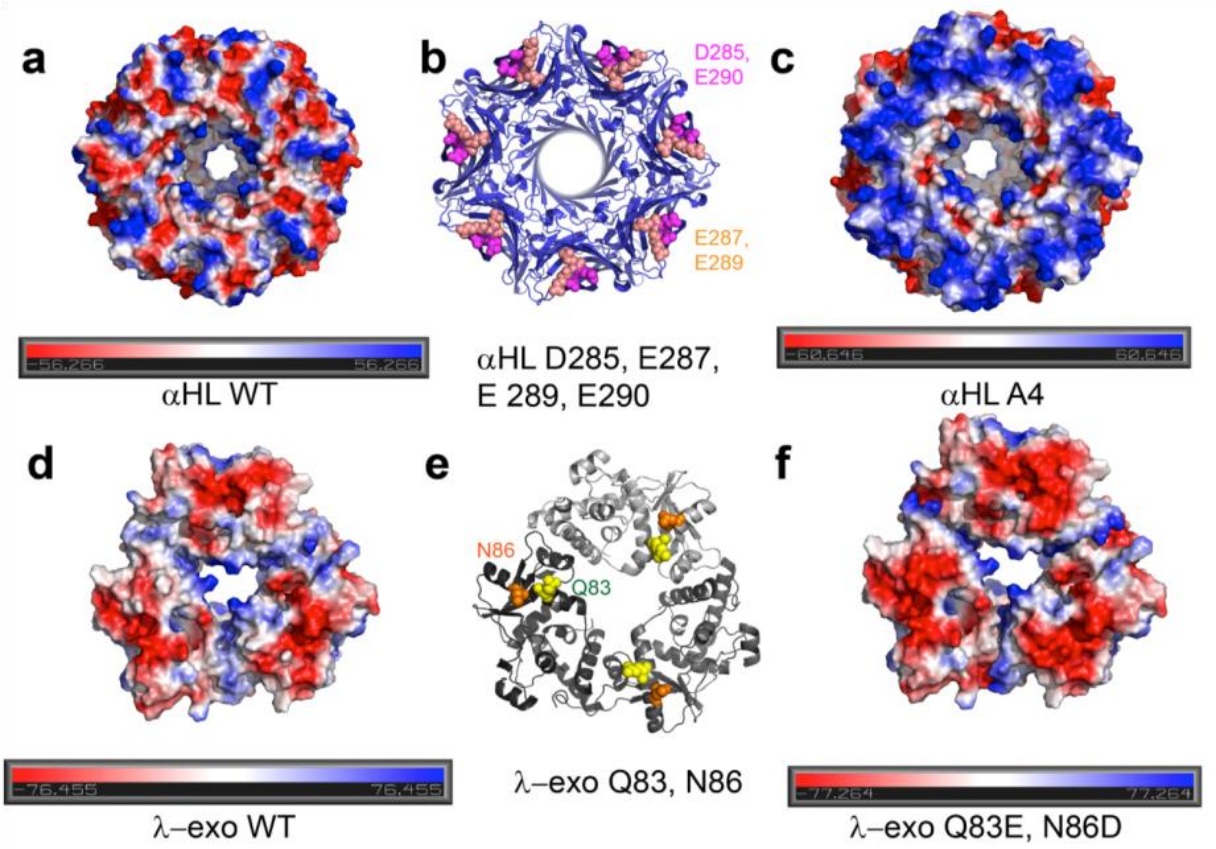



**Figure 3.20. Electrostatic maps of WT αHL and λ-exo and their charged mutants.** To introduce electrostatic attraction between the protein-protein interface of αHL and λ-exo, their charged mutants were made. Vacuum electrostatic maps of αHL and λ-exo were analyzed to determine the charges on the protein surfaces, where red denotes negative charge, blue denotes positive charge and white denotes neutral charge. (**a**) Electrostatic surface of WT αHL. (**b**) Four negatively charged cap residues (pink and wheat), D285 (pink), E287 (wheat), E289 (wheat) and E290 (pink) were mutated to alanine (αHL4A$^{+}$ mutant) to impart an additional +28 charges on the α7 cap surface. Figure shows the electrostatic surface of (**c**) αHL after the tetra-alanine mutation and (**d**) WT λ-exo. To impart a negative charge on the bottom surface of λ-exo, two neutral residues, Q83 (yellow) and N86 (orange), were modified to their negatively charged analogues, Glu and Asp, respectively. (**d**) Electrostatic map of λ-exo after the Q83E/N86D mutations.

The vacuum electrostatic surface of λ-exo shows a positively charged central channel and a negatively charged bottom surface that needs to be linked to αHL (Fig. 3.20d). The λ-exo bottom surface was made more negatively charged by changing two neutral residues, Q83 and N86, to their charged amino acids analogues, Glu and Asp, respectively (Fig. 3.20d-f, 3.21). The residues Q83 and N86 were chosen because of: (i) their surface accessibility; and (ii) according to the co-crystal structure of dsDNA and λ-exo (PDB ID 3SM4) these residues do not interact with the DNA (Fig. 3.9); and (iii) their side chains only minimally interact with the rest of the λ-exo structure (Fig. 3.21). The backbone carbonyl and amide atoms of Q83 and N86, respectively, hydrogen bond with a bridging water molecule (Fig. 3.21). The backbone amide of Q83 interacts with the backbone carbonyl oxygen atoms of A79 and W80. The backbone amide atoms of N86 associate with the side chain hydroxyl and backbone amide of T90. The side chain carbonyl of N86, a structural feature that would be present even after modification to Asp, interacts with the guanidino side chain of R89. Hence, modification of Q83 and N86 residues seemed a good choice with no serious consequences on the structural stability or activity of λ-exo.

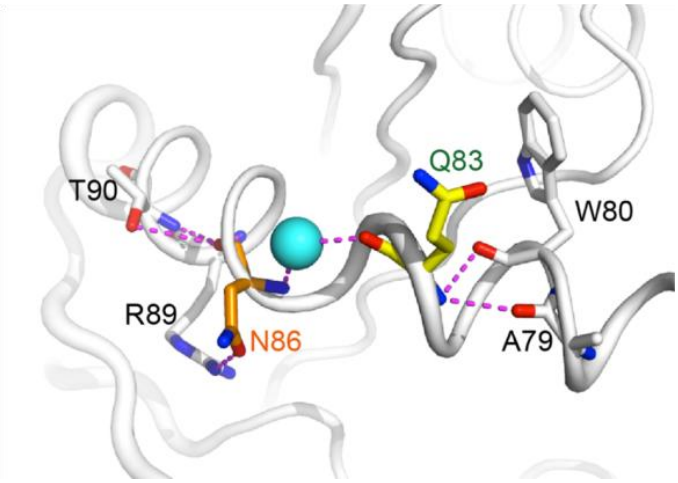



**Figure 3.21. Intramolecular interactions made by the λ-exo residues Q83 and N86.** Figure showing the intramolecular polar contacts (pink dashed lines)

made by two residues, Q83 (yellow) and N86 (orange) in the bottom face of λ-exo. Only the backbone atoms of Q83 and N86 are involved in hydrogen bonding with A79 and W80, and T90, respectively while bridging a water molecule (cyan). The side chain carbonyl of N86 makes hydrogen bonds with the side chain guanidinium nitrogen of R89. Nitrogen atoms are coloured blue and oxygen red. Figure was made using the λ-exo structure by Zhang *et al.* (PDB ID 3SM4) and rendered in PyMOL.

As compared to λ-exo, the electrostatic map of the αHL cap surface showed evenly distributed patches of negative and positive charges (Fig. 3.20a). To make the αHL cap complementarily charged to λ-exo, four negatively charged residues close to the αHL C-terminus, D285, E287, E289 and E290 were mutated to alanine and the mutant was termed αHL4A$^{+}$ (Fig. 3.20b, c). These residues were chosen on the basis of earlier mutagenesis work of Walker *et al.* where each charged residue of αHL was changed to cysteine and its effect on activity noted (Walker and Bayley, 1995). Walker *et al.* showed that D285, E287, E289 or E290 could be individually mutated to cysteine and modified with a bulky chemical group without loss in activity.

Like other αHL and λ-exo mutants, the tetra-mutant αHL4A$^{+}$ was expressed *in vitro* using the *E.coli* T7 S30 extract IVTT system (Promega) and the λ-exo Q83E/N86D mutant was expressed in *E.coli.* The protein monomer of αHL4A$^{+}$ was not stable, i.e., it precipitated during *in vitro* expression and did not possess hemolytic activity (Fig. 3.22a, b). αHL4A$^{+}$ could not form heptamers on rRBCm (Fig. 3.12c) or insert in planar lipid bilayers. Almost all the λ-exo Q83E/N86D mutant protein was found in the insoluble pellet of the cell lysate and exonuclease activity was not observed in the purified lysate supernatant (Fig. 3.22c).

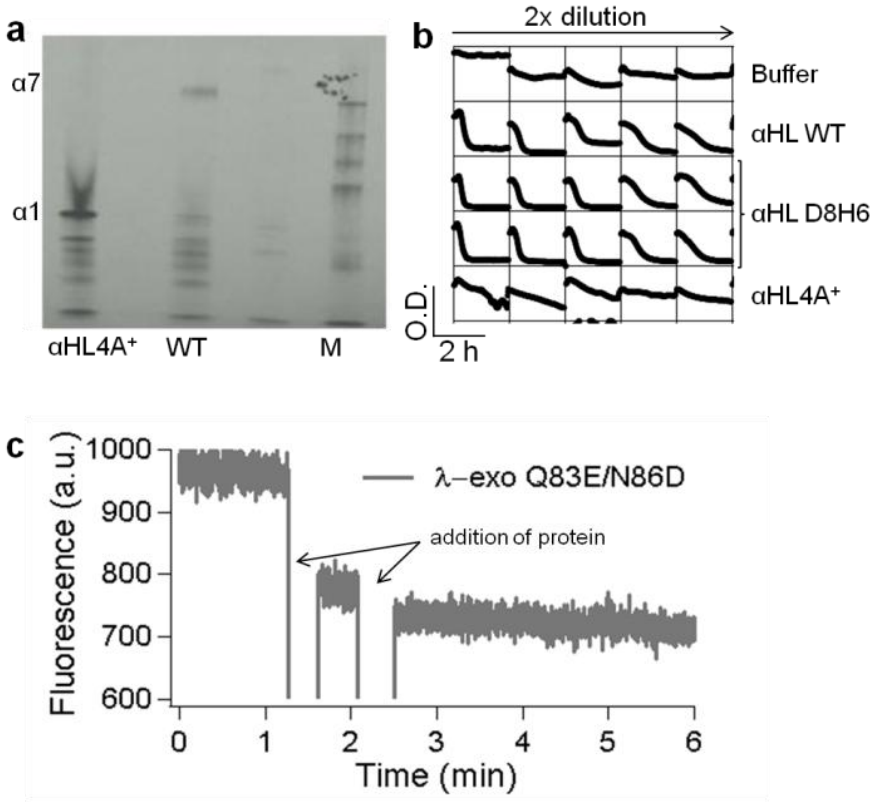



**Figure 3.22. Activity assays of the tetra-alanine mutant (αHL4A$^+$) of αHL and λ-exo Q83E/N86D mutant.** (**a**) The αHL4A$^+$ mutant precipitated while *in vitro* expression. Autoradiograph shows $^{35}$S methionine labeled proteins. (**b**) αHL4A$^+$ monomer also did not show hemolytic activity. (**c**) PicoGreen based fluorescence assay of the λ-exo Q83E/N86D mutant expressed in *E.coli* showed that the double mutations led to total loss in the exonuclease activity of λ-exo.

Electrostatics can be a powerful way of making protein complexes with high affinity constants (Buckle *et al.*, 1994; Worsdorfer *et al.*, 2011) (Chapter 1, section 1.6). However, electrostatics also plays a crucial role in protein folding (Dill, 1990) and the effect of charged mutations could be deleterious to the protein stability as demonstrated in the case of the αHL4A$^+$ mutant of αHL and the Q83E/N86D mutant of λ-exo. A more straightforward method of genetic fusion was adopted next to link αHL and λ-exo.

## 3.8. Genetic fusion of λ-exo and αHL

### 3.8.1. Genetic fusion constructs of λ-exo and αHL monomers

Genetic fusion is a routinely used method for linking two or more non-interacting proteins covalently (Mao *et al.*, 2004; Moll *et al.*, 2002; Moreau *et al.*, 2008; Sinclair *et al.*, 2011). The genes of the proteins of interest are put in tandem, usually connected by a linker sequence, which is then transcribed into a single mRNA encoding for the fusion protein. Genetically linking λ-exo and αHL had several advantages over other protein engineering methods. Genetic fusion did not require any mutagenesis screening of λ-exo and αHL. High concentrations of proteins were also not needed for covalent linkage, as required for disulfide cross-linking. The orientation of proteins could also be controlled by changing the length of the linker between λ-exo and αHL.

Genetic fusion constructs of λ-exo and αHL were prepared in several different combinations. In one set of constructs, termed as AE constructs, the C-terminus of αHL was attached to the N-terminus of λ-exo (Fig. 3.23b). The C-terminus of λ-exo was linked to the N-terminus of αHL in another set of constructs, which were termed EA (Fig. 3.23a).

To link the λ-exo and αHL termini, appropriately long linkers were needed. The distance between the λ-exo and αHL termini was determined by computational docking studies using PyMOL. The λ-exo trimer was mounted on the cap of αHL such that the pores of the two proteins aligned. The distance between the C-terminus of λ-exo and N-terminus of αHL (EA constructs) was determined to be 4.0 nm. This was measured in two steps, first from the αHL N-terminus to

the residue N17 near the pore opening of αHL (2.2 nm), and then from αHL N17 to the C-terminus of λ-exo (1.8 nm) (Fig. 3.23a). Similarly, the distance between the N-terminus of λ-exo and the C-terminus of αHL in the AE constructs was measured to be 3.7 nm (Fig. 3.23b).

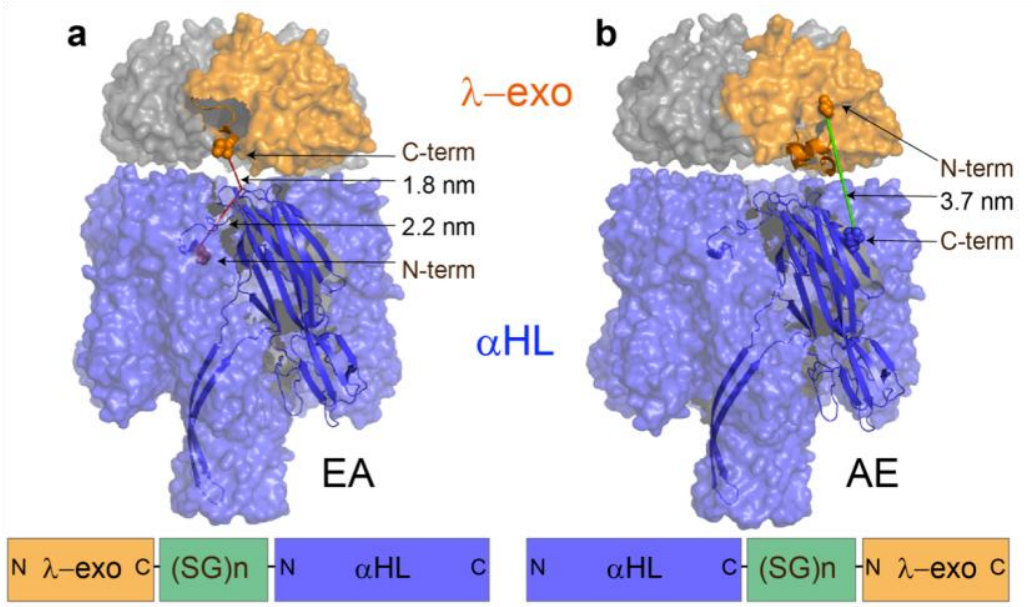



**Figure 3.23. Models of the genetic fusion constructs of αHL and λ-exo.** Initially two kinds of genetic fusion constructs of αHL and λ-exo were made. The N-terminus of αHL was linked to the C-terminus of λ-exo (EA construct). The order of genes was reversed in another set of constructs where the C-terminus of αHL was linked to the N-terminus of λ-exo (AE construct). To determine the length of linkers required between αHL and λ-exo, the distances between the N and C-termini of the proteins were measured in PyMOL. (**a**) Space filling models of the desired αHL–λ-exo complex showing that the distance between the N-terminus of αHL and the C-terminus of λ-exo measured in two steps of 2.2 nm and 1.8 nm (explained in the text). (**b**) Similarly, the distance between the C-terminus of αHL and the N-terminus of λ-exo for the AE construct was measured to be 3.7 nm. In both the constructs, the αHL and λ-exo proteins were connected by serine-glycine linkers (SG) composed of 'n' (n=10 and 13) amino acids.

Linkers that are rich in serine and glycine (SG linkers) are innocuously used for linking genetic fused proteins (Hammerstein *et al.*, 2011; Regan, 1999; Robinson and Sauer, 1998). The Ramachandran space for glycine is the largest amongst all amino acids (Voet and Voet, 2011) and the short hydrophilic side chain of serine also provides flexibility to loops (Fellouse *et al.*, 2005).

It was recently shown that the monomers of αHL (α2_10) could be genetically linked with a 10 amino acid long linker (SG10; ~3.3 nm) while retaining αHL activity (Hammerstein *et al.*, 2011). As the SG10 linker was found to be sufficient to link the N-terminus and C-terminus of adjacent αHL monomers, we surmised an SG10 linker with the same sequence as that used for α2_10 could also be used to also link the N- and C-termini of αHL and λ-exo, termed as EA10 (Fig. 3.23a) and AE10 constructs (Fig. 3.23b). Two more constructs were made with 13 amino acid long linkers (SG13; ~4.4 nm) and termed AE13 and EA13. However, it should be noted that a subsequent computational modeling suggested that the SG10 linker was not sufficiently long to link the N-terminus of αHL to the C-terminus of λ-exo, i.e. EA10 (Fig. 3.47).

### 3.8.2. Fusion construct of αHL dimer and λ-exo (EAA10)

Czajkowsky *et al.* proposed that a small fraction of αHL could exist as hexamers based on AFM and SDS-PAGE studies (Czajkowsky *et al.*, 1998). To determine whether αHL could oligomerize in stoichiometries other than a heptamer, α2_10 was used (Hammerstein *et al.*, 2011). Although α2_10 was able to form channels and lyse rRBC, one α1 subunit was excluded from forming the heptamer. It was also shown that hetero-heptamers with different ratios of α2_10 and α1 could be made.

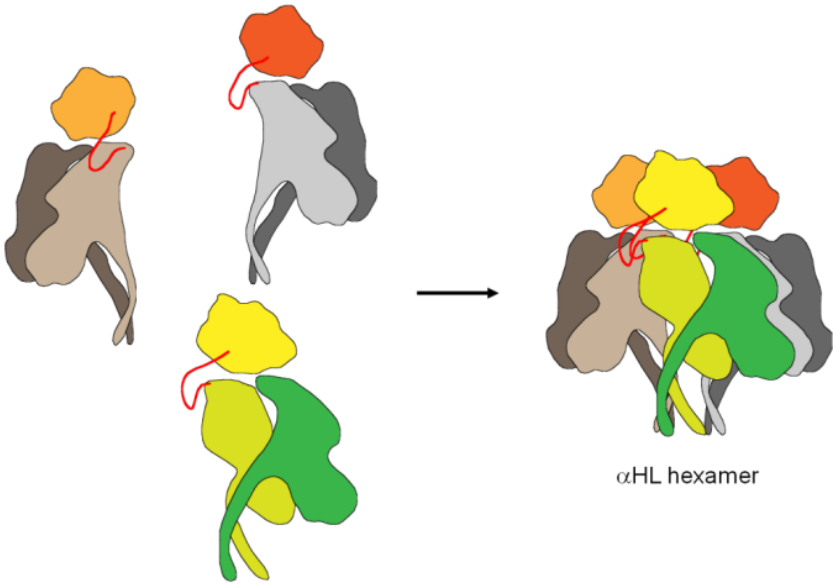



**Figure 3.24. Concept behind the EAA10 fusion construct of αHL and λ-exo.** λ-exo was linked to the N-terminus of a α2_10 with a SG10 linker (EAA10). The λ-exo part of EAA10 should form trimers, which would in turn lead to the assembly of the αHL parts as a hexamer. Adding a α1 subunit should then lead to the formation of α7.

If λ-exo monomer were to be linked to α2_10, the formation of the λ-exo trimer would lead to the assembly of the αHL dimers as hexamers (Fig. 3.24). If the αHL hexamer was not functional, contrary to what previous studies suggest, (Czajkowsky *et al.*, 1998), WT α1 could be added in order to form heptamers. Thus another set of λ-exo αHL fusion constructs (termed EAA10) was designed where the C-terminus of λ-exo was linked to the N-terminus of α2_10 with a SG10 linker (Fig. 3.25).

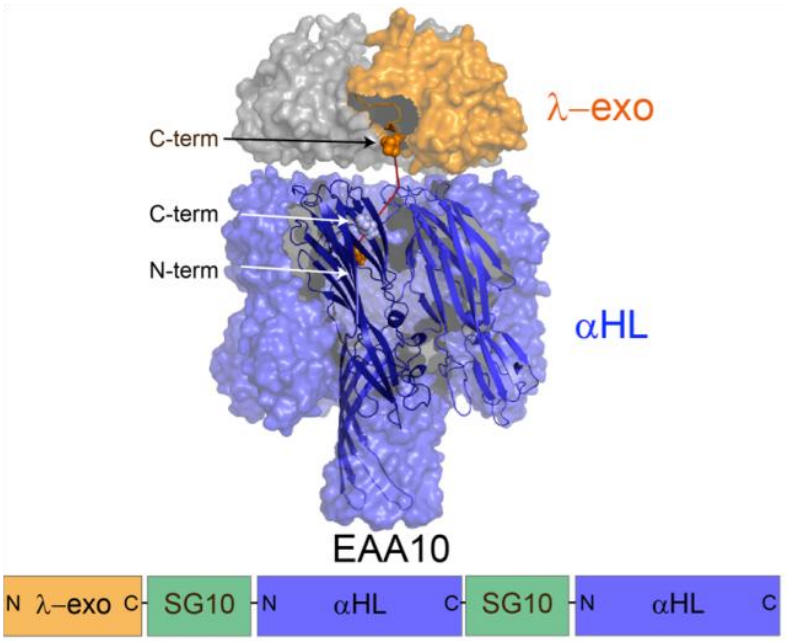



**Figure 3.25. Fusion construct of αHL dimer and λ-exo.** λ-exo was fused to a previously studied functional monomer dimer (α2_10; blue cartoons) of αHL (see Fig. 3.24 for the design principle of EAA10). The N-terminus of the α2_10 with a SG10 linker in between was joined to the C-terminus of λ-exo with another SG10 linker.

### 3.8.3. Expression, purification and characterization of the fusion constructs

One of the applications of the λ-exo-αHL pore complex is to function as a molecular device for DNA sequencing (section 3.4.2). Stoddart *et al.* have showed that DNA base recognition in an immobilized strand within the αHL pore is more pronounced when certain mutations are introduced in the αHL β barrel (Stoddart *et al.*, 2009; Stoddart *et al.*, 2010). On replacing the salt bridge forming pair of Glu111 and Lys147 with Asn, the constriction of the αHL pore becomes wider by 5 Å (Stoddart, Heron *et al.* 2009; Stoddart, Heron *et al.* 2010). It was found that although the conductance of the E111N/K147N mutant (NN αHL) pore remained similar to WT αHL, its residual conductance was higher than WT αHL when an immobilized strand of ssDNA occupied the pore. Thus NN αHL was capable of distinguishing between all four DNA bases in

immobilized homo- or hetero-polymeric DNA strands (Stoddart, Heron *et al.* 2009). Hence, NN αHL was used to make the AE, EA and EAA10 constructs. All the αHL-λ-exo fusion plasmid constructs were prepared by homologous recombination and expressed in different IVTT systems and in *E. coli.*

To optimize the yield of protein expression, two different IVTT systems were tried at different temperatures of 0°C, 25°C, 30°C and 37°C. One of the IVTT systems used was the previously described *E.coli* T7 S30 IVTT system (section 3.5) and another was the rabbit reticulocyte lysate translation (TNT quick-coupled) system (Promega). The TNT quick-coupled system contains purified rabbit reticulocytes and all the necessary components for protein synthesis such as amino acids, tRNA, ribosomes, etc.

The IVTT products of AE10, AE13, EA10, EA13 and NN αHL (control) were analyzed on SDS-PAGE. Majorly one protein band was observed in the case of AE10, AE13, EA10 and EA13 with $M_{app}$≈55,000, close to the expected molecular weight ($M_w$) of the fusion constructs' monomers of ~61 kDa (Fig. 3.26). A higher oligomer band was also observed above the 97 kDa $M_w$ marker in each case, predominantly in the lanes of NN αHL, AE10 and AE13 (Fig. 3.26). Like some barrel mutants of αHL (Cheley *et al.*, 1999; Cheley *et al.*, 1997), a fraction of NN αHL also pre-oligomerized, i.e., monomers oligomerized into heptamers even in the absence of any oligomerization agent (Fig. 3.26a). Thus the $M_{app}$≈180,000 band in the NN αHL lane was attributed to the NN α7. By analogy, the $M_{app}$≈230,000 band seen with the AE10 and AE13 monomers were attributed to their heptamers formed presumably by their NN αHL parts.

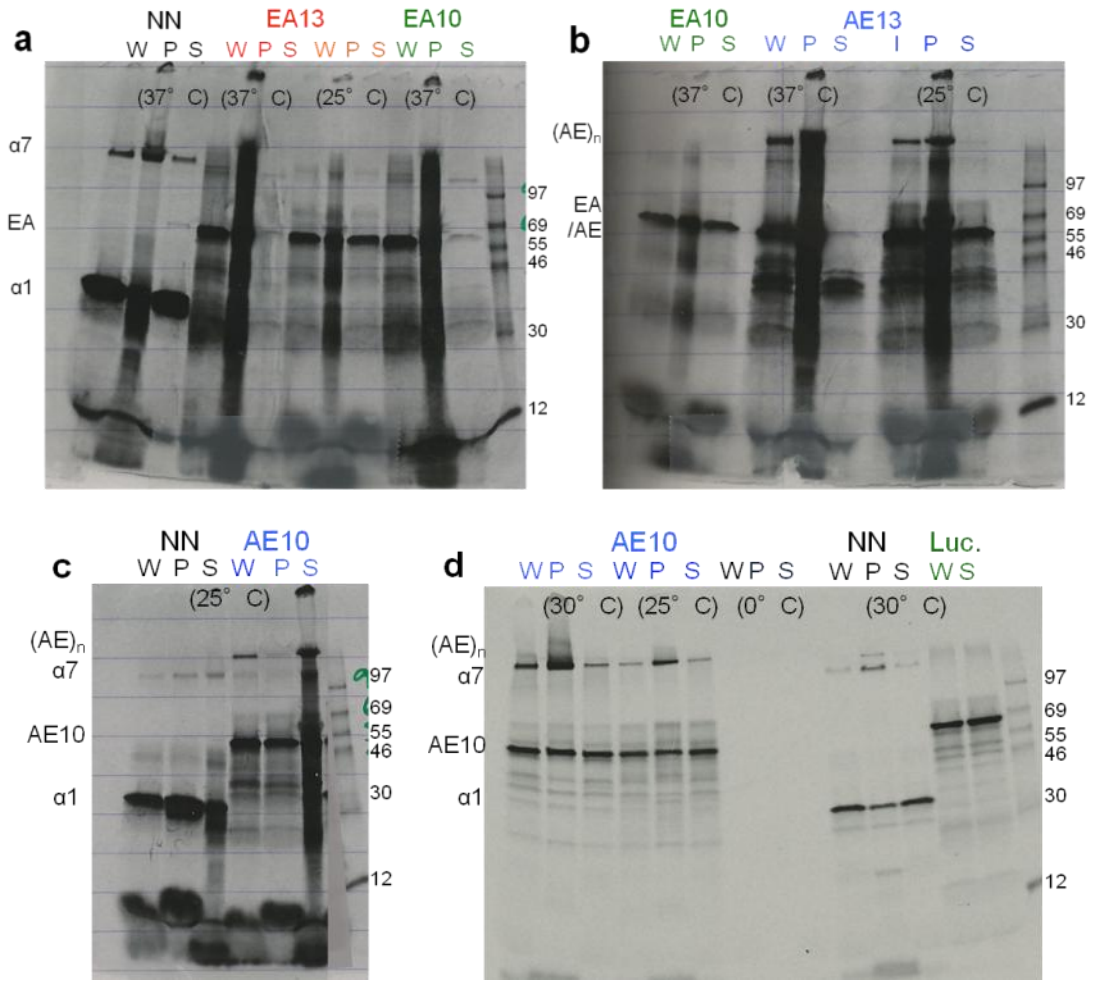



**Figure 3.26. *In vitro* expression of the αHL-λ-exo fusion constructs under different conditions.** αHL-λ-exo fusion constructs were expressed and labeled with $^{35}$S methionine *in vitro* using the S30 extract IVTT system at 25°C and 37°C. AE10 was also expressed using the TNT quick coupled IVTT system at 0°C, 25°C and 30°C. Figures show the autoradiographs of SDS-PAGE of the IVTT reaction products: whole IVTT (W), supernatant (S) and pellet (P). IVTT products of (**a**-**c**) NN αHL (NN), EA13, EA10, AE13 and AE10 at 25°C and 37°C (T7 S30 extract system). (**d**) Expression of AE10 using the TNT quick coupled system at 30°C, 25°C and 0°C. NN and luciferase (Luc.) were expressed as controls at 30°C. AE10 was not expressed at 0°C during the incubation time of 1 h. The relative amount of soluble AE10 monomer (S) as compared to insoluble monomer (P) was more in the TNT quick coupled system than the S30 IVTT at 25°C. A higher band of AE10 and AE13 oligomers ($(AE)_n$; $M_{app}$≈230,000) other than the monomer band ($M_{app}$≈55,000) was observed during expression in both IVTT systems and was attributed to homo-heptamers of AE10 and AE13 (section 3.8.8).

The yield of soluble protein was assessed by its amount in the IVTT supernatant. With the exception of NN αHL, the yield of soluble proteins expressed using the *E.coli* T7 S30 extract IVTT system was greater at 25°C

than at 37°C. The yield of soluble monomer and oligomer of AE10 was greater when the TNT quick-coupled system was used in comparison to the S30 IVTT system (Fig. 3.26d). At 0°C, AE10 proteins were not expressed at all (TNT quick-coupled system). The yields of soluble AE10 were comparable when the TNT quick-coupled system was used at 25°C or 30°C (Fig. 3.26d).

### 3.8.4. Hemolytic activity assays of the αHL-λ-exo fusion constructs

The soluble fractions of all the IVTT expressed fusion constructs were assayed for hemolytic activity (Fig. 3.27). Both AE10 and AE13 showed hemolytic activity. However, EA10, EA13 and EAA10 constructs were not hemolytically active. In EA10, EA13 and EAA10, the SG linkers held by the αHL N-terminus would be crowded within its vestibule and would not be long enough to accommodate seven λ-exo monomers on the αHL cap. This could explain the lack of hemolytic activity in these constructs. As the linkers were on the C-terminus of αHL in the AE constructs (AE10 and AE13), steric crowding of seven λ-exo monomers was prevented and these constructs possessed WT-like hemolytic activity.

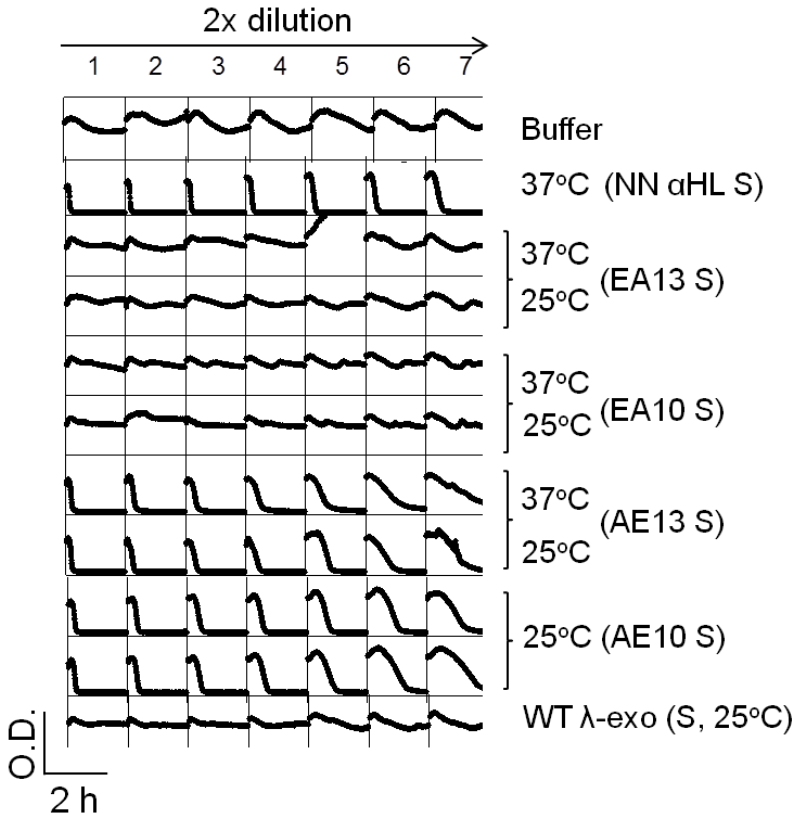



**Figure 3.27. Hemolytic activity assays of the αHL-λ-exo fusion constructs.** Hemolytic activity of the soluble fractions (S) of αHL-λ-exo fusion constructs expressed with the *E.coli* T7 S30 extract IVTT system were tested on rRBC. AE10 and AE13 showed hemolytic activity similar to NN αHL whereas EA10 and EA13 failed to lyse the rRBC. The supernatant fraction of WT λ-exo was used as a negative control. The lysis of rRBC was monitored as a decrease in light scattering at 595 nm.

### 3.8.5. Exonuclease activity assays of the αHL-λ-exo fusion constructs

#### 3.8.5.1. Exonuclease activity assays of AE10

As a control for the αHL-λ-exo fusion constructs, WT λ-exo expressed in IVTT (T7 S30 extract) was assayed for its exonuclease activity. The soluble fraction of the IVTT product of WT λ-exo was purified by immobilized metal affinity chromatography (IMAC) using Co-TALON magnetic beads and subsequently desalted for activity assays (Fig. 3.28a). In the PicoGreen fluorescence assays (section 3.3.4), exonuclease activity WT λ-exo was not detected. This result

was attributed to the low concentration (~10 nM) of purified IVTT WT λ-exo. WT λ-exo protein, which was expressed in much higher concentration (~6 µM) in *E.coli* demonstrated exonuclease activity (Fig. 3.28b).

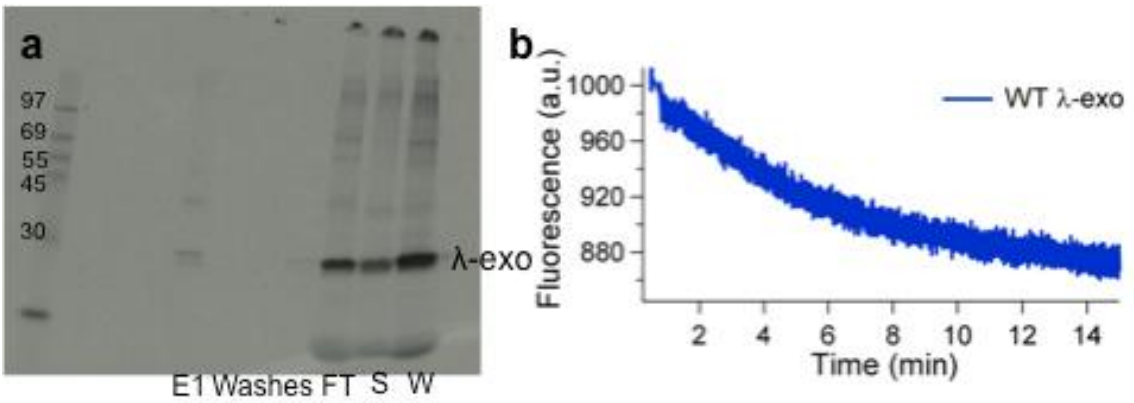



**Figure 3.28. Purification of WT λ-exo expressed in IVTT and activity of *E.coli* expressed protein.** (**a**) SDS-PAGE showing purification of WT λ-exo expressed and labelled with $^{35}$S methionine in IVTT using IMAC. The supernatant (S) of the whole IVTT product (W) was loaded on Co TALON magnetic beads. Most WT λ-exo protein was found in the flow through (FT) and <5% could be recovered in the elution (E1). The desalted eluate did not show exonuclease activity. (**b**) WT λ-exo expressed in much higher concentrations in *E.coli* (~6 µM) displayed exonuclease activity in fluorescence assays (section 3.3.4).

Hence, to assay their exonuclease activity, the AE, EA and EAA10 constructs were expressed in *E. coli*. To optimize the protein yield, AE10 was expressed in two different *E. coli* strains, Rosetta (DE3) pLysS and NiCo21 (DE3) and grown under two different temperature protocols. In one case, the expression cultures were grown at 20°C for 24 h (termed 20°C) and in the other case cultures were grown at 30°C for 8 h and then further grown at 20°C for 16 h (termed 30/20°C) (Methods). Soluble expression of AE10 was achieved under all four growth conditions (Fig. 3.29).

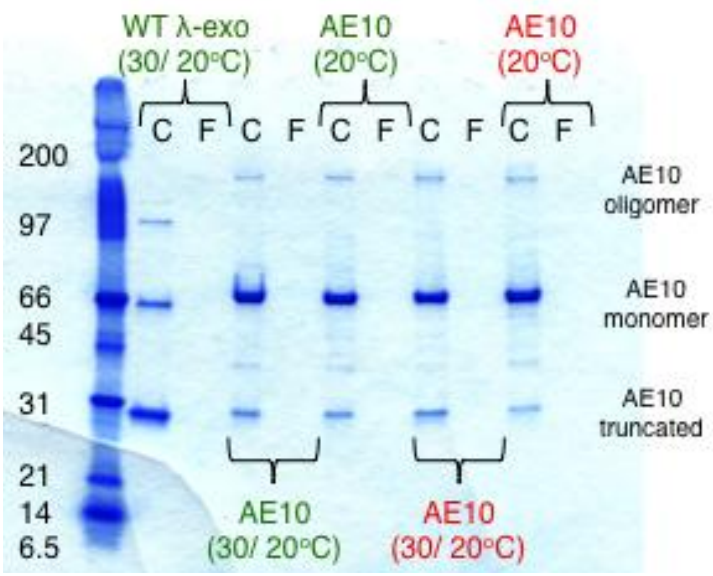



**Figure 3.29. Expression of AE10 under different conditions.** AE10 was expressed in the Rosetta (DE3) pLysS (green) and NiCo21 (DE3) (red) *E.coli* strains under two different growth temperature protocols (30/20°C and 20°C, Methods). The proteins were purified on Co TALON columns and concentrated using spin filters with 10 kDa cut off (Millipore) (Methods). SDS-PAGE comparing the yields of the concentrated protein expressed under different conditions. C and F denote the concentrated (C) and the flow-through (F) fractions. Apart from the monomer band of AE10 ($M_{app}$≈67,000; expected $M_w$ 61 kDa), a smaller truncated protein band ($M_{app}$≈29,000), which migrated above WT λ-exo ($M_{app}$≈26,000) (labeled as AE10 truncated) was also observed (see section 3.8.7 for discussion).

Activity assays of AE10 protein fractions, expressed under different conditions, showed that the exonuclease activity of the λ-exo components of AE10 was preserved (Fig. 3.30).

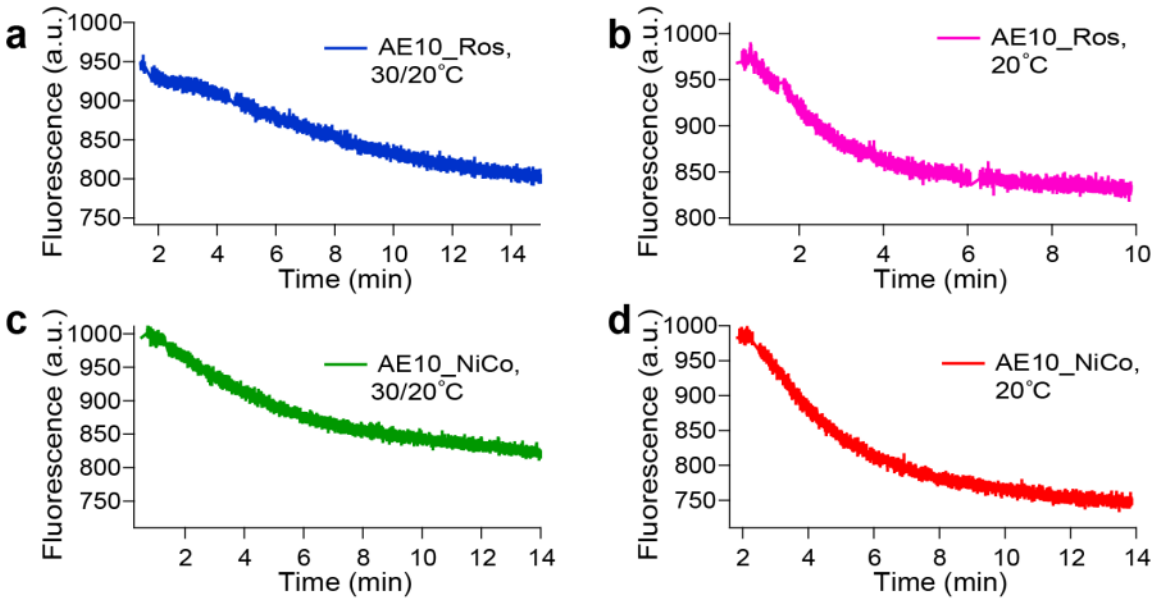



**Figure 3.30. Exonuclease activity assays of AE10 grown under different conditions.** PicoGreen based fluorescence assays of AE10 expressed in (**a**,**b**) Rosetta pLysS DE3 cells grown (**a**) under a dual temperature protocol, 30/20°C and (**b**) at 20°C. (**c**, **d**) NiCo21 cells grown (**c**) under a dual temperature protocol (30/20°C) and (**d**) at 20°C. The final concentration of the proteins in the assay were (**a**) 13 nM, (**b**) 17 nM, (**c**) 15 nM and (**d**) 20 nM. All the assays were done at 25°C, with $\lambda_{ex}$=485 nm and $\lambda_{em}$=520 nm (Methods).

### 3.8.5.2. Activity assays of EA10, EA13, AE13 and EAA10

The AE13, EA10, EA13 and EAA10 constructs were expressed under the same conditions as AE10, employing NiCo21 (DE3) cells and the 30/20°C temperature protocol. All the constructs, AE13, EA10, EA13 and EAA10, showed exonuclease activity (Fig. 3.31a-d). In hemolytic activity assays, the *E.coli* expressed proteins gave the same results as the IVTT proteins. AE13 was found to be hemolytically active like AE10 whereas, EA10, EA13 and EAA10 were not hemolytically active (Fig. 3.31e).

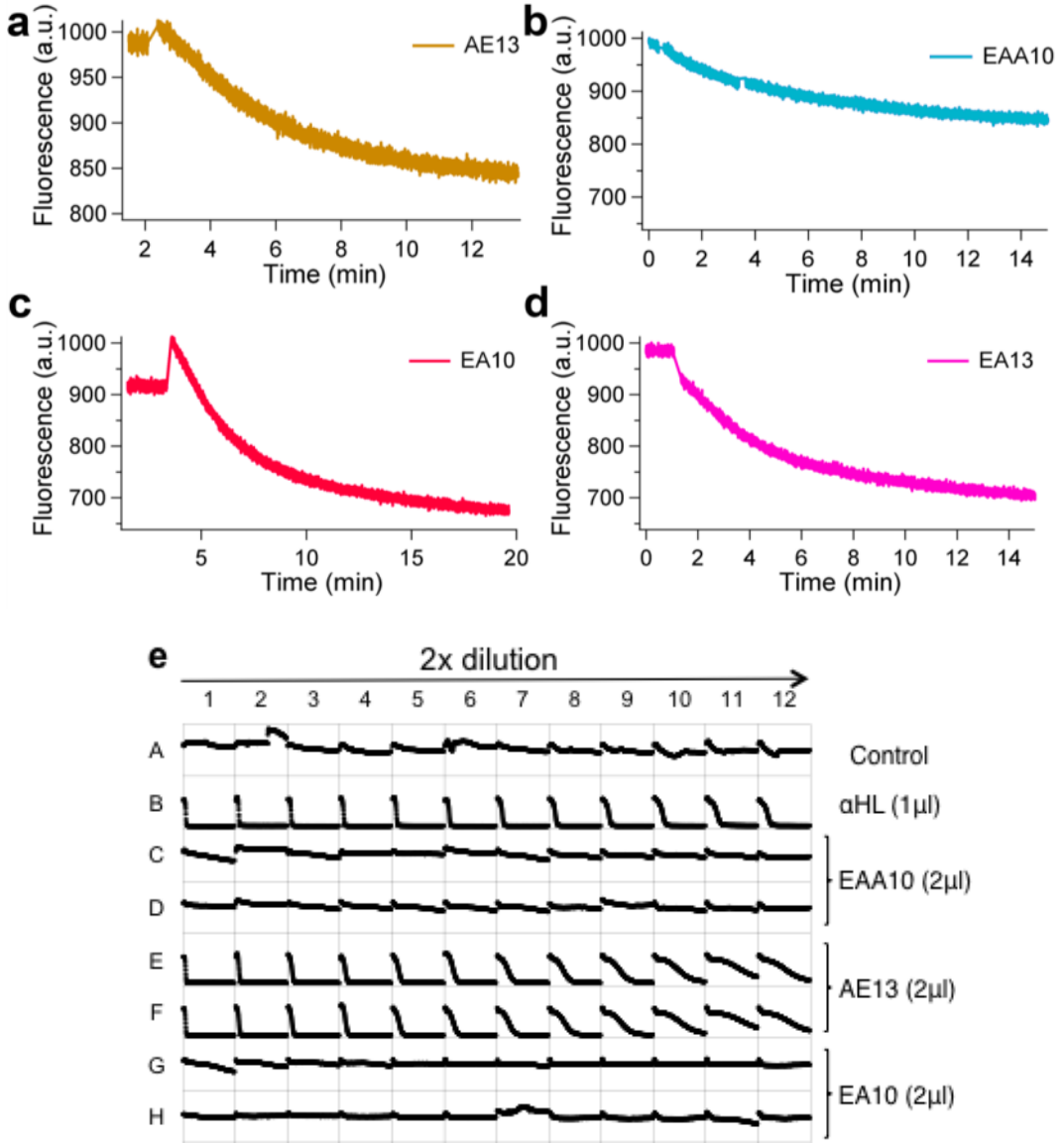



**Figure 3.31. Activity assays of αHL-λ-exo fusion constructs expressed in *E.coli*.** Fluorescence based activity assays of (**a**) AE13, (**b**) EAA10, (**c**) EA10, and (**d**) EA13. All the αHL-λ-exo fusion constructs showed exonuclease activity like WT λ-exo (also Fig. 3.30). The final concentration of the protein monomers in the assay were (**a**) 14 nM, (**b**) 10 nM, (**c**) 45 nM and (**d**) 16 nM. (**e**) Hemolytic activity assay of the *E.coli* expressed EAA10, AE13 and EA10 proteins. Only AE13 was hemolytically active. Volumes in parantheses denote the amount of protein added to the first well of the microtiter plate.

### 3.8.6. Full length AE10 protein shows hemolytic and exonuclease activities

AE10 showed hemolytic activity comparable to WT α1 (Fig. 3.32b). α1 was not observed in the IMAC and SEC purified AE10 fractions by staining the gels with

Coomassie stain or silver stain, which has a much higher sensitivity (~1 ng) (Rabilloud *et al.*, 1994) (data not shown). SDS-PAGE showed IMAC purified AE10 proteins contained a smaller protein, which migrated just below the 31 kDa $M_w$ marker with $M_{app}$≈29,000 (Fig. 3.29). The $M_{app}$≈29,000 protein in AE10 fractions migrated in between the bands of WT α1 ($M_{app}$≈34,000) and WT λ-exo ($M_{app}$≈24,000) in SDS-PAGE (Fig. 3.32a) (section 3.8.7).

EA13, which did not possess a small protein band unlike AE10, did not show hemolytic activity (Fig. 3.32a, b). Thus, to assess whether the activity of AE10 was due to the small protein ($M_{app}$≈29,000), the activity and concentrations of WT λ-exo, AE10 and the small protein seen in the case of AE10 were compared. The concentrations of the AE10 monomer ($M_{app}$≈67,000) and the small protein ($M_{app}$≈29,000) were estimated to be 7 and 2 μM, respectively based on the band intensities in SDS-PAGE. The concentration of the small protein ($M_{app}$≈29,000) was much lower than that of the WT proteins, WT α1: 60 μM and WT λ-exo: 6 μM.Based on the fluorescence decay slopes of the dsDNA digestion in exonuclease activity assays, it was concluded that similar quantities of AE10 and WT λ-exo showed equivalent exonuclease activities (Fig. 3.32c). The different concentrations of the small protein ($M_{app}$≈29,000) as compared to the WT proteins could not account for similar activities of AE10 and WT λ-exo and WT αHL. These experiments indicated that the full length AE10 protein ($M_{app}$≈67,000) showed both exonuclease and hemolytic activities.

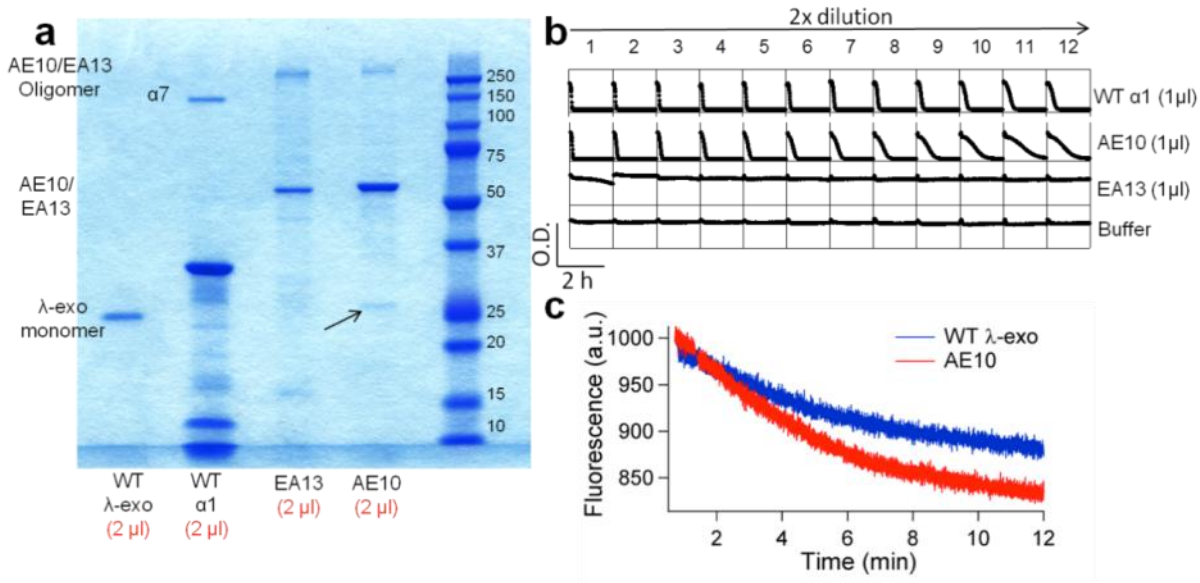



**Figure 3.32. Comparison of the hemolytic and exonuclease activities of AE10 and the WT α1 and WT λ-exo proteins.** (**a)** SDS-PAGE of the purified and concentrated fractions of AE10 and EA13, WT α1 and WT λ-exo. In the AE10 fraction, a very light band of a small protein (black arrow; $M_{app}$≈29,000) was observed between the protein bands of WT α1 and WT λ-exo. (**b**) AE10 showed hemolytic activity, however EA13 did not. EA13 did not contain a smaller protein band ($M_{app}$≈29,000) unlike AE10. Thus, to ensure that the exonuclease and hemolytic activity of AE10 was not due to the small protein band ($M_{app}$≈29,000), the concentrations of full length AE10, small protein in AE10 ($M_{app}$≈29,000), WT α1 and WT λ-exo were estimated from SDS-PAGE. Compared to the intensity of the truncated AE10 band ($M_{app}$≈29,000), the WT λ-exo band was 3 times more intense and the WT α1 band was 10 times more intense. Volumes in parentheses denote the amount of protein loaded on SDS-PAGE and used in column 1 of the microtiter plate for the hemolytic assay. (**c**) Fluorescence traces depicting comparable activities of 2 µl of AE10 (total 10 µM) and WT λ-exo (6 µM).

### 3.8.7. AE10 gets truncated into its constituent αHL and λ-exo parts

During the expression of genetically fused proteins, truncated proteins could also form (Hammerstein *et al.*, 2011; Nicke *et al.*, 2003). Nicke *et al.* reported the formation of monomer and dimer units of their concatenated trimer construct of a ligand gated ion channel, P2X1 (Nicke *et al.*, 2003). Incomplete translation of the αHL-λ-exo fusion constructs genes could give rise to a smaller protein consisting of a part of the first protein, i.e., NN αHL in the case of AE10 and AE13. Limited proteolysis at the connecting SG linker between NN αHL and λ-

exo in the fusion constructs would lead to the formation of the truncated products consisting of the NN αHL and λ-exo parts. Only the λ-exo part in the fusion constructs contains a His-6 tag. Hence, if AE10 were fragmented into its αHL and λ-exo parts before purification, only the λ-exo segment would remain after IMAC purification. Hence, the small protein ($M_{app}$≈29,000) observed in SDS-PAGE could be a truncated form of AE10 consisting of λ-exo (Fig. 3.32a).

To test this possibility, a western blot of AE10 using an anti-His antibody was performed. AE10 protein, which had been purified by IMAC and SEC (termed AE10f10) was run on a 4–15% gradient gel and then transferred onto a nitrocellulose membrane. The membrane was treated with an anti-His antibody conjugated with alkaline phosphatase. On adding a phosphorylated substrate, the His-6 tag containing protein bands turned violet. The bands of AE10 monomer and the truncated product were stained (Fig.3.33). A WT α1 sample treated with sodium deoxycholate (DOC) to yield α7 and run as a negative control on the same gel was not stained in the western blot.

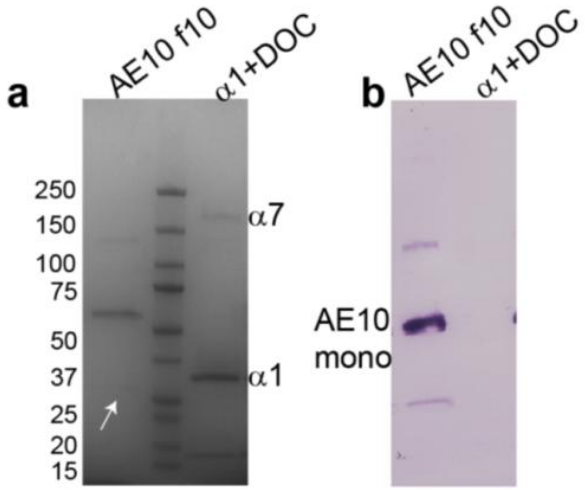



**Figure 3.33. Western blot of AE10.** To ascertain the nature of small protein band ($M_{app}$≈29,000) that was observed with the full length AE10 monomer purified using IMAC and SEC (termed AE10f10) ($M_{app}$≈67,000), a western blot

was performed. (**a**) SDS-PAGE showing a Coomassie blue stained bands of AE10 where the small protein band ($M_{app}$≈29,000) is just discernible (white arrow). A WT α1 sample treated with DOC (WT α1+DOC) to yield WT α7 was also loaded on the same gel. (**b**) Western blot of the same protein samples blotted using an anti-His antibody. The His-antibody failed to bind to the αHL species and WT α1+DOC is seen as an empty lane. On the other hand, the anti-His antibody bound to the His-6 tag on the λ-exo part of AE10 and also to the small protein ($M_{app}$≈29,000). As the small protein ($M_{app}$≈29,000) ran above the level of WT λ-exo and below WT α1 in SDS-PAGE, we concluded that it contained the λ-exo part with the His-6 tag and perhaps some part of the preceding αHL protein of AE10.

This implied that the truncated protein was composed of λ-exo with the His-6 tag and possibly a part of the preceding αHL protein as it migrated between the levels of WT λ-exo ($M_{app}$≈24,000) and WT α1 ($M_{app}$≈34,000) in SDS-PAGE (Fig. 3.32a). The truncated AE10 protein could be contributing to the observed exonuclease activity of AE10 (Fig. 3.32c) (section 3.8.6). Like AE10, a smaller protein ($M_{app}$≈26,000) was also observed in the AE13 protein fractions (Fig. 3.34) and was attributed to a truncated AE13 protein consisting of λ-exo.

### 3.8.8. Spontaneously formed oligomers of αHL-λ-exo fusion constructs

Some αHL barrel mutants spontaneously assemble into α7 even in the absence of lipid membranes (Cheley *et al.*, 1999; Cheley *et al.*, 1997). The NN mutant of αHL, used as the background of the αHL part of the constructs also has a propensity to oligomerize into α7 even in the absence of membranes (Fig. 3.26a). A higher $M_w$ band was observed in all the αHL-λ-exo fusion protein fractions in SDS-PAGE, which migrated above the WT α7 ($M_{app}$≈200,000) band (Fig. 3.32, 3.34). Based on the $R_f$ values in SDS-PAGE, the $M_{app}$ of higher $M_w$ bands of the αHL-λ-exo fusion proteins were estimated to be: AE10 ($M_{app}$≈330,000), AE13 ($M_{app}$≈390,000), EA10 ($M_{app}$≈400,000), EA13

($M_{app}$≈300,000), and EAA10 ($M_{app}$≈320,000). If the high $M_w$ bands were composed of the αHL-λ-exo fusion constructs, these bands could be attributed to the heptamers of AE10, AE13, EA10 and EA13 ($M_w$ of monomer=61 kDa, expected $M_w$ of heptamer=427 kDa), and a trimer of EAA10 ($M_w$ of monomer=94 kDa, expected $M_w$ of trimer=282 kDa). The αHL parts of AE10, AE13, EA10, and EA13 could be involved in heptamer formation, which is stable in SDS up to 65°C (Cheley *et al.*, 1999). Under the conditions of SDS-PAGE, λ-exo is denatured and the inter-subunit interactions of λ-exo are broken (section 3.8.11). Hence, the EAA10 trimers should also be associated with their αHL parts (Czajkowsky *et al.*, 1998). To test this possibility, the purified fractions of AE13, EAA10 and EA10 were heated at 95°C for 5 min and subsequently analyzed on SDS-PAGE (Fig. 3.34). The high $M_w$ bands ($M_{app}$≈300,000-400,000) dissociated into the fusion protein monomer bands in each case implying that they were formed of non-covalently bound fusion protein monomers.

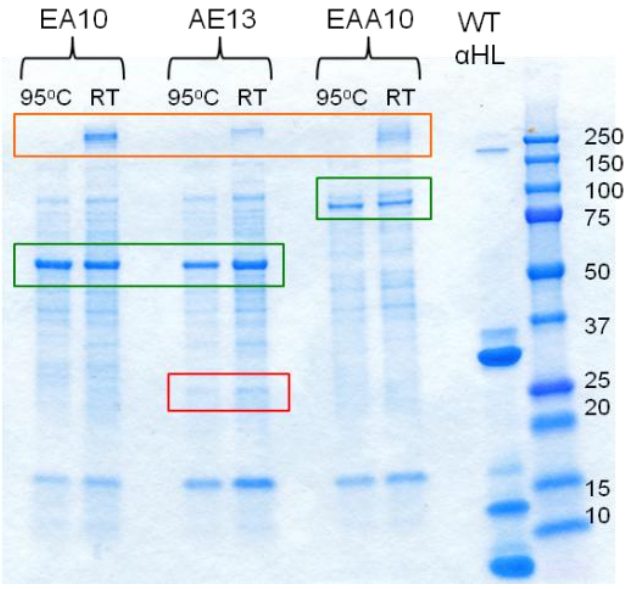



**Figure 3.34. Heating analysis of AE13, EA10 and EAA10 oligomers.** SDS-PAGE showed higher oligomer bands (orange box) of EA10 ($M_{app}$≈400,000), AE13 ($M_{app}$≈390,000) and, EAA10 ($M_{app}$≈320,000) above the level of WT α7.

These higher bands dissociated into the monomer bands (green boxes) upon heating at 95°C for 5 min. These bands were attributed to the heptamers of EA10 and AE13 and the trimer of EAA10 based on their $M_{app}$. The dimer bands of EA10 and AE13 ($M_{app}$≈108,000 and 104,000) were also observed formed presumably because of cross-linking between the native cysteines of their λ-exo components. Like AE10, a faint band ($M_{app}$≈26,000) was observed with the AE13 protein (red box).

### 3.8.9. Single channel electrical recordings

AE10, AE13, EA10, EA13 and EAA10 formed oligomers in solution that migrated as putative heptamers in SDS-PAGE. To assess whether the fusion constructs could form oligomeric pores upon encountering planar lipid bilayers, single channel electrical measurements were performed (Chapter 2, section 2.10). λ-exo is active in only low concentrations of salt ([KCl] > 200 mM, [NaCl] > 100 mM) (Little *et al.*, 1967). λ-exo has been reported to be inhibited by 31%, 83% and 99% in the presence of 50 mM, 100 mM and 200 mM KCl compared to its exonuclease activity in no salt (Little *et al.*, 1967). The effect of NaCl on exonuclease activity is even more drastic. Hence, for electrical measurements of the αHL-λ-exo fusion constructs, an electrolyte was needed such that its concentration was low enough to not compromise the activity of λ-exo units of the αHL-λ-exo fusion constructs, but high enough to allow detection of the unitary conductance of αHL in the planar lipid bilayer set-up. A low salt buffer (100 mM KCl, 25 mM Tris.HCl, 50 µM EDTA, pH 7.8) was thus used for the electrical recordings. The bias potential in all electrical recordings was applied as $V_{trans}$-$V_{cis}$, where *cis* is the grounded side and where the protein was added (Chapter 2, Methods).

#### 3.8.9.1. Single channel behaviour of EA10, EA13 and EAA10

EA10, EA13 and EAA10 failed to lyse rRBC (section 3.8.4) but could insert and

form pores in planar lipid bilayers. The pore insertion frequency of EA10, EA13 and EAA10 was extremely low as compared to WT αHL. Moreover, the EA10, EA13 and EAA10 pores were very noisy (Fig. 3.35a-c). As a representative example, the open pore current ($I_o$) histogram of EA10 has been plotted which showed a wide distribution of channels (*n*=87; where *n* denotes the number of single pore insertions) (Fig. 3.35d). At +50 mV, the mean $I_o$ of majority of the channels was 2.7±1.0 pA (mean±standard deviation (S.D.)) (peak 2, Fig. 3.35d). Another smaller peak (peak 1, Fig. 3.35d) was also observed with a mean $I_o$ of 1.4±0.1 pA (mean±S.D.).

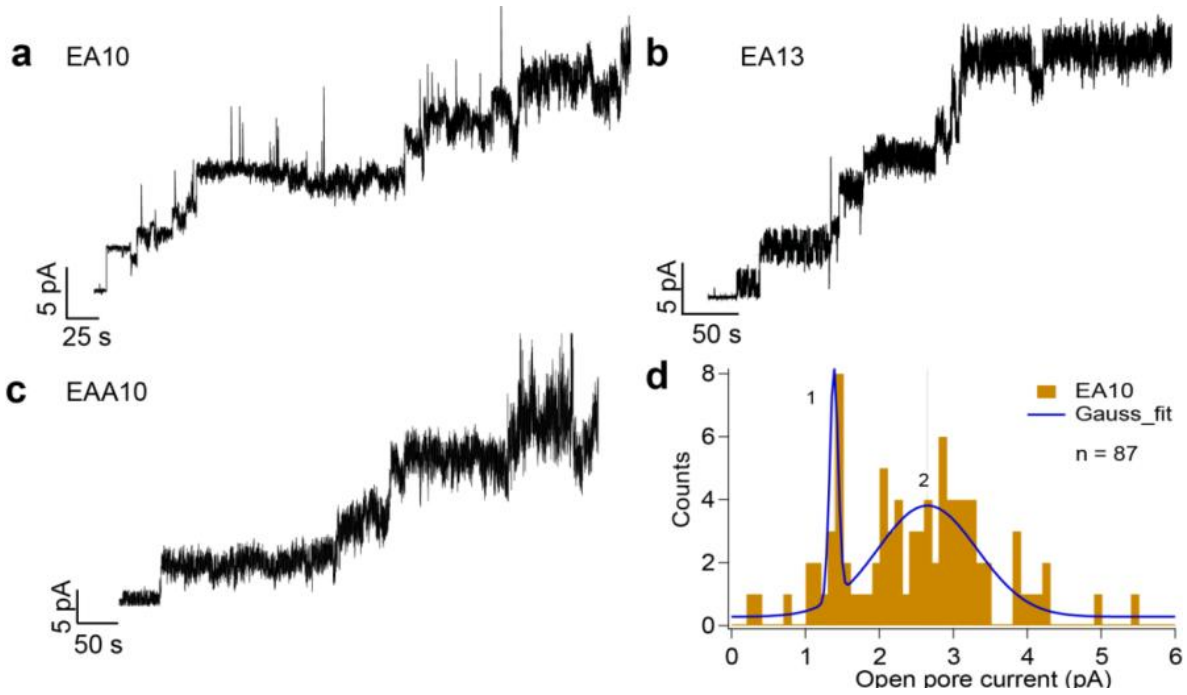



**Figure 3.35. Single channel measurements of EA10, EA13 and EAA10 in planar lipid bilayers.** Current traces showing step-wise insertions of (**a**) EA10, (**b**) EA13 and (**c**) EAA10 channels in planar lipid bilayers at +50 mV. These fusion constructs showed only few and very noisy channels. (**d**) As a representative example of EA10, EA13 and EAA10, the $I_o$ histogram of EA10 channels at +50 mV was plotted. EA10 channels displayed a broad range of channels with mean $I_o$ of 1.4±0.1 pA (peak 1) and 2.7±1.0 pA (peak 2) (mean±S.D.). Current measurements were done in 100 mM KCl, 25 mM Tris.HCl, 50 µM EDTA, pH 7.8.

### 3.8.9.2. Single channel recordings of AE10 and AE13

In comparison to EA10, EA13 and EAA10, AE10 and AE13 readily formed pores in planar lipid bilayers (Fig. 3.36a, b). Although all the proteins were expressed and purified in the same way, the noisy pores of the EA constructs were distinctly different than the quiet AE10 and AE13 pores. This implied that the channel activity was due to the αHL-λ-exo fusion constructs and not some contaminant left over from IMAC purification (section 3.8.9.3). The $I_o$ histograms of AE10 and AE13 pores showed two distinct peaks (Fig. 3.36c, d). The mean currents of the two channel populations of AE13 (red), termed AE13 (1) and AE13 (2), were 1.9±1.0 pA and 4.2±0.8 pA (mean±S.D.; *n*=115, *N*=4; where *n* denotes the number of single pore insertions in *N* independent experiments), respectively (Fig. 3.36c). AE10 channels (AE10 (1) and AE10 (2)) were divided into less distinct populations with mean currents of 2.8±0.5 pA and 4.9±1.2 pA (mean±S.D.; *n*=137, *N*=3) respectively (Fig. 3.36d) (see section 3.8.9.6 for discussion).

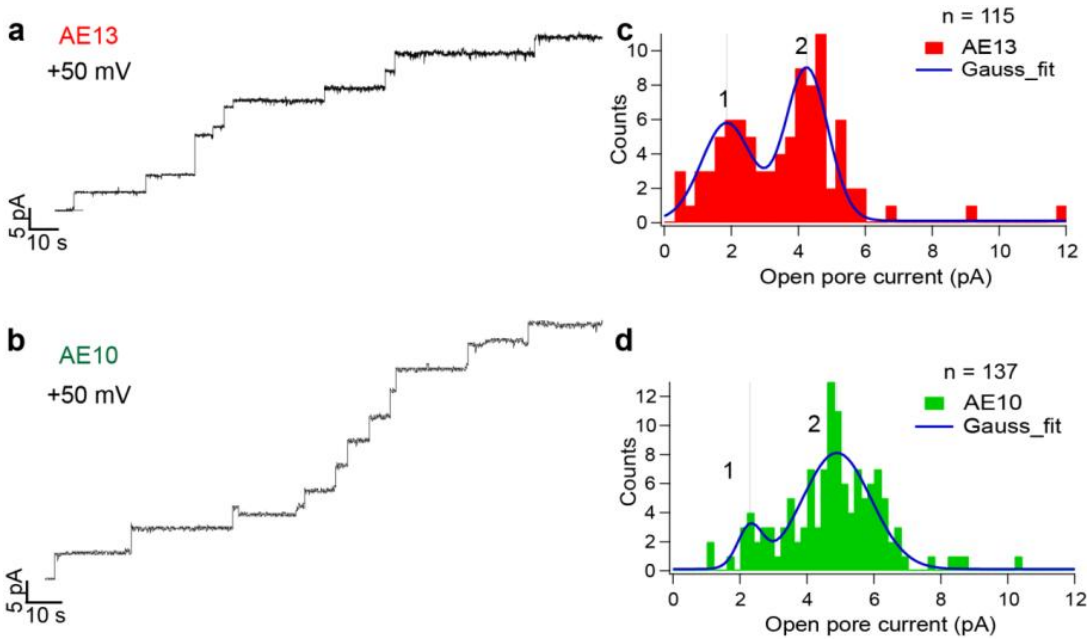



**Figure 3.36. Single channel recordings of AE10 and AE13.** AE10 and AE13 formed channels in planar lipid bilayer. Current races showing rapid insertion of

(**a**) AE13 and (**b**) AE10 at +50 mV. (**c**, **d**) Frequency histograms of the $I_o$ values, determined from the channel steps, showed two distinct peaks and were fitted with a double Gaussian function. (**c**) The mean currents of the two channel populations of AE13 (red) were 1.9±1.0 pA and 4.2±0.8 pA (*n*=115, *N*=4). (**d**) AE10 channels were divided into less distinct populations of mean $I_o$ of 2.8±0.5 pA and 4.9±1.2 pA (*n*=137, *N*=3). Current values are given as mean±S.D.. Experiments were performed in 100 mM KCl, 25 mM Tris.HCl, 50 µM ETDA, pH 7.8.

The mean $I_o$ and electrical characteristics of AE10 (2) and AE13 (2) pores (peak 2, Fig. 3.36c, d) were very similar to the mean $I_o$ of NN α7 of 5.2±1.0 pA (mean±S.D.; *n*=25, *N*=3). About 9% of the AE10 pores (AE10 (1); peak 1, Fig. 3.36d) and ~43% of the AE13 channels (AE13 (1)) (peak 1, Fig. 3.36c) had a lower conductance than NN α7. This could be due to a longer pore length (Chapter 2, section 2.10) or a difference in the charges at the pore mouths of AE10 (1) or AE13 (1) as compared to α7 (section 3.8.9.7).

#### 3.8.9.3. Electrical measurements of AE13 in asymmetric conditions

To increase the ionic flow through the AE10 and AE13 pores and thereby improve their signal to noise ratio, the concentration of salt was increased in the *trans* compartment (asymmetric salt conditions). 550 mM KCl was added to the *trans* compartment to increase the conductance of the AE10 and AE13 pores. But in the *cis* compartment, the [KCl] was kept constant at 100 mM to maintain the exonuclease activity of the AE10 and AE13 proteins (section 3.8.9).

Similar to the $I_o$ populations observed in experiments with symmetric salt conditions (100 mM KCl in both *cis* and *trans*) (section 3.8.9.2, Fig. 3.36d), the $I_o$ values of AE13 under asymmetric salt conditions (100 mM KCl in *cis* and 550

mM KCl in *trans*) were also divided into two populations denoting the existence of two different types of channels (Fig. 3.37). The separation between the mean $I_o$ of AE13 (1) and AE13 (2) channels was more pronounced in the asymmetric salt conditions. The mean $I_o$ of the two major current peaks at +100 mV were 21±3 pA (AE13 (1); peak 1) and 27±3 pA (AE13 (2); peak 2) (mean±S.D.; *n*=127, *N*=4) (Fig. 3.37). The mean $I_o$ of AE13 (2) channels was close to the mean $I_o$ of NN α7 of 28±6 pA (mean±S.D.; *n*=23, *N*=3) under the same salt conditions at +100 mV.

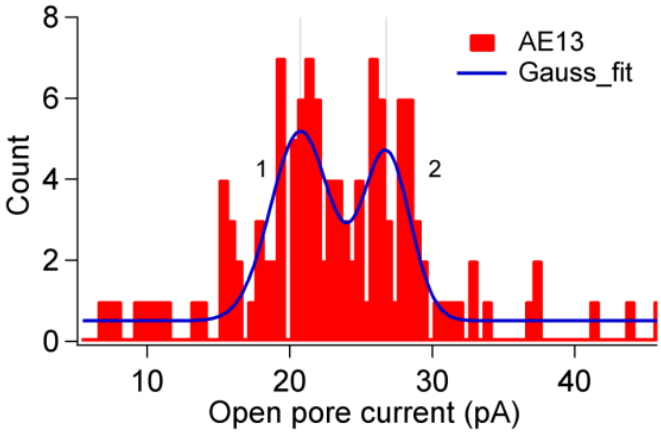



**Figure 3.37. Single channel recordings of AE13 in asymmetric salt conditions.** Histogram of $I_o$ of AE13 in asymmetric salt conditions (100 mM KCl (*cis*) and 550 mM KCl (*trans*)) at +100 mV. Mainly two populations of channels were observed with mean $I_o$ of 21±3 pA (peak 1) and 27±3 pA (peak 2). Current values are denoted as mean±S.D..

### 3.8.9.4. Single channel behaviour of SEC purified AE10

In the IMAC purified AE10 protein fractions, only the AE10 monomer, a truncated AE10 protein consisting of λ-exo (section 3.8.7) and a higher oligomeric band of $(AE10)_7$ presumably, were observed in SDS-PAGE (Fig. 3.29). However, it has been shown that trace quantities of indigenous *E.coli* proteins can bind to the IMAC column and elute with the over-expressed target proteins (Bolanos-Garcia and Davies, 2006; Robichon *et al.*, 2011). To ensure

that the channel activity of the IMAC purified AE10 and AE13 (section 3.8.9.2) was not due to the contaminating *E.coli* proteins or trace quantities of α1 before SEC purification, single channel electrical measurements were repeated with the SEC purified AE10 fraction (AE10_SEC).

AE10_SEC showed mainly two populations of pores with unitary conductance of 3.1±0.6 pA and 4.8±0.5 pA (mean±S.D.; *n*=128) which closely matched with the $I_o$ of AE10 before SEC purification, of 2.8±0.5 pA and 4.9±1.2 pA (mean±S.D.; *n*=137), in 100 mM KCl, 25 mM Tris.HCl, 50 µM EDTA, pH 7.8, at +50 mV (Fig. 3.38). A population with mean $I_o$ of 4.8±0.5 pA, similar to the mean $I_o$ of NN α7 of 5.2±1 pA (mean±S.D.; *n*=25) was observed in the SEC purified AE10 protein as well. As discussed before, the AE10_SEC sample did not contain any α1 (section 3.8.6); thus the channels with conductance similar to NN α7 should also be formed of only AE10 monomers. This result suggested AE10 could form heptamers upon encountering bilayers and form two kinds of channels with different mean $I_o$. The different conductance could be owed to different arrangements of the λ-exo subunits on the αHL cap (section 3.8.9.6).

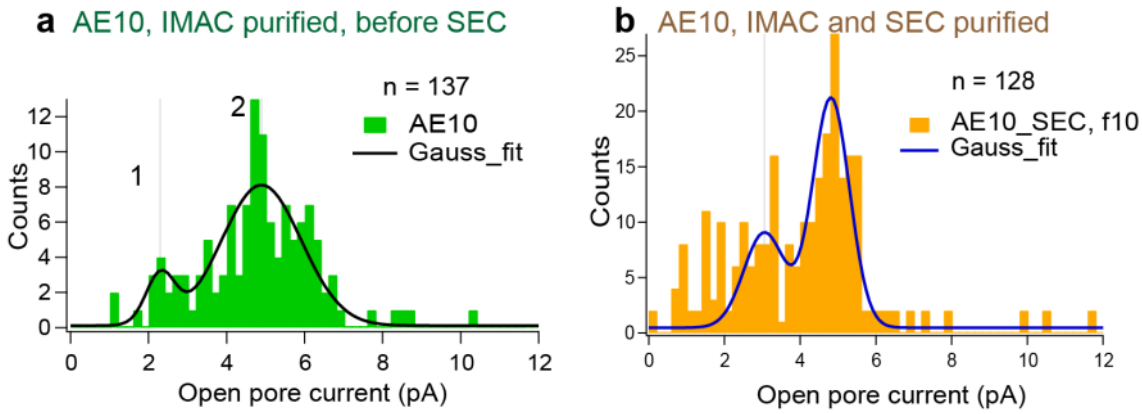



**Figure 3.38. $I_o$ histograms of AE10 before and after SEC.** (**a**) IMAC purified AE10 before SEC showed two kinds of pores. (**b**) IMAC purified AE10 protein after SEC (AE10_SEC) showed a similar distribution of pores confirming that AE10 formed two kinds of pores in planar lipid bilayers.

### 3.8.9.5. γ cyclodextrin (γCD) binding confirms an α7-like barrel of AE13

Although a fraction of AE13 and AE10 formed pores on planar lipid bilayers with open pore conductance like α7, the stoichiometry of AE10 and AE13 monomers in the pores was not obvious from conductance measurements (section 3.8.9.2). To confirm that the membrane inserting or pore forming part of AE10 and AE13 was like the β barrel of α7, γCD, a molecular pore adapter of α7 was used (Gu *et al.*, 1999). As the difference in AE10 and AE13 is just in the linker lengths between αHL and λ-exo, the stoichiometry of AE10 and AE13 pores was assumed to be similar. Thus as a representative example, γCD experiments were done using AE13.

γCD binds near the constriction of the α7 barrel, which can be detected as short blockades in $I_o$ in current recordings (Chapter 2, section 2.11). γCD bound to 18 out of 20 AE13 pores when added to the *trans* side of the bilayer (*N*=2). At +50 mV, in 100 mM KCl salt on both *cis* and *trans* sides, γCD binding caused 51±1% (mean±s.e.m.; *n*=25, *N*=2; where *n* denotes the number of channels studied in *N* independent experiments) blockades in the $I_o$ amplitude of AE13 (Fig. 3.39). The dissociation rate constant ($k_{off}$) of γCD binding to AE13 was calculated from the mean dwell time of γCD within the pore (Chapter 2, section 2.12) to be 3.5±0.4 $s^{-1}$ (mean±s.e.m.; *n*=678, *N*=11; where *n* denotes the number of binding events in *N* number of channels). The $k_{off}$ values of γCD binding to AE13 were similar to the values observed for γCD binding to the α7 pore (Chapter 2, section 2.12). This implies that the γCD binding site in AE13 was similar to that of α7, and that seven αHL units of AE13 formed pores in

planar lipid bilayers. However, it is not clear how the seven λ-exo components of AE10 and AE13 were arranged on the αHL cap in the AE10 or AE13 heptamers (section 3.8.9.6).

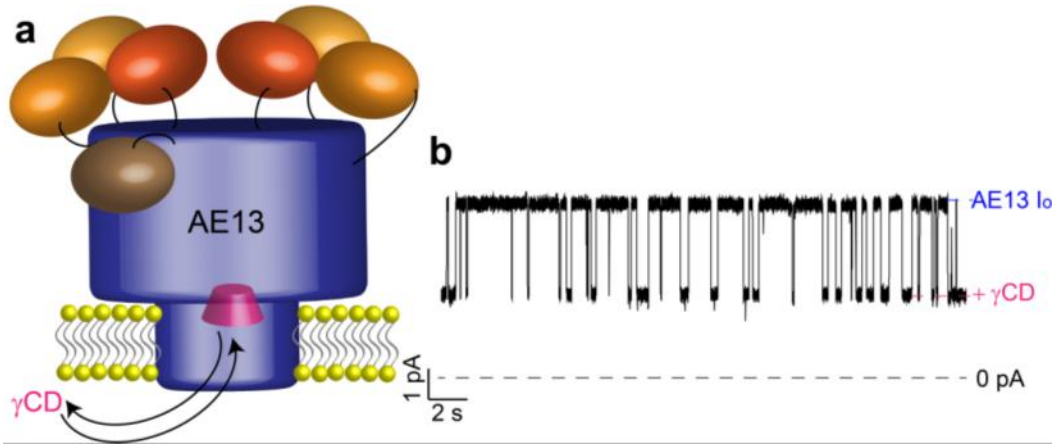



**Figure 3.39. γCD binding to AE13**. To ensure that the pore forming part of AE13 pores was a heptamer formed of its αHL units, the pore adapter of α7, γCD, was employed (Gu *et al.*, 1999). (**a**) Cartoon showing γCD binding near the constriction in the β barrel formed by the αHL units. The orientation of the λ-exo units of AE13 was not evident from single channel measurements (section 3.8.9.6). (**b**) Single channel current trace of AE13 showing blockages due to γCD (40 μM; *trans*) binding at +50 mV in 100 mM KCl, 25 mM Tris.HCl, 50 μM ETDA, pH 7.8.

### 3.8.9.6. Difference between AE10 and AE13 channels

Heptamers of AE10 and AE13 formed two kinds of pores in planar lipid bilayers (sections 3.8.9.2-5). ~91% AE10 pores (AE10 (2)) and ~57% of AE13 (AE13 (2)) pores had a mean open pore conductance like NN α7 (peak 2, Fig. 3.36c, d). The AE10 pores (AE10 (1) and AE10 (2)) were divided into less distinct populations (Fig. 3.36d) as compared to the AE13 pores (AE13 (1) and AE13 (2)) (Fig. 3.36c). The difference between the AE10 and AE13 pores could be attributed to the difference in the linker lengths between αHL and λ-exo monomers in AE10 and AE13. The linker lengths would govern the orientation of the seven λ-exo units of the AE10/ AE13 on the α7 cap. Due to the short linker between αHL and λ-exo in AE10 as compared to AE13, the λ-exo units in

AE10 would be constrained to occupy positions away from the αHL cap pore entrance, primarily forming pores with α7 type conductance (AE10 (2)) (Fig. 3.40c). In AE13, owing of the longer linker length, the λ-exo units will be free to associate with neighbouring subunits, close to the pore entrance of the αHL cap and form pores with conductance different from α7 (AE13 (1)) (Fig. 3.40a, b).

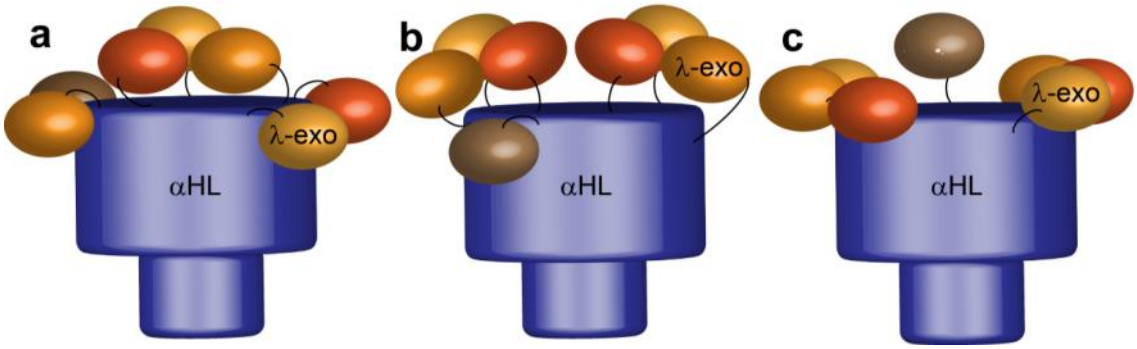



**Figure 3.40. Models of possible AE10 and AE13 pores.** AE10 and AE13 pores are formed by heptamers of AE10 and AE13 monomers, respectively (section 3.8.9.5). The orientation of the seven λ-exo units of the AE10 and AE13 on the α7 cap would determine the open pore conductance characteristics of these pores. In structure **'a'**, three λ-exo units could form a trimer whose central pore aligns with that of the α7 pore. In structure **'b'**, λ-exo units could associate with neighbouring λ-exo units and form two trimers side-by-side. Structure **'c'** is like **'b'** but with the two λ-exo trimers away from the central pore of α7. Hence, the conductance of structure **'c'** would be like α7 whereas the conductance of structures **'a'** and **'b'** would be different from α7 because of a longer pore length, and difference in charges at the pore openings of λ-exo and α7. If the inter-subunit interactions between the λ-exo units were broken upon pore formation, they would also keep away from the α7 cap yielding pores with unitary conductance like α7. Owing to the short linker length in AE10, pores would primarily form with structure **'c'**. AE13, which has a longer linker between λ-exo and αHL, could form pores of structure **'b'**. This would explain the α7 type channels in AE10 and AE13 and the low population of AE10 (1) pores as compared to AE13 (1).

### 3.8.9.7. Current-voltage (I-V) characteristics of AE13 channels

As the AE13 pores were divided into distinct populations of AE13 (1) and AE13(2) pores (Fig. 3.36c), the AE13 pores were further characterized. The I-V characteristics of AE13 channels were compared to NN α7 in symmetric and asymmetric salt conditions. The AE13 (1) and AE13 (2) pores could be

distinguished from their characteristic I-V curves. The I-V characteristics of AE13 (2) were exactly similar to NN α7.

In symmetric salt conditions (100 mM KCl in both *cis* and *trans*), the I-V plots of AE13 (1) and AE13 (2) differed at positive potentials. The difference increased at higher positive voltages (+50–+100 mV) (Fig. 3.41a). At negative potentials, the I-V curves of all three AE13 (1), AE13 (2) and NN α7 pores, were similar. If the pore length of AE13 (1) channels was different than that of AE13 (2) and NN α7, the conductance of AE13 (1) would have differed from AE13 (2) and NN α7 at both positive and negative potentials.

The difference between the I-V characteristics of AE13 (1) and AE13 (2) were more pronounced in asymmetric salt conditions (100 mM KCl in *cis* and 550 mM KCl in *trans*) (Fig. 3.41b). Unlike the symmetric conditions, at negative potentials the $I_o$ values of AE13 (1) significantly differed from those of AE13 (2) and NN α7.

The rectification of the AE (1), AE13 (2) and NN α7 pores was visualized by plotting the rectification ratios (RR) ($-I_{+V} / I_{-V}$) from 20 -100 mV (Fig. 3.41c, d). AE13 (2) and NN α7 rectified slightly with increasing potentials in symmetric salt conditions (Fig. 3.41c). The RR of NN α7 was 1.02±0.06 at 20 mV and increased to 1.19±0.10 at 100 mV in symmetric salt conditions (mean±s.e.m.; *n*=3, *N*=2), which is similar to values reported for WT α7 (Walker *et al.*, 1992b). Similar to the values of NN α7, the RR of AE13 (2) increased from 1.10±0.03 at 20 mV to 1.29±0.11 at 100 mV under the same conditions (mean±s.e.m.; *n*=3,

*N*=3). In comparison to the pores of AE13 (2) and NN α7, AE13 (1) channels displayed an opposite trend with its RR decreasing with increasing applied potentials. For example, $(-)I_{+20}$ / $I_{-20}$ of AE (1) was 0.94±0.01 and $(-)I_{+100}$ / $I_{-100}$ was 0.79±0.01 (mean±s.e.m.; *n*=3, *N*=2) in symmetric salt conditions.

Under asymmetric salt conditions, the rectification of AE13 (2) and NN α7 became much less in asymmetric salt conditions and the difference between AE13 (1) and AE13 (2) was even more pronounced (Fig. 3.41d). $(-)I_{+20}$ / $I_{-20}$ of AE13 (2) was 0.98±0.02 and $(-)I_{+100}$ / $I_{-100}$ was 0.99±0.01 (mean±s.e.m.; *n*=11, *N*=3). On the other hand, AE13 (1) channels rectified appreciably with increasing potentials with its RR d from 0.94±0.02 ($-I_{+20}$ / $I_{-20}$) to 0.85±0.09 ($-I_{+100}$ / $I_{-100}$) (mean±s.e.m.; *n*=8, *N*=3).

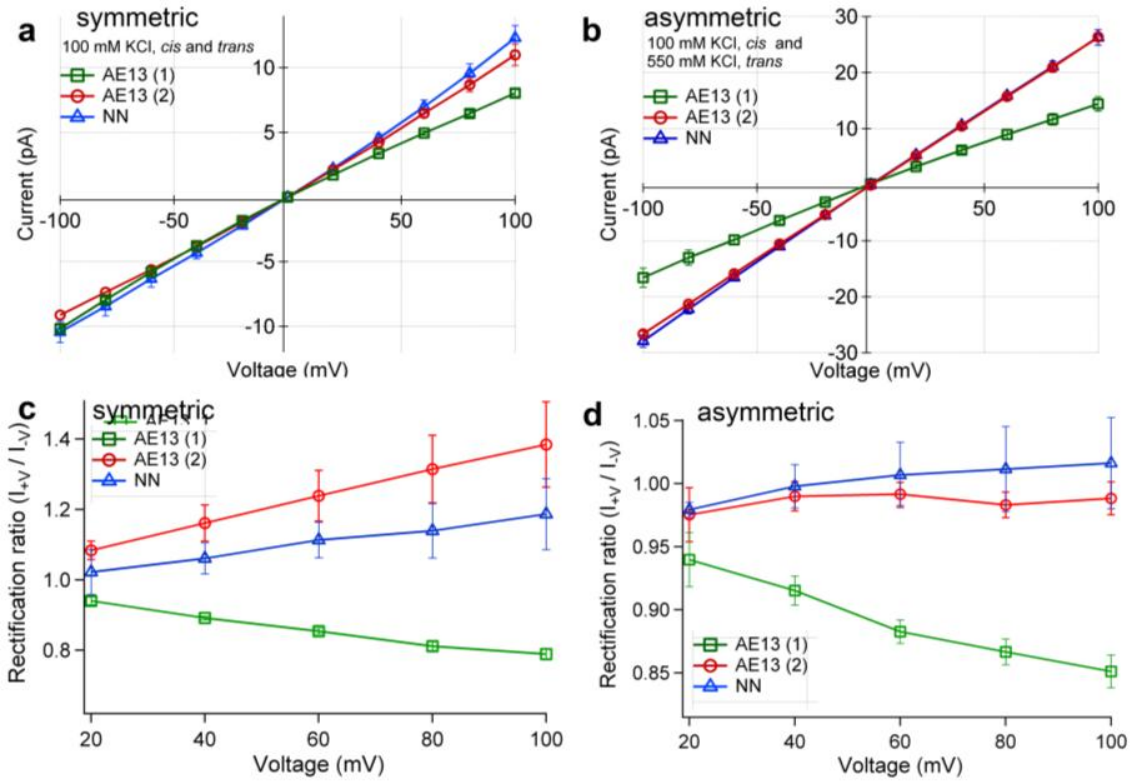



**Figure 3.41. I-V characteristics of AE13 pores as compared to NN α7.** I-V plots of AE13 were made in (**a**) symmetric (100 mM KCl in both *cis* and *trans*) and (**b**) asymmetric (100 mM KCl in *cis* and 550 mM KCl in *trans*) salt conditions and compared to the I-V characteristics of NN α7. The two kinds of

AE13 channels, AE13 (1) and AE13 (2), had different I-V properties; with the latter being very similar to NN α7. Rectification ratios (RR) were calculated as the absolute value of $I_{+V}$ / $I_{-V}$ from 20-100 mV under (**c**) symmetric and (**d**) asymmetric salt conditions. In symmetric salt conditions, the RR of AE13 (2) and NN α7 were increased with increasing applied potential, lying between 1.10 ($-I_{+20}$ / $I_{-20}$) to 1.29 ($-I_{+100}$ / $I_{-100}$). AE13 (1) channels showed an entirely opposite trend, i.e., with a decreasing RR from 0.94 ($-I_{+20}$ / $I_{-20}$) to 0.79 ($-I_{+100}$ / $I_{-100}$) with increasing applied potential. With different salt concentrations in the *cis* and *trans* chambers (asymmetric salt conditions), the rectification of AE13 (2) and NN α7 was much subdued lying from 0.98 ($-I_{+20}$ / $I_{-20}$) to 1.01 ($-I_{+100}$ / $I_{-100}$), whereas the rectification of AE13 (1) remained the same decreasing from 0.94 ($-I_{+20}$ / $I_{-20}$) to 0.85 ($-I_{+100}$ / $I_{-100}$).

To explain the physical origin of the difference in rectification behaviour of AE13 (1) and, AE13 (2) and NN α7, the vacuum electrostatic maps of αHL (PDB ID 7AHL) and λ-exo (PDB ID 1AVQ) were analysed (Fig. 3.42a, b). The rectification of α7 in symmetric salt conditions is owed to the difference in the charges at its cap (*cis* entrance) and the β barrel (*trans* entrance) (Chapter 2, section 2.10) (Fig. 3.42a). The *cis* entrance of α7 is lined by lysine residues (Lys 8); positively charged in the experimental conditions of pH 7.8. The *trans* entrance, on the other hand, is formed of the negatively charged residues, Asp 127 and Asp 128, at pH 7.8. The negatively charged pore mouth of the β barrel at pH 7.8, presents a repulsive barrier to negatively charged ions flowing from the *trans* chamber to the *cis* chamber at negative applied potentials. The effect of differential flow of ions through α7 at positive and negative potentials is seen more pronounced at higher potentials. A higher salt concentration in the *trans* chamber (asymmetric salt conditions) offsets the smaller flow of negative ions from *trans* to *cis* at negative potentials. Thus, the rectification of NN α7 and AE13 (2) was less in asymmetric salt conditions than in symmetric salt conditions.

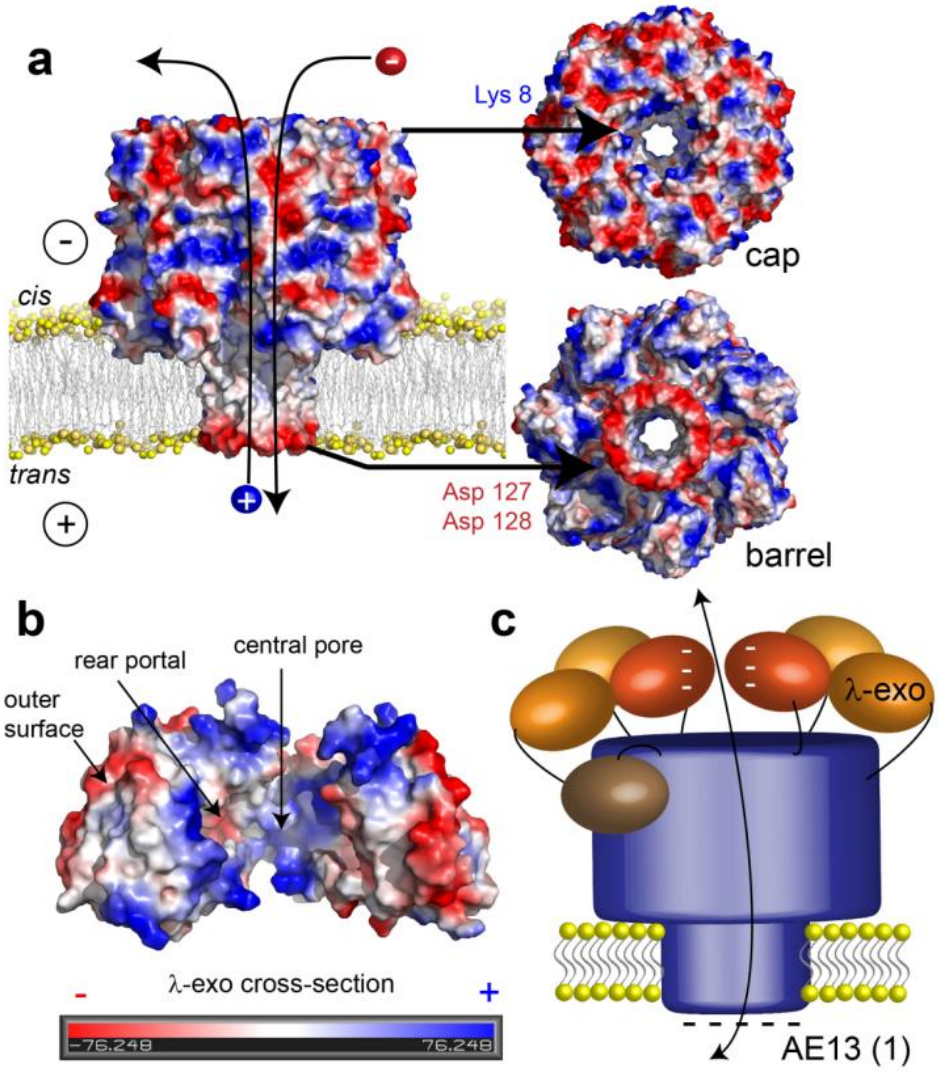



**Figure 3.42. Molecular models explaining RR of AE13 (1) and AE13 (2) pores.** As compared to NN α7 and AE13 (2), the RR of AE13 (1) channels followed an opposite trend. This could be explained by considering the charges on the λ-exo and αHL proteins. (**a**) The vacuum electrostatics map of αHL (PDB ID 7AHL) showed a positively charged cap pore mouth (*cis* entrance) formed by Lys 8. The pore mouth at the β barrel of αHL (*trans* entrance) is negatively charged due to the residues, Asp127 and Asp128. Due to the charges at the pore mouth of αHL, the current is less at negative potentials as compared to positive potentials and hence, RR increases with increasing potential. The difference in AE13 (1) and AE13 (2) pores was in their *cis* entrances formed by λ-exo subunits. (**b**) The vacuum electrostatics map of WT λ-exo showed a highly positive central pore (blue) and a negatively charged outer surface and rear portal (red) (Fig. 3.9). (**c**) Cartoon showing the possible arrangement of the λ-exo units of AE13 (1). Six λ-exo units could form two trimers side by side. The negative charges on the outer surface of the λ-exo units of AE13 (1) would make its *cis* entrance negatively charged, which could explain its rectification behaviour.

The *trans* entrances of AE13 (1) and AE13 (2) channels would be similar as they are formed of αHL units (section 3.8.9.5). However, the *cis* entrances of AE13 (1) and AE13 (2) could be different because of different orientations of the λ-exo units. The opposite trend of rectification seen in the case of AE13 (1) as compared to AE13 (2) implied that the charge of *cis* entrance of these pores was oppositely charged to the AE13 (2) channels. If the *cis* entrance of AE13 (1) pores was more negatively charged than its *trans* entrance, the net flow of negative ions would be higher at negative potentials (from *trans*→*cis*) than at positive potentials (from *cis* →*trans*). This hypothesis was strengthened by the experimental observation that RR of AE13 (1) pores was <1 and decreased further with increasing potentials. Thus, negatively charged regions of λ-exo were examined. Crystal structure of λ-exo revealed that its pore is positively charged to let the DNA substrate in and its outer surface is negatively charged (Fig. 3.42b). AE13 (1) channels could be formed with six of its seven λ-exo units forming two trimers on top of αHL cap (Fig. 3.42c). Side-by-side arrangement of the two trimers on the αHL cap would lead to a negatively charged *cis* entrance and could be a possible explanation for the rectification behaviour of AE13 (1).

### 3.8.10. Summary of the αHL- λ-exo fusion constructs' activities

All the αHL-λ-exo fusion constructs retained their exonuclease activity. The AE10 and AE13 constructs were also hemolytically active and rapidly inserted in planar lipid bilayers. Hence, our results indicate that the AE10 and AE13 constructs are essentially a new kind of processive exonuclease that could also insert in bilayers. Table 3.1 lists a summary of the activity studies of the αHL- λ-exo fusion constructs.

**Table 3.1. Summary of the fusion construct characterization experiments.** The hemolytic, exonuclease and channel activity of the different αHL-λ-exo fusion constructs are marked as '+' if activity was seen and '-' if activity was not seen. In the case of EA10, EA13 and EAA10, '+/-' denotes that only a few and very noisy pores were observed as compared to the proliferous pore insertions of the AE10 and AE13 constructs.

| Construct | Hemolytic | Exonuclease | Channels in bilayers |
|---|---|---|---|
| AE10 | + | + | + |
| AE13 | + | + | + |
| EA10 | - | + | +/- |
| EA13 | - | + | +/- |
| EAA10 | - | + | +/- |

### 3.8.11. Determining the oligomeric state of AE10 using SEC

The active oligomeric state of αHL is a heptamer (Song *et al.*, 1996) and that of λ-exo is a trimer (Kovall and Matthews, 1997b; Subramanian *et al.*, 2003; Zhang *et al.*, 2011). The AE10 and AE13 proteins formed heptamers and lysed rabbit blood cells and formed active trimers upon encountering suitable dsDNA substrates (Figs. 3.30, 3.31a). Hence, AE10 and AE13 could exist as heptamers or trimers in solution depending on the relative strength of the inter-subunit interactions in λ-exo and α7 (Fig. 3.43).

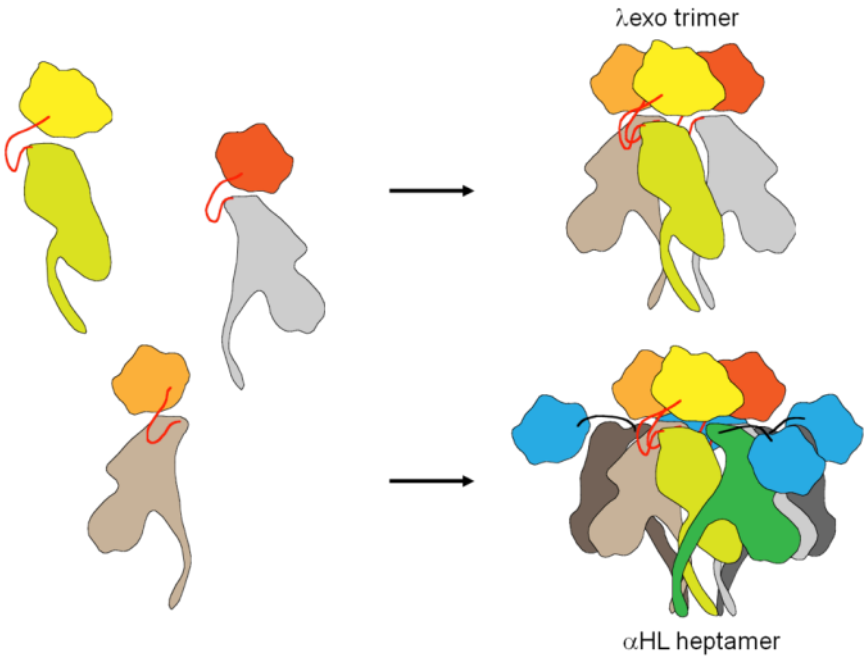



**Figure 3.43. Possible oligomeric states of AE10 and AE13.** Cartoon showing the two possible oligomeric states of the AE fusion proteins. The λ-exo part would oligomerize as a trimer and the αHL part would prefer a heptameric state. Hence, depending upon the subunit interactions, the fusion proteins could either be a heptamer or a trimer.

Like λ-exo, AE10 samples migrated as monomers ($M_{app}$≈67,000) in denaturing SDS-PAGE gels (Fig. 3.32). Hence, the exact stoichiometry of the AE10 could not be determined by the preliminary SDS-PAGE analysis. SEC was thus used to determine the oligomeric state of AE10 in solution.

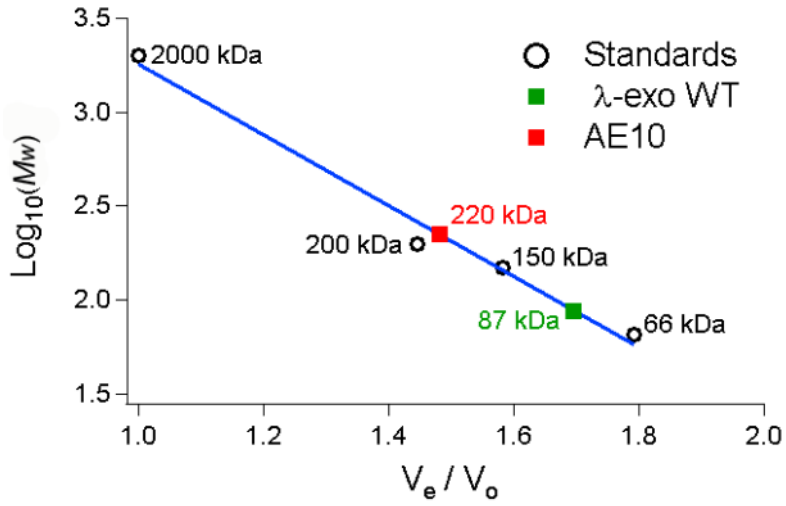



**Figure 3.44. Calibration of the size exclusion column for $M_w$ determination of λ-exo and AE10.** The void volume ($V_o$) of a Superdex S200 26/10 column

was measured using blue dextran (2000 kDa). The column was calibrated with three other $M_w$ standards of 66 kDa (albumin), 150 kDa (alcohol dehydrogenase) and 200 kDa (β amylase) by noting their volume of elution ($V_e$). The plot of $\log_{10}$ MW of the three $M_w$ standards as a function of $V_e$ / $V_o$ was used to compute the $M_w$ of λ-exo and AE10 from their respective $V_e$ from the same column. The flow rate was kept at a constant 0.4 ml $min^{-1}$ in each case.

A Superdex 200 column (column volume 24 ml) was calibrated with four protein standards (66 kDa, 150 kDa, 200 kDa and 2000 kDa) (Fig. 3.44). SEC was performed on IMAC purified protein fractions. AE10 (Rosetta 20°C, section 3.8.3) eluted as a single peak (11-12 ml) from the Superdex 200 column (Fig. 3.45a). WT λ-exo (Fermentas), run as a control, also eluted as a single peak (14 ml) (Fig. 3.45c). All the elution fractions from the SEC column were analysed under denaturing conditions (2% SDS) using SDS-PAGE (Fig. 3.45b, d).

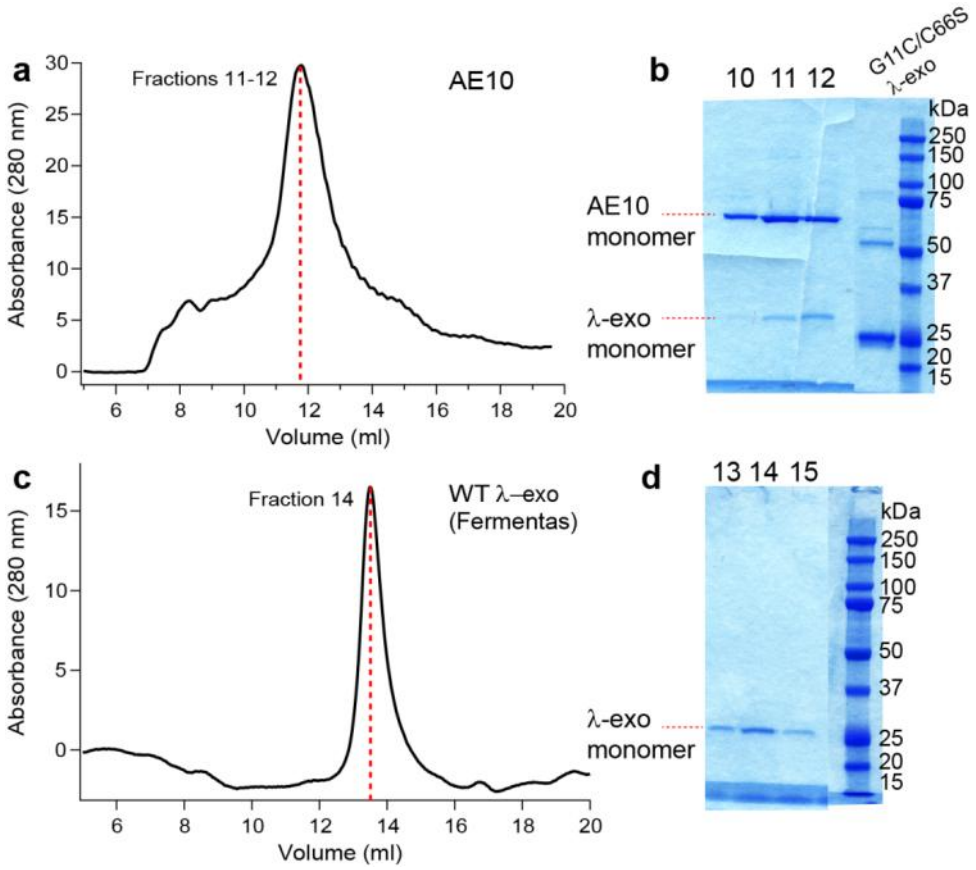



**Figure 3.45. SEC of AE10 and WT λ-exo.** SEC of IMAC purified AE10 and WT λ-exo (Fermentas) was performed in order to deduce the oligomerization state

of AE10. (**a**) Figure shows elution profile of AE10 from a 24 ml Superdex 200 column with 8.0 ml void volume. The elution peak of AE10 centred at the 11-12 ml fraction. (**b**) SDS-PAGE showing that the elution fractions of 10-12 ml contained the AE10 species. The monomer band of AE10 and its truncated protein band (section 3.8.7) were observed. As a control, WT λ-exo was also run on the same column. (**c**) Chromatogram showing the SEC elution of profile WT λ-exo. (**d**) SDS-PAGE analysis of the SEC fractions of WT λ-exo that were collected. WT λ-exo was detected in the elution fractions of 13-15 ml.

AE10 co-eluted with its truncated product ($M_{app}$≈29,000) (Fig. 3.32), which migrated just above the G11C/C66S λ-exo protein ($M_{app}$≈25,000) purified under the same conditions as AE10 on SDS-PAGE (Fig.3.45b). This confirmed that the truncated AE10 protein was formed of the λ-exo part (section 3.8.7). As SEC could not separate the truncated AE10 protein from full length AE10 proteins, it was concluded that the truncated proteins were associated with the AE10 oligomers.

The sharp elution peaks of AE10 and WT λ-exo in SEC implied the proteins majorly formed just one kind of oligomer (Fig. 3.45). From the calibration curve of the SEC column, the accurate weight of AE10 and WT λ-exo proteins were calculated to be 220 kDa and 87 kDa (Fig. 3.44). Each full length AE10 monomer should have a $M_w$ of ~61 kDa and the expected $M_w$ of WT λ-exo monomer is 24 kDa. This implied that the AE10 oligomer was composed of 3.7 AE10 monomers and WT λ-exo consisted of 3.2 monomers in solution (150 mM NaCl, 25 mM Tris.HCl, pH 8.0). AE10 was thus assumed to exist as trimers in solution.

EA13 and AE13 also eluted with a similar profile as AE10 in SEC (11-12 ml fraction) implying that EA13 and AE13 also existed as trimers in solution.

### 3.8.12. Oligomerization of AE constructs

The αHL units of the αHL-λ-exo fusion constructs spontaneously formed a small quantity of heptamers (section 3.8.8). AE10 and AE13 could be induced to form a higher quantity of heptamers on adding a suitable oligomerization reagent such as rRBCm or detergent (DOC ≥ 5 mM (c.m.c.)) (Bhakdi *et al.*, 1981; Walker *et al.*, 1992a) (Figs. 3.49, 3.50).

The heptamers formed by AE10 and AE13 on rRBCm, detergent micelles, and planar lipid bilayers (sections 3.8.9) consisted of seven λ-exo monomers on top of the αHL heptamer. To study the single molecule activity of λ-exo and to use the λ-exo-αHL pore complex for DNA sequencing, only one active molecule of λ-exo, i.e., a λ-exo trimer is required on the top of the α7 cap (Fig. 3.46a, b) (section 3.4). SEC revealed that the fusion proteins formed trimers in solution, which should be formed by the λ-exo parts (section 3.8.11). Thus, to construct the desired λ-exo-αHL pore complex, WT α1 could be added to the fusion proteins trimers to fill in the gaps between the αHL parts of the fusion proteins and form active heptamers (Fig. 3.46).

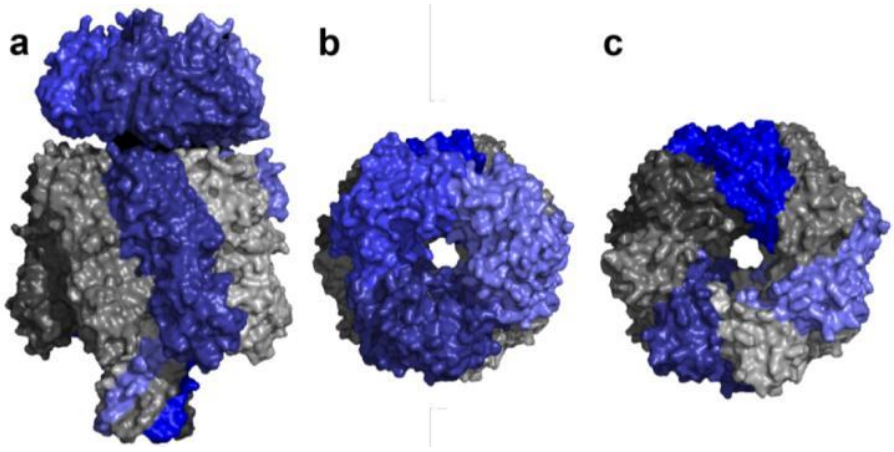



**Figure 3.46. Hetero-heptamers of fusion constructs to form the αHL-λ-exo complex.** To achieve an active αHL-λ-exo complex, valid for DNA sequencing and single molecule studies, a trimer of λ-exo is needed on the αHL cap. Three

αHL-λ-exo fusion monomers (blue) could be mixed with αHL WT monomers (α1) (grey) to form hetero-heptamers. Space-filling models of the hetero-heptamers of αHL-λ-exo fusion monomers and α1 showing the (**a**) side-view and (**b**) the top-view. (**c**) Cross-section of the hetero-heptamer only showing the αHL parts of the αHL-λ-exo fusion monomers, which need to integrate with α1.

#### 3.8.12.1. Models of hetero-oligomers of αHL-λ-exo fusion constructs with WT α1

The feasibility of forming hetero-heptamers of the fusion constructs with WT α1 was assessed computationally in several steps (Fig. 3.48a i-iv).

(i) To model the trimers of the fusion constructs, the monomer models of AE10, AE13, EA10 and EA13 were made using MODELLER, a protein structure modeling software (Sali *et al.*, 1995). MODELLER predicts protein structure from its sequence based on a known structure with a similar sequence. The software then refines the predicted protein structure using an inbuilt function for energy minimization. The structures of αHL and λ-exo monomers, extracted from the crystal structures of WT α7 (PDB ID 7AHL) and λ-exo (PDB ID 1AVQ), respectively, were used as the input structural files for modeling AE10, AE13, EA10 and EA13. In all the cases, the membrane inserting β strand of the αHL parts was remodeled significantly because it is formed of hydrophobic residues on one side. The rest of the protein structures of αHL and λ-exo monomers remained intact in AE10, AE13, and EA13 (Fig. 3.47). This implied that the linker lengths in AE10, AE13, and EA13 were sufficiently long to join the protein termini without affecting the individual protein structure. On the other hand, the length of linker in EA10 was not adequate and the N-terminus of αHL deviated by 7.9 Å with respect to WT αHL. This could explain the inactivity of EA10 in hemolytic assays (section 3.8.4).

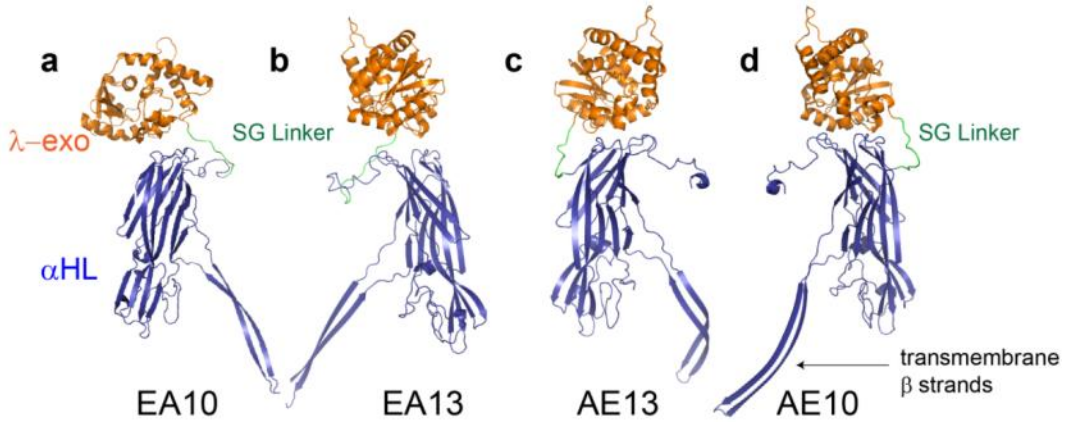



**Figure 3.47. Monomer models of AE10, AE13, EA10 and EA13.** Monomer models of (**a**) EA10 and (**b**) EA13, (**c**) AE13 and (**d**) AE10, generated using MODELLER (Sali *et al.*, 1995). The transmembrane spanning β strands of the αHL parts (blue) were remodeled significantly. The λ-exo parts are shown in orange and the connecting SG10 and SG13 linkers in green.

(ii) Next, the trimer models of the fusion constructs were created by aligning three αHL-λ-exo fusion monomer models with the three monomers of λ-exo (PDB ID 1AVQ) (Fig.3.48a ii).

(iii) To model the hetero-heptamers, the trimer models of the fusion constructs were placed on the crystal structure of WT α7 (PDB ID 7AHL) such that the central pore of λ-exo components coincided with the WT α7 pore; as required in the desired λ-exo-αHL pore complex (Figs. 3.46, 3.48a iii).

(iv) The α1 subunits with which the αHL components of the fusion constructs overlapped were removed. The resulting hetero-heptamer made by the trimer of the fusion constructs and the rest of the α1 subunits was analysed (Fig.3.48a iv).

It was found that the αHL units of EA10 and EA13 trimers were shifted outwards by a significant distance from other WT α1 units and mixed heptamers did not

seem feasible (Fig. 3.48b, d). For example, the N-terminus of the αHL component of one of the EA13 monomers deviated by ~2 nm and its C-terminus deviated by ~1.4 nm from the respective termini of α1, which the EA13 monomer had replaced. On the other hand, AE10 and AE13 trimers could be integrated with WT α1 to form hetero-heptamers (Fig. 3.48a, c).

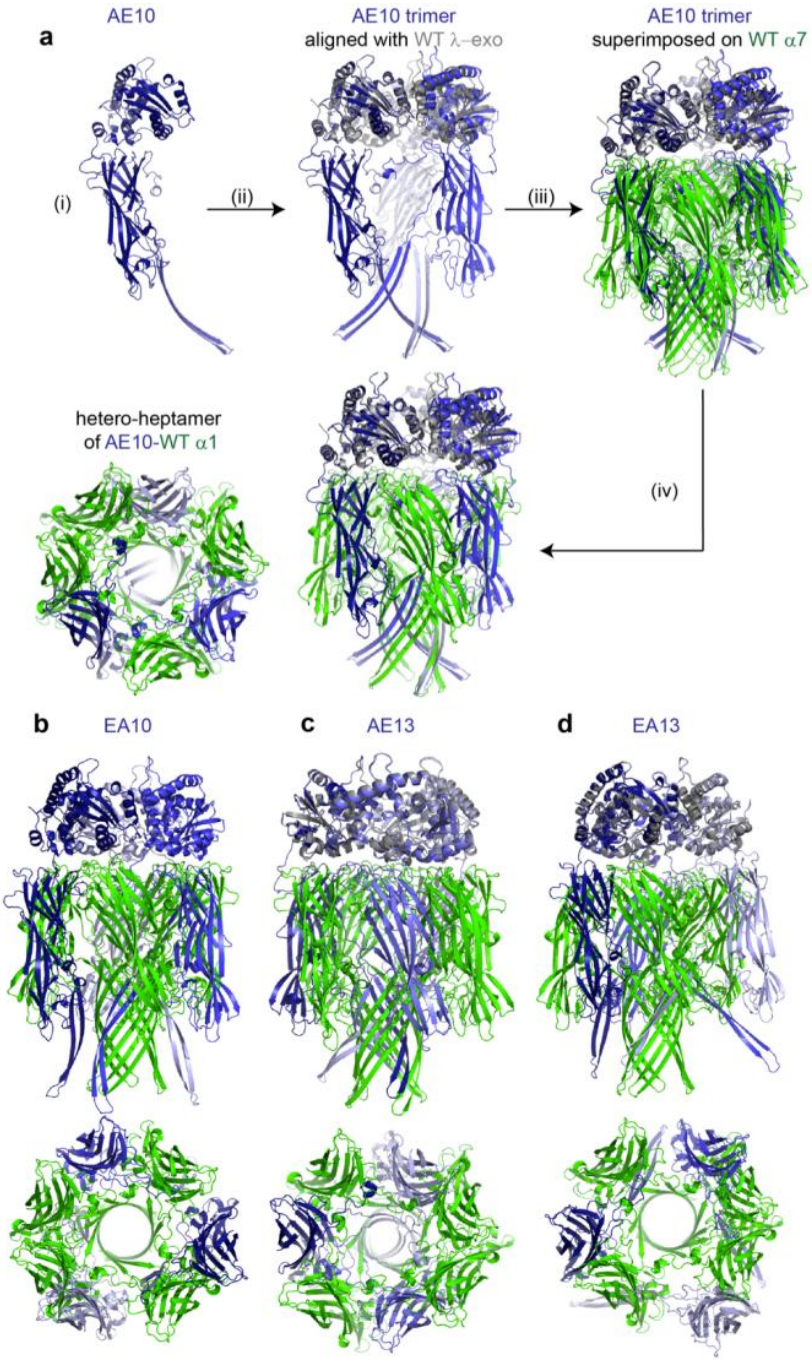



**Figure 3.48. Hetero-heptamer models of the fusion proteins with WT α1.** The feasibility of forming hetero-heptamers of the αHL-λ-exo fusion constructs

was assessed by computational studies. (**a**) Schematic showing the steps involved in modelling the hetero-heptamer of AE10-WT α1. (i) The monomer of AE10 (blue) was modelled in MODELLER (Sali *et al.*, 1995). (ii) To make the trimer model, AE10 monomers were aligned with the crystal structure of the WT λ-exo (PDB ID 1AVQ) (grey) in PyMOL. (iii) The AE10 trimer was superimposed on the crystal structure of WT α7 (PDB ID 7AHL) (green) such that the central pores were aligned. (iv) The α1 subunits overlapping with the αHL components of the AE10 trimers were removed. The resulting structure was the desired AE10-WT α1 hetero-heptamer (left: top-view, right: side-view). A similar protocol was followed to model the hetero-heptamers of (**b)** EA10, (**c)** AE13 and (**d)** EA13. Top and bottom panels in each figure show the side and the top-views of the hetero-heptamers. The λ-exo parts of the fusion proteins are omitted from the top-views for clarity. WT α1 units could be integrated with AE10 and AE13 to form heptamers. In comparison, the αHL parts of EA10 and EA13 deviated significantly from what would be required to form hetero-heptamers.

#### 3.8.12.2. Hetero-heptamers of AE constructs with WT α1

Methods to make hetero-heptamers of αHL containing engineered α1 subunits (α1*) in any desired stoichiometry are well established (Braha *et al.*, 1997; Cheley *et al.*, 1999). The α1* proteins are mixed with WT α1 with an oligomerization agent, which results in hetero-heptamers with different number of α1* subunits. Alternatively, the DNA of α1* and α1 can be co-transcribed and co-translated in the presence of oligomerization agents such as rRBCm, or the proteins of α1* and α1 could be added to rRBCm at a later step. If the engineered subunit is tagged, for instance, with an Asp-8 tail, depending upon the number of α1* in the hetero-heptamers, they would have different overall charges and can be separated on SDS-PAGE (Miles *et al.*, 2002).

α1 was added to coassemble with the αHL units of AE10 and AE13 and form hetero-heptamers (Fig.3.48). Equal concentrations of AE10 and NN α1 DNA were mixed and co-translated *in vitro* in the presence of rRBCm. In another experiment, IVTT proteins of AE10 and NN α1 were mixed and then added to

rRBCm. As controls, AE10 and NN α1 proteins were incubated with rRBCm separately (*lanes 5, 8*, Fig. 3.49a). The oligomerization products were analysed by SDS-PAGE (Tris.HCl 4-15% w/v gradient polyacrylamide gel) (Fig. 3.49a). A ladder of four bands of bands was observed between the homo-heptamers of AE10 and NN α1 when AE10 and NN were co-translated in the presence of rRBCm suggesting that they had intermediate $M_w$ (*lane 4*, Fig. 3.49a). This was a clear indication that AE10 formed hetero-heptamers with NN α1 in different ratios. Similar experiments as above were performed with EA13 (Fig. 3.49a). Protein bands were observed above the NN α7 level when EA13 and NN α1 were co-expressed in IVTT and could be attributed to the hetero-heptamers of EA13 and NN α1 (*lane 7*, Fig. 3.49a). However, the yield of the oligomer bands of EA13-NN α1 was much lower than the hetero-heptamers of AE10 with NN α1 (*lane 4*, Fig. 3.49a).

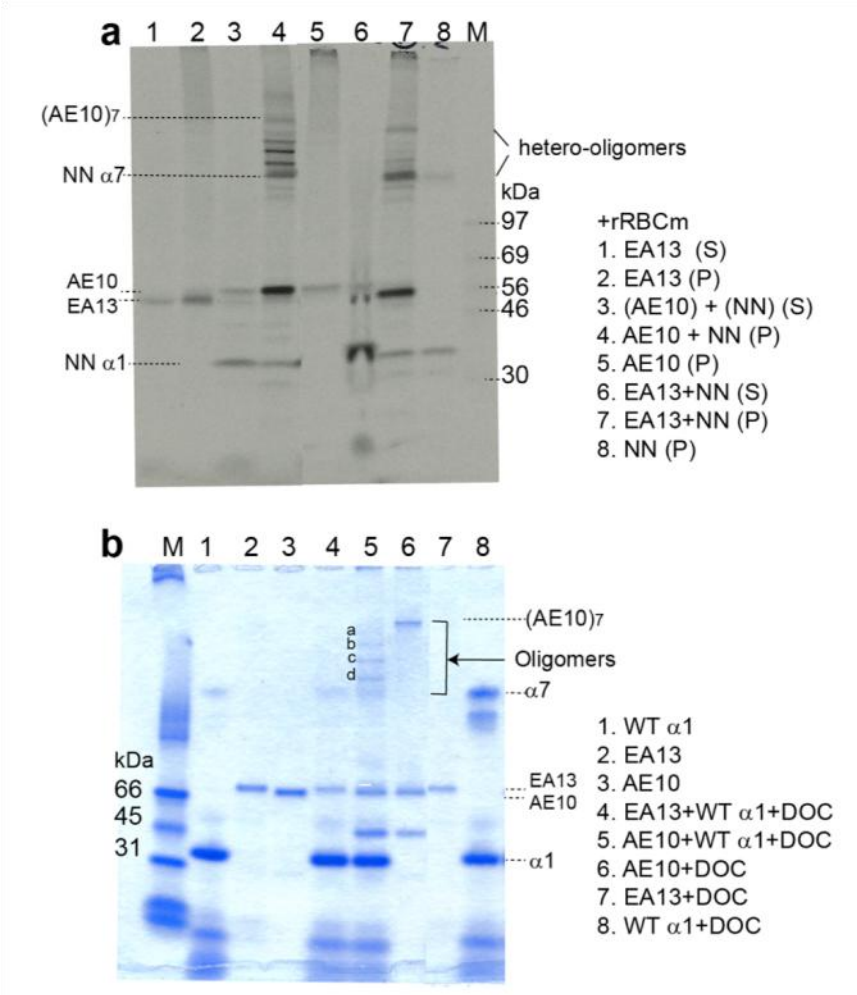



**Figure 3.49. Formation of hetero-heptamers of AE10 and EA13 using rRBCm and detergent.** (**a**) To form hetero-heptamers, AE10 and EA13 proteins were expressed with NN αHL and rRBCm. SDS-PAGE shows the AE10 hetero-oligomers with NN α1 as a ladder of bands between the homo-heptamers of NN α1 and AE10 (*lane 4*). The ladder of bands was not seen as distinctly as in the case of EA13 (*lane 7*). Oligomers bound to the rRBCm were observed in the pellets (P) and not in the supernatant (S). (**b**) Oligomerization could also be induced using a detergent such as DOC. AE10 and EA13 were mixed with WT α1 in a ratio of 4:1 v/v and DOC was added in 10 steps to a final concentration of 10 mM. As controls, AE10, EA13 and α1 were oligomerized separately as well. SDS-PAGE shows homo-heptamers of AE10 ($(AE10)_7$, *lane 6*) and α1 (*lane 8*), and four intermediate bands between $(AE10)_7$ and α7, when mixed together (*lane 5*, labeled a-d). EA13 failed to form homo-heptamers or hetero-heptamers (*lanes 4, 7*). For SDS-PAGE, (**a**) Bis-tris 4-12% (w/v) polyacryalmide and (**b**) Tris.HCl 4-15% (w/v) polyacryalmide gels were used.

The concentration of proteins expressed using the IVTT system (sub-micromolar) is orders of magnitude less than proteins expressed in *E.coli.* Thus, to obtain a higher yield of the hetero-oligomers, proteins expressed in *E.coli*

were used. As indigenous proteins of rRBCm are also detected on Coomassie stained gels, they could co-migrate with the monomer and hetero-heptamer protein bands and interfere with SDS-PAGE analysis (Füssle *et al.*, 1981). Hence, another oligomerization agent suggested by Bhakdi *et al.*, DOC, was used (Bhakdi *et al.*, 1981). WT α1 forms heptamers in presence of DOC concentrations above its c.m.c. of 5 mM.

IMAC purified AE10 (~10 µM) was mixed with WT α1 (~60 µM) in a ratio of 4:1 v/v and DOC was added in 10 equal installments over 30 min to a final concentration of 10 mM. The oligomerization products showed a ladder of bands starting with WT α7 and going up to the homo-heptamer of AE10 ($(AE10)_7$) in SDS-PAGE (*lanes 5* and *6*, Fig. 3.49b). The intermediate bands could be attributed to the AE10-α1 heptamers formed with different number of AE10 subunits. EA13 was also treated with 10 mM DOC in presence and absence of WT α1. Unlike AE10, the homo-heptamer of EA13 and hetero-heptamers of EA13 and WT α1 were not observed on SDS-PAGE (*lanes 4* and *7*, Fig. 3.49b). This suggested that the structure of EA13 monomers was not suited geometrically to form heptamers as discussed before (section 3.8.12.1).

Similar to the results observed in the case of AE10, when AE13 (~15 µM) was mixed with WT α1 (~60 µM) in a ratio of 4:1 (v/v) and oligomerized using 10 mM DOC, a ladder of four hetero-oligomer bands appeared on SDS-PAGE (labeled a-d, *lane 8*, Fig. 3.50). To selectively form a hetero-heptamer with only one λ-exo trimer on top of α7, a 5’ phosphorylated dsDNA substrate was used to form a complex with AE13. In this manner, the λ-exo units of AE13 were forced to

form a trimer. AE13 was pre-incubated with the dsDNA before adding WT α1 and/ or DOC (*lanes 3, 6, 7, 9*, Fig. 3.50). However, there was no observable effect of adding the dsDNA on $(AE13)_7$ and the pattern or yield of the hetero-heptamers of AE13 with WT α1.

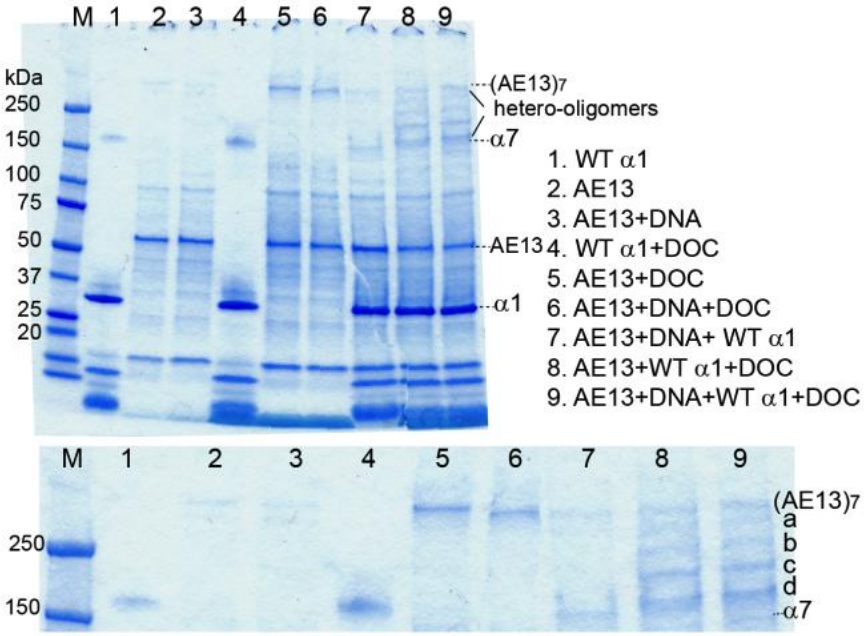



**Figure 3.50. Formation of hetero-heptamers of AE13 with WT α1 using DOC.** AE13 was oligomerized with α1 using DOC to yield hetero-heptamers. The effect of dsDNA on AE13 was also analyzed (*lanes 3, 6, 7, 9*). SDS-PAGE shows that the homo-heptamer of AE13 ($(AE13)_7$) becomes more prominent on treatment with DOC (*lanes 2, 5*). Zoomed inset shows the four bands that are observed between α7 and AE7 when AE13 and WT α1 were mixed in a ratio of 4:1 v/v and 10 mM DOC (*lane 8*). There was no effect of dsDNA on the homo- and hetero-heptamer forming characteristics of AE13 (*lane 9*).

#### 3.8.12.3. Making only one AE10-WT α1 hetero-heptamer selectively

Statistically, altering the molar ratio of AE10 or AE13 to WT α1 could change the relative yields of the different hetero-oligomers. If all possible combinations of AE10-WT α1 and AE13-WT α1 hetero-heptamers were possible, the probability ($P_n$) of *n* AE10 or AE13 subunits in a hetero-heptamer can be given by equation (3.1).

$P_n = C^n \times (1 - C)^{7-n}$, such that, $\sum_{n=0}^{7} P_n = 1$ (3.1)

Where C is the fraction of the concentrations of AE10 or AE13 monomer as compared to WT α1.

This implied that if WT α1 were in excess over the AE10 or AE13 protein, WT α7 would be the major product. The concentration of the next product (band 'd', Figs. 3.49, 3.50) will be the second highest. To confirm this possibility, AE10 (~10 µM) was mixed with WT α1 (~60 µM) in different v/v ratios, 1:3.3, 1:2 and 5:1 and DOC was added to a final concentration of 10 mM. The oligomers were analysed on SDS-PAGE by Coomassie blue staining and later by silver staining. It was observed that in the first two cases, with AE10 and α1 ratios of 1:3.3 and 1:2 (v/v), i.e., where WT α1 was in excess, only one band was observed above the WT α7 band (band *d* in Fig.3.49) (Fig. 3.51). Increasing the amount of AE10 with respect to α1 (5:1 v/v), led to a ladder of four hetero-heptamer bands, only faintly visible in the Coomassie blue stained gel but clearly prominent after silver staining (Fig. 3.51b). These observations followed the trend suggested by the theoretical model described above (equation (3.1)). Interestingly, a multi-ladder was more prominent when AE10 and WT α1 mixed in a ratio of 1:2 and oligomerized using DPhPC liposomes than when DOC was used (data not shown).

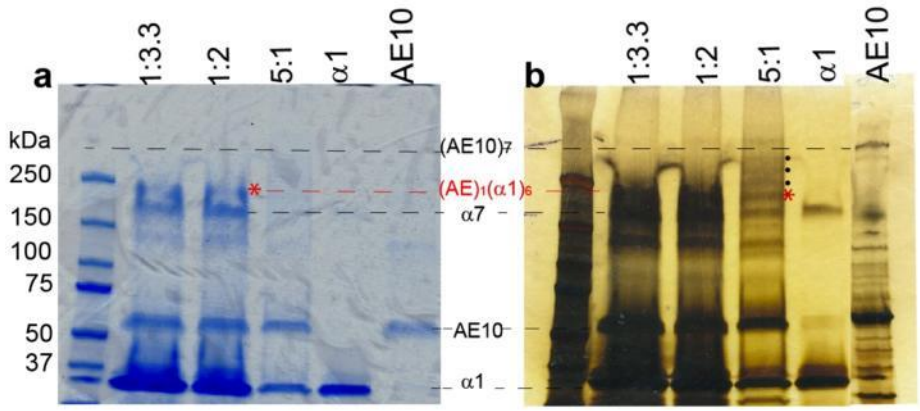



**Figure 3.51. Optimization of hetero-heptamer yields by varying the ratios of α1 to AE10.** Statistically, only one hetero-heptamer of α1 and AE10 or AE13 protein should be formed if an excess of α1 is added. AE10 was mixed with α1 in a ratio of 1:3.3, 1:2, and 5:1 (AE10:WT α1, v/v), oligomerized using DOC and analyzed on SDS-PAGE. (**a**) Coomassie blue stained gel showed only one hetero-heptamer band (marked with a red star and as $(AE)_1(α1)_6$, section 3.8.12.4) above α7 when α1 was in excess, i.e., in the lanes of 1:3.3 and 1:2 (AE10:WT α1, v/v). (**b**) Silver staining of the same gel showed a multiple ladder of hetero-heptamer bands (black dots and red asterisk) only in the *lane* of 5:1 (AE10:WT α1, v/v) where α1 was not in excess.

Hence, altering the ratio of AE10 or AE13 and WT α1 could be a possible method of selectively forming only one hetero-oligomer. If the triplet valency of λ-exo was fulfilled in the hetero-heptamer protein (Fig. 3.51), it could function as the desired λ-exo-αHL protein complex (Fig. 3.47). The number of AE subunits in each of the AE-WT α1 hetero-heptamers was probed from their $M_{app}$ on SDS-PAGE next (section 3.8.12.4).

### 3.8.12.4. Number of AE10 monomers in the AE10-WT α1 hetero-heptamers

Theoretically, six hetero-heptamers of AE10 or AE13 with WT α1 are possible, containing one to six AE10 or AE13 monomer units. If all the possible hetero-heptamers of AE10 or AE13 with WT α1 could form, six bands should have been observed on SDS-PAGE between WT α7 and $(AE10)_7$ or $(AE13)_7$.

However, no more than four hetero-heptamers of AE10 or AE13 monomer and WT α1 were observed (*lane 5,* Fig. 3.49, *lane 8,* Fig. 3.50).

The $M_{app}$ of the WT α7 and $(AE10)_7$ was 190,000 and 390,000, respectively (*lanes 6, 8*, Fig. 3.49b). WT α7 migrates lower than its $M_w$ (Chapter 2, section 2.5). Hence, the estimated $M_{app}$ of the hetero-heptamers of 220,000 (band 'd'), 260,000 (band 'c'), 300,000 (band 'b') and 350,000 (band 'a') may not be a precise reflection of their $M_w$ (*lane 5*, Fig. 3.49b). However, the number of AE10 subunits in the hetero-heptamers could be deduced from the difference in their $M_{app}$ and that of WT α7. Thus, AE10-WT α1 hetero-heptamers were attributed to $AE_1WT_6$ (band 'd'), $AE_3WT_4$ (band 'c'), $AE_4WT_3$ (band 'b') and $AE_6WT_1$ (band 'a') (*lane 5*, Fig. 3.49b) with 1, 3, 4 and 6 AE10 monomer subunits, respectively.

Possibly, in $AE_3WT_4$ and $AE_6WT_1$, the λ-exo units of AE10 could be forming trimers on the α7 cap similar to Fig. 3.40a, c. The formation of $AE_1WT_6$ and $AE_4WT_3$ is difficult to explain because the triplet valency of the λ-exo units is not fulfilled. A plausible explanation would be that two truncated AE10 protein subunits consisting of λ-exo (section 3.8.7) associate with the full length AE10 monomers in $AE_1WT_6$ and $AE_4WT_3$ to form λ-exo trimers. The truncated AE10 subunits (λ-exo) would dissociate from the hetero-heptamers on SDS-PAGE (section 3.8.7). This explanation however, does not explain why hetero-heptamers with other combinations of AE10 or AE13 and WT α1, viz. $AE_2WT_5$ and $AE_5WT_2$, do not form. The hetero-heptamers could also be associated to one another through their λ-exo units in solution, i.e., $AE_1WT_6$, $AE_4WT_3$ and $(AE10)_7$ were linked to each other such that the λ-exo components of all the

AE10 subunits formed trimers. This supramolecular complex of hetero-heptamers and $(AE10)_7$ would break apart on SDS-PAGE leading to three discreet bands of $AE_1WT_6$, $AE_4WT_3$ and $(AE10)_7$. Formation of AE13-WT α1 hetero-heptamers could also be explained by the same analogy.

### 3.8.13. Purification strategies for AE10 and AE13 hetero-heptamers

On mixing AE10 or AE13 with WT α1 with an oligomerization agent, protein bands of AE10 or AE13 monomers and WT α1 which had not oligomerized, and the homo- and hetero-heptamers of AE10 or AE13 and WT α1 were observed on SDS-PAGE (Figs. 3.49-3.51). The homo-oligomers of AE10 and AE13 could be eliminated if WT α1 was added in excess over AE10 and AE13 proteins (section 3.8.12.3). Thus, to characterize the hetero-oligomers of AE10 or AE13 with WT α1, they needed to be separated from WT α1, WT α7 and AE10 or AE13 trimers.

#### i. Gel extraction from SDS-PAGE: activity of AE10 in SDS

In conventional SDS-PAGE, the SDS in the sample buffer binds to proteins and denatures them (Laemmli, 1970). As α7 is SDS-stable, hetero-heptamers can be separated on SDS-PAGE and extracted directly from SDS containing preparative polyacrylamide gels, eliminating the need of any other purification step (Cheley *et al.*, 1999). However, this method could not be directly adopted for the purification of AE10-WT α1 and AE13-WT α1 hetero-oligomers because the λ-exo part of AE10 and AE13 denatures in SDS (Subramanian *et al.*, 2003). AE10-WT α1 and AE13-WT α1 hetero-heptamers could be gel-extracted purified from low-SDS containing gels such that the exonuclease activity of

AE10 and AE13 was not lost. To estimate the maximum amount of SDS that could be used in the preparative polyacrylamide gels for gel extraction purification of the AE10-WT α1 and AE13-WT α1 hetero-oligomers, the activity of AE10 in varying amounts of SDS was measured.

In a preliminary experiment, SDS was added to the reaction mix containing exonuclease activity buffer (67 mM Glycine-KOH, 2.5 mM $MgCl_2$, pH 9.4), PicoGreen and the dsDNA substrate. The addition of SDS to a final concentration of 0.05% (w/v) led to a decrease in the dsDNA bound PicoGreen fluorescence by >80%. Hence, the AE10 protein was pretreated with 0.01 to 0.10% (w/v) SDS before addition to 1 ml of the reaction mix, such that the final concentration of SDS in the reaction mix was <0.001%.

From the slopes of the fluorescence plots of dsDNA degradation, it was determined that AE10 showed only ~60% exonuclease activity in 0.01% SDS and lost all activity in SDS ≥ 0.02% (Fig. 3.52). This experiment also implied that just ~100x dilution of SDS (0.02% to 0.0002% (w/v)) from the SDS-treated AE10 sample in the reaction buffer was not enough to renature the protein.

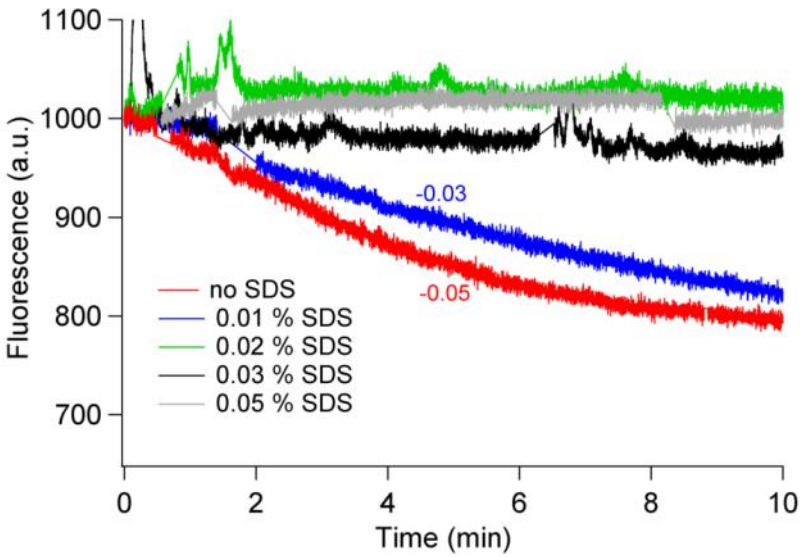



**Figure 3.52. Activity of AE10 in SDS.** To study the effect of SDS on the AE proteins, AE10 was treated in 1:1 v/v with 0.02-0.20% (w/v) SDS. Activity assays of AE10 pretreated with SDS showed that the exonuclease activity was lost beyond 0.01% (w/v) SDS. The numbers written next to the fluorescence traces report the slope of the linear region of the fluorescence decay.

The oligomer species of AE10 failed to migrate efficiently on SDS-PAGE with running buffer (25 mM Tris.HCl, 192 mM Glycine, 0.01% (w/v) SDS, pH 8.8) and the polyacrylamide gel containing 0.01% (w/v) SDS (data not shown). Hence, methods to renature AE10 and AE13 proteins after SDS treatment were sought.

Mild non-ionic detergents such as Triton X-100 sequester SDS molecules in micelles where upon SDS denatured protein samples can refold (Ossipow *et al.*, 1993). Using this method, gel extracted DNA binding proteins from nuclear and liver extracts were shown to renature and regain activity (Ossipow *et al.*, 1993). To test whether removal of SDS from AE10 could be used to renature the protein, 1% (v/v) Triton X-100 was added to denatured AE10 samples in 0.05% (w/v) SDS, and incubated for different amounts of time before measuring the exonuclease activity. It was found that SDS denatured AE10 did not regain

activity even after Triton X-100 treatment for 2 h at 25°C and a further 24 h at 4°C (Fig. 3.53). Hence, it was concluded that Triton X-100 could not be used to renature the hetero-heptamer of AE10 extracted from a SDS containing gel.

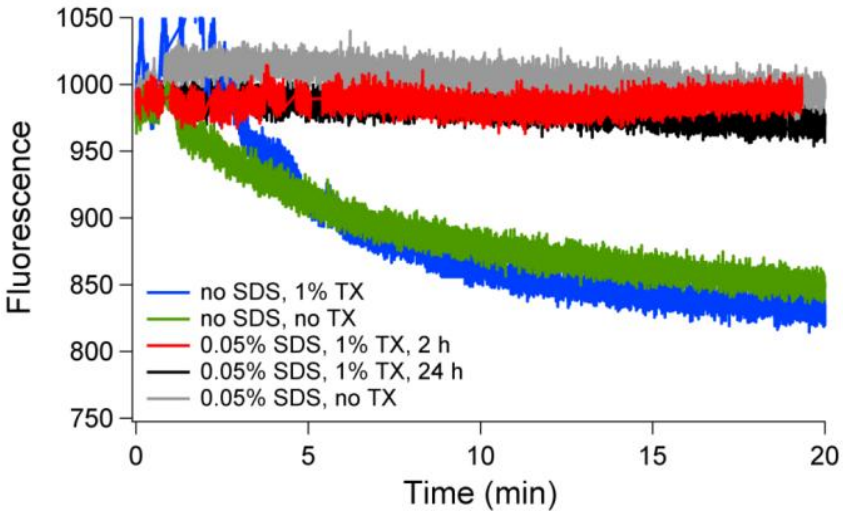



**Figure 3.53. Activity of AE10 after denaturation in SDS and subsequent treatment with Triton X-100.** Triton X-100 (TX) can be used to remove SDS from gel extracted protein samples, upon which the proteins can refold (Ossipow *et al.*, 1993). To check the feasibility of extracting denatured AE10 hetero-oligomers from SDS containing polyacrylamide gels, AE10 was first treated with 0.05% (w/v) SDS and then with 1% (v/v) TX for 2 h at 25°C (2 h) and for a further 24 h at 4°C. AE10 which had not been treated with SDS was active in 1% (v/v) TX. However, SDS treated AE10 did not regain exonuclease activity upon TX treatment.

### ii. IMAC and SEC purification

AE10 or AE13-α1 hetero-oligomers could be separated from α7 utilizing the DNA binding capacity of AE10 or AE13 trimers by using heparin columns or the His-6 tag on AE10 or AE13 proteins using IMAC. Another purification step such as SEC would be required to separate AE10-α1 hetero-oligomer and AE10 trimer. But due to the low concentration of AE-α1 hetero-heptamers, SEC was not performed. Work on increasing the yield and subsequent purification of the hetero-oligomers is currently underway.

### 3.9. Construction of a fusion trimer of λ-exo with concatenated monomers

In the AE10 and AE13 constructs, one λ-exo monomer was fused to α1. These constructs could be mixed with WT α1 to assemble into hetero-heptamers with different number of λ-exo monomers on top of α7 (section 3.8.12.4). However, the final yield of hetero-heptamers was very low and purification not so straightforward (section 3.8.13). To attach an active λ-exo molecule to α7, another strategy is being pursued. Three monomer subunits of λ-exo have been genetically linked with sufficiently long flexible linkers such that it is expressed as a monomer protein and retains exonuclease activity (exo3) (Chapter 4). The active exo3 variant will be genetically fused to α1 (exo3-α1). Hetero-heptamers of exo3-α1 and WT α1, with one subunit of exo3-α1 and six WT α1 subunits could then be selectively made by adding an excess of WT α1 (section 3.8.12.3). The hetero-heptamer, $(\text{exo3-}\alpha1)_1(\alpha1)_6$ is the desired complex of λ-exo and αHL.

## Conclusions and outlook

The successful construction of the αHL heptamer dimer presented an encouraging prospect of attaching other proteins to αHL. Attaching proteins to nanopores is still an unexplored field of study. Engineered nanopores with DNA processing enzymes would have direct applications in the field of nanopore based DNA sequencing (Cherf *et al.*, 2012; Deamer, 2010) and to measure the activity of single enzyme molecules (Benner *et al.*, 2007; Dahl *et al.*, 2012; Hornblower *et al.*, 2007). This thesis has explored the possibility of attaching a

functional λ-exo to the cap of the protein nanopore, αHL.

Several methods have been explored in order to link λ-exo to αHL ranging from covalent cross-linking to genetic fusion. Active cysteine mutants of αHL and λ-exo were made and screened for activity. However, the native cysteines of λ-exo were also found to be reactive to small thiol-reactive molecules. The orientation of the proteins could also be not controlled with thiol-specific cross-linking. Hence, genetic fusion constructs of λ-exo and αHL were made in different combinations. It was found that αHL could be attached to λ-exo to both its C and N-termini without loss of exonuclease activity. However, hemolytic activity was lost when λ-exo was attached to the N-terminus of αHL, perhaps due to steric hindrance of the linkers in the mouth of the αHL pore. This study demonstrates that other proteins could be linked to the C-terminus of αHL by a suitably long linker without affecting the activity of αHL. It also demonstrates that λ-exo can be modified at both termini, which is an important piece of information for other functionalization studies.

By linking the monomers of αHL and λ-exo, we have synthesized a new kind of processive exonuclease that can also form pores upon encountering bilayers. Lipid demarcated aqueous vesicles embedded with α7 pores have been used for studying single molecule activity of biomolecules entrapped inside them (Okumus *et al.*, 2009). The α7 pores allow small molecules such as enzyme co-factors/ analytes to diffuse from the external buffer into the nano-reactor chamber of the vesicles. Similarly, AE10 or AE13 pores could be used to feed-in nucleotides at a constant rate inside vesicles if dsDNA were added to the

bulk solution (Fig. 3.54). Long ssDNA, which cannot be introduced inside the vesicles through nanopores due to their high secondary structure, could be delivered in a continuous fashion across the vesicle bilayer using AE10 and AE13 pores.

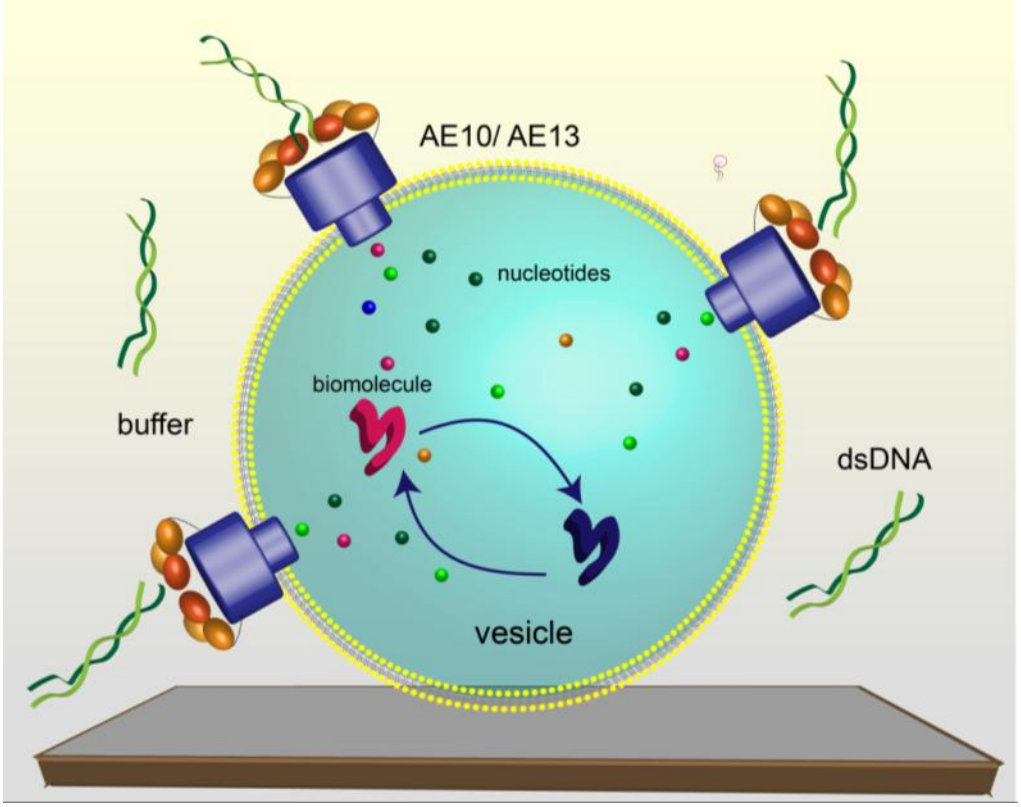



**Figure 3.54. Application of αHL-λ-exo monomer fusion constructs.** The AE10 and AE13 constructs have exonuclease activity and can also form stable pores in bilayers. Conceptual cartoon (not to scale) shows the use of AE10/ AE13 constructs to feed in nucleotides and ssDNA (if a potential is applied) at a constant rate across the bilayer of an immobilized vesicle. Such a set-up could be used to study reactions of entrapped biomolecules at the single molecule level (Okumus *et al.*, 2009).

However, for DNA sequencing or studying the single molecule activity of λ-exo, only a single λ-exo molecule needs to be linked to αHL. Assembly of multimeric proteins with different oligomer stoichiometries, a trimer and a heptamer in the case of λ-exo and αHL, into a bimolecular complex has not been shown previously. This problem has been tackled by constructing hetero-oligomers of αHL-λ-exo fusion constructs with WT α1. However, the hetero-heptamers could

not be successfully purified from a mixture of WT α1, WT α7 and the fusion construct trimers. Methods to increase the yield and to separate the hetero-heptamers still need to be investigated.

A fusion trimer construct of λ-exo is also being prepared where three monomers of λ-exo have been genetically fused spaced with flexible linkers. The fusion trimer will also be genetically fused to αHL. The hetero-heptamer with one unit of the fusion trimer-αHL construct and six WT α1 units will have only one λ-exo trimer molecule on the αHL cap, which is our desired protein complex.

Once the λ-exo-αHL pore complex is isolated, electrical recordings of the protein complex would be done to study the activity of single λ-exo molecules. The pore through λ-exo could be modified to study its effect on DNA binding and processivity electrically. The interactions of DNA with the central pore could also be scrutinized in much finer detail. As the λ-exo cleaves the DNA base-by-base into the nanopore, the sequence dependent pausing behaviour of λ-exo (Perkins *et al.*, 2003) could be studied as well.

αHL-λ-exo constructs with λ-exo mutants that cleave dsDNA at a lower speed (Zhang et al., 2011) could be used to slow down the translocation speed of ssDNA through the αHL pore; a crucial prerequisite for DNA sequencing through nanopores (Bayley, 2006). The αHL-λ-exo pore complex would thus function as a DNA sequencing molecular machine with a tunable speed.

# Methods

## 3.10. Mutagenesis of αHL and λ-exo

### 3.10.1. Point mutagenesis of αHL and λ-exo

All the single residue mutants (point mutants) of αHL and λ-exo mentioned in this thesis were prepared by polymerase chain reaction (PCR)-based site-directed mutagenesis and subsequent homologous recombination in *E.coli.* Both αHL and λ-exo genes were cloned between NdeI and HindIII restriction sites in a pT7 vector (Walker *et al.*, 1992b), which also encoded for β-lactamase protein which confers ampicillin resistance (Amp-r gene). To introduce point mutations in the αHL and λ-exo genes, the plasmid was amplified between Amp-r gene using primers amp F and amp R (Appendix Table 1) and the region to be mutated using primers containing the desired mutation as listed in Appendix Table 1. For each mutagenesis reaction, two PCRs were set-up, 'forward' (f-PCR) and 'reverse' (r-PCR) where the primers containing the desired mutation bound to anti-sense (F primer) and sense strands (R primer), respectively. The two halves of the plasmid generated in the f- and r-PCR, had homologous ends, which could be recombined in *E.coli*.

PCRs (50 µl) were carried out with 10 ng of template plasmid digested with NdeI (f-PCR) or HindIII (r-PCR), 1 µM of primers (F primer and amp-47 for f-PCR; R primer and amp-46 for r-PCR) and 2x Phusion® Flash High-Fidelity PCR Master Mix (Finnzymes) (Table 3.2). The PCR protocol involved an initial denaturation step at 98°C for 30 s, followed by 25-30 cycles of denaturation at 98°C for 30 s, annealing at 62-70°C (~$T_m$ of primers + 3°C) for 30 s and extension at 72°C for 50 s. At the end of the cycles, a final extension step was

performed at 72°C for 2 min.

**Table 3.2.** Table shows the components of f- and r-PCR.

| Reagent | f-PCR | r-PCR |
|---|---|---|
| DNA | pT7-gene/ NdeI | pT7-gene/ HindIII |
| Primer with mutation | F primer | R primer |
| Amp-primer | amp R | amp F |

To circularize the two plasmid fragments amplified in the f- and r-PCRs by homologous recombination, 5 µl of each PCR product was transformed in 150 µl of XL-10 gold ultra-competent cells (Agilent). The transformed XL-10 cells were plated on LB-agar plates containing ampicillin (100 µg $ml^{-1}$) at 37°C for 16 h. The plasmid sequences of 3-4 colonies were screened for the mutant gene. LB cultures (10 mL) containing ampicillin (100 µg $ml^{-1}$) (LB-amp) were inoculated with single colonies and grown for 16 h at 37°C. The plasmids were extracted using Qiagen Spin Miniprep kits and sequenced using the universal T7F and T7R primers (Appendix Table 1) (Source BioScience Lifesciences, Oxford). Plasmids containing the desired mutation were extracted and purified from 100 ml LB-amp cultures using the Qiagen Plasmid Maxi kit and stock solutions of the plasmids (400 ng $µl^{-1}$) were stored at -20°C.

The αHLΔC5 mutant of αHL was made by introducing a stop codon after residue 288. The exo+C5 mutant of λ-exo was made by inserting the nucleotide (nt) sequence of the 5 C-terminal amino acids of WT αHL between residue 226 (C-terminus of WT λ-exo) and the stop codon (Appendix Table 1).

### 3.10.2. Construction of αHL-λ-exo fusion constructs.

A double barrel mutant (E111N and K147N) of αHL (Ellina Mikhailova), termed NN αHL, was used for the constructing the genetic fusion constructs. The genes of NN αHL (A) and λ-exo (E) were arranged in tandem spaced by a flexible 10 or 13 residue linker within a pT7 vector.

In one set of constructs, the λ-exo gene (gene 1) preceded the NN αHL gene (gene 2) and was termed EA10 and EA13, where the number denotes the number of the amino acids in the linker between the two genes. In another set of constructs, termed AE10 and AE13, the order of genes was reversed, i.e., the N-terminus of λ-exo (gene 2) was linked to the C-terminus of NN αHL (gene 1) (Table 3.3). The nt sequences of the linkers were SG10: GGAGGCTCGAGTGGAGGTAGTGGTGGATCA and SG13: GGAGGCTCGAGTGGAGGTAGTGGTGGATCAGGCTCCTCA.

**Table 3.3.** Table shows the order of genes in the various αHL-λ-exo fusion constructs. The C-terminus of gene 1 and the N-terminus of gene 2 were extended with SG10 and SG13 in two PCRs.

| Construct | Gene 1 | Gene 2 |
|---|---|---|
| AE10/ AE13 | NN αHL | λ-exo |
| EA10/ EA13 | λ-exo | NN αHL |
| EAA10 | EA10 | NN αHL |

Primers were designed to extend the C-terminus of gene 1 and the N-terminus of gene 2 by SG10 and SG13 (Appendix table 1). Plasmids containing gene 1 and gene 2 were linearized with HindIII and NdeI, respectively. 10 ng of linearized plasmid was used for carrying out PCR (50 µl) supplemented with 1

μM of primers and 2x Phusion® Flash High-Fidelity PCR Master Mix (Finnzymes) As the primers for the SG linkers had high GC% and hence a high $T_m$, a touchdown PCR protocol was used to amplify the plasmids between the amp-r gene and genes 1 and 2 with the SG linker overhang (Table 3.4). The two PCR products (5 μl) were then mixed in XL-10 Gold cells for homologous recombination into AE10, AE13, EA10, EA13 or EAA10 (Fig. 3.55). The sequence of each construct was verified by sequencing between the T7 promoter and T7 terminator regions.

**Table 3.4.** Table shows the PCR temperature protocol used to make the αHL-λ-exo fusion constructs.

| Temperature ($^oC$) | Time (s) | Cycles |
|---|---|---|
| 98 | 30 | |
| 98 | 30 | |
| 85-60 | 30 | 25 |
| 72 | 50 | |
| 98 | 30 | |
| 60 | 30 | 30 |
| 72 | 50 | |
| 72 | 2 min | |

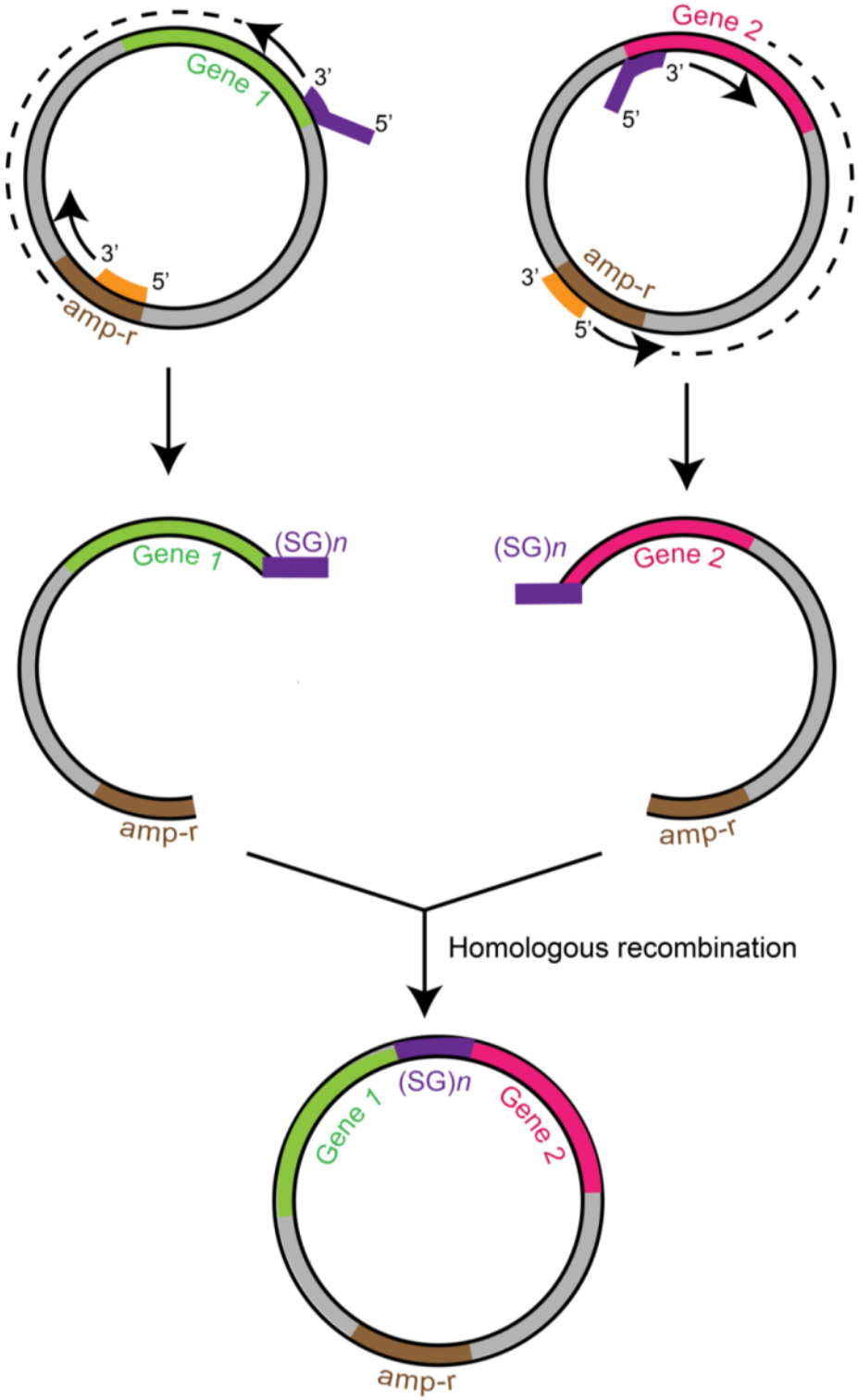



**Figure 3.55. PCR strategy for the construction of αHL–λ-exo fusion constructs.** To link αHL and λ-exo genes, linkers that encoded for 10 (SG10) and 13 (SG13) amino acids were used. These linkers were added to the N- and C-termini of gene 1 and gene 2 in two PCRs, where, gene 1=NN αHL and gene 2=λ-exo (AE constructs); gene 1=λ-exo and gene 2=NN αHL (EA constructs); and gene 1=EA10 and gene 2=NN αHL (EAA10 construct). The PCR products of the two PCR reactions were subsequently mixed for homologous recombination in *E.coli.*

## 3.11. Expression and purification of αHL–λ-exo fusion constructs

### 3.11.1. *In vitro* expression

DNA can be transcribed and translated into proteins *in vitro* (IVTT) using the isolated transcription-translation machinery of *E.coli* and rabbit reticulocyte lysate. Proteins were expressed *in vitro* using the *E.coli* T7 S30 extract system for circular DNA (Promega). For a 50 µl reaction, 3.2 µg of DNA was mixed with an amino acid mixture minus methionine (100 µM final concentration) and 2 µl of radiolabeled $^{35}$S methionine (MP Biomedicals, 10 mCi ml$^{-1}$). 20 µl of S30 premix and 6 µl of S30 extract which contained 1 µl of 500 µg ml$^{-1}$ rifampicin in 150 µl of extract, are added to the above mixture and incubated at 37°C or other temperatures as specified for 1 h.

TNT quick-coupled IVTT system, which utilizes rabbit reticulocyte lysate was also used to express the αHL-λ-exo fusion constructs to assess if soluble protein expression could be increased as compared to T7 S30 IVTT. 50 µl reactions were carried out with 3.2 µg of DNA, 2 µl of radiolabeled $^{35}$S methionine (10 mCi ml$^{-1}$) and water was added to adjust the volume to 25 µl. To the above mixture, 25 µl of 2x TNT master mix was added and incubated for 1 h at 25°C or 37°C.

To analyze the amount of soluble and insoluble protein expression, the IVTT reactions were spun at 25,000 g for 10 min at 4°C. The products were mixed with 4x sample buffer (Bio-Rad) and analyzed on SDS-PAGE.

### 3.11.2. *In vivo* expression

All the WT and mutants proteins of αHL and λ-exo, and the αHL-λ-exo genetic fusion constructs were expressed in Rosetta DE3 pLysS strain of *E. coli* (Novagen), a derivative of BL21 DE3 cells. As compared to BL21 DE3 cells, Rosetta DE3 pLysS cells can also encode for six rare codons. The DE3 lysogen encodes for T7 RNA polymerase under a lac promoter. Rosetta (DE3) pLysS cells contain a pLysS plasmid that encodes for T7 lysozyme in low quantities sufficient to inhibit any basal activities of T7 RNA polymerase. Upon induction with IPTG, T7 RNA polymerase is expressed in much higher levels, enough to overcome the inhibition by T7 lysozyme. As the T7 lysozyme gene is under a weaker promoter than the T7 promoter of the target protein, upon induction, almost all mRNA transcription and hence protein production is directed towards the target protein.

Rosetta (DE3) pLysS (150 µl) were transformed with plasmid DNA (400 ng). After incubation on ice for 30 min, the cells were subjected to heat shock (10 s at 42°C) and incubated further on ice for 5 min. The cells were plated on warm LB agar plates containing 100 µg $ml^{-1}$ ampicillin and incubated overnight at 37°C. For protein expression, a single colony was inoculated in self-inducing media (MagicMedia, Invitrogen) (100 ml) and grown for 7 h at 30°C. The temperature was decreased to 20°C and the cultures were grown further for an additional 20-22 h (dual temperature protocol, section 3.8.3). In the single temperature protocol (section 3.8.3), the expression culture was inoculated with a single colony and grown at 20°C for 24 h. The cell culture was spun down at 13,000 g for 30 min and the supernatant was discarded. The cell pellets were

lysed with BugBuster master mix (10 ml) (Merck) and incubated on ice for 30 min. The cell debris was removed by centrifugation at 13,000 g for 30 min.

The lysate supernatants from Rosetta cells were loaded on a TALON metal affinity resin (2 ml bed volume (b/v)) (Clontech). The TALON metal affinity resin, charged with $Co^{2+}$ ions has a higher specificity for His-tagged proteins than $Ni^{2+}$-NTA resin, resulting in higher purity of the purified proteins. The TALON metal affinity column was pre-washed with deionized water and equilibrated with 25 mM HEPES, 500 mM NaCl, 30 mM imidazole, pH 8.0. The columns were placed on a shaker and incubated for 1 h at 4°C. The columns were then washed 3 times with 10 ml 25 mM HEPES, 500 mM NaCl, 30 mM imidazole, pH 8.0 (wash buffer). The proteins were eluted with 25 mM HEPES, 500 mM NaCl, 250 mM imidazole, pH 8.0 in five 2.5 ml fractions.

Another *E. coli* strain, NiCo21 (DE3) (NEB) was also used for the expression of αHL-λ-exo fusion constructs. A few of the most common *E.coli* proteins that can bind to IMAC columns in an unspecific manner and elute with the protein of interest have been removed or modified in NiCo21 (DE3) cells (Bolanos-Garcia and Davies, 2006; Robichon *et al.*, 2011). SlyD, ArnA and Can are tagged to enable removal by chitin affinity chromatography and GlmS is mutated to eliminate binding to IMAC resins. This results in higher purity of the recombinant His-tagged target protein after IMAC.

The protocols followed for DNA transformation and protein expression in NiCo21 (DE3) cells was similar to that used for Rosetta (DE3) PLyS cells

except for a few additional steps. After transforming the DNA, the NiCo21 (DE3) cells were outgrown in SOC media (950 µl) for 1 h at 37°C of which 50 µl was plated on the LB-agar/amp plates. For protein purification, the cells were grown under two different temperature conditions (dual temperature and 20°C protocol, section 3.8.3). After lysing the cells with 10x BugBuster buffer, the supernatant of the cell lysate of NiCo21 (DE3) cells was loaded on a chitin-binding column (2 ml) equilibrated in wash buffer. The columns were placed on a shaker at 4°C and incubated for 1 h. The flow through was loaded on TALON affinity columns (2 ml b/v), incubated at 4°C for 1 h, and purified as mentioned above.

The lysate pellets of Rosetta (DE3) PLyS and NiCo21 (DE3) cells were resuspended in 8 M urea (1 ml) and used for estimating the amount of protein expressed as inclusion bodies. The purification fractions and the lysate pellets were analyzed using precast Any kD gels (Bio-Rad) in SDS-PAGE. The eluate fractions containing the protein were pooled and concentrated ~10x using micro-spin centrifugation columns (Microcon) with a 10 kDa cut-off for λ-exo mutants or a 50 kDa cut-off in the case of αHL-λ-exo fusion constructs. The concentrated λ-exo mutants and αHL-λ-exo fusion constructs were mixed with an equal volume of cold glycerol and stored at -20°C.

The concentration of the proteins were estimated from their UV absorbance using equation (3.2).

$$A = \varepsilon l C \tag{3.2}$$

where, *A* is the absorbance, ***ε*** is the molar absorption coefficient, *l* is the path

length and *C* is the concentration of the protein. The values *ε* for the αHL-λ-exo fusion constructs and the λ-exo mutants was calculated using the online Expasy ProtParam tool to be 111,270 $M^{-1}$ $cm^{-1}$ and 46,410 $M^{-1}$ $cm^{-1}$, respectively at 280 nm.

### 3.12. Estimation of $M_{app}$

The $M_{app}$ of the protein bands in SDS-PAGE were estimated using the Quantity One® 1-D Analysis Software (Bio-Rad). The software uses the migration profile of known $M_w$ standards on the same gel to calculate the $M_{app}$ of the protein bands.

### 3.13. Determination of number of active cysteines using PEG-OPSS

To determine the number of reactive cysteines in the λ-exo cysteine mutants, a monomethoxy poly(ethylene glycol)-*o*-pyridyl disulfide (PEG-OPSS) reagent ($M_w$≈5 kDa) (Creative PEGworks) was used along with the λ-exo R226C (RC) mutant.

The buffer of RC protein (125 µl) expressed in *E.coli* was exchanged with another buffer containing 200 mM NaCl, 20 mM Sodium phosphate, 500 µM EDTA and 1 mM DTT (pH 7.5), using a desalting P6 column (Bio-Rad). 10 µl of 100 mM 5 kDa PEG o-pyridyl disulfide (PEG-OPSS) reagent was added to 90 µl of the buffer exchanged RC protein at 25°C (t=0). Aliquots of the reaction mixture were taken at 15 min, 30 min and 45 min and stored on ice. The unreacted PEG-OPSS was removed from each aliquot using P6 columns (Bio-Rad) at 4°C. The products of the PEG-OPSS reaction were analyzed on a 10% bis-tris gel using SDS-PAGE.

## 3.14. Activity assays

### 3.14.1. Hemolytic activity assay

αHL monomers can assemble as heptamers on red blood cells and cause lysis, which leads to a decrease in the light scattering (Hildebrand *et al.*, 1991). Hemolytic activity assays were performed using rabbit blood and a microplate reader for measuring optical density (O.D.).

450 µl of fresh rabbit blood was washed 3-4 times with 900 µl buffer containing 10 mM MOPS, 150 mM NaCl, 0.1% BSA (w/v), pH 7.4 (MBSA) until the supernatant was clear. The washed blood was diluted 20X in MBSA. The soluble fraction of proteins expressed in the T7 S30 IVTT system (unless specified otherwise) was used for the hemolytic assays. The protein (2 µl typically of 25 µl IVTT reaction) was diluted in MBSA to a final volume of 100 µl in the first well of a microtitre plate. The protein was then serially diluted twofold in MBSA in each well of that plate row, such that the final volume of each well was 50 µl. To each well, 50 µl of the washed and diluted blood sample was added, beginning with the well containing the most diluted protein. Hemolysis was monitored by measuring the O.D. of the blood in each well, every 15-30 s over 2 h at 25°C, using a microplate reader.

### 3.14.2. Exonuclease activity assay

λ-exo activity on dsDNA was studied using a fluorescence-based assay described by Tolun *et al.* (section 3.3.4) (Tolun and Myers, 2003). The DNA substrate for exo activity assays was a 1 kb DNA with phosphorylated 5’ ends. It was prepared by amplifying a 1 kb region in NN αHL plasmid between NdeI and

HindIII sites using two phosphorylated primers (E_act_f and E_act_R) in PCR (Table 3.5). The dsDNA substrate was purified using a PCR purification kit (Qiagen) to remove the primers, nucleotides and enzymes in the PCR master mix to a final concentration of 70 ng $\mu l^{-1}$, i.e., ~120 nM.

**Table 3.5.** PCR protocol used preparing the 1 kb dsDNA substrate for exonuclease activity assays

| Temperature ($^{o}$C) | Time (s) | Cycles |
|---|---|---|
| 98 | 1 min | |
| 98 | 30 | |
| 68 | 30 | 25 |
| 72 | 50 | |
| 72 | 2 min | |

The emission spectrum of the dsDNA binding dye, PicoGreen (Invitrogen), was measured with different excitation wavelengths ($\lambda_{ex}$). The excitation of PicoGreen was optimum at 485 nm where it fluoresced most strongly at 522 nm ($\lambda_{em}$). $\lambda_{ex}$ and $\lambda_{em}$ were fixed at 485 nm and 520 nm, respectively for all activity assay experiments.

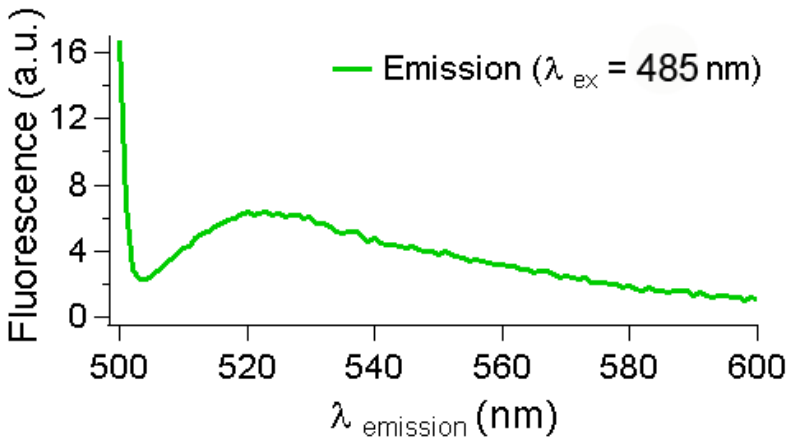



**Figure 3.56. Emission of PicoGreen.** Trace shows emission spectrum, fluorescence measured in absorbance units (a.u.), of the dsDNA binding dye, PicoGreen when excited at 485 nm ($\lambda_{ex}$).

In a 1 ml quartz cuvette with transparent sides, 100 µl of 10x λ-exo activity buffer (670 mM Glycine-KOH, 25 mM $MgCl_2$, pH 9.4) (NEB) was added to 900 µl of water. 0.5 µl of PicoGreen (Invitrogen) was added, followed by mixing 1–2 µl of the DNA substrate (240 nM phosphorylated ends). The activity mix was incubated in the dark for 5 min at 25°C. 1–2 µl of the protein (2-20 µM) was subsequently added and the fluorescence output was measured over time at 25°C, until it neared a constant value. The starting fluorescence value of the dsDNA bound PicoGreen was normalized to 1000 a.u. in each activity assay experiment. Slopes of the linear portion of the fluorescence decay curves were used to compare protein activities.

### 3.14.3. Exonuclease activity measurements in presence of SDS

Solutions of SDS with concentrations ranging from 0.02-0.20% (w/v) were prepared in water. AE10 was mixed in 1:1 v/v with 0.02 to 0.2% w/v SDS to achieve the final concentrations of SDS of 0.01, 0.02, 0.03, 0.05 and 0.10% (w/v) and incubated for 5 min at 25°C. The AE10-SDS mixture was added in aliquots of 2 µl while measuring the activity in the real-time fluorescence assay.

## 3.15. Electrical measurements in planar lipid bilayers

Electrical recordings of the αHL-λ-exo fusion constructs were done using planar lipid bilayers as described in Chapter 2, section 2.10. The electrical recordings were done using a Delrin chamber that had two buffer reservoirs separated by a 25 µm thick polytetrafluoroethylene film (PTFE, Teflon) (Goodfellow) which contained a small aperture (~125 µm). The aperture was pretreated with

hexadecane (10% v/v) in pentane solution on each side and the reservoirs were filled with 1 ml of recording buffer (100 mM KCl, 25 mM Tris, 50 µM EDTA, pH 7.8). ~20 µl of 1, 2-diphytanoyl phosphatidylcholine (DPhPC) (Avanti Polar Lipids) was added to both the reservoirs. A vertical planar bilayer was formed on the aperture by sequentially raising the buffer on each side. Ag/AgCl electrodes were used to apply a biased potential. The chamber with the active Ag/AgCl electrode was been conventionally termed '*trans*' and grounded side was labeled '*cis*'. 1–2 µl of the protein was added to the *cis* side. The current was amplified with a patch-clamp amplifier (Axopatch 200B, Axon Instruments) and digitized with a Digidata 1320 A/D convertor (Axon Instruments). Electrical data was filtered at 1 kHz and acquired at a sampling frequency of 5 kHz. For electrical recordings in asymmetric salt conditions, after the formation of a stable bilayer, 50 µl of the buffer in the *trans* chamber was exchanged with 2 M KCl and the current offset at 0 mV was adjusted to 0 pA again.

### 3.16. Western blot

To determine the nature of the truncated AE10 protein that could not be separated from full length AE10, the presence of a His-6 tag on its λ-exo part only, was utilized. A western blot was performed on the SEC purified AE10 fraction using an anti-his tag antibody.

The peak elution fraction of AE10 (AE10f10) in SEC and a sample of WT α1 treated with DOC were run on a 4–15% gradient gel. The proteins were transferred from the polyacrylamide gel onto a nitrocellulose membrane using the iBlot device (Invitrogen) (program P3 for 7 minutes). The protein containing

membrane was then washed with 50 mL of 10 mM Tris-HCl, 150 mM NaCl, 0.05% Tween (TBST) solution containing 2.5 g BSA for 1 h at 25°C in a covered container on a gel rocker. The membrane was further washed thrice with 20 ml of TBST. The membrane was then treated with 16 ml of monoclonal anti-polyhistidine-alkaline phosphatase antibody produced in mouse and a solution containing 2.5 g BSA in 50 ml TBST for 2 h at 25°C. The excess BSA was washed away by rinsing the membrane three times with 20 ml TBST. The protein bands were then visualized by adding ~10 µl BCIP/NBT Alkaline Phosphatase substrate (Sigma) drop by drop. The bands of proteins containing His-6 tags turned violet. The reaction was quenched by adding an excess of deionized water and thence air-dried.

### 3.17. Size exclusion chromatography (SEC)

To separate AE10 from its truncated protein and to determine its oligomeric state of AE10, SEC was performed on the IMAC purified protein fraction.

#### 3.17.1. Calibration of the size exclusion column

A 24 ml Superdex 200 10/300 column (GE healthcare) was equilibrated with 60 ml of wash buffer (150 mM NaCl, 25 mM Tris.HCl, pH 8.0) at 25°C. The column was calibrated with four protein standards with $M_w$ of 66 kDa, 150 kDa, 200 kDa and 2000 kDa (Sigma). The void volume ($V_o$) was determined from the peak elution volume of 600 µl of blue dextran (2000 kDa; 2 mg $ml^{-1}$) at 25°C. In three separate experiments, 500 µl of the other $M_w$ standards, albumin (66 kDa; 10 mg $ml^{-1}$), alcohol dehydrogenase (150 kDa; 5 mg $ml^{-1}$) and β amylase (200 kDa; 4 mg $ml^{-1}$) were loaded on the column and eluted with wash buffer at 25°C. The

flow rate was kept fixed at 0.4 ml $min^{-1}$ in each case. The peak elution volume ($V_e$) was noted in each case, and the calibration curve of $\log_{10}(M_w)$ *vs.* $V_e/V_o$ was plotted (Fig. 3.44).

### 3.17.2. SEC of the αHL-λ-exo fusion constructs

500 µl of AE10 (Rosetta 20°C) was loaded on the pre-equilibrated column and washed with 30 ml of wash buffer at 4°C and 25°C. The eluate was collected as 1 ml fractions. The elution profile of AE10 at 4°C and 25°C was similar. 500 µl WT λ-exo (Fermentas) was also run on the same column at 4°C. The flow rate was kept fixed at 0.4 ml $min^{-1}$ in each case. Eluate fractions of AE10 and WT λ-exo were analysed under denaturing conditions (2% SDS) using SDS-PAGE. SEC of 300 µl of AE13 and EA13 was performed under the same conditions as AE10 at 25°C.

## 3.18. Oligomerization of αHL-λ-exo fusion constructs.

### 3.18.1. Oligomerization using rabbit red blood cell membranes (rRBCm)

IVTT proteins of AE10 and EA13 were oligomerized using rRBCm. rRBCm (4 µl) was washed with MBSA buffer (500 µl) and spun at 25,000 g for 15 min at 4°C. The supernatant was discarded and rRBCm pellets were used for oligomerization of IVTT protein products.

Co-translational IVTT reactions (50 µl; T7 S30 extract system) of AE10 and NN α1 were performed in the presence of rRBCm (4 µl) by adding equal amounts of NN αHL and AE10 DNA (1.6 µg each) in the IVTT reaction, which was carried

out for 1 h at 30°C. In other cases, rRBCm were added later to the IVTT products and incubated for a further 1.5 h at 30°C. Oligomerization products were analyzed on a 4-15% gradient pre-cast tris-glycine gel (Bio-Rad).

### 3.18.2. Oligomerization using detergent

AE10 and AE13 proteins, expressed in *E.coli*, were oligomerized using sodium deoxycholate (DOC) following protocols described before (Bhakdi *et al.*, 1981; Walker *et al.*, 1992a). For forming hetero-heptamers, AE10 or AE13 was mixed with WT α1 in a ratio of 4:1 (v/v) (unless stated otherwise) and the volume was made up to 32 µl with water. Water was added instead of AE10/ AE13 for making WT α7 and, instead of the WT α1 for making homo-heptamers of AE10 or AE13. DOC (50 mM) was added in ten equal installments with constant stirring over a course of 30 min, such that final concentration of DOC was 10 mM. The proteins treated with DOC were further incubated for 1 h at 25°C. The oligomerization products were analyzed on 4-15% gradient pre-cast tris-glycine gels (Bio-Rad).

# References


Astier, Y., Braha, O., and Bayley, H. (2006). Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. Journal of the American Chemical Society *128*, 1705-1710.

Astier, Y., Kainov, D.E., Bayley, H., Tuma, R., and Howorka, S. (2007). Stochastic detection of motor protein-RNA complexes by single-channel current recording. ChemPhysChem *8*, 2189-2194.

Bayley, H. (2006). Sequencing single molecules of DNA. Current opinion in chemical biology *10*, 628-637.

Bayley, H., Braha, O., and Gu, L.-Q. (2000). Stochastic sensing with protein pores. AdvMater *12*, 139-142.

Beese, L.S., and Steitz, T.A. (1991). Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. The EMBO journal *10*, 25-33.

Benner, S., Chen, R.J.A., Wilson, N.A., Abu-Shumays, R., Hurt, N., Lieberman, K.R., Deamer, D.W., Dunbar, W.B., and Akeson, M. (2007). Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. Nature nanotechnology *2*, 718-724.

Bhakdi, S., Füssle, R., and Tranum-Jensen, J. (1981). Staphylococcal a-toxin: oligomerization of hydrophilic monomers to form amphiphilic hexamers induced through contact with deoxycholate micelles. ProcNatlAcadSciUSA *78*, 5475-5479.

Bolanos-Garcia, V.M., and Davies, O.R. (2006). Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. Biochimica et biophysica acta *1760*, 1304-1313.

Borgia, A., Williams, P.M., and Clarke, J. (2008). Single-molecule studies of protein folding. Annual review of biochemistry *77*, 101-125.

Boyer, H.W. (1971). DNA restriction and modification mechanisms in bacteria. Annual review of microbiology *25*, 153-176.

Braha, O., Walker, B., Cheley, S., Kasianowicz, J.J., Song, L., Gouaux, J.E., and Bayley, H. (1997). Designed protein pores as components for biosensors. ChemBiol *4*, 497-505.

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X.*, et al.* (2008). The potential and challenges of nanopore sequencing. Nature biotechnology *26*, 1146-1153.

Buckle, A.M., Schreiber, G., and Fersht, A.R. (1994). Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. Biochemistry *33*, 8878-8889.

Buisson, M., Geoui, T., Flot, D., Tarbouriech, N., Ressing, M.E., Wiertz, E.J., and Burmeister, W.P. (2009). A bridge crosses the active-site canyon of the Epstein-Barr virus nuclease with DNase and RNase activities. Journal of molecular biology *391*, 717-728.

Carter, D.M., and Radding, C.M. (1971). The role of exonuclease and beta protein of phage lambda in genetic recombination. II. Substrate specificity and the mode of action of lambda exonuclease. The Journal of biological chemistry *246*, 2502-2512.

Cassuto, E., Lash, T., Sriprakash, K.S., and Radding, C.M. (1971). Role of exonuclease and beta protein of phage lambda in genetic recombination. V. Recombination of lambda DNA in vitro. Proceedings of the National Academy of Sciences of the United States of America *68*, 1639-1643.

Cassuto, E., and Radding, C.M. (1971). Mechanism for the action of lambda exonuclease in genetic recombination. Nature: New biology *229*, 13-16.

Chan, E.Y. (2005). Advances in sequencing techniques. Mutation Res *573*, 13-40.

Cheley, S., Braha, O., Lu, X., Conlan, S., and Bayley, H. (1999). A functional protein pore with a "retro" transmembrane domain. Protein Sci *8*, 1257-1267.

Cheley, S., Malghani, M.S., Song, L., Hobaugh, M., Gouaux, J.E., Yang, J., and Bayley, H. (1997). Spontaneous oligomerization of a staphylococcal a-hemolysin conformationally constrained by removal of residues that form the transmembrane b barrel. Protein Eng *10*, 1433-1443.

Cheley, S., Xie, H., and Bayley, H. (2006). A genetically-encoded pore for the stochastic detection of a protein kinase. Chembiochem : a European journal of chemical biology *7*, 1923-1927.

Cherf, G.M., Lieberman, K.R., Rashid, H., Lam, C.E., Karplus, K., and Akeson, M. (2012). Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. Nature biotechnology *30*, 344-348.

Claessen, V.I., Engelkamp, H., Christianen, P.C., Maan, J.C., Nolte, R.J., Blank, K., and Rowan, A.E. (2010). Single-biomolecule kinetics: the art of studying a single enzyme. Annu Rev Anal Chem (Palo Alto Calif) *3*, 319-340.

Clarke, J., Wu, H., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. Nature Nanotechnology *4*, 265-270.

Copeland, N.G., Jenkins, N.A., and Court, D.L. (2001). Recombineering: a powerful new tool for mouse functional genomics. Nature reviews Genetics *2*, 769-779.

Crampton, D.J., Mukherjee, S., and Richardson, C.C. (2006). DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. Molecular cell *21*, 165-174.

Czajkowsky, D.M., Sheng, S., and Shao, Z. (1998). Staphylococcal a-hemolysin can form hexamers in phospholipid bilayers. JMolBiol *276*, 325-330.

Dahl, J.M., Mai, A.H., Cherf, G.M., Jetha, N.N., Garalde, D.R., Marziali, A., Akeson, M., Wang, H., and Lieberman, K.R. (2012). Direct observation of translocation in individual DNA polymerase complexes. The Journal of biological chemistry *287*, 13407-13421.

Dahlroth, S.L., Gurmu, D., Schmitzberger, F., Engman, H., Haas, J., Erlandsen, H., and Nordlund, P. (2009). Crystal structure of the shutoff and exonuclease protein from the oncogenic Kaposi's sarcoma-associated herpesvirus. The FEBS journal *276*, 6636-6645.

Dapprich, J. (1999). Single-molecule DNA digestion by lambda-exonuclease. Cytometry *36*, 163-168.

Davis, L.M., Fairfield, F.R., Harger, C.A., Jett, J.H., Keller, R.A., Hahn, J.H., Krakowski, L.A., Marrone, B.L., Martin, J.C., Nutter, H.L.*, et al.* (1991). Rapid

DNA sequencing based upon single molecule detection. Genetic analysis, techniques and applications *8*, 1-7.

Deamer, D. (2010). Nanopore analysis of nucleic acids bound to exonucleases and polymerases. Annual review of biophysics *39*, 79-90.

Dill, K.A. (1990). Dominant forces in protein folding. Biochemistry *29*, 7133-7155.

Dudko, O.K., Mathe, J., and Meller, A. (2010). Nanopore force spectroscopy tools for analyzing single biomolecular complexes. Methods in enzymology *475*, 565-589.

Fellouse, F.A., Li, B., Compaan, D.M., Peden, A.A., Hymowitz, S.G., and Sidhu, S.S. (2005). Molecular recognition by a binary code. Journal of molecular biology *348*, 1153-1162.

Finkelstein, I.J., and Greene, E.C. (2008). Single molecule studies of homologous recombination. Molecular bioSystems *4*, 1094-1104.

Foldes-Papp, Z., Angerer, B., Thyberg, P., Hinz, M., Wennmalm, S., Ankenbauer, W., Seliger, H., Holmgren, A., and Rigler, R. (2001). Fluorescently labeled model DNA sequences for exonucleolytic sequencing. J Biotechnol *86*, 203-224.

Füssle, R., Bhakdi, S., Sziegoleit, A., Tranum-Jensen, J., Kranz, T., and Wellensiek, H.-J. (1981). On the mechanism of membrane damage by Staphylococcus aureus a-toxin. JCell Biol *91*, 83-94.

Gu, L.-Q., Braha, O., Conlan, S., Cheley, S., and Bayley, H. (1999). Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. Nature *398*, 686-690.

Ha, T., Enderle, T., Ogletree, D.F., Chemla, D.S., Selvin, P.R., and Weiss, S. (1996). Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. Proceedings of the National Academy of Sciences of the United States of America *93*, 6264-6268.

Ha, T., Kozlov, A.G., and Lohman, T.M. (2012). Single-molecule views of protein movement on single-stranded DNA. Annual review of biophysics *41*, 295-319.

Hammerstein, A.F., Jayasinghe, L., and Bayley, H. (2011). Subunit dimers of α-hemolysin expand the engineering toolbox for protein nanopores. J Biol Chem, in press.

Hendrix, R.W. (2002). Bacteriophages: evolution of the majority. Theoretical population biology *61*, 471-480.

Herbert, K.M., La Porta, A., Wong, B.J., Mooney, R.A., Neuman, K.C., Landick, R., and Block, S.M. (2006). Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell *125*, 1083-1094.

Hildebrand, A., Pohl, M., and Bhakdi, S. (1991). Staphylococcus aureus a-toxin. Dual mechanism of binding to target cells. JBiolChem *266*, 17195-17200.

Hornblower, B., Coombs, A., Whitaker, R.D., Kolomeisky, A., Picone, S.J., Meller, A., and Akeson, M. (2007). Single-molecule analysis of DNA-protein complexes using nanopores. Nature methods *4*, 315-317.

Howorka, S., Movileanu, L., Braha, O., and Bayley, H. (2001). Kinetics of duplex formation for individual DNA strands within a single protein nanopore. ProcNatlAcadSciUSA *98*, 12996-13001.

Jett, J.H., Keller, R.A., Martin, J.C., Marrone, B.L., Moyzis, R.K., Ratliff, R.L., Seitzinger, N.K., Shera, E.B., and Stewart, C.C. (1989). High-speed DNA sequencing: an approach based upon fluorescence detection of single molecules. Journal of biomolecular structure & dynamics *7*, 301-309.

Johnson, D.S., Bai, L., Smith, B.Y., Patel, S.S., and Wang, M.D. (2007). Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. Cell *129*, 1299-1309.

Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., and Hendrix, R.W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. Journal of molecular biology *299*, 27-51.

Kasianowicz, J.J., Brandin, E., Branton, D., and Deamer, D.W. (1996). Characterization of individual polynucleotide molecules using a membrane channel. ProcNatlAcadSciUSA *93*, 13770-13773.

Kelman, Z., and O'Donnell, M. (1995). Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. Nucleic acids research *23*, 3613-3620.

Kovall, R., and Matthews, B.W. (1997a). Toroidal structure of l-exonuclease. Science *277*, 1824-1827.

Kovall, R., and Matthews, B.W. (1997b). Toroidal structure of lambda-exonuclease. Science *277*, 1824-1827.

Kovall, R.A., and Matthews, B.W. (1998). Structural, functional, and evolutionary relationships between lambda-exonuclease and the type II restriction endonucleases. Proceedings of the National Academy of Sciences of the United States of America *95*, 7893-7897.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Lee, G., Yoo, J., Leslie, B.J., and Ha, T. (2011). Single-molecule analysis reveals three phases of DNA degradation by an exonuclease. Nature chemical biology *7*, 367-374.

Lee, J.Y., Chang, J., Joseph, N., Ghirlando, R., Rao, D.N., and Yang, W. (2005). MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage. Molecular cell *20*, 155-166.

Li, Z., Karakousis, G., Chiu, S.K., Reddy, G., and Radding, C.M. (1998). The beta protein of phage lambda promotes strand exchange. Journal of molecular biology *276*, 733-744.

Lieberman, K.R., Cherf, G.M., Doody, M.J., Olasagasti, F., Kolodji, Y., and Akeson, M. (2010). Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase. Journal of the American Chemical Society *132*, 17961-17972.

Little, J.W. (1967). An exonuclease induced by bacteriophage lambda. II. Nature of the enzymatic reaction. The Journal of biological chemistry *242*, 679-686.

Little, J.W., Lehman, I.R., and Kaiser, A.D. (1967). An exonuclease induced by bacteriophage lambda. I. Preparation of the crystalline enzyme. The Journal of biological chemistry *242*, 672-678.

Manrao, E.A., Derrington, I.M., Laszlo, A.H., Langford, K.W., Hopper, M.K., Gillgren, N., Pavlenok, M., Niederweis, M., and Gundlach, J.H. (2012). Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. Nature biotechnology *30*, 349-353.

Mao, C., Solis, D.J., Reiss, B.D., Kottmann, S.T., Sweeney, R.Y., Hayhurst, A., Georgiou, G., Iverson, B., and Belcher, A.M. (2004). Virus-based toolkit for the directed synthesis of magnetic and semiconducting nanowires. Science *303*, 213-217.

Marti, T.M., and Fleck, O. (2004). DNA repair nucleases. Cellular and molecular life sciences : CMLS *61*, 336-354.

Matsuura, S., Komatsu, J., Hirano, K., Yasuda, H., Takashima, K., Katsura, S., and Mizuno, A. (2001). Real-time observation of a single DNA digestion by lambda exonuclease under a fluorescence microscope field. Nucleic acids research *29*, E79.

Mehta, A.D., Rief, M., Spudich, J.A., Smith, D.A., and Simmons, R.M. (1999). Single-molecule biomechanics with optical methods. Science *283*, 1689-1695.

Miles, G., Bayley, H., and Cheley, S. (2002). Properties of Bacillus cereus hemolysin II: a heptameric transmembrane pore. Protein Sci *11*, 1813-1824.

Min, W., English, B.P., Luo, G., Cherayil, B.J., Kou, S.C., and Xie, X.S. (2005). Fluctuating enzymes: lessons from single-molecule studies. Accounts of chemical research *38*, 923-931.

Mitsis, P.G., and Kwagh, J.G. (1999). Characterization of the interaction of lambda exonuclease with the ends of DNA. Nucleic acids research *27*, 3057-3063.

Moll, D., Huber, C., Schlegel, B., Pum, D., Sleytr, U.B., and Sára, M. (2002). S-layer-streptavidin fusion proteins as template for nanopatterned molecular arrays. ProcNatlAcadSciUSA *99*, 14646-14651.

Moreau, C.J., Dupuis, J.P., Revilloud, J., Arumugam, K., and Vivaudou, M. (2008). Coupling ion channels to receptors for biomolecule sensing. Nature nanotechnology *3*, 620-625.

Mosig, G. (1998). Recombination and recombination-dependent DNA replication in bacteriophage T4. Annual review of genetics *32*, 379-413.

Movileanu, L., Cheley, S., and Bayley, H. (2003). Partitioning of individual flexible polymers into a nanoscopic protein pore. BiophysJ *85*, 897-910.

Movileanu, L., Howorka, S., Braha, O., and Bayley, H. (2000). Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore. Nature biotechnology *18*, 1091-1095.

Nicke, A., Rettinger, J., and Schmalzing, G. (2003). Monomeric and dimeric byproducts are the principal functional elements of higher order P2X1 concatamers. Molecular pharmacology *63*, 243-252.

Nowotny, M., Gaidamakov, S.A., Crouch, R.J., and Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell *121*, 1005-1016.

Okumus, B., Arslan, S., Fengler, S.M., Myong, S., and Ha, T. (2009). Single molecule nanocontainers made porous using a bacterial toxin. Journal of the American Chemical Society *131*, 14844-14849.

Olasagasti, F., Lieberman, K.R., Benner, S., Cherf, G.M., Dahl, J.M., Deamer, D.W., and Akeson, M. (2010). Replication of individual DNA molecules under electronic control using a protein nanopore. Nature nanotechnology *5*, 798-806.

Ossipow, V., Laemmli, U.K., and Schibler, U. (1993). A simple method to renature DNA-binding proteins separated by SDS-polyacrylamide gel electrophoresis. Nucleic acids research *21*, 6040-6041.

Perkins, T.T., Dalal, R.V., Mitsis, P.G., and Block, S.M. (2003). Sequence-dependent pausing of single lambda exonuclease molecules. Science *301*, 1914-1918.

Perkins, T.T., Li, H.W., Dalal, R.V., Gelles, J., and Block, S.M. (2004). Forward and reverse motion of single RecBCD molecules on DNA. Biophysical journal *86*, 1640-1648.

Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. Cellular and molecular life sciences : CMLS *62*, 685-707.

Poteete, A.R. (2001). What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. FEMS microbiology letters *201*, 9-14.

Rabilloud, T., Vuillard, L., Gilly, C., and Lawrence, J.J. (1994). Silver-staining of proteins in polyacrylamide gels: a general overview. Cell Mol Biol (Noisy-le-grand) *40*, 57-75.

Radding, C.M., and Carter, D.M. (1971). The role of exonuclease and beta protein of phage lambda in genetic recombination. 3. Binding to deoxyribonucleic acid. The Journal of biological chemistry *246*, 2513-2518.

Ray, S., Mehta, G., and Srivastava, S. (2010). Label-free detection techniques for protein microarrays: prospects, merits and challenges. Proteomics *10*, 731-748.

Regan, L. (1999). Protein redesign. CurrOpStructural Biol *9*, 494-499.

Roberts, J.W. (2006). Biochemistry. RNA polymerase, a scrunching machine. Science *314*, 1097-1098.

Robichon, C., Luo, J., Causey, T.B., Benner, J.S., and Samuelson, J.C. (2011). Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography. Applied and environmental microbiology *77*, 4634-4646.

Robinson, C.R., and Sauer, R.T. (1998). Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. Proceedings of the National Academy of Sciences of the United States of America *95*, 5929-5934.

Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995). Evaluation of comparative protein modeling by MODELLER. Proteins *23*, 318-326.

Sangha, N., Kaur, S., Sharma, V., and Krishnasastry, M.V. (1999). Importance of the carboxyl terminus in the folding and function of a-hemolysin of Staphylococcus aureus. JBiolChem *274*, 9193-9199.

Schafer, D.A., Gelles, J., Sheetz, M.P., and Landick, R. (1991). Transcription by single molecules of RNA polymerase observed by light microscopy. Nature *352*, 444-448.

Sheetz, M.P. (1998). Laser tweezers in cell biology. Introduction. Methods in cell biology *55*, xi-xii.

Shulman, M.J., Hallick, L.M., Echols, H., and Signer, E.R. (1970). Properties of recombination-deficient mutants of bacteriophage lambda. Journal of molecular biology *52*, 501-520.

Sinclair, J.C., Davies, K.M., Venien-Bryan, C., and Noble, M.E. (2011). Generation of protein lattices by fusing proteins with matching rotational symmetry. Nature nanotechnology *6*, 558-562.

Singleton, M.R., Dillingham, M.S., Gaudier, M., Kowalczykowski, S.C., and Wigley, D.B. (2004). Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. Nature *432*, 187-193.

Singleton, M.R., Sawaya, M.R., Ellenberger, T., and Wigley, D.B. (2000). Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. Cell *101*, 589-600.

Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., and Gouaux, J.E. (1996). Structure of staphylococcal a-hemolysin, a heptameric transmembrane pore. Science *274*, 1859-1865.

Sriprakash, K.S., Lundh, N., Huh, M.-O., and Radding, C.M. (1975). The specificity of lambda exonuclease. Interactions with single-stranded DNA. The Journal of biological chemistry *250*, 5438-5445.

Stoddart, D., Heron, A., Mikhailova, E., Maglia, G., and Bayley, H. (2009). Single nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc Natl Acad Sci USA *106*, 7702-7707.

Stoddart, D., Heron, A.J., Klingelhoefer, J., Mikhailova, E., Maglia, G., and Bayley, H. (2010). Nucleobase recognition in ssDNA at the central constriction of the alpha-hemolysin pore. Nano Lett *10*, 3633-3637.

Subramanian, K., Rutvisuttinunt, W., Scott, W., and Myers, R.S. (2003). The enzymatic basis of processivity in lambda exonuclease. Nucleic acids research *31*, 1585-1596.

Tachiiri, Y., Ishikawa, M., and Hirano, K. (2000). Investigation of the hydrolysis of single DNA molecules using fluorescence video microscopy. Analytical chemistry *72*, 1649-1656.

Takahashi, N., and Kobayashi, I. (1990). Evidence for the double-strand break repair model of bacteriophage lambda recombination. Proceedings of the National Academy of Sciences of the United States of America *87*, 2790-2794.

Tolun, G., and Myers, R.S. (2003). A real-time DNase assay (ReDA) based on PicoGreen fluorescence. Nucleic acids research *31*, e111.

van Oijen, A.M., Blainey, P.C., Crampton, D.J., Richardson, C.C., Ellenberger, T., and Xie, X.S. (2003). Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. Science *301*, 1235-1238.

Voet, D., and Voet, J.G. (2011). Biochemistry, 4th edn (Hoboken, NJ: John Wiley & Sons).

Walker, B., and Bayley, H. (1995). Key residues for membrane binding, oligomerization, and pore-forming activity of staphylococcal a-hemolysin identified by cysteine scanning mutagenesis and targeted chemical modification. JBiolChem *270*, 23065-23071.

Walker, B.J., Krishnasastry, M., Zorn, L., and Bayley, H. (1992a). Assembly of the oligomeric membrane pore formed by staphylococcal a-hemolysin examined by truncation mutagenesis. JBiolChem *267*, 21782-21786.

Walker, B.J., Krishnasastry, M., Zorn, L., Kasianowicz, J.J., and Bayley, H. (1992b). Functional expression of the a-hemolysin of Staphylococcus aureus in intact Escherichia coli and in cell lysates. JBiolChem *267*, 10902-10909.

Wang, M.D., Schnitzer, M.J., Yin, H., Landick, R., Gelles, J., and Block, S.M. (1998). Force and velocity measured for single molecules of RNA polymerase. Science *282*, 902-907.

Werner, J.H., Cai, H., Jett, J.H., Reha-Krantz, L., Keller, R.A., and Goodwin, P.M. (2003). Progress towards single-molecule DNA sequencing: a one color demonstration. JBiotechnol *102*, 1-14.

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

Yang, W. (2011). Nucleases: diversity of structure, function and mechanism. Quarterly reviews of biophysics *44*, 1-93.

Yang, W., Chen, W.Y., Wang, H., Ho, J.W., Huang, J.D., Woo, P.C., Lau, S.K., Yuen, K.Y., Zhang, Q., Zhou, W.*, et al.* (2011). Structural and functional insight into the mechanism of an alkaline exonuclease from Laribacter hongkongensis. Nucleic acids research *39*, 9803-9819.

Yang, W., Lee, J.Y., and Nowotny, M. (2006). Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. Molecular cell *22*, 5-13.

Yuan, R. (1981). Structure and mechanism of multifunctional restriction endonucleases. Annual review of biochemistry *50*, 285-319.

Zhang, J., McCabe, K.A., and Bell, C.E. (2011). Crystal structures of lambda exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. Proceedings of the National Academy of Sciences of the United States of America *108*, 11872-11877.

Zhang, J., Xing, X., Herr, A.B., and Bell, C.E. (2009). Crystal structure of E. coli RecE protein reveals a toroidal tetramer for processing double-stranded DNA breaks. Structure *17*, 690-702.

Zhou, R., Kozlov, A.G., Roy, R., Zhang, J., Korolev, S., Lohman, T.M., and Ha, T. (2011). SSB functions as a sliding platform that migrates on DNA via reptation. Cell *146*, 222-232.

# CHAPTER 4

## DESIGNING A CONCATENATED FUSION CONSTRUCT OF λ-EXONUCLEASE

## Introduction

### 4.1. Constructing a concatemer of λ-exo (exo3) to make the λ-exo-αHL pore complex

Nature has designed many proteins to function as homo- and hetero-multimers (Griffin and Gerrard, 2012; Kovall and Matthews, 1997; Martin *et al.*, 2005; Yoshida *et al.*, 2001). αHL functions as heptamers and λ-exo as a trimer (Song et al., 1996; Zhang et al., 2011). αHL and λ-exo could be coupled in order to build a molecular device for DNA sequencing and single molecule enzymatic studies (Chapter 3, section 3.4). Fusion constructs of αHL and λ-exo monomers were synthesized, expressed and characterized (Chapter 3, section 3.8). The fusion constructs (AE constructs) with the λ-exo monomer attached to the C-terminus of αHL monomer (α1), showed exonuclease activity and could also form pores in rabbit erythrocytes and planar lipid bilayers. It was also found that hetero-heptamers with only one subunit of the AE construct with six WT α1 subunits could be selectively made by adding an excess of α1. Hence, instead of the λ-exo monomer as in the AE constructs, if α1 were linked to a fusion trimer of λ-exo with concatenated λ-exo monomers, we could achieve a hetero-heptamer with a λ-exo trimer mounted on the αHL cap.

### 4.2. Fusion constructs of other multimeric proteins

Previously other multimeric proteins such as $K^+$ channels (Isacoff *et al.*, 1990), GABA receptors (Baumann *et al.*, 2002), nACh receptors (Zhou *et al.*, 2003) and mechanosensitive channels (Folgering *et al.*, 2005) have been investigated by genetically linking their monomer subunits. Liman *et al.* synthesized hetero-multimers of the mammalian $K^+$ channel $K_V1.1$ by genetically fusing 2 to 5 monomers and studied the effect of single residue mutations (Liman *et al.*, 1992). To determine the stoichiometry of the mechanosensitive channel MscL from *E.coli in vivo*, fusion constructs with 2 to 6 monomer units were made (Folgering *et al.*, 2005). From this study, Folgering *et al.* found that only the fusion pentamers of MscL displayed similar activity to the WT protein *in vivo*, which implied that the functional form of MscL was a pentamer.

Concatenated monomer fusion constructs of the hexameric ATPase, ClpX, involved in unfolding proteins *in vivo*, revealed the mechanism of ClpX in a new light (Martin et al., 2005, 2007). Fusion constructs were made with 2, 3 and 6 linked monomer subunits (Martin et al., 2005). By replacing WT monomer subunits with inactive variants of ClpX monomers in the fusion constructs, hetero-hexamers were made. It was found that not all monomer subunits needed to be functional. For example, ClpX hexamers with two inactive subunits were ~60% as active as the protein with all active subunits. Fusion hetero-hexamers of ClpX with the same number of inactive subunits but with different arrangements of the active and inactive monomer subunits were functionally equivalent. A new mechanism of ClpX activity was thus proposed where there is no obligatory order of the monomer subunits of ClpX to bind and

hydrolyze ATP. Similarly, the trimer fusion construct of λ-exo could also be used to study the mechanistic bases of its enzymatic action.

### 4.3. Exo3 variants for mechanistic studies of λ-exo

The co-crystal structure of λ-exo with dsDNA shows that the active form of λ-exo is a trimer (Chapter 3, sections 3.2, 3.3). However, it still remains to be known whether all three subunits of λ-exo are required for activity and whether hetero-trimers could be formed. The hetero-oligomers of λ-exo could prove to be pivotal in answering many outstanding mechanistic questions regarding its exonuclease activity (Chapter 3, section 3.3.6). If a hetero-trimer of λ-exo containing one or more inactive subunits retains a part of its activity, it will be a clear indication that the subunits do not work in a fixed order, rotating each time a base is cleaved from a dsDNA substrate. It will also imply that once λ-exo has clamped onto a dsDNA substrate, only one subunit is employed in progressively processing 1000s of nt. Nanopore based single molecule studies (Chapter 3, section 3.4.1) could be used to study the processivity and speed of λ-exo heterotrimers. Hence, methods to make hetero-trimers and link it to αHL were sought.

## Results

### 4.4. Synthesis of the exo3 construct

In order to design the concatenated construct of λ-exo, a tandem gene of three λ-exo monomer genes was prepared. The distance between the N- and C-termini of two adjacent λ-exo monomers was found to be ~60 Å when measured in a clockwise direction (C->N) and ~90 Å when measured in an anti-clockwise

direction (C->N) (Fig. 4.1). These distances were measured in multiple steps. Hence, linkers of 15-20 amino acids (Chapter 3, section 3.8.1) (Martin *et al.*, 2005) were required between the monomer proteins. The linkers also need to be flexible and hence were designed to be composed of only serine, glycine and threonine. The exo3 fusion gene could be made in many different ways. We opted to make the construct by two rounds of homologous recombination in *E.coli*.

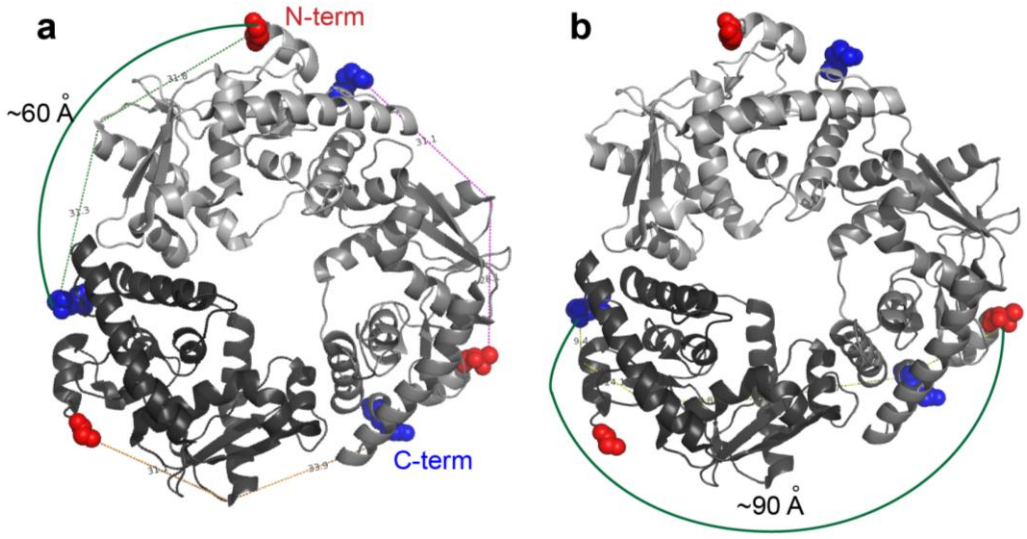



**Figure 4.1. Distances between the N- and C-termini of λ-exo.** To construct a genetically fused trimer construct of λ-exo, linkers were required between the N- and C-termini of adjacent subunits of λ-exo. To determine the length of the linkers required, the distances between the N- and C-termini of λ-exo were measured in PyMOL. (**a**) Going in a clockwise direction (C->N), the termini distance was measured in two steps to be ~60 Å. (**b**) Moving in the anticlockwise direction (C->N), the termini distance was measured in multiple steps to be ~90 Å.

First, two λ-exo monomer genes were placed next to one another spaced by a short SG linker and two unique restriction sites of SpeI and XhoI, in a pT7 vector (Walker *et al.*, 1992) (Fig. 4.2).

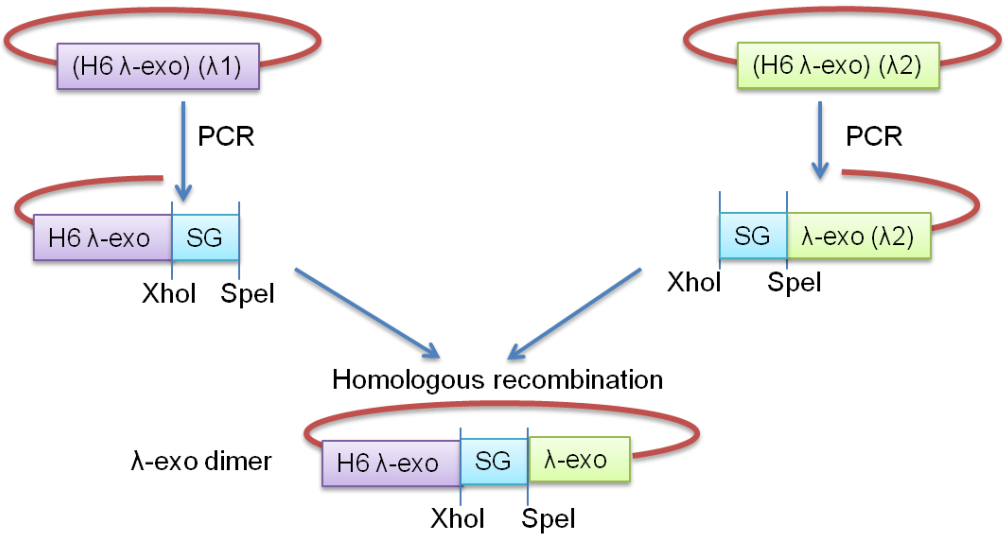



**Figure 4.2. Construction of λ-exo dimer.** Two simultaneous PCRs were done to introduce two restriction sites for XhoI and SpeI at the N- and C- termini of λ-exo monomer genes. The restriction sites were spaced by 6 nucleotides encoding for serine and glycine. The two PCR reactions were mixed for homologous recombination in *E.coli* to place two λ-exo genes in tandem within the same pT7 vector. These two λ-exo genes would act as the 1$^{st}$ (λ1, violet) and the 2$^{nd}$ (λ2, green) genes in the λ-exo trimer construct.

A third λ-exo gene was added by homologous recombination to the λ-exo dimer (Fig. 4.3). The sequence of the trimer fusion construct of λ-exo was then modified such that each pair of λ-exo genes was separated by a unique pair of restriction sites (plasmid termed exo3_2), where varying lengths of SG linkers could be inserted (Fig. 4.3). As each λ-exo monomer gene was placed between two unique restriction sites, the exo3_2 plasmid design would also allow us to easily replace one or more λ-exo monomers with an inactive λ-exo mutant to make engineered variants of exo3.

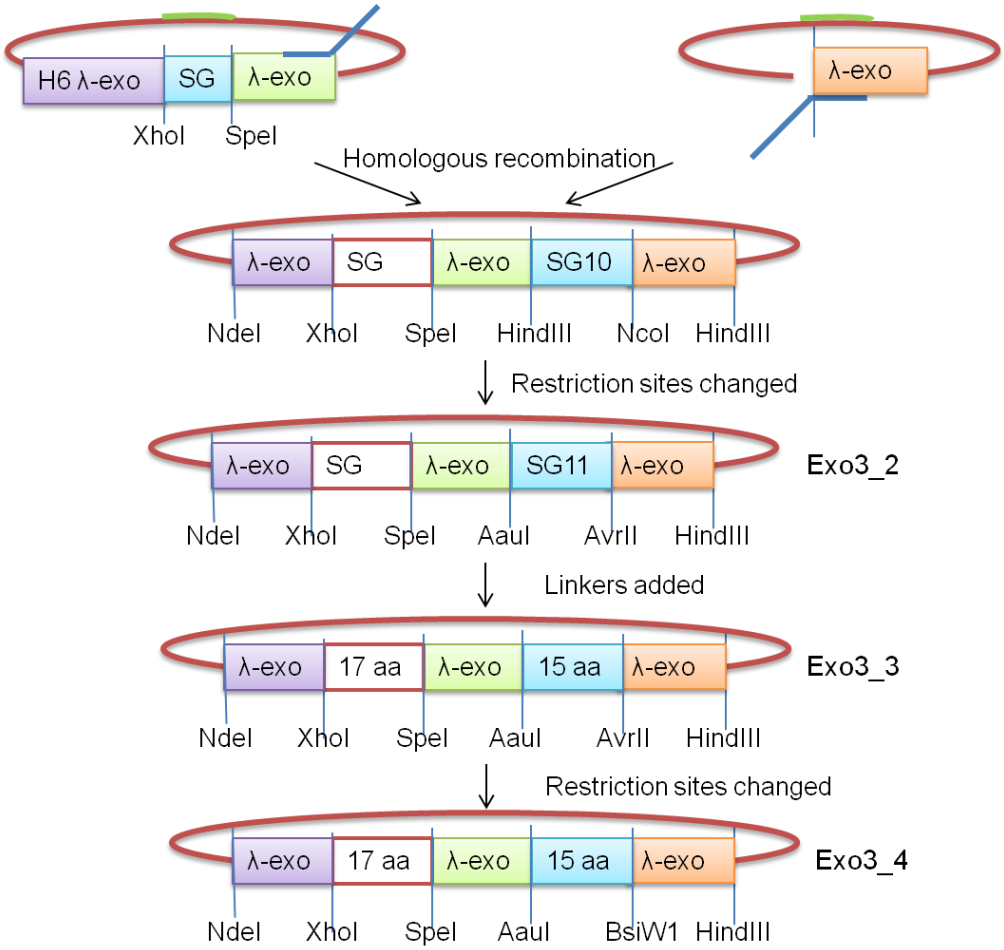



**Figure 4.3. Construction of λ-exo trimer.** The third λ-exo monomer gene (λ3, orange) was added to the λ-exo dimer gene by following a similar PCR and homologous recombination strategy as for the dimer λ-exo construct (Fig. 4.2). Two PCRs were done to introduce SG10 at the C-terminus of the dimer gene and at the N-terminus of the λ-exo monomer gene and transformed together in *E.coli* to yield the trimer construct. The restriction sites were further changed such that each λ-exo gene was between unique restriction sites (exo3_2). A 14 amino acid long linker (SG14) was ligated between λ1 and λ2 (exo3_3) and subsequently restriction sites changed (exo3_4).

In exo3_2, the first two λ-exo genes (λ1 and λ2) are spaced by XhoI and SpeI sites, and λ2 and the third λ-exo (λ3) gene are separated by AauI and AvrII restriction sites (Fig. 4.3). λ1 and λ2 were connected by only 6 amino acids and λ2 and λ3 were linked by 15 amino acids. The shortest distance between the termini of λ-exo monomers was determined to be 60 Å (Fig. 4.1a). Hence, a 14 amino acid long SG linker was then inserted between λ1 and λ2, such that the

C-terminus of λ1 and the N-terminus of λ2 were now separated by 17 amino acids (plasmid termed exo3_3) (Fig. 4.3).

### 4.5. Expression of the exo3 protein

The exo3_3 plasmid with 17 and 15 amino-acid long linkers between the λ-exo monomer genes was then expressed in recombination deficient BLR (DE3) pLysS cells at different temperatures. Expression of the trimer protein (exo3) was optimum at 30°C. SDS-PAGE revealed that the IMAC purified exo3_3 protein fractions contained two proteins. The protein band ($M_{app}$≈108,000) migrating close to the 100 kDa $M_w$ standard was attributed to the fusion trimer the λ-exo protein ($(\lambda\text{-exo})_3$; expected $M_w$=84 kDa). The other protein band ($M_{app}$≈66,000) migrated between the 50 kDa and 75 kDa $M_w$ markers and was attributed to a dimer species of λ-exo ($(\lambda\text{-exo})_2$; expected $M_w$=56 kDa). To determine the nature of the smaller protein ($M_{app}$≈66,000), the IMAC purified exo3_3 fractions were heated at 95°C for 5 min and reduced with βME. The smaller protein ($M_{app}$≈66,000) was intact even after heating and reducing treatments as determined by SDS-PAGE. Hence, it was concluded that the smaller protein ($M_{app}$≈66,000) was $(\lambda\text{-exo})_2$ which bound to the IMAC column.

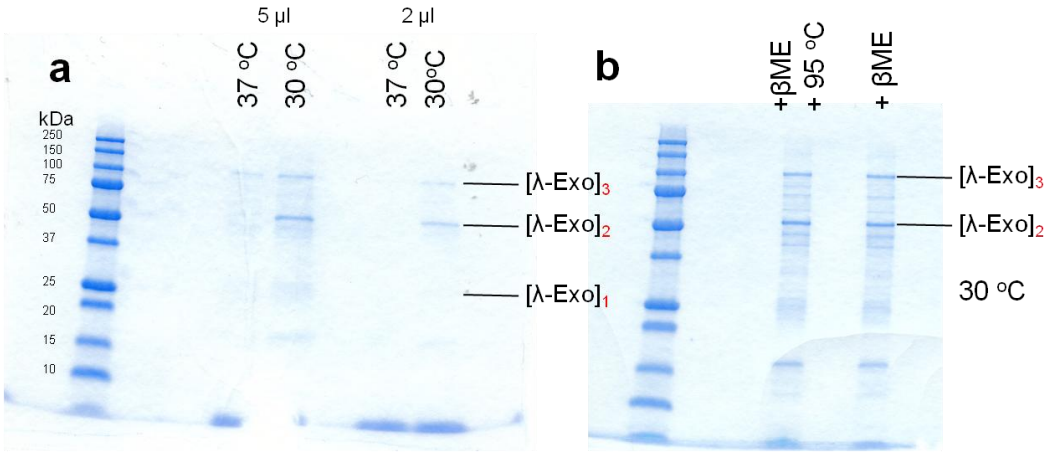



**Figure 4.4. Expression of exo3_3 under different conditions.** The exo3_3

plasmid with 17 and 15 amino-acid long linkers between the λ-exo genes was expressed in recombination deficient BLR (DE3) pLysS cells at 30°C and 37°C. The IMAC purified exo3_3 proteins showed the presence of two protein bands, which was attributed to the dimer ($M_{app}$≈66,000; $(\lambda\text{-exo})_2$) and trimer species ($M_{app}$≈108,000; $(\lambda\text{-exo})_3$) of λ-exo. To determine the nature of the $(\lambda\text{-exo})_2$ species ($M_{app}$≈66,000), the exo3_3 protein fractions were heated at 95°C and reduced with βME. The $(\lambda\text{-exo})_2$ band remained intact suggesting that it was a recombined or proteolysed product of the $(\lambda\text{-exo})_3$ protein ($M_{app}$≈108,000).

### 4.6. Exonuclease activity of exo3_3

The exonuclease activity of the exo3_3 proteins was assayed using the fluorescence-based method described in Chapter 3 (section 3.3.4). The dsDNA binding dye, PicoGreen, was used as the reporter of the extent of dsDNA hydrolysis. Addition of the exo3_3 protein expressed at 30°C (exo3 (30°C)) led to a sharp decrease in the fluorescence of dsDNA substrate bound dye, implying a rapid conversion of dsDNA to ssDNA. The exo3_3 protein expressed at 37°C (exo3 (37°C)) also hydrolyzed dsDNA but at a slower rate, presumably because of a lower protein concentration. As the monomer band of λ-exo was not observed in the exo3 protein fractions on SDS-PAGE (Fig. 4.4), the exonuclease activity of exo3_3 was attributed to $(\lambda\text{-exo})_2$ (see next paragraph).

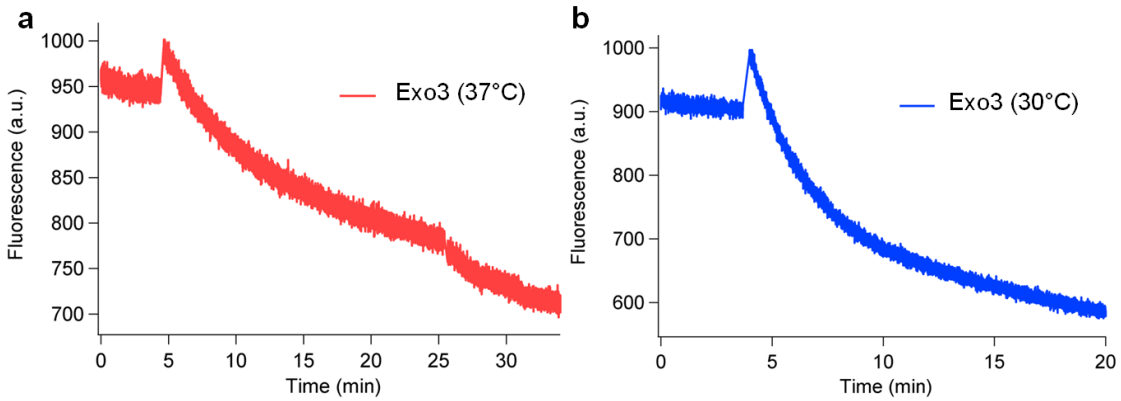



**Figure 4.5. Exonuclease activity assay of exo3_3 proteins.** The exonuclease activity of exo3 protein fractions was assayed using the fluorescence-based method described previously (Chapter 3, section 3.3.4). Exo3 proteins expressed using BLR (DE3) pLysS cells at (**a**) 37°C (exo3 (37°C)) and (**b**) 30°C (exo3 (30°C)), both showed exonuclease activity. Addition of WT λ-exo to the reaction mixture containing exo3 (37°C) did not lead to an

appreciable decrease in the fluorescence confirming that undigested dsDNA did not remain after 20 min of the reaction.

Subsequently, we realized that a stop codon had been included in the exo3_3 sequence. The stop codon was removed by changing the AvrII restriction site to BsiW1 and termed as the exo3_4 plasmid (Fig. 4.3). Exo3_4 was expressed using BLR (DE3) pLysS cells at 30°C and was found to possess exonuclease activity (data not shown). Current work is focused on improving the protein expression yield and characterization of the fusion trimer protein.

## Future work and outlook

To make the λ-exo-αHL pore complex (Chapter 3, section 3.4.3), a fusion construct of exo3_4 protein and α1 is currently being prepared using the AE13 plasmid. The λ-exo part in AE13, contained between NcoI and HindIII sites, can be easily replaced by the exo3 gene. To allow IMAC purification of the exo3-α1 fusion construct, an exo3_4 variant (exo3_H6), which has a His-6 tag at its C-terminus is being used (work in progress).

The recombinant exo3-α1 fusion construct with a His-6 tag at the C-terminus will be used to make hetero-heptamers with WT α1 (Chapter 3, section 3.8.12.2) (Fig. 4.6). Preferentially only one subunit of the exo3-αHL subunit could be incorporated in the hetero-heptamers by adding WT α1 in excess over exo3-α1 (section 3.8.12.3). The $(exo3\text{-}\alpha1)_1(\alpha1)_6$ hetero-heptamer of exo3-α1 and WT α1 would have only one exo3 molecule mounted on the αHL cap, which is the desired λ-exo-αHL pore complex (Chapter 3, section 3.4.3) (Fig. 4.6). The

$(exo3\text{-}\alpha1)_1(\alpha1)_6$ protein complex could then be used for nanopore-based single molecule activity studies of λ-exo and DNA sequencing.

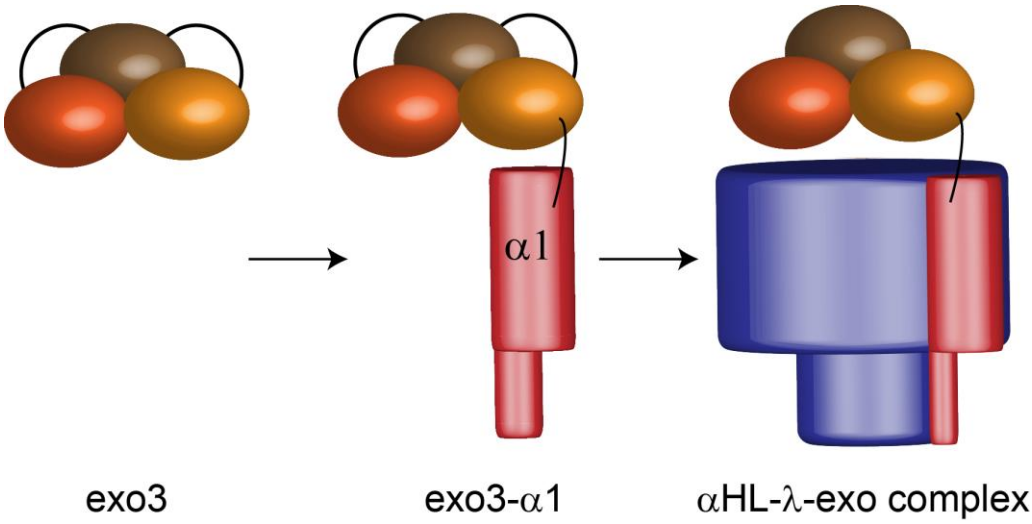



**Figure 4.6. Using the λ-exo trimer fusion construct to make αHL-λ-exo complex.** A concatemer construct of λ-exo has been prepared by linking the monomers of λ-exo with flexible linkers (exo3; orange). Exo3 could then be genetically linked to α1 (red) (exo3-α1). Hetero-heptamers containing just one unit of exo3-α1 amongst six α1 units (blue) would be desired αHL-λ-exo protein complex. Such hetero-heptamers have been successfully made in other contexts (Chapter 3, section 3.8.12.3).

The exo3_4 plasmid will also be used for designing hetero-trimers of λ-exo. Utilizing the restriction sites of the exo3_4 plasmid, one or more active λ-exo monomers could be easily substituted by another inactive or functionalized λ-exo monomer unit. The concatenated λ-exo trimer also holds promise for a detailed study of effect of point mutations on λ-exo activity. The hetero-trimers of the exo3 protein could also be decorated with chemical groups or non-natural amino acids to allow site-specific functionalization of one of the subunits. These exo3 mutants would throw light specifically on two outstanding questions regarding λ-exo activity: (i) do the λ-exo monomers work in a sequential manner or is just one monomer employed while the enzyme processively degrades dsDNA? (ii) Do the residues of the central pore of λ-exo interact with the ssDNA

generated as a result of dsDNA hydrolysis and contribute to the processivity of λ-exo? It will be intriguing if one of the λ-exo monomers could be replaced by another protein to create hybrid λ-exo pores, opening new frontiers in nuclease design.

# Methods

## 4.7. Construction of the concatenated λ-exo construct

The λ-exo dimer and trimer constructs were made by homologous recombination in *E.coli* using a PCR-based strategy very similar to the one shown in Fig. 3.55 (Chapter 3, section 3.10.2).

In two separate restriction digest reactions, the plasmid containing the WT λ-exo monomer gene with a His-6 tag at the N-terminus (pT7-H6 λ-exo) was linearized with NcoI and HindIII. In the f-PCR reaction, the N-terminus of NcoI digested pT7-H6 λ-exo was extended with a sequence containing the restriction site sequences of XhoI and SpeI spaced by 6 nucleotides encoding for serine and glycine using a primer termed E_XS F (Appendix Table 1). In the r-PCR reaction, the same sequence extension was added to the C-terminus of HindIII digested pT7-H6 λ-exo using the primer termed E_XS R (Appendix Table 1). PCR (50 μl) were performed using 1 μM primers (E_XS F or R and amp F or R, Appendix Table 1), 10 ng of linearized plasmid and a PCR protocol as described in Chapter 3, section 3.10.1. Equal amounts of PCR products (5 μl) were mixed and transformed in XL-10 Gold cells to circularize the plasmid. The cells were incubated on ice for 30 min, followed by a heat-shock step at 42°C for 30 s, and again incubation on ice for 5 min. The cells were then plated on ampicillin selective plates and incubated at 37°C overnight (~16 h).

To construct the λ-exo trimer, the dimer construct of λ-exo was used along with pT7-H6 λ-exo. In this case, the plasmid contained the λ-exo dimer gene was linearized with HindIII and the pT7-H6 λ-exo plasmid was cut with NcoI. In two

simultaneous PCRs, the C-terminus of the λ-exo dimer and the N-terminus of λ-exo monomer was extended with a sequence encoding a SG10 linker flanked by HindIII and NcoI restriction sites. The two halves of the λ-exo trimer plasmid thus created were recombined in XL-10 as described before (Chapter 3, section 3.10.2). The sequence of the λ-exo trimer was confirmed *via* sequencing reactions with T7F and T7R primers (Source Bioscience Lifesciences, Oxford). The presence of three λ-exo monomer genes in tandem was also confirmed by restriction digest analysis using NdeI and KpnI (data not shown).

Next, the following restriction sites between λ2 and λ3 were changed by PCR-based mutagenesis: the HindIII (AAGCTT) and NcoI (CCATGG) sites were changed to AauI (TGTACA) and AvrII (CCTAGG) respectively. The primers (NA_F for f-PCR and HAXB R1L for r-PCR) and the PCR protocol used are listed in Appendix Table 1 and Table 4.1 respectively. The f- and r- PCR products were mixed and transformed in XL-10 to yield the trimer plasmid termed exo3_2.

**Table 4.1.** Table shows the PCR temperature protocol used to make the αHL-λ-exo fusion constructs.

| Temperature ($^{o}$C) | Time (s) | Cycles |
|---|---|---|
| 98 | 1 min | |
| 98 | 5 | |
| 65 | 5 | 30 |
| 72 | 45 | |
| 72 | 5 min | |

A 14 amino acids long linker encoding for Ser, Gly and Thr (SG14) and flanked by SpeI restriction site sequences was ligated between λ1 and λ2. The sense and antisense strands of the SG14 linker were annealed to make the dsDNA insert sequence. 100 μM of the sense- and anti-sense strands of SG14 were mixed to a total volume of 100 μl and heated at 95°C for 5 min followed by slow cooling at 25°C for 2.5 h. 1 μl of SG14 dsDNA linker (50 μM) and ~1 μg of exo3_2 plasmid were digested with SpeI at 37°C for 30 min in a total reaction volume of 20 μl and 30 μl respectively. The digestion reactions were heated at 80°C for 20 min to heat inactivate the SpeI enzyme. The digested exo3_2 plasmid was dephosphorylated to minimize background colonies after ligation. ~450 ng of exo3_2 linearized plasmid with sticky ends was treated with 0.5 μl Antarctic phosphatase (NEB), supplemented with Antarctic phosphatase reaction buffer (50 mM Bis-Tris-Propane-HCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, pH 6.0) (NEB) in a total reaction volume of 20 μl at 37°C for 30 min. This was followed by heat inactivation of the Antarctic phosphatase enzyme at 65°C for 5 min. The digested SG14 linker was inserted into the digested and dephosphorylated exo3_2 plasmid using the Quick Ligation kit (NEB). The SpeI digested SG14 linker was diluted to a final concentration of ~15 nM and mixed in a 1:1 v/v ratio with the exo3_2 plasmid (~5 nM), followed by addition of the 2x Quick Ligase buffer (132 mM Tris.HCl, 20 mM $MgCl_2$, 2 mM DTT, 2 mM ATP, 15% PEG 6000, pH 7.6) (NEB) and 1 μl of Quick Ligase (NEB). The ligation reaction mix was incubated at 25°C for 5 min before transformation in 75 μl XL10 gold cells. The transformed colonies were then screened for exo3_3, i.e., plasmids where the SG14 linker was successfully ligated.

### 4.8. Protein expression of exo3

Exo3_3 was transformed in BL21 (DE3) pLysS and recombination free BLR (DE3) pLysS cells. For protein expression using BL21 (DE3) pLysS and BLR (DE3) pLysS cells, a streak of colonies was inoculated in 100 ml Magic Media (Invitrogen) (+100 μg/ml ampicillin) and grown at different temperatures for different lengths of time: at 37°C for ~16 h, at 20°C for 40 h and at 30°C for ~21 h. The expression cultures were centrifuged at 14,000 rpm for 30-45 min. The culture pellets were resuspended in 10 ml ($1/10^{th}$ volume of the total culture) 10x BugBuster Mastermix solution (Merck) supplemented with two protease inhibitor cocktail tablets (Roche). The lysed cells were centrifuged at 14,000 rpm for 45 min again and the supernatant was loaded on Co TALON columns. The exo3 protein was purified further in a similar manner as the IMAC purification described in Chapter 3, section 3.11.2.

### 4.9. Exonuclease activity assay

The concentrated IMAC purified exo3 protein fractions expressed in BLR (DE3) pLysS cells at 30°C and 37°C were assayed for exonuclease activity (Chapter 3, section 3.14.2). 100 µl of 10x λ-exo activity buffer (670 mM Glycine-KOH, 25 mM $MgCl_2$, pH 9.4) (NEB) was added to 900 µl water in a 1 ml quartz cuvette with transparent sides. 2 µl of the 1 kb dsDNA substrate with phosphorylated ends was added next followed by addition of 0.5 µl of PicoGreen (Invitrogen). 5 µl of exo3 protein was added to the reaction mixture and the fluorescence was monitored for 20-30 min.

# References


Baumann, S.W., Baur, R., and Sigel, E. (2002). Forced subunit assembly in alpha1beta2gamma2 GABAA receptors. Insight into the absolute arrangement. The Journal of biological chemistry *277*, 46020-46025.

Folgering, J.H., Wolters, J.C., and Poolman, B. (2005). Engineering covalent oligomers of the mechanosensitive channel of large conductance from Escherichia coli with native conductance and gating characteristics. Protein Sci *14*, 2947-2954.

Griffin, M.D., and Gerrard, J.A. (2012). The relationship between oligomeric state and protein function. Advances in experimental medicine and biology *747*, 74-90.

Isacoff, E.Y., Jan, Y.N., and Jan, L.Y. (1990). Evidence for the formation of heteromultimeric potassium channels in Xenopus oocytes. Nature *345*, 530-534.

Kovall, R., and Matthews, B.W. (1997). Toroidal structure of l-exonuclease. Science *277*, 1824-1827.

Liman, E.R., Tytgat, J., and Hess, P. (1992). Subunit stoichiometry of a mammalian K+ channel determined by construction of multimeric cDNAs. Neuron *9*, 861-871.

Martin, A., Baker, T.A., and Sauer, R.T. (2005). Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. Nature *437*, 1115-1120.

Martin, A., Baker, T.A., and Sauer, R.T. (2007). Distinct static and dynamic interactions control ATPase-peptidase communication in a AAA+ protease. Molecular cell *27*, 41-52.

Nicke, A., Rettinger, J., and Schmalzing, G. (2003). Monomeric and dimeric byproducts are the principal functional elements of higher order P2X1 concatamers. Molecular pharmacology *63*, 243-252.

Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., and Gouaux, J.E. (1996). Structure of staphylococcal a-hemolysin, a heptameric transmembrane pore. Science *274*, 1859-1865.

Walker, B.J., Krishnasastry, M., Zorn, L., Kasianowicz, J.J., and Bayley, H. (1992). Functional expression of the a-hemolysin of Staphylococcus aureus in intact Escherichia coli and in cell lysates. JBiolChem *267*, 10902-10909.

Yoshida, M., Muneyuki, E., and Hisabori, T. (2001). ATP synthase--a marvellous rotary engine of the cell. Nature reviews Molecular cell biology *2*, 669-677.

Zhang, J., McCabe, K.A., and Bell, C.E. (2011). Crystal structures of lambda exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. Proceedings of the National Academy of Sciences of the United States of America *108*, 11872-11877.

Zhou, Y., Nelson, M.E., Kuryatov, A., Choi, C., Cooper, J., and Lindstrom, J. (2003). Human alpha4beta2 acetylcholine receptors formed from linked subunits. The Journal of neuroscience : the official journal of the Society for Neuroscience *23*, 9004-9015.

# CHAPTER 5

## CONCLUSIONS AND OUTLOOK

Nature has become adept at protein engineering over the course of evolution, employing many different means to build protein assemblies (Arnold, 1998; Jones and Thornton, 1996; Keskin et al., 2008). These protein assemblies work as nano-machines within the cell and are involved in a variety of tasks such as maintaining the structure of the cell, signal transduction, protein translation, etc. (Alberts, 1998; Finley, 2009; Kadonaga, 1998; Nooren and Thornton, 2003). Employing the design principles found in Nature, synthetic protein assemblies have been constructed (Kim et al., 1996; Moreau et al., 2008; Padilla et al., 2001). The new classes of proteins thus created are more than just the sum of the constituent proteins; while the activity of each protein involved is retained, these protein complexes can be functional for altogether new applications. Following cue, if one assumes that there are only 2000 proteins with unique functions (Govindarajan et al., 1999; Liu et al., 2004), different combinations of any of two proteins will yield ~$10^6$ unique bifunctional proteins. If any three of the 2000 available proteins were combined, ~$10^9$ new proteins could be made. Thus, Nature offers vast opportunities for the design of custom-made protein complexes for any predetermined function.

The transmembrane αHL pore has been extensively studied and its robust structure offers an ideal scaffold for protein engineering. This thesis reports the engineering of two protein complexes of αHL, using two different protein

engineering methodologies- disulfide cross-linking and genetic fusion. A αHL dimer complex, $(\alpha7)_2$ was made by cross-linking two αHL heptamers (α7) *via* disulfide bridges in a cap-to-cap orientation. Molecular dynamics simulations of $(\alpha7)_2$ suggest that the cap-to-cap structure of $(\alpha7)_2$, formed with seven disulfides between two α7 units, is energetically favourable. The biochemical and electrical characterization studies of $(\alpha7)_2$ are consistent with a pore that is longer than α7. As compared to the molecular adapter γCD binding to the α7 β-barrel when added to the *trans* side of the lipid bilayer only (Gu *et al.*, 1999), γCD binds to $(\alpha7)_2$ from both the *cis* and *trans* sides of the bilayer confirming the presence of two functional β-barrels in $(\alpha7)_2$. The structure of $(\alpha7)_2$ was convincingly confirmed by TEM. Further, we have shown that the two transmembrane barrels of $(\alpha7)_2$ can insert into two bilayers simultaneously, and small molecules can be transferred through $(\alpha7)_2$ from an attolitre liposome across a planar lipid bilayer. A theoretical model of $(\alpha7)_2$, which considered the α7 units as series resistors and a leak resistor is in agreement with the experimental values of current blockades due to γCD binding and insertion into a liposome. The electrical model thus confirmed the presence of a leak in the $(\alpha7)_2$ structure. The leak current could arise from a flow of ions between the cap-domains of the two α7 units or between the α1 units of the *cis*-α7 unit.

We propose that the double bilayer penetrating $(\alpha7)_2$ pores could be used a nano-conduit between vesicles (minimal cells) to form large functional networks (minimal tissues) (Fig. 2.2). $(\alpha7)_2$ pores could also be used to link minimal synthetic tissues with biological tissues (Fig. 2.34). Further applications could be extended by engineering $(\alpha7)_2$ pores as size-exclusion filters for the selective

translocation of analytes or for chemical/ optical triggered drug delivery (Chang *et al.*, 1995; Eroglu *et al.*, 2000; Russo *et al.* 1997).

Encouraged by the promise of $(\alpha7)_2$, we then tried to couple a processive exonuclease to the αHL pore so that the construct could be used as a molecular machine for DNA sequencing. The trimeric exonuclease of bacteriophage λ, λ-exo, was chosen for coupling with the αHL pore because of several factors- (i) λ-exo and αHL have a cyclic rotational symmetry (C3 and C7, respectively) around a central pore; (ii) λ-exo is a highly processive exonuclease, which acts on dsDNA, necessitating minimal sample preparation for sequencing long dsDNA substrates; (iii) λ-exo generates a ssDNA strand and nucleotides as products of dsDNA hydrolysis. The λ-exo-αHL pore complex might be used for strand sequencing of the ssDNA product or base sequencing of the nucleotides, which would be fed into the αHL pore in a single-file sequential order (Fig. 3.11) (Astier et al., 2006; Clarke et al., 2009). The construct might also be used to study the activity of λ-exo at the single molecule level.

Several fusion constructs of λ-exo monomer and αHL monomer were made and assayed for activity. The fusion construct with λ-exo at the C-terminus of αHL was found to retain the exonuclease activity of λ-exo, and the hemolytic and pore forming activities of αHL. In the desired complex of λ-exo and αHL, only one λ-exo trimer is required on top of the αHL heptamer. Hence, the major challenge in using genetically fused monomers of λ-exo and αHL was to achieve the asymmetry in the protein stoichiometries (a trimer and a heptamer) in the final complex. This issue was addressed by co-assembling WT αHL

monomers with the λ-exo-αHL fusion constructs to form hetero-heptamers. However, hetero-heptamers containing different numbers of λ-exo-αHL fusion constructs with WT α1 could not be successfully purified.

Therefore, another approach was adopted to build the λ-exo trimer-αHL heptamer complex. A concatemeric construct of λ-exo, exo3, with three λ-exo monomers genetically linked to each other by long flexible linkers was constructed. The exo3 fusion protein with 17 and 15 amino acids long linkers between the λ-exo monomers had full exonuclease activity.

Work on creating a genetic fusion construct of the exo3 protein and αHL monomer (exo3-α1) is in progress. Next, we aim to generate hetero-heptamers of the exo3-α1 construct and WT αHL monomers with only one λ-exo fusion trimer on the cap of αHL. Such hetero-heptamers have been successfully made in other contexts. The λ-exo-αHL pore complex thus created will offer promising prospects for DNA sequencing and single molecule enzyme studies. The exo3 plasmid design permits effortless synthesis of hetero-trimers of λ-exo with one or more inactive λ-exo subunits. Hetero-trimers of λ-exo have not been reported previously and could prove pivotal in understanding the mechanistic basis of the high processivity of λ-exo.

The synthesis of functional αHL hybrid pores has deepened our understanding about the nanopore. Using similar strategies to those described in this thesis, other proteins could be linked to αHL. These novel protein-nanopore complexes could then be used for single molecule enzymatic studies, DNA/ RNA

sequencing, protein translocation and sequencing, and other applications, limited only by the creativity of the experimenter.

# References


Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell *92*, 291-294.

Arnold, F.H. (1998). When blind is better: protein design by evolution. Nature biotechnology *16*, 617-618.

Astier, Y., Braha, O., and Bayley, H. (2006). Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. Journal of the American Chemical Society *128*, 1705-1710.

Clarke, J., Wu, H., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. Nature nanotechnology *4*, 265-270.

Finley, D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. Annual review of biochemistry *78*, 477-513.

Govindarajan, S., Recabarren, R., and Goldstein, R.A. (1999). Estimating the total number of protein folds. Proteins *35*, 408-414.

Jones, S., and Thornton, J.M. (1996). Principles of protein-protein interactions. Proceedings of the National Academy of Sciences of the United States of America *93*, 13-20.

Kadonaga, J.T. (1998). Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. Cell *92*, 307-313.

Keskin, O., Gursoy, A., Ma, B., and Nussinov, R. (2008). Principles of protein-protein interactions: what are the preferred ways for proteins to interact? Chemical reviews *108*, 1225-1244.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences of the United States of America *93*, 1156-1160.

Liu, X., Fan, K., and Wang, W. (2004). The number of protein folds and their distribution over families in nature. Proteins *54*, 491-499.

Moreau, C.J., Dupuis, J.P., Revilloud, J., Arumugam, K., and Vivaudou, M. (2008). Coupling ion channels to receptors for biomolecule sensing. Nature nanotechnology *3*, 620-625.

Nooren, I.M., and Thornton, J.M. (2003). Diversity of protein-protein interactions. The EMBO journal *22*, 3486-3492.

Padilla, J.E., Colovos, C., and Yeates, T.O. (2001). Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proceedings of the National Academy of Sciences of the United States of America *98*, 2217-2221.

# APPENDIX

## pT7 αHL WT

TTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGA
CGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATG
TATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACA
TTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGA
AAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATC
CTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATT
ATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACT
CACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGT
GATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACAT
GGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACA
CCACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGG
CAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTG
GTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTA
AGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
GAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTT
AAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAAC
GTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTG
CGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACC
AACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGT
TAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCT
GCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGG
CTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACA
GGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTG
ACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTT
TACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAAC
CGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGA
GGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTG
CACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGT
CATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTT
ACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGG
CAGCGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGC

CGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGC

CGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCA

GCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTAGCCCGCCTA

ATGAGCGGGCTTTTTTTTAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTC

CCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCAGATTCTGATATTAATATTAAAACC

GGTACTACAGATATTGGAAGCAATACTACAGTAAAAACAGGTGATTTAGTCACTTATGATAAAGAAAATGGCAT

GCACAAAAAAGTATTTTATAGTTTTATCGATGATAAAAATCACAATAAAAAACTGCTAGTTATTAGAACAAAAG

GTACCATTGCTGGTCAATATAGAGTTTATAGCGAAGAAGGTGCTAACAAAAGTGGTTTAGCCTGGCCTTCAGCC

TTTAAGGTACAGTTGCAACTACCTGATAATGAAGTAGCTCAAATATCTGATTACTATCCAAGAAATTCGATTGA

TACAAAAGAGTATATGAGTACTTTAACTTATGGATTCAACGGTAATGTTACTGGTGATGATACAGGAAAAATTG

GCGGCCTTATTGGTGCAAATGTTTCGATTGGTCATACACTGAAATATGTTCAACCTGATTTCAAAACAATTTTA

GAGAGCCCAACTGATAAAAAAGTAGGCTGGAAAGTGATATTTAACAATATGGTGAATCAAAATTGGGGACCATA

CGATCGAGATTCTTGGAACCCGGTATATGGCAATCAACTTTTCATGAAAACTAGAAATGGTTCTATGAAAGCAG

CAGATAACTTCCTTGATCCTAACAAAGCAAGTTCTCTATTATCTTCAGGGTTTTCACCAGACTTCGCTACAGTT

ATTACTATGGATAGAAAAGCATCCAAACAACAAACAAATATAGATGTAATATACGAACGAGTTCGTGATGATTA

CCAATTGCATTGGACTTCAACAAATTGGAAAGGTACCAATACTAAAGATAAATGGACAGATCGTTCTTCAGAAA

GATATAAAATCGATTGGGAAAAAGAAGAAATGACAAATTAATGTAAATTATTTGTACATGTACAAATAAATATA

ATTTATAACTTTAGCCGAAAGCTTGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCA

CCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGA

ACTATATCCGGATAATTCGAGCTCGGTACCCACCCCGGTTGATAATCAGAAAAGCCCCAAAAACAGGAAGATTG

TATAAGCAAATATTTAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCA

TTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGT

TGTTCCAGTTTGGAACAAGAGTCCAGTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATC

AGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAAT

CGGAACCCTAAAGGGATGCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAA

GAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCG

CGCTTAATGCGCCGCTACAGGGCGCGTGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTATCCCGCAAG

AGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAG

CGCATTGTTAGATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCT

AGCTTATCGATGATAAGCTGTCAAA

The cloning sites, *NdeI* and *HindIII*, are in red bold type and underlined. αHL gene is shown in blue. Stop codon is shown in green. Binding sites of T7F and T7R primers are highlighted in green and binding sites of SC46/ 47 in cyan.

# pT7 H6-λ-exo

TTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGA
CGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATG
TATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACA
TTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGA
AAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATC
CTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATT
ATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACT
CACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGT
GATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACAT
GGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACA
CCACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGG
CAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTG
GTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTA
AGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
GAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTT
AAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAAC
GTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTG
CGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACC
AACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGT
TAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCT
GCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGG
CTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACA
GGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTG
ACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTT
TACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAAC
CGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGA
GGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTG
CACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGT
CATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTT
ACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGG
CAGCGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGC
CGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGC
CGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCA
GCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTAGCCCGCCTA
ATGAGCGGGCTTTTTTTTAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTC
CCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATA**CATATG**CACCATCACCACCATCAT**T**CCATGGGA
acaccggacattatcctgcagcgtaccgggatcgatgtgagagctgtcgaacaggggatgatgcgtggcacaa
attacggctcggcgtcatcaccgcttcagaagttcacaacgtgatagcaaaaccccgctccggaaagaagtggc
ctgacatgaaaatgtcctacttccacaccctgcttgctgaggtttgcaccggtgtggctccggaagttaacgct
aaagcactggcctggggaaaacagtacgagaacgacgccagaaccctgtttgaattcacttccggcgtgaatgt
tactgaatccccgatcatctatcgcgacgaaagtatgcgtaccgcctgctctcccgatggtttatgcagtgacg
gcaacggccttgaactgaaatgcccgtttacctcccgggatttcatgaagttccggctcggtggtttcgaggcc
ataaagtcagcttacatggcccaggtgcagtacagcatgtgggtgacgcgaaaaaatgcctggtactttgccaa
ctatgacccgcgtatgaagcgtgaaggcctgcattatgtcgtgattgagcgggatgaaaagtacatggcgagtt
tttgacgagatcgtgccggagttcatcgaaaaaatggacgaggcactggctgaaattggttttgtatttggggag
caatggcgatg**AAGCTT**GGATCCTGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCA
CCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGA
ACTATATCCGGATAATTCGAGCTCGGTACCCACCCCGGTTGATAATCAGAAAAGCCCCAAAAACAGGAAGATTG
TATAAGCAAATATTTAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCA
TTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGT
TGTTCCAGTTTGGAACAAGAGTCCAGTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATC
AGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAAT
CGGAACCCTAAAGGGATGCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAA
GAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCG

CGCTTAATGCGCCGCTACAGGGCGCGTGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTATCCCGCAAG
AGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAG
CGCATTGTTAGATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCT
AGCTTATCGATGATAAGCTGTCAAACATGAGAA

The cloning sites, *NdeI* and *HindIII*, are in red bold type and underlined. *NcoI* site is also underlined. His-6 tag at N-terminus is shown in **pink.** Exo gene is shown in violet.

## Exo3_3 gene

Exo3_3 gene is contained within the pT7 vector between NdeI and HindIII restriction sites.

CATATGCACCATCACCACCATCATTCCATGGGAacaccggacattatcctgcagcgtaccgggatcgatgtgag
agctgtcgaacaggggatgatgcgtggcacaaattacggctcggcgtcatcaccgcttcagaagttcacaacg
tgatagcaaaaccccgctccggaaagaagtggcctgacatgaaaatgtcctacttccacaccctgcttgctgag
gtttgcaccggtgtggctccggaagttaacgctaaagcactggcctggggaaaacagtacgagaacgacgccag
aaccctgtttgaattcacttccggcgtgaatgttactgaatccccgatcatctatcgcgacgaaagtatgcgta
ccgcctgctctcccgatggtttatgcagtgacggcaacggccttgaactgaaatgcccgtttacctcccgggat
ttcatgaagttccggctcggtggtttcgaggccataaagtcagcttacatggcccaggtgcagtacagcatgtg
ggtgacgcgaaaaaatgcctggtactttgccaactatgacccgcgtatgaagcgtgaaggcctgcattatgtcg
tgattgagcgggatgaaaagtacatggcgagttttgacgagatcgtgccggagttcatcgaaaaaatggacgag
gcactggctgaaattggttttgtatttggggagcaatggcgaCTCGAGtCtaccGGaagctcgggtactatgtc
tAgtggaagtaccACTAGTGGAacaccggacattatcctgcagcgtaccgggatcgatgtgagagctgtcgaac
agggggatgatgcgtggcacaaattacggctcggcgtcatcaccgcttcagaagttcacaacgtgatagcaaaa
ccccgctccggaaagaagtggcctgacatgaaaatgtcctacttccacaccctgcttgctgaggtttgcaccgg
tgtggctccggaagttaacgctaaagcactggcctggggaaaacagtacgagaacgacgccagaaccctgtttg
aattcacttccggcgtgaatgttactgaatccccgatcatctatcgcgacgaaagtatgcgtaccgcctgctct
cccgatggtttatgcagtgacggcaacggccttgaactgaaatgcccgtttacctcccgggatttcatgaagtt
ccggctcggtggtttcgaggccataaagtcagcttacatggcccaggtgcagtacagcatgtgggtgacgcgaa
aaaatgcctggtactttgccaactatgacccgcgtatgaagcgtgaaggcctgcattatgtcgtgattgagcgg
gatgaaaagtacatggcgagttttgacgagatcgtgccggagttcatcgaaaaaatggacgaggcactggctga
aattggttttgtatttggggagcaatggcgaTGTACACTTGGAGGCACCAGTGGAGGTAGTGGTGGATCATCCT
AGGGAacaccggacattatcctgcagcgtaccgggatcgatgtgagagctgtcgaacaggggatgatgcgtgg
cacaaattacggctcggcgtcatcaccgcttcagaagttcacaacgtgatagcaaaaccccgctccggaaagaa
gtggcctgacatgaaaatgtcctacttccacaccctgcttgctgaggtttgcaccggtgtggctccggaagtta
acgctaaagcactggcctggggaaaacagtacgagaacgacgccagaaccctgtttgaattcacttccggcgtg
aatgttactgaatccccgatcatctatcgcgacgaaagtatgcgtaccgcctgctctcccgatggtttatgcag
tgacggcaacggccttgaactgaaatgcccgtttacctcccgggatttcatgaagttccggctcggtggtttcg
aggccataaagtcagcttacatggcccaggtgcagtacagcatgtgggtgacgcgaaaaaatgcctggtacttt
gccaactatgacccgcgtatgaagcgtgaaggcctgcattatgtcgtgattgagcgggatgaaaagtacatggc
gagttttgacgagatcgtgccggagttcatcgaaaaaatggacgaggcactggctgaaattggttttgtatttg
gggagcaatggcgatgAAGCTT

## Exo3_4 gene

CATATGCACCATCACCACCATCATTCCATGGGAacaccggacattatcctgcagcgtaccgggatcgatgtgag
agctgtcgaacaggggatgatgcgtggcacaaattacggctcggcgtcatcaccgcttcagaagttcacaacg
tgatagcaaaaccccgctccggaaagaagtggcctgacatgaaaatgtcctacttccacaccctgcttgctgag
gtttgcaccggtgtggctccggaagttaacgctaaagcactggcctggggaaaacagtacgagaacgacgccag
aaccctgtttgaattcacttccggcgtgaatgttactgaatccccgatcatctatcgcgacgaaagtatgcgta
ccgcctgctctcccgatggtttatgcagtgacggcaacggccttgaactgaaatgcccgtttacctcccgggat
ttcatgaagttccggctcggtggtttcgaggccataaagtcagcttacatggcccaggtgcagtacagcatgtg
ggtgacgcgaaaaaatgcctggtactttgccaactatgacccgcgtatgaagcgtgaaggcctgcattatgtcg
tgattgagcgggatgaaaagtacatggcgagttttgacgagatcgtgccggagttcatcgaaaaaatggacgag
gcactggctgaaattggttttgtatttggggagcaatggcgaCTCGAGtCtaccGGaagctcgggtactatgtc
tAgtggaagtaccACTAGTGGAacaccggacattatcctgcagcgtaccgggatcgatgtgagagctgtcgaac
agggggatgatgcgtggcacaaattacggctcggcgtcatcaccgcttcagaagttcacaacgtgatagcaaaa
ccccgctccggaaagaagtggcctgacatgaaaatgtcctacttccacaccctgcttgctgaggtttgcaccgg
tgtggctccggaagttaacgctaaagcactggcctggggaaaacagtacgagaacgacgccagaaccctgtttg
aattcacttccggcgtgaatgttactgaatccccgatcatctatcgcgacgaaagtatgcgtaccgcctgctct
cccgatggtttatgcagtgacggcaacggccttgaactgaaatgcccgtttacctcccgggatttcatgaagtt
ccggctcggtggtttcgaggccataaagtcagcttacatggcccaggtgcagtacagcatgtgggtgacgcgaa

aaaatgcctggtactttgccaactatgacccgcgtatgaagcgtgaaggcctgcattatgtcgtgattgagcgg
gatgaaaagtacatggcgagttttgacgagatcgtgccggagttcatcgaaaaaatggacgaggcactggctga
aattggttttgtatttggggagcaatggcgaTGTACACTTGGAGGCACCAGTGGAGGTAGTGGTGGATCATCGT
ACGGAacaccggacattatcctgcagcgtaccgggatcgatgtgagagctgtcgaacaggggatgatgcgtgg
cacaaattacggctcggcgtcatcaccgcttcagaagttcacaacgtgatagcaaaaccccgctccggaaagaa
gtggcctgacatgaaaatgtcctacttccacaccctgcttgctgaggtttgcaccggtgtggctccggaagtta
acgctaaagcactggcctggggaaaacagtacgagaacgacgccagaaccctgtttgaattcacttccggcgtg
aatgttactgaatccccgatcatctatcgcgacgaaagtatgcgtaccgcctgctctcccgatggtttatgcag
tgacggcaacggccttgaactgaaatgcccgtttacctcccgggatttcatgaagttccggctcggtggtttcg
aggccataaagtcagcttacatggcccaggtgcagtacagcatgtgggtgacgcgaaaaaatgcctggtacttt
gccaactatgacccgcgtatgaagcgtgaaggcctgcattatgtcgtgattgagcgggatgaaaagtacatggc
gagttttgacgagatcgtgccggagttcatcgaaaaaatggacgaggcactggctgaaattggttttgtatttg
gggagcaatggcgatgAAGCTT

Restriction sites are colour coded as:
**NdeI** **NcoI** **XhoI** **SpeI** **AauI/ BsrGI** **AvrII** **BsiWI**
The WT exo3_3 contains a His-6 tag at the N-terminus. The first λ-exo gene is highlighted in yellow. The SG14 linker inserted in between the first and the second λ-exo genes is underlined.

**Table 1.** Table lists the primer sequences used for site-directed mutagenesis and for synthesizing the genetic fusion constructs of λ-exo and αHL.

| Mutant | Forward primer (F) | Reverse primer (R) |
|---|---|---|
| | | |
| Amp | ATAAAGTTGCAGGACCACTTCTG | CAGAAGTGGTCCTGCAACTTTAT |
| | | |
| T7 | GTAATACGACTCACTATAGGG | GCTAGTTATTGCTCAGCGG |
| | | |
| **λ-exo** | | |
| | | |
| R226C | GGGGAGCAATGGTGCTGAAGCTT GGATCCTGGATCCGGC | GCCGGATCCAGGATCCAAGCTT CAGCACCATTGCTCCCC |
| | | |
| E+C5 | GAAGAAATGACAAATTGAAGCTTG GATCCTGGATCCGGCTGC | GCTTCAATTTGTCATTTCTTCTCGC CATTGCTCCCCAAATACAAAACC |
| | | |
| C116S | ccgccTCCtctcccgatggtttatgcagtgacg | cgtcactgcataaaccatcgggagaGGAggcgg |
| | | |
| C132S | ctgaaaTCCccgtttacctcccggg | cccgggaggtaaacggGGAtttcag |
| | | |
| Q83E/ N86D | GAAAAGAGTACGAGGACGACGCC AGAACCCTGTTTGAATTCAC | GTGAATTCAAACAGGGTTCTGGCGT CGTCCTCGTACTCTTTTC |
| | | |
| E_H6 | CACCATCACCACCATCATTGAAGC TTGGATCCTGGATCCGGCTGC | CTTCAATGATGGTGGTGATGGTGTC GCCATTGCTCCCCAAATACAAAACC |
| | | |
| E_act | 5’ P CATATGGCAGATTCTGATATTAATA TTAAAACCGGTACTAC | 5’ P AGCGGTGGCAGCAGCCAACTCAGC TTC |

| | | |
|---|---|---|
| **αHL** | | |
| | | |
| K237C | CTATGGATAGATGCGCATCCAAACAACAAACAAATATAG | CTATATTTGTTTGTTGTTTGGATGCGCATCTATCCATAG |
| | | |
| E287C | GATTGGTGCAAAGAAGAAATGACAAATGATGACG | CGTCATCATTTGTCATTTCTTCTTTGCACCAATC |
| | | |
| αHL4A$^{+}$ | GCGTGGGCGAAAGCGGCGATGACAAATGATGACGATGATGACGAC | CGCCGCTTTCGCCCACGCGATTTTATATCTTTCTGAAGAACGATCT |
| | | |
| αHLΔC5 | CGATTGGGAAAAATAAGAAATGACAAATTAATGTAAATTATTTGTACATGTAC | GTACATGTACAAATAATTTACATTAATTTGTCATTTCTTATTTTTCCCAATCG |
| | | |
| **αHL-λ-exo** | | |
| | | |
| AE10 | GGAGGCTCGAGTsGGAGGTAGTGGTGGATCATCCATGGGAACACCG | TGATCCACCACTACCTCCACTCGAGCCTCCATTTGTCATTTCTTCTTTTTCCC |
| | | |
| AE13 | GGAGGCTCGAGTGGAGGTAGTGGTGGATCAGGCTCCTCATCCATGGGAACACCG | TGAGGAGCCTGATCCACCACTACCTCCACTCGAGCCTCCATTTGTCATTTCTTCTTTTTCCC |
| | | |
| EA10 | GGAGGCTCGAGTGGAGGTAGTGGTGGATCAGCAGATTCTGATATTAATATTAAAACCGG | TGATCCACCACTACCTCCACTCGAGCCTCCTCGCCATTGCTCCCC |
| | | |
| EA13 | GGAGGCTCGAGTGGAGGTAGTGGTGGATCAGGCTCCTCAGCAGATTCTGATATTAATATTAAAACCGG | TGAGGAGCCTGATCCACCACTACCTCCACTCGAGCCTCCTCGCCATTGCTCCCC |
| | | |
| Exo3 | | |
| | | |
| E_XS | CTCGAGGGTTCAactagtGGAacaccggacattatcctg | actagtTGAACCCTCGAGtcgccattgctcccccaaatac |
| | | |
| SG10Hind | | TGATCCACCACTACCTCCACTCGAG |

| | | |
|---|---|---|
| 3R | | CCTCCAAGCTTCCTCGCCATTGCTC |
| | | |
| NA_F | TGGAGGTAGTGGTGGATCATCCTAGGGAACAC | |
| | | |
| HAXB R1L | | CCACCACTACCTCCACTGGTGCCTCCAAGCTGTACACGCC |
| | | |
| SG14 linker | ACTCACactagtggtacttcctcaacatatagtaccggatcttcaggtacatctactagtAaTaAt | aTtAtTactagtagatgtacctgaagatccggtactatatgttgaggaagtaccactagtGTGAGT |
| | | |

*The road to wisdom?*

*Well, it's plain and simple to express:*

*Err and err and err again*

*but less and less and less.*

-Piet Hein