HAL
open science

Protein symmetrization as a novel tool in structural biology

Francesca Coscia

To cite this version:

Francesca Coscia. Protein symmetrization as a novel tool in structural biology. Structural Biology [q-bio.BM]. Université de Grenoble, 2014. English. NNT : 2014GRENV066 . tel-01344632

HAL Id: tel-01344632

https://theses.hal.science/tel-01344632

Submitted on 12 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 7 août 2006

Présentée par

Francesca COSCIA

Thèse dirigée par Dr **Carlo PETOSA**
codirigée par Dr **Guy SCHOEHN**

Préparée au sein de l'Institut de Biologie Structurale JP Ebel
dans **l'École Doctorale de chimie et science du vivant**

Protein symmetrization as a novel tool in structural biology

(La symétrisation des protéines: un nouvel outil pour la biologie structurale)

Thèse soutenue publiquement le **4 Décembre 2014**

devant le jury composé de :

Prof Elena, ORLOVA
Birkbeck, University of London, London (Rapporteur)
Prof Anthony, WATTS
University of Oxford, Oxford (Rapporteur)
Dr Stephane, BRESSANELLI
Laboratoire de Virologie Moléculaire et Structurale, Gif-sur-Yvette (Examinateur)
Dr Irina, GUTSCHE
Unit for Virus Host-Cell Interactions, Grenoble (Examinateur)
Dr Carlo, PETOSA
Institut de Biologie Structurale JP Ebel, Grenoble (Examinateur)

Université Joseph Fourier / Université Pierre Mendès France /
Université Stendhal / Université de Savoie / Grenoble INP

On fait la science avec des faits comme une maison avec des pierres ; mais une accumulation de faits n'est pas plus une science qu'un tas de pierres n'est une maison.

Science is built up with facts, as a house is with stones.
But a collection of facts is no more a science than a heap of stones is a house.

Jules Henri Poincaré

La science, mon garçon, est faite d'erreurs, mais d'erreurs qu'il est bon de commettre, car elles mènent peu à peu à la vérité.

Science, my lad, has been built upon many errors;
but they are errors which it was good to fall into, for they led to the truth.

Jules Verne

Table of contents

LIST OF ABBREVIATIONS **10**

Abstract: 12

Résumé 14

1. INTRODUCTION **17**

Abstract 19

Résumé 19

1.1 Current challenges in structural biology 20

1.2 Use of Redundancy to acquire noise-free averages : X-ray crystallography and cryoEM 23

1.3 Improvement of resolution achieved by single particle cryoEM over the last decades 28

1.4 Interplay between protein engineering and structural biology 30

1.5 Macromolecular engineering methods to overcome size limits in cryoEM 34

1.6 The protein symmetrization method: aim and overview of the thesis 37

2. RESULTS **41**

Abstract 43

Résumé 43

2.1 PROTEIN SYMMETRIZATION: STRATEGY AND BUILDING BLOCKS **44**

2.1.1 Templates 47

2.1.2 Targets: 51

2.2 SCREENING OF DIFFERENT TARGET-TEMPLATE COMBINATIONS **53**

2.2.1 Importin-β - GS fusions 53

2.2.2 GS fusions with globular targets 64

2.2.3 Mbp-E2 fusions 71

2.3 OPTIMIZATION OF THE LINKER LENGTH IN MBP-GS FUSIONS **76**

2.3.1 Biophysical characterization 77

2.3.2 Negative stain EM analysis 81

2.4 CRYOEM ANALYSIS OF MAGΔ5 AND MAGΔ8 ..... **83**

2.4.1 CryoEM sample optimization 83
2.4.2 CryoEM reconstruction of Mbp-GS chimeras 85
2.4.3 Refinement of the cryoEM volumes 89
2.3.4 Local resolution estimation of cryoEM maps 94
2.5 CRYSTALLOGRAPHIC STUDIES OF MAGΔ5 98

3. DISCUSSION AND CONCLUDING REMARKS 107
Abstract 109
Résumé 109
3.1 PROTEIN SYMMETRIZATION FEASIBILITY STUDIES: EXPLORATORY SCREENING 110
3.2 LINKER OPTIMIZATION IN THE HELIX-BASED STRATEGY 113
3.2.1 Biophysical analysis of MagΔN constructs 113
3.2.2 Validation of protein symmetrisation 114
3.2.3 Experimental data rationalized by *in silico* modelling 115
3.3 LINKER OPTIMIZATION IN UNCONSTRAINED CONNECTION STRATEGY 119
3.4 PERSPECTIVES FOR FUTURE APPLICATIONS 122
3.5 CONCLUSIONS 123

4. EXPERIMENTAL AND COMPUTATIONAL PROCEDURES 125
Abstract 127
Résumé 127
4.1 *IN SILICO* MODELLING OF HELIX-BASED FUSIONS 128
4.2 CLONING 130
4.2.1 Cloning strategies and recombinant expression in *E.coli* 130
4.2.2 Importin-β – GS cloning 132
4.3 PROTEIN EXPRESSION AND PURIFICATION 132
4.3.1 Recombinant expression and purification of templates and of globular protein fusions 132
4.3.2 Importin-β – GS expression and purification in *E.coli* 133
4.4 BIOPHYSICAL CHARACTERIZATION OF CHIMERIC CONSTRUCTS 134
4.4.1 Size exclusion chromatography 134
4.4.2 Native polyacrylamide gel electrophoresis 136

4.4.3 Thermal shift assay 136
4.4.4 Dynamic light scattering 137
4.4.5 Fluorescence polarization binding assay 138
4.5 ELECTRON MICROSCOPY 139
4.5.1 Sample preparation and data collection 139
4.5.2 Image processing 142
4.5.3 Resolution estimation and map visualization 151
ACKNOWLEDGEMENTS 154
REFERENCES 155

LIST OF ABBREVIATIONS

BOG	β-D-Octyl glucoside
CC	Cross correlation
CCD	Charged couped device
CryoEM	Cryo-electron microscopy
CTF	Contrast transfer function
DLS	Dynamic light scattering
DTT	Dithiothreitol
E.coli	Escherichia coli
E2	Dihydrolipoyl transacetylase of pyruvate dehydrogenase complex
EDTA	Ethylenediaminetetra-acetic acid
EM	Electron microscopy
FP	Fluorescence polarization
FSC	Fourier shell correlation
GFP	Green fluorescent protein
GS	Glutamine Synthetase
Gsat	Glutamine Synthetase adenylyltransferase
His-tag	histidine tag
Impα	Importin alpha
Impβ	Importin beta
IPTG	Isopropil-β-D-1-tiogalactopyranoside
Kpr	Ketopanthoate reductase
L9	helical linker from the homonimous ribosomal protein (PDB ID code 1div)
LB	Lysogeny broth
mae	Mbp-E2 fusion
mag	Mbp-GS fusion
Mbp	Maltose binding protein
MW	Molecular weight
Ni-NTA	Nichel (II) - Nitrilotriacetic acid
NLS	Nuclear localization signal
NMR	Nuclear magnetic resonance
NPC	Nuclear pore complex
PCR	Polymerase chain reaction
PD	Polydispersity index
PDB	Protein data bank
RF cloning	Restriction free cloning

R_h	Hydrodynamic radius
SDS-PAGE	Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis
SEC	Size exclution chromatography
SER	Surface entropy reduction
SNR	Signal to noise ratio
SST	$Na_4O_4OSiW_{12}$ Sodium Silico Tungstate
TEM	Transmission electron microscopy
TLS	Translation/Libration/Screw
Trea	Trealase
TRIS	tris(hydroxymethyl)aminomethane
TSA	Thermal shift assay

ABSTRACT:

Structural determination of proteins at atomic level resolution is crucial for unravelling their function. X-ray crystallography has successfully been used to determine macromolecular structures with sizes ranging from kDa to MDa, and currently remains the most efficient method for the high-resolution structure determination of monomeric proteins within the 40-100 kDa range. However, this method is limited by the ability to grow well diffracting crystals, which is problematic for several targets, such as membrane proteins. Single particle cryo electron microscopy (cryoEM) allows near atomic (3-4Å) resolution structural determination of large, preferably symmetric, assemblies in solution. Biological molecules scatter electrons weakly and, to avoid radiation damage, only low electron doses can be used during imaging. Consequently, raw cryoEM images are extremely noisy. However, averaging many molecular images aligned in the same orientation permits one to increase the signal-to-noise ratio, ultimately allowing the achievement of a 3D electron density map of the molecule of interest. Nevertheless, as the molecular size and degree of symmetry decrease, the individual images loose adequate features for accurate alignment. Currently, cryoEM analysis is practically impossible for monomeric proteins below ~100 kDa in mass. We propose to circumvent this obstacle by fusing such monomeric target proteins to a homo-oligomeric protein (template), thereby generating a self-assembling particle whose large size and symmetry should facilitate cryoEM analysis. In the present thesis we seek to test and demonstrate the feasibility of this 'protein symmetrization' approach and to evaluate its usefulness for protein structure determination. To set up the pilot study we combined selected targets of known structure with two templates: Glutamine Synthetase (GS), a 12-mer with D6 symmetry and a helical N-terminus, and the E2 subunit of the pyruvate dehydrogenase complex, a 60-mer with icosahedral symmetry and an unstructured N-terminus. After recombinant production in E.coli we identified by negative stain EM a promising dodecameric chimera for structural analysis, comprising maltose binding protein (Mbp) connected to GS by a tri-alanine linker (denoted "Mag"). In order to optimize sample homogeneity we produced a panel of Mag deletion constructs by sequentially truncating the 17 residues between the Mbp and GS domains. A combination of biophysical techniques (thermal shift assay, dynamic light scattering, size exclusion chromatography) and negative stain EM allowed us to select the best candidate for cryoEM analysis,

MagΔ5. By enforcing D6 symmetry we obtained a cryoEM map with a resolution of 10Å (FSC 0.5 criterion). The electron density of the symmetrized 40 kDa Mbp presents shape and features corresponding to the known atomic structure. In particular, the catalytic pocket and specific α-helical elements are distinguishable. The cryoEM map is additionally validated by a 7Å crystal structure of the MagΔ5 oligomer. The presence of a continuous helical connection between target (Mbp) and template (GS) likely contributed to the conformational homogeneity of MagΔ5. Moreover, comparing MagΔ5 with other chimeras studied in this work suggests that a large buried surface area and favorable interactions between the target and template limit the flexibility of the chimera and improve its resolution by cryoEM. For the symmetrization of a target of unknown structure, we envisage proceeding by a trial and error approach by fusing it to a panel of templates with helical termini and different surface properties, and subsequently selecting the best ones using biophysical assays.

In conclusion, the present work establishes the proof-of-concept that protein symmetrization can be used for the structure determination of monomeric proteins below 100 kDa by cryoEM, thereby providing a promising new tool for analyzing targets resistant to conventional structural analysis.

RÉSUMÉ

La détermination de la structure des protéines à une résolution atomique est cruciale pour la compréhension de leur fonction cellulaire. Actuellement, la cristallographie aux rayons X est la méthode la plus efficace pour la détermination à haute résolution de la structure de protéines monomériques allant 40 et 100 kDa. Par contre, elle est limitée par la croissance de cristaux de bonne qualité, qui est problématique pour nombreuses cibles. La cryo-microscopie électronique (cryoME) permet la détermination structurale à résolution quasi-atomique de larges structures protéiques, de préférence symétrique et en solution. Cependant, les images de cryoME sont très bruitées, car une faible dose d'électrons est appliquée de manière à limiter les dommages d'irradiation. En moyennant des dizaines d'images correspondant à la même orientation moléculaire, le rapport signal sur bruit est amélioré. La combinaison des images moyennées de plusieurs orientations permet l'obtention d'une carte de densité électronique 3D de la molécule d'intérêt. Si la taille et la symétrie de la molécule diminuent, l'analyse cryoME devient de moins en moins précise, il est alors impossible d'analyser des protéines monomériques de taille inférieure à 100 kDa. Le but de ce travail a été de développer une nouvelle approche pour réduire cette limite de poids moléculaire. Elle consiste à fusionner la protéine d'intérêt (cible) à une matrice homo-oligomérique, générant une particule symétrique et de taille importante adaptée à l'analyse par cryoME. Dans cette thèse, nous avons cherché à tester et démontrer la faisabilité de cette approche de symétrisation en utilisant des protéines cibles de structure connue.

Pour mettre en place notre étude pilote, nous avons choisi différentes combinaisons de cibles et de matrices connectées par des peptides de liaison (linker) de longueur différentes. Nous avons caractérisé les fusions exprimées en bactéries par microscopie électronique après coloration négative et par plusieurs techniques biophysiques. Grace à ces techniques, nous avons trouvé que la meilleure combinaison est la fusion entre la protéine matrice glutamine synthétase (GS), un 12-mer de symétrie D6 et la cible *maltose binding protein* (Mbp), connectées par un linker contenant trois alanines, que nous avons appelée « Mag ». En jouant sur la longueur du linker nous avons ensuite sélectionné la fusion la plus compacte pour l'analyse cryoME: MagΔ5. Nous avons obtenu la carte cryoME à 10 Å de MagΔ5, qui présente une bonne corrélation avec les modèles atomiques de Mbp et GS. Plus particulièrement, le site catalytique et quelques hélices α

sont identifiables. Ces résultats sont confirmés par l'étude cristallographique que nous avons conduite sur MagΔ5. L'ensemble de ce travail souligne que la présence d'une grande interface d'interactions cible-matrice stabilise la fusion et améliore la résolution en cryoME. Pour la symétrisation d'une cible inconnue, nous envisageons la même procédure expérimentale que celle développée pour MagΔ5. La matrice et le linker les plus adaptés devront être identifiés en utilisant les mêmes méthodes biophysiques.

En conclusion, ce travail établit la preuve de concept que la méthode de symétrisation des protéines permet la détermination de la structure de protéines de poids moléculaire inférieur à 100 kDa par cryoME. Cette méthode a le potentiel d'être un nouvel outil prometteur, qui faciliterait l'analyse de cibles résistantes à l'analyse structurale conventionnelle.

1. INTRODUCTION

ABSTRACT

The structural determination of proteins at atomic level is crucial for understanding their cellular function. Nuclear magnetic resonance allows one to solve the structures of proteins below ~40 kDa in size. X-ray crystallography covers a wider range, from a few kDa to MDa. However, the growth of well diffracting crystals is problematic for many targets, such as membrane proteins, although several methods for protein engineering have been developed to improve crystallogenesis. Cryo-electron microscopy (cryoEM) allows the structure determination at quasi-atomic resolution of large, preferably symmetrical, macromolecular assemblies in solution. However, it is not suitable for most monomeric proteins of biomedical interest, that are below ~ 100 kDa and asymmetrical. To overcome this hurdle we want to fuse the monomeric target of interest to a homo-oligomeric template, thereby increasing the size and symmetry of the imaged particle. The aim of the thesis is to demonstrate a proof-of-concept of protein symmetrization.

RÉSUMÉ

La détermination de la structure des protéines à résolution atomique est cruciale pour la compréhension de leur fonction. La résonance magnétique nucléaire permet de résoudre la structure des protéines de taille inférieure à 40 kDa. La cristallographie aux rayons X couvre une large gamme de tailles. Même si plusieurs méthodes d'ingénierie des protéines ont été développées pour favoriser la cristallogénèse, la croissance des cristaux de bonne qualité reste problématique pour de nombreuses cibles, telles que les protéines membranaires. La cryo-microscopie électronique (cryoME) permet la détermination au niveau quasi-atomique de gros complexes macromoléculaires, de préférence symétriques. Par contre, elle n'est pas adaptée à la plupart des protéines monomériques qui ont des tailles inférieures à environ 100 kDa et sont asymétriques. Pour résoudre ce problème, nous envisageons la fusion d'une cible monomérique d'intérêt à une matrice homo-oligomérique, ce qui va augmenter sa taille et symétrie.

1.1 CURRENT CHALLENGES IN STRUCTURAL BIOLOGY

Cellular complexity is mainly determined by the interplay of sophisticated protein machineries with other cellular components, dictated by the nucleotide sequence of the encoding genes (Alberts, 1998). Interestingly, the diversity of biochemical processes results from the combination of a handful of twenty evolutionarily selected amino acid building blocks. Polypeptide chains fold into higher level structures (helices and sheets), that are arranged in a modular fashion by generating an extraordinary variety of objects of different size (mostly 1-100 nm) and shapes, mainly asymmetrical (Goodsell and Olson, 1993). Pockets, clefts, hinges, pores and surfaces with different physico-chemical properties determine specific binding sites that carry on catalysis, movement, cell adhesion, electron and ion transport, cellular response, control of nucleic acid synthesis and host pathogen interactions, to mention just a few protein functions. An important way to tackle biological problems, fight disease and have a deeper understanding of cellular mechanisms is to decipher how such functions are determined by three-dimensional arrangements at the atomic level, i.e., structural biology.

The early days of structural biology were essentially the realm of X-ray crystallography. However, since the first crystallographic structures of myoglobin and hemoglobin, rewarded by the Nobel prize in Chemistry in 1962 (Kendrew, 1959; Kendrew et al., 1960; Pellam and Harker, 1962; Perutz et al., 1960), the whole field has made a considerable leap forward: the number of protein structures deposited in the Protein Data Bank (PDB) has increased exponentially since this database was established. The over 100,000 entries currently present in the PDB have elucidated a number of biological mechanisms and allowed for the design of diagnostic tools and structure-based drugs (Zheng et al., 2014). This have been possible thanks to methodological and technological advances, that have improved crystallography and made new techniques valuable tools for obtaining detailed structural models. Nowadays, mainly three high resolution techniques exist in structural biology, covering different molecular sizes. Nuclear Magnetic Resonance (NMR) allows atomic structural determination from samples either in highly concentrated solutions (~mM) or in solid state. However, as the number of amino acid residues increases, the resonance peaks become copious and overlap, complicating the structure determination of proteins

above ~40 kDa, although huge developments are currently ongoing in the field (Luca et al., 2003; Yu, 1999). Single particle cryo-electron microscopy (cryoEM) allow structural analysis of macromolecules and complexes larger than ~100 kDa, starting from relatively dilute solutions (~50 nM). X-ray crystallography covers in principle an unlimited range of sizes from small molecules to assemblies of several megadaltons, and generally speaking (although there are some exceptions) remains the most efficient method for the high-resolution structure determination of monomeric proteins within the 40-100 kDa range. However, this method is limited by the growth of high quality crystals, which is a stochastic process and can be an insurmountable problem.

One group of proteins particularly resistant to crystallographic analysis are membrane proteins. These proteins are involved in a number of important cellular processes, such as signal transduction, bacterial secretion and virulence, and the budding of viral capsids from host cell membrane, to name only a few. Because membrane proteins have large hydrophobic surfaces, their isolation and purification requires the use of detergents or other amphiphilic compounds that often hinder the achievement of well-packed crystals. Although recently developed crystallization techniques (*in meso* crystallization with lipidic and sponge phases) have been successfully applied to a class of small flexible membrane proteins (G-protein coupled receptors, GPCRs), the crystallization of this class of proteins remains nevertheless challenging and their structures remain under-represented in the PDB (Caffrey and Cherezov, 2009; Cherezov, 2011).

Another group of proteins that are challenging to crystallize and are of particular interest to the lab where I carried out the present PhD project are members of the importin β family of nuclear transport factors. The prototypical member of this family is Importin-β (Impβ), which together with its binding partner Importin-α (Impα) mediates the nuclear import of proteins bearing a classical nuclear localization signal (NLS). Impβ is a ~97 kDa protein composed of 19 tandem helical hairpin motifs called HEAT repeats, which adopt a superhelical or "solenoid" structure (Cingolani et al., 1999) exposing hydrophobic patches on the surface. This feature allows for the adaptive binding of a variety of cargos upon substantial conformational changes, and their shuttling through the Nuclear Pore Complex (NPC) (Stewart, 2007). First, Impα

tightly associates with Impβ through the N-terminal Importin-β binding (IBB) domain via numerous electrostatic interactions. Subsequently, the Impβ-Impα heterodimer interacts with cytosolic proteins bearing a nuclear localization signal to mediate their transport into the nucleus (Cingolani et al., 1999). In contrast, the HIV proteins Tat and Rev, whose import into the nucleus is required for viral replication, are imported by directly binding to Impβ (Truant and Cullen, 1999) (Figure 1.1). Therefore, the Impβ/Tat and Impβ/Rev complexes are potential targets for the rational design of drugs against HIV.

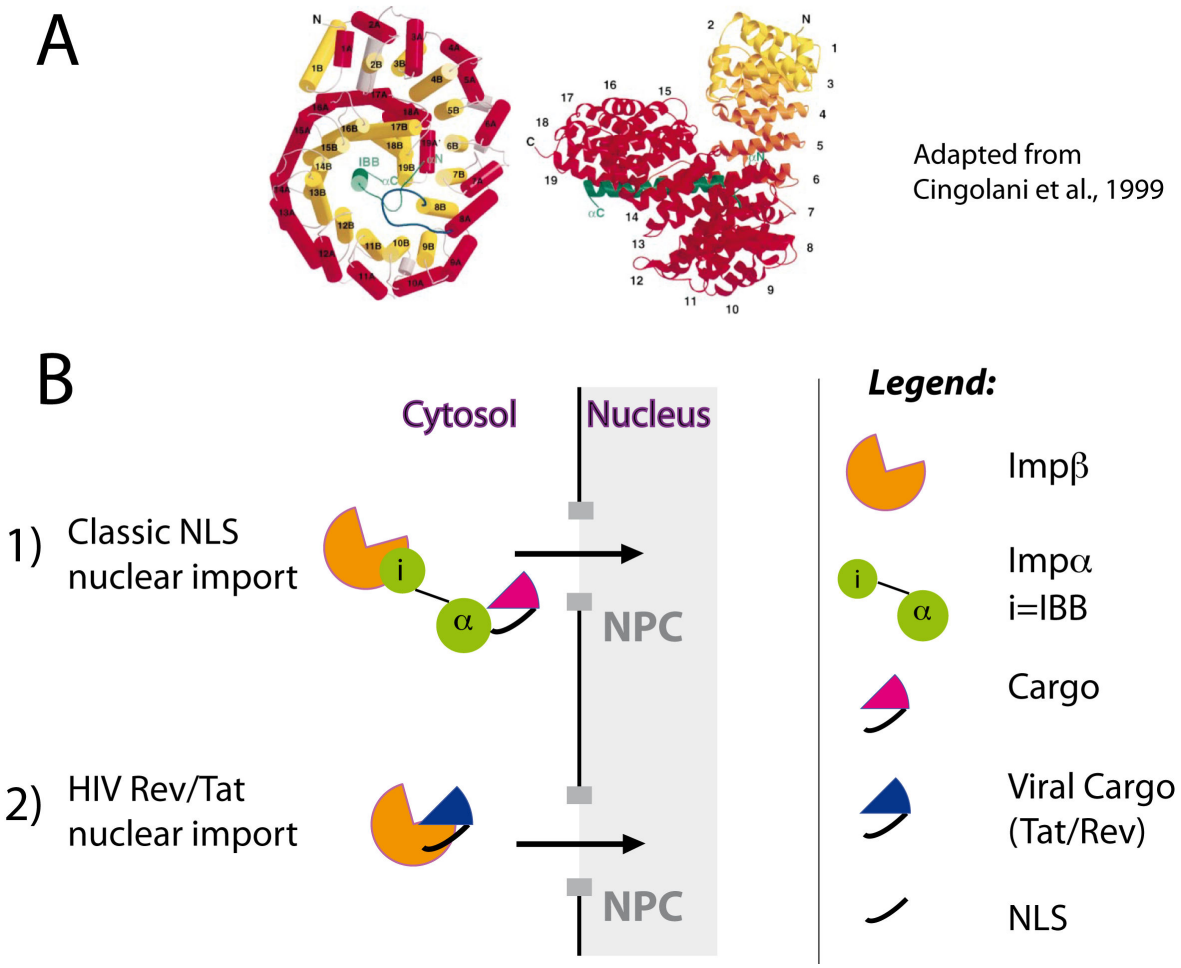


Figure 1.1: Importin-β nuclear transport factor A) Ribbon diagram of Impβ (19 helical HEAT repeats) complex with the Impα IBB domain (green) (PDB ID code 1QGK). B) Nuclear import mechanism for endogenous proteins bearing a classic NLS (1), which bind to the Impβ/Impα heterodimer, and for HIV Tat and Rev (2), which bind directly to Impβ before being translocated through the Nuclear pore complex, NPC.

However, due to the intrinsic flexibility of Impβ that relates to its adaptor function, complexes of Impβ are difficult to isolate at a high degree of purity and are recalcitrant to forming well ordered crystals that diffract at atomic resolution. On the other hand, such complexes are too large to be readily analyzed by NMR and, in most cases, too small to be imaged by cryoEM; therefore, their structural characterization remains challenging. Other proteins that are difficult to crystallize and in general challenging to study at atomic detail are those with many large loops or disordered regions.

Despite extensive current knowledge of protein fold space and the development of advanced bioinformatic tools, accurately predicting the 3D structure of a protein remains challenging when its sequence identity to a protein of known structure is low (<20%). Therefore, the comprehension of structure-function relationship for many proteins resistant to classical analysis, relies on the development of existing and new experimental methods. In this sense, a considerable breakthrough has recently happened in the field of single particle cryoEM, as described below. In this thesis work we intend to develop a new protein engineering tool that can further broaden the enormous potential of cryoEM.

1.2 Use of Redundancy to acquire noise-free averages [1]: X-ray crystallography and cryoEM

The diffraction of electromagnetic radiation allows one to deduce structural information about the diffracting sample. In order to reach atomic resolution, the wavelength of the diffracted radiation must be comparable to the lengths of the chemical bonds, typically ~1 Å, which corresponds to X-rays (McPherson, 2009). As biological molecules are composed of light atoms (mainly C, H, O, and N) their scattering power is extremely low. At the same time biological molecules are highly susceptible to X-ray radiation damage, due to the generation of free radicals (inelastic scattering). In X-ray crystallography, the signal to noise ratio (SNR) is amplified by exploiting the simultaneous scattering from many unit cells within the crystal, which

[1] Frank, J. (2006). Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State (2nd edition).

can be measured under cryogenic conditions to limit the effects of radiation damage. A simplified scheme of an X-ray diffraction experiment is shown in Figure 1.2A. A macromolecular crystal is tilted in an X-ray beam and the intensities of discontinuously diffracted waves are recorded on a detector, although the phase information is lost (phase problem). The Fourier synthesis of directly measured intensities (high SNR) combined with directly or indirectly estimated phases produces the electron density map of the molecule.

The resolution of this map is related to the diffraction quality of the crystal and determines the accuracy with which one can trace the polypeptide backbone and position the side chains, thereby determining the reliability of the atomic model (Figure 1.2 A). Iterative refinement and validation of the map and model are carried out by minimizing the difference between calculated and experimental structure factor amplitudes, measured using a suitable discrepancy index such as the crystallographic R-factor. Nowadays this process is quite straightforward thanks to the development of efficient phasing techniques and modern computing power.

Notwithstanding, as mentioned above, the main bottleneck is the growth of a well diffracting crystal, which is usually obtained by a trial-and-error procedure that may not always succeed. The chances of a successful crystallization are reduced when the sample presents a high compositional or conformational heterogeneity (McPherson, 2009) or when it presents low solubility and cannot be sufficiently concentrated to perform crystallization trials. Furthermore, many proteins do not exist as isolated entities, but in multi-subunit complexes with other biomolecules (such as nucleic acids, sugars, lipids and small effectors), arranged as functional modules that catalyze essential cellular processes (Nie et al., 2009). The production of such complexes (especially those from eukaryotic systems) in the amount and degree of homogeneity required for crystallization is often prohibitive. Therefore, single particle (crystal-free) techniques, such as Electron Microscopy, are emerging as a powerful alternative or a complementary method for imaging biological molecules (Lander et al., 2012).

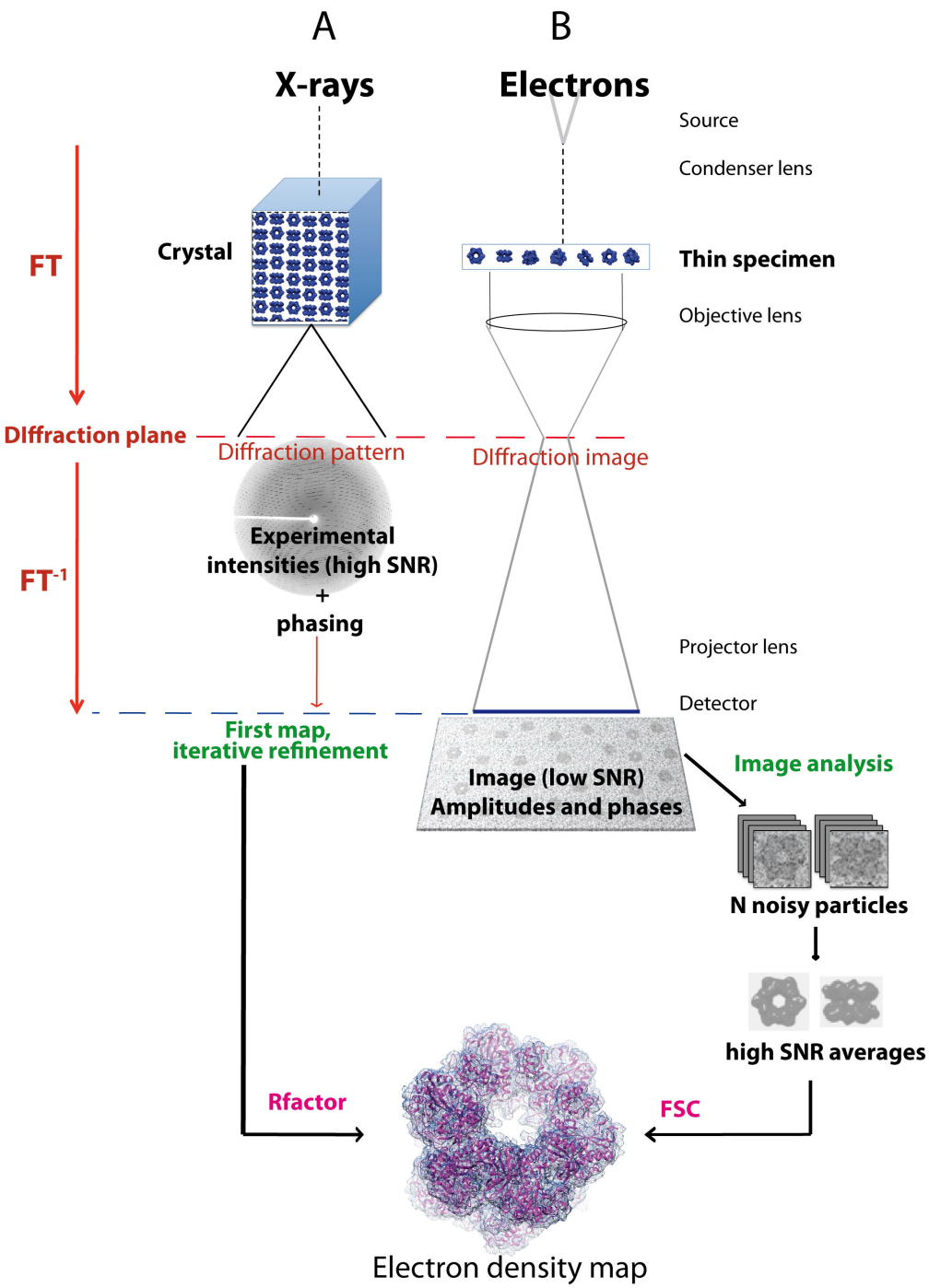


Figure 1.2: Simplified diagram of X-ray crystallography (A) and cryoEM methods (B). Both techniques are based on weak scattering from biological samples, composed of light atoms. To overcome this hurdle, redundancy is exploited to get a noise-free average. In X-ray crystallography redundancy is achieved by the crystalline nature of the sample, producing high SNR spots on the diffraction pattern, with loss, however, of the phase information. In cryoEM the redundancy is obtained by averaging *in silico* many noisy particles having similar features through image processing. More details are described in the text. As an example, the Glutamine synthetase structure is used to illustrate the two techniques. The elements of the electron microscope are analogous to those of a light microscope.

Single particle cryo electron microscopy (cryoEM) is a technique based on the electron scattering from a vitrified biomolecular solution specimen kept at liquid nitrogen temperature (see § 4.5.1 for more details on preparation). A simplified diagram of an electron microscope is shown in Figure 1.2B. As for X-ray crystallography, a diffraction image of the specimen is produced (in the back focal plane); however, electrons can be projected by magnetic lenses to recompose the magnified image of the specimen on a detector, thereby recovering the phase information (Saibil, 2000). Being biological specimens composed of light atoms, their density is very similar to the aqueous medium, thereby presenting almost no amplitude contrast and only weak phase contrast. This can be enhanced by exploiting defocus and spherical aberration of the microscope to enable detection of macromolecular particles on the image (Contrast transfer function, CTF § 4.5.2). On the other hand, because the scattering cross section is much larger for electrons than for X-rays, a fraction of the irradiating electrons strongly interact with the specimen (inelastic scattering), causing serious radiation damage and deterioration of high resolution features (Henderson, 1995). This implies that the sample must be very thin to avoid multiple scattering events, the whole imaging procedure must be performed under vacuum (to prevent electron scattering by air) and the electron dose must be extremely low (~$10e^{-}$/Å) to preserve the sample during imaging (Orlova and Saibil, 2011). A trick to overcome these hurdles is to stain and protect the molecule in a strongly scattering heavy atom salt solution, providing, however, non-native, low resolution information (negative stain EM, § 4.5.1). Conversely, by vitrifying the sample solution at cryogenic temperatures (CryoEM), macromolecules are imaged in their native hydrated state, overcoming evaporation inside the microscope and reducing radiation damage at the same time (§ 4.5.1). Hence, the raw cryo micrographs consist of a collection of extremely noisy projections of the inherent electron density of the molecules, which are ideally randomly oriented. Once the orientation of individual images is determined, it is possible to align and average them by enhancing common features and averaging out the random noise, ultimately yielding a 3D electron density map equivalent to that achieved by X-ray crystallography (Orlova and Saibil, 2011; Saibil, 2000; van Heel et al., 2000). In fact, this procedure is equivalent to recreating a pseudo-crystal *in silico* (for more details on image

processing see § 4.5.2). The determination of angular and positional parameters is a challenging task for extremely noisy particles. In general, given a certain level of noise, the procedure is easier for large and symmetrical particles, because they can be easily identified on the electron micrograph and their symmetry provides more constraints for alignment and 3D reconstruction (Orlova and Saibil, 2011). As the size and symmetry of the sample decrease one loses adequate features for accurate alignment, lowering the resolution and reliability of the reconstruction. Icosahedral viruses, with 60 asymmetric units in a single particle, are the ideal and most popular specimen in cryoEM (Crowther et al., 1970), while asymmetric low molecular weight particles below 100 kDa are unsuitable for cryoEM analysis (Henderson, 1995, 2004). Often, map selection and refinement are achieved through correlation with projections of a starting model. However, due to the high noise level, there is the risk of selecting and averaging noise, thereby reproducing the starting model (model bias), or of interpreting noise as high resolution features (overfitting) (Henderson, 2013a; Scheres and Chen, 2012). A standard procedure for validating and estimating the resolution of the final map, as well as a figure of merit analogous to the crystallographic R-factor, is still under debate. At the moment, the average resolution is estimated through Fourier Shell Correlation (FSC) (§ 4.5.3) and by visualizing the effective resolvability of structural features in the map, such as secondary structure elements (Lander et al., 2012; Saxton and Baumeister, 1982; van Heel and Schatz, 2005).

A typical single particle cryoEM analysis requires only 4-5 µL of a ~50 nM solution of the target macromolecule, which corresponds to one thousand times less material than that required to grow a protein crystal (Frank, 2006). This represents a big advantage in the analysis of heterogeneous and poorly abundant macromolecular complexes. In fact, in solution macromolecules adopting a wide range of physiologically relevant conformations, that are sometimes limited by crystal packing. By taking advantage of computational sorting techniques it is not only possible to exclude aggregates, misassembled molecules and impurities from the particle dataset (an *in silico* purification), but also to resolve more than one conformational state at the same time, thereby providing a powerful tool for studying structural rearrangements upon the binding of ligands. Outstanding examples include the study of the GroEL chaperonin folding mechanism and of ribosomes during translocation (Clare et al.,

2012; Lander et al., 2012; Stark et al., 2000). When the sample is too unstable or flexible it is possible to lock it in a limited number of preferential conformations by using a gradient of a cross-linking reagent, a procedure named as *Grafix* (Kastner et al., 2008).

1.3 IMPROVEMENT OF RESOLUTION ACHIEVED BY SINGLE PARTICLE CRYOEM OVER THE LAST DECADES

By referring to the experimental X-ray map it is possible to correlate the resolvability of protein structural elements to the resolution of an EM map, as shown in Figure 1.3 (Zhou, 2008). At low (20-10 Å) resolutions it is possible to fit pre-existing models by recognizing their shape. At a subnanometer resolution (<10 Å) α-helices can be defined due to their tubular and rigid nature, while thin β-strands can be recognized only at a resolution of ~5 Å. Between 3 Å and 4 Å (near-atomic resolution) *de novo* tracing of the backbone conformation and identification of bulky residues is possible. At 2 Å resolution almost all the side chains conformations are defined. In X-ray crystallography, for a well packed crystal the typical resolution is rather high (~2 Å). Despite the great potential of electron microscopy due to the use of very short wavelengths (typically ~0.02 Å), a number of instrumental and physical limitations cause the loss of high resolution information in the image.

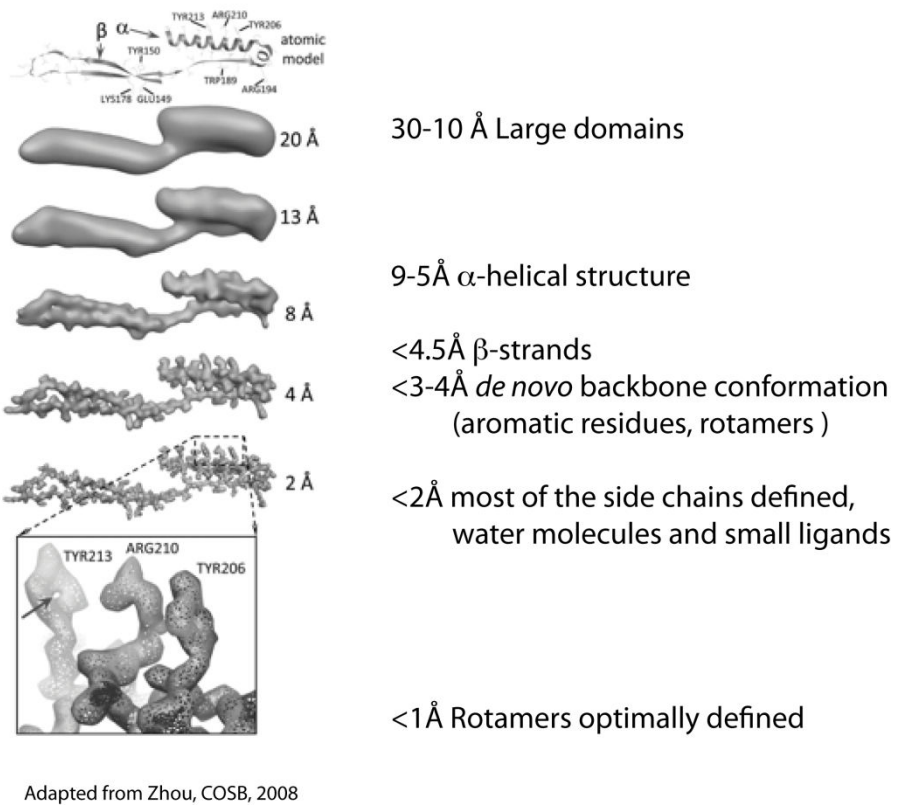


Figure 1.3: Visual appearance of EM map at different resolutions: Electron density of an α/β domain at different resolutions and corresponding resolvability of protein features. Figure adapted from Zhou (Zhou, 2008)

The degradation of resolution depends in part on the microscope and on imaging system imperfections: lens aberrations, current fluctuations, incoherent illumination, stage stability, etc. A major cause of noise is radiation damage: the interactions of electrons with matter imply a big energy transfer with ionization phenomena that deteriorate the chemical structure and cause movement of the molecules in vitreous ice, therefore blurring the image. Moreover, due to defocus, the frequencies from the objet are transferred to the image with alternating contrast. This needs compensation and accurate computational correction prior to determining the 3D orientation and position of each particle (for more details see § 4.5.2). Low image quality can substantially limit how accurately these parameters are computationally determined, thereby affecting the final resolution of the map (Saad et al., 2001). Finally, sample preparation (ice thickness) and heterogeneity can also decrease the final resolution. Hence, in the last century, EM was confined to the exploration of molecular morphology and nicknamed 'blob-ology' because of the low resolution (30-15 Å) maps

obtained compared to those obtained by crystallography (Smith, 2014). A notable exception was the study of icosahedral viruses, for which substantially higher resolutions (~7Å) could be obtained, allowing the definition of the backbone fold by the fitting of atomic models (Bottcher et al., 1997). In the last decade a series of developments leading to the use of higher coherence sources, better stability of EM stages and lenses, and improved sample preparation and computational procedures have allowed the near atomic resolution of several viral structures with *de novo* determination of the protein fold (Yu et al., 2008; Zhang et al., 2010; Zhang et al., 2008). Conversely, the resolution attained for objects of lower mass and symmetry has been in the 5-10 Å range, which is sufficient for docking crystal structures but not for *de novo* structure determination. Another leap forward seen in the last couple of years has been the development of direct electron detectors, which reduce the noise introduced by the recording process and yield a higher efficiency in detecting high-resolution information (Brilot et al., 2012; Faruqi and Henderson, 2007; Li et al., 2013). With these extremely sensitive detectors it is now possible to correct for motion induced by the beam or due to stage instability by aligning consecutive short-exposure frames, thereby increasing the inherent resolution and SNR of the resulting averaged micrograph. This breakthrough has extended the field of near-atomic resolution determination from icosahedral viruses to much smaller objects of lower symmetry, such as a four-fold symmetric 300 kDa ion channel (Liao et al., 2013), and even allowed the detection of an antibiotic bound to a ribosome (Wong et al., 2014). This "resolution revolution" (Kuhlbrandt, 2014) obviates the need for prior structural information in the interpretation of EM maps and makes single particle cryoEM a viable alternative for analysing proteins resistant to crystallographic analysis.

1.4 INTERPLAY BETWEEN PROTEIN ENGINEERING AND STRUCTURAL BIOLOGY

In the last two decades, structures of increasingly large protein complexes were obtained, including symmetrical assemblies (viral capsids, clathrin coats, microtubules, bacterial S-layers), as well as nanomachines such as ribosomes, chaperonins, photosynthetic systems, bacterial secretion systems and ATP synthases (Abrahams et al., 1994; Braig et al., 1994; Deisenhofer et al., 1985; Low et al., 2014; Zhang et al., 2010; Zhao, 2011). Such architecturally sophisticated assemblies have inspired

methods of protein engineering to duplicate, modify or expand upon what Nature has achieved, for biotechnological applications (Channon et al., 2008; Fischlechner and Donath, 2007). Examples include the introduction of point mutations to increase thermal stability (Zhang et al., 1995), to modulate metal binding and catalytic activity (Wilcox et al., 1998), and to modify molecular interfaces so as to tune the oligomerization propensity of a target (Ogihara et al., 1997). The idea of using macromolecular building blocks to create bio-nano-materials with desired properties was initially developed in the 1980's. This idea was first applied to DNA, whose structure is more predictable than that of proteins, to design so-called "DNA origami" (Rothemund, 2006; Seeman, 1982). Following these DNA studies, biotechnologists were inspired by the vast repertory of shapes revealed by structural biology to design auto-assembling protein nanomaterials. Different approaches include the use of coiled-coil peptides to form tetrahedral cages and protein fusions of oligomeric proteins (Gradisar et al., 2013; Gradisar and Jerala, 2011; Ringler and Schulz, 2003; Sinclair et al., 2011; Whitesides et al., 1991; Woolfson and Alber, 1995).

A notable advance in the field of protein-based materials design was achieved by Todd Yeates and colleagues at the University of California, Los Angeles (UCLA). These researchers connected pairs of naturally symmetric proteins (homo-oligomers) to generate self-assembling particles with specific shapes (Yeates and Padilla, 2002). Indeed, a fusion between two homo-oligomeric proteins tends form an ordered structure, whose geometry depends on the relative orientation of the individual symmetry axes (Figure 1.4). For example, by fusing a dimer and a trimer with parallel, non-coincident 3- and 2-fold axes, a hexagonal layer is produced. On the other hand, if the axes intersect at angles of (ideally) 54.7° or 35.3°, tetrahedral or octahedral particles will result, respectively. These construction rules were applied by genetically fusing a dimeric and a trimeric protein [bromoperoxidase and M1 matrix protein of influenza virus, (Hecht et al., 1994; Sha and Luo, 1997)], having a C- and N-terminal α-helix, respectively. These terminal helices were connected by a 9 residue linker with a strong helical propensity, taken from a portion of the L9 ribosomal protein (Hoffman et al., 1994). As the angle between the 3-fold and 2-fold axes was close to 54˚, the fusion generated a dodecameric tetrahedral cage of 16 nm in diameter (EM measurements), as predicted. Different cage geometries were successfully produced via this approach

and, by refining the contacts between oligomerization subunits, it was possible to obtain monodisperse cages that could be crystallized (Lai et al., 2012a; Lai et al., 2013). The advantage of having a continuous α–helix extending from one oligomerization domain to the other is two-fold: first, it provides rigidity and directionality and increases the likelihood of proper folding of independent domains with helical termini; second, because there are 3.6 residues and a rise of 5.4 Å per helical turn, the deletion of one residue results in a shift of 1.5 Å and a 100˚ rotation, allowing one to control the relative orientation of domains by changing the linker length. Despite the fact that such changes in orientation are discrete and cannot be changed arbitrarily, the chimeric construct can be appropriately designed to match the construction rules, allowing a level of control upon the three-dimensional arrangement of the particle that would otherwise be impossible to achieve with an arbitrary connection. Recently, Yeates and co-workers have formulated general algorithms to accurately design multi-component nanohedra having different shapes (King et al., 2014).

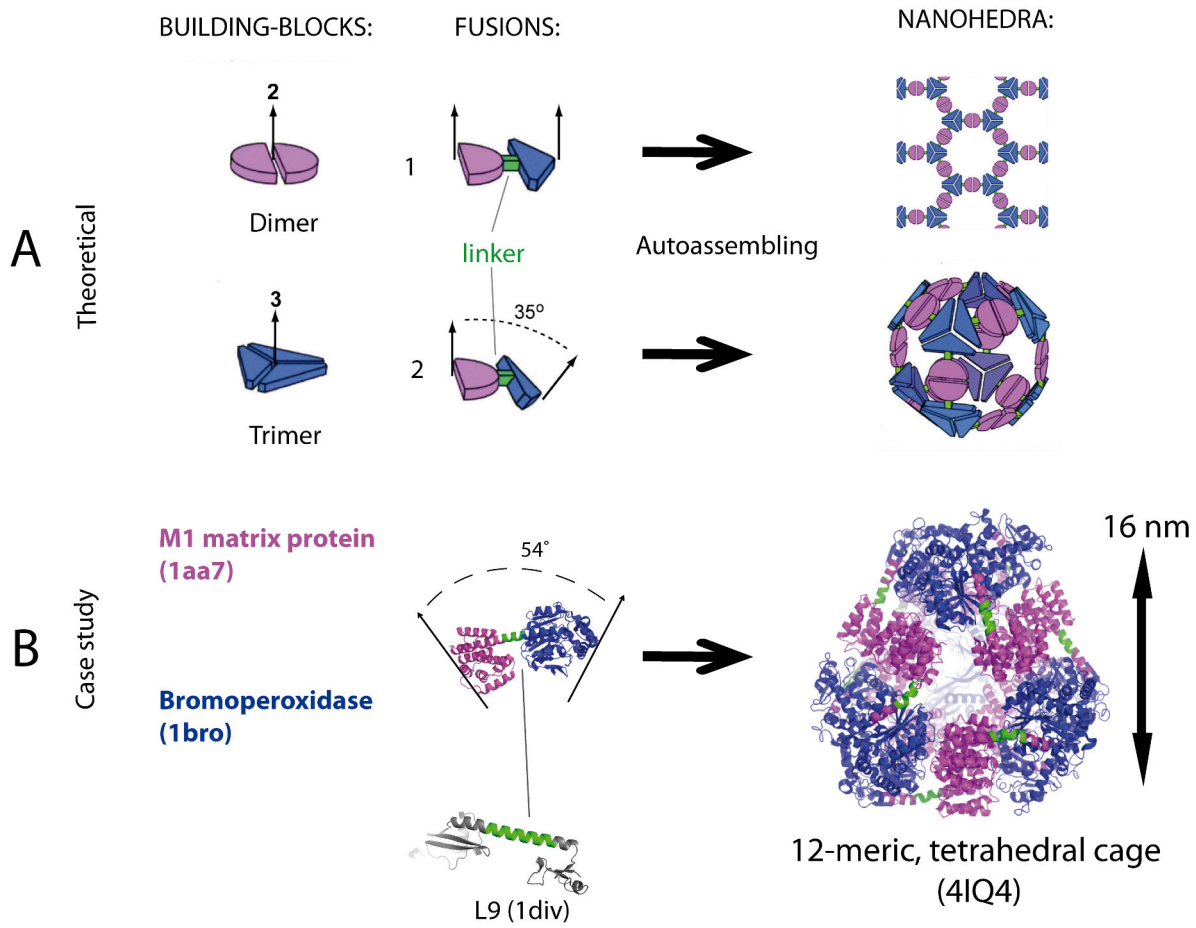


Figure 1.4: Design of protein nanohedra by Padilla et al. 2001. A) construction rules for a 2D layer and octahedral cage by fusing a 3-fold and a 2-fold object. B) Application of construction rules of a tetrahedron by fusing the M1 matrix protein (dimer, in magenta) and bromoperoxidase (trimer, in blue) having symmetry axes intersecting each other at an angle of 54˚. The helical linker extracted from the stable isolated helix of the L9 protein joins their C and N helical termini, thereby generating a rigid dodecameric 16 nm protein cage.

Nanomaterials present a number of useful applications. Protein layers may have applications as biosensors, detectors (Sara and Sleytr, 1996) or molecular sieves with precise cut-off values unreachable by inorganic materials. Hollow cages, with a precisely defined internal volume and diameter of the solvent accessible gates, can be used as as drug delivery systems or as nanocompartments with specific physico-chemical properties to control enzymatic reactions. A recent study demonstrated how protein cages can be used to sequester HIV protease *in vivo* (Worsdorfer et al., 2011). The increasing number of available protein 'building blocks' uncovered by structural

biologists, combined with a deeper knowledge of their cellular functions, will certainly expand the field of bionanotechnology.

Conversely, protein engineering has made important contributions to the development of structural biology, for instance in facilitating the crystallogenesis of challenging targets. For example, Surface Entropy Reduction (SER) (Goldschmidt et al., 2007), consists of replacing surface residues characterized by high conformational entropy (e.g. lysines) with alanines, thereby increasing the chances of hydrophobic and stable interactions among molecules during crystallogenesis (Cooper et al., 2007). Another approach is co-crystallization with antibodies or Designed Ankyrin Repeat Proteins (DARPins). The latter are fast folding, stable helical scaffold proteins whose specificity can be tuned towards different binding partners (Sennhauser and Grutter, 2008). DARPins can be readily produced in bacteria and have been successfully used as an alternative to antibodies to facilitate the crystallization of difficult protein targets (Boersma and Pluckthun, 2011).

Another valuable crystallization strategy is to genetically fuse a protein to an oligomeric protein to create intermolecular symmetry and favour lattice formation (Banatao et al., 2006). Alternatevely, the protein can be fused to a monomeric soluble and easily crystallizing protein, such as Glutathione S-Transferase (GST), Lysozyme or Maltose Binding Protein (Mbp) (Moon et al., 2010). Indeed, one of the most successful approaches to promote crystallization is to combine the latter approach with SER. The idea is to fuse a target protein to the C-terminus of Mbp, whose sequence is modified to reduce surface entropy (Moon et al., 2010). First, N-terminal Mbp facilitates folding, increase solubility and purification of the C-terminal target. Second, due to its improved surface properties Mbp increases crystallizability. The latter strategy has permitted the crystallographic analysis of several challenging proteins and protein complexes (Bethea et al., 2008; Ullah et al., 2008). Similarly, as for crystallography, several macromolecular engineering techniques have been devised to overcome the size limitation problem in cryoEM, as reviewed below.

1.5 MACROMOLECULAR ENGINEERING METHODS TO OVERCOME SIZE LIMITS IN CRYOEM

Single particle cryoEM promises to provide near-atomic resolution of targets resistant to crystallographic analysis. However, most monomeric proteins are below

the cryoEM size limit (100 kDa). A way to overcome this hurdle is to increase the size (and possibly symmetry) of the target by scaffolding it onto a template. In recent years, several such scaffolding techniques have been introduced (Figure 1.5).

In 2011 DNA layers functionalized with NTA-Ni^{2+} were used as a template to bind and image 40 kDa histidine-tagged GPCR, bearing a long linker to allow the protein to adopt a wide range of orientations (Figure 1.5A). The 2D class averages of the scaffolded randomly oriented GPCR matched quite well with reprojections of the crystal structure (Selmi et al., 2011). This technique could possibly be used in the future to obtain high resolution 3D reconstructions. In 2012, mFabs (monoclonal Fragments Antigen Binding) were proposed as an affinity scaffold to image small proteins by cryoEM (Figure 1.5B). The specific binding of the 50 kDa mFab to a certain target not only increases the mass of the molecule, but provides a fiducial marker with specific features to facilitate image alignment and to validate the final 3D reconstruction. The authors screened a series of different available antibodies for different targets and selected, by negative stain EM, the most suitable to perform cryoEM analysis. By this procedure they managed to obtain a ~13 Å cryoEM structure of a 65 kDa HIV-integrase dimer, bound to two mFabs, that increased its molecular weight by 100 kDa (Wu et al., 2012). However, in case of a monomeric protein the size of the protein would only be increased by 50 kDa, and so the size of the complex would probably remain too small for accurate structure determination. In 2013, a group of researchers developed a general strategy to decorate nanoscaffolds with a given protein via an antibody binding tag called Z33 (Pille et al., 2013). They demonstrated the principle by showing the efficient decoration of TMV (Tobacco mosaic virus) with an anti-capsid antibody, binding in turn to a Z33 tagged protein (Figure 1.5C).

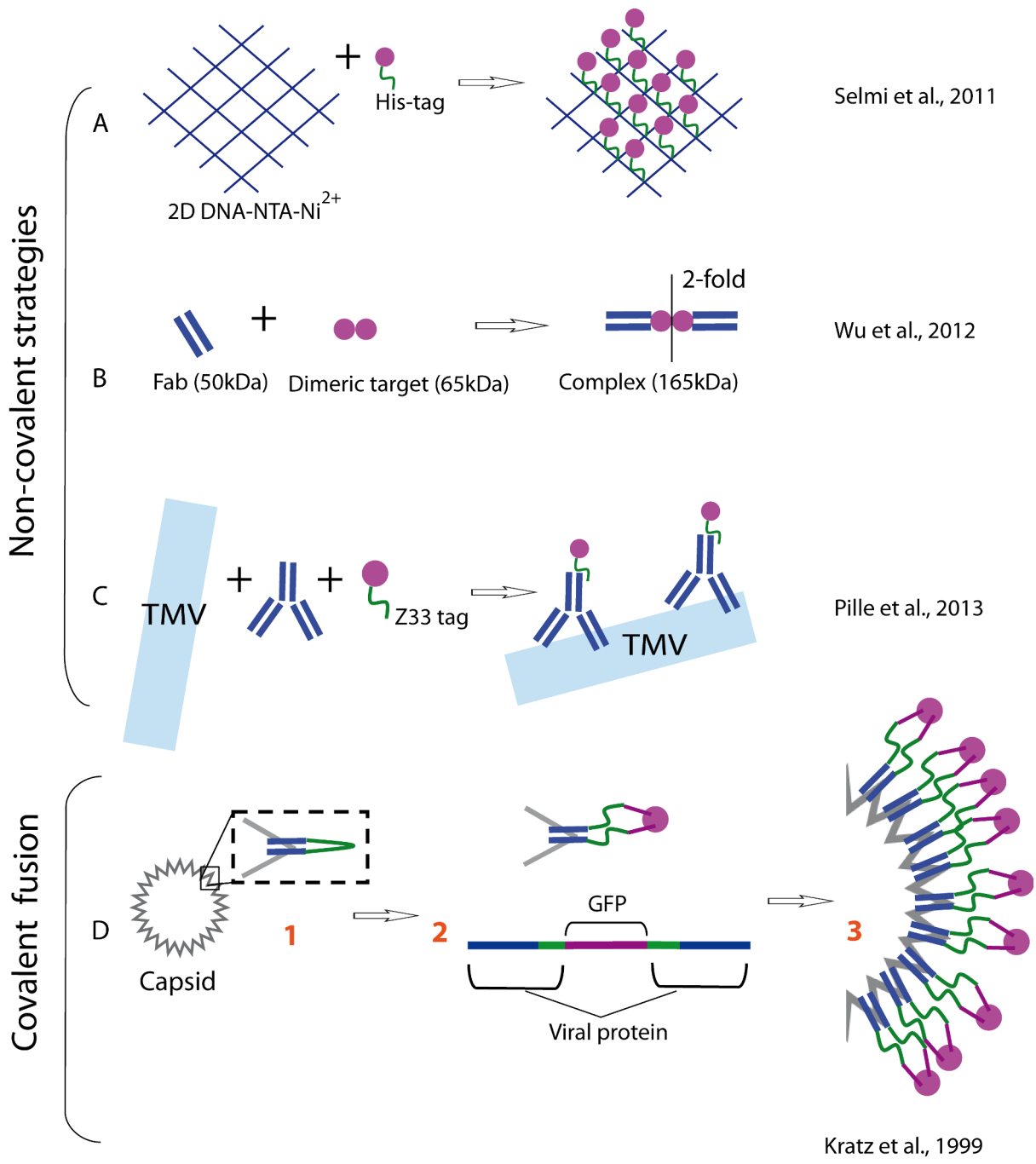


Figure 1.5: Different scaffolding strategies to allow imaging of small particles by cryoEM. The circle in magenta represents the small target, the blue components the templates that serve as scaffolds and the green stretches the linker. Three non-covalent strategies and one covelent strategy are illustrated. A) DNA-NTA-Ni^{2+} layers as a template to bind His-tagged proteins. B) Two Fabs binding a dimeric protein and increasing its size. C) TMV-antibody complex binds to a Z33-tagged target, which in turn binds the constant antibody part. D) Genetic fusion of both GFP termini on a viral capsid. 1) The capsid presents an exposed loop (green) joining the monomers of a stable dimer (blue). 2)The exposed loop is interrupted by inserting a GFP sequence, with long linker sequences at the termini. 3) The dimer-GFP fusion auto-assembles onto a well-formed capsid decorated with flexibly bound GFP.

The drawback of non-covalent scaffolding techniques is the possibility of generating a heterogeneous mixture of fully, partially or non- decorated scaffolds that might complicate the particle sorting procedure by decreasing symmetry and ultimately affecting the quality of the map. In 1999 Kratz and colleagues tried to image the green fluorescent protein (GFP) covalently linked by genetic fusion with the hepatitis B capsid (Kratz et al., 1999). This capsid is formed by multiple dimeric units separated by a solvent exposed loop, which is not involved in dimer formation (Figure 1.5D1). Into this loop the authors inserted GFP (which has both N- and C-termini free) via two flexible linkers, thereby facilitating its proper folding within the chimera (Figure 1.5D2). Biochemical data relative to the hepatitis B core-GFP chimera showed that the core was well formed and fully decorated by GFP, although its electron density was mostly averaged out during the symmetry restrained map calculation, due to the flexibility of the long linkers (Figure 1.5D3). By reducing the linker length and bringing GFP closer to the capsid surface, one could conceivably obtain a better cryoEM map for GFP. However, optimizing both the N- and C-terminal linkers, while retaining proper folding of the GFP and correct assembly of the capsid would be quite difficult and time consuming.

From this brief overview it is clear that many valuable ideas have been envisaged to overcome the size limitation in cryoEM. The present PhD work seeks to reach this goal using a different approach, described next.

1.6 THE PROTEIN SYMMETRIZATION METHOD: AIM AND OVERVIEW OF THE THESIS

The present thesis proposes a new method to achieve the structure determination of small proteins (<100 kDa) by cryoEM, named protein symmetrization. Our idea is to engineer a monomeric target protein by genetically fusing it to an homo-oligomeric template (Figure 1.6). This chimera, triggered by the oligomerization of the template, self-assembles into a large and symmetrical particle suitable for cryoEM analysis. The protein symmetrization approach is inspired by the protein nanohedra design (Padilla et al., 2001) and combines several advantages of the scaffolding approaches reviewed in § 1.5. First, full decoration of the template is ensured by the genetic fusion and occurs via a single rather than a double linkage. Second, the symmetry of the scaffold provides useful constraints in detection, alignment and image reconstruction of the

chimeric fusion. Third, the template and linker length can be easily modified by routine molecular biology techniques and the protein expressed in bacteria, thereby allowing an eventual automatization of the process. Ultimately, this approach could in principle be applicable to proteins of any size.

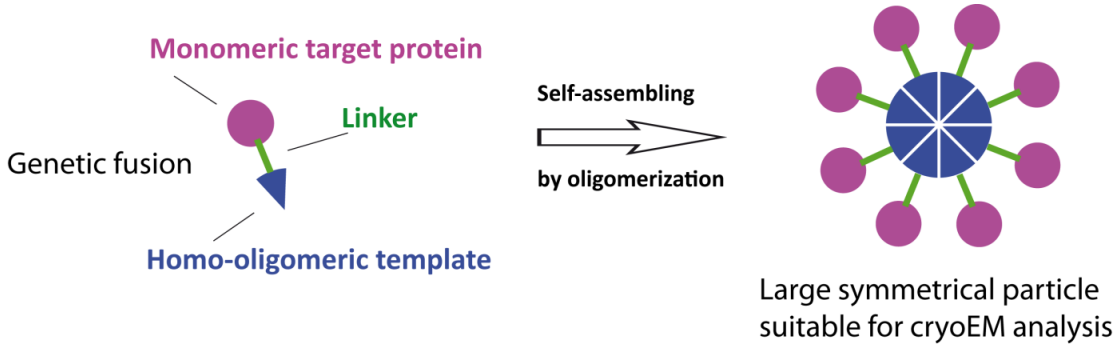


Figure 1.6: Concept of protein symmetrization: By fusing the monomeric target to a homo-oligomeric template, the chimera is designed to self-assemble as a large and symmetrical particle suitable for high resolution cryoEM analysis.

Ideally, one would like the symmetrized target to form a highly compact and rigid particle for which a near-atomic resolution cryoEM structure could be determined, allowing for *de novo* modelling of the target backbone. It is clear that the linker length plays a crucial role in defining the conformational homogeneity of the particle. If the linker is sufficiently long the domains have the best chance of folding independently, but this may produce a flexible assembly. If the linker is too short, sterical hindrance might compromise the folding of the target or the template, or hinder oligomerization. The ideal linker should maximize interactions between the target and template without compromising their ability to fold or oligomerize.

The general questions we ask are: Is protein symmetrization and cryoEM a feasible alternative to crystallographic analysis of a monomeric protein? Given a certain target, what are the optimal linker sequence and template protein one should use to obtain a reliable cryoEM resolution? What resolution is attainable by this approach? We address these questions by performing a feasibility study in which target proteins of known structure are combined with specific oligomeric template proteins of bacterial

origin. We devise two linkage strategies and use these to symmetrize several globular proteins as well as the superhelical protein Importin-β (§ 2.1). Subsequently, we evaluate these constructs using biophysical techniques as well as negative stain and cryoEM (§ 2.2). The best target-template combination is then chosen to determine the optimal linker length. During this procedure, a general protocol for identifying the optimal constructs for cryoEM analysis is formulated (§ 2.3). A few well behaved constructs are quantitatively analysed by cryoEM (§ 2.4). Results for the best oligomeric chimera are validated by crystallographic data (§ 2.5). Chapter 3 discusses the limitations of protein symmetrization and suggests guidelines for the symmetrization of a new target of unknown structure. Finally, Chapter 4 details the experimental and computational procedures used.

2. RESULTS

ABSTRACT

To establish whether protein symmetrization can feasibly facilitate the cryoEM analysis of small monomeric proteins, we fused several proteins of known structure to two oligomeric template proteins: glutamine synthetase (GS) and the E2 subunit of pyruvate dehydrogenase, which have D6 and icosahedral symmetry, respectively. As judged by the visual appearance of particles by EM, a fusion of maltose binding protein (Mbp) with GS (Mag) was determined to be the most promising chimera for further investigation. We produced a panel of Mag constructs with different linker lengths and used biophysical assays (TSA, DLS, native PAGE, SEC) and negative stain EM to select the best candidate for cryoEM analysis, MagΔ5. We determined the cryoEM structure of MagΔ5 at 10 Å resolution (FSC 0.5 criterion) by enforcing D6 symmetry. Comparison to the Mbp crystal structure indicates that the catalytic pocket and specific α-helices are well defined in the cryoEM map. The crystal structure of MagΔ5 was solved at 7 Å resolution and further validates the structure determined by cryoEM.

RÉSUMÉ

Pour étudier la faisabilité de la symétrisation nous avons fusionné plusieurs protéines de structure connue à deux matrices de symétrie D6 (GS) et icosaédrique (E2). L'oligomérisation et le degré de décoration de la matrice montrent que la fusion Mbp-GS (Mag) est la chimère la plus prometteuse. Nous avons produit une collection de fusions Mag contenant des peptides de liaison de longueurs différentes. Une série de techniques biophysiques et structurales (TSA, DLS, gel natif, SEC) a été utilisée pour sélectionner MagΔ5, la chimère la plus homogène pour l'analyse structurale. La carte cryoME de MagΔ5 a permis de résoudre la densité de la cible de 40 kDa à 10 Å (FSC 0,5). Le recalage du modèle atomique indique que la poche catalytique et quelques α-hélices sont bien définies. Le modèle cristallographique à 7 Å confirme les résultats de cryoME et révèle la présence de contacts polaires à l'interface modèle-cible, qui peuvent contribuer à réduire la flexibilité de la position de la chimère.

2.1 PROTEIN SYMMETRIZATION: STRATEGY AND BUILDING BLOCKS

The previous chapter introduced the idea of fusing a protein of interest (the target) to a homo-oligomeric protein (the template) to facilitate structure determination by cryoEM. The present chapter aims to validate this concept using a target protein whose atomic structure is already known. The approach involves symmetrizing the target, subjecting it to cryoEM imaging and single-particle analysis, and comparing the resulting reconstruction with the known atomic structure.

Critical to the success of this project is the initial choice of protein building blocks: target, template and linker. The minimal requirement for the target and template is that the N-terminus of one and the C-terminus of the other should both be solvent accessible so that they can be physically connected. Moreover, the linker must be chosen so as to allow proper folding of both the target and template, as well as proper oligomerization of the template. Initially, the linker can be made relatively long and then progressively shortened to optimize the rigidity and compactness of the resulting particle. We examined the Protein Databank (PDB) for suitable template and target proteins in the context of two possible connection strategies, using either: i) a continuous helix or ii) an unstructured linker (Figure 2.1)

Strategy i: The target and template have α-helical termini which are connected by an α-helical linker. This strategy generates a continuous α-helix which extends from one domain into the other, and is inspired by the design of protein nanohedra mentioned in § 1.4 (Figure 1.4). Assuming no distortion of the connected helices, the length of the linker determines the spatial relationship between the target and template, hence determining the shape of the particle. This strategy is admittedly only possible when the target protein contains a terminal α-helix, which would not necessarily be the general case for a target of unknown structure. However, the setup is particularly convenient for a pilot study because it allows one to predict *in silico* the 3D structure of the symmetrized target, which can inform data interpretation (For more details see § 4.1). As a starting linker we used the same sequence used in the protein nanohedra study (Padilla et al., 2001), a long helix in ribosomal protein L9 (PDB ID code 1DIV). Assuming no helical distortion, the target samples specific orientations relative to the template due to the geometrical parameters of an α-helix (3.6 residues per turn), varying by a 100˚ rotation and 1.5 Å shift as the linker length is increased or decreased by one residue.

Strategy ii) The target and template have termini with arbitrary secondary structure connected by an unstructured linker. In this case, the linker is presumably more flexible than in a helix-based connection, and so the relative orientation of the target cannot reliably be predicted. However, the sampling of the target orientation relative to the template is less restricted and may fortuitously lead to the formation of a rigid and highly symmetric particle in case a stable interaction arises between protein surfaces (Figure 2.1). This setup simulates real cases in which a target of unknown structure may not necessarily have a terminal helix. As a starting linker we used a stretch of three alanines. This choice is based on a study which investigated the fusion of various carrier proteins as a method to facilitate crystallization of a protein of interest (Moon et al., 2010). Moon and colleagues identified the tri-alanine peptide as one of the most successful linker sequences for enhancing crystallization, irrespective of the secondary structure of the protein termini connected.

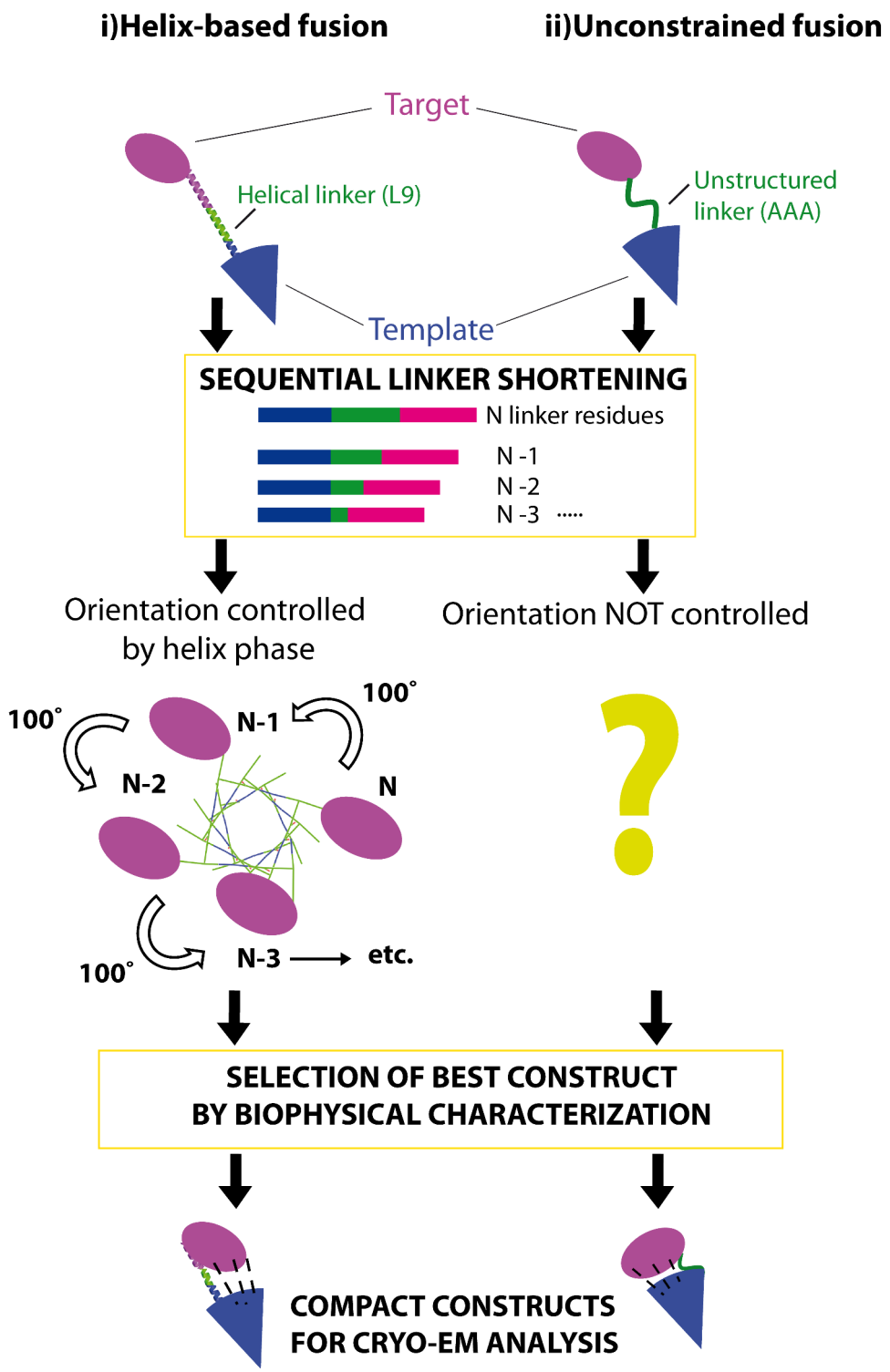


Figure 2.1: Illustration of the two fusion strategies tested for protein symmetrization. The target and template are connected either by a helical or unstructured linker. Several chimeras are produced by sequentially shortening the linker. The selection of a compact chimera for cryoEM studies is made through biophysical and negative stain EM characterization. In the case of a helix-based fusion, the geometry of the particle is determined by the number of residues in the linker (assuming helical integrity of the linker and no other distortion-inducing steric constraints), thereby allowing results to be interpreted in the light of the corresponding predicted structural models.

2.1.1 TEMPLATES

Following the criteria described above, we examined the Protein Data Bank (PDB) and selected three proteins to investigate as homo-oligomeric templates: glutamine synthetase (GS), lumazine synthase (LS) and the transacetylase subunit E2 of pyruvate dehydrogenase (PDH). As a preliminary control, we expressed, purified and examined each of these by negative-stain EM to verify their ability to yield homogeneous, symmetric particles. A brief description of each template protein is provided below.

Glutamine synthetase type I (**GS**) forms a large (~600 kDa) dodecamer with dihedral (D6) symmetry. This enzyme is highly conserved across prokaryotes, catalysing the condensation of ammonium and glutamate to from glutamine, an essential precursor for the biosynthesis of many metabolites (Eisenberg et al., 2000). Being an abundant, stable and symmetrical enzyme its quaternary structure was one of the first to be analyzed by electron microscopy (Streicher and Tyler, 1980; Valentine et al., 1968). The highest resolution (2.5 Å) X-ray structure was determined from the *Salmonella typhimurium* ortholog (PDB ID code 1F52) and shows that GS consists of two stacked hexameric rings with an inner and outer diameter of ~4 nm and ~14 nm, respectively, and a height of ~11 nm (Figure 2.2A). When viewed along the six-fold axis, the monomeric subunits of one ring eclipse those of the other. The quaternary structure is stabilized by Mg^{2+} or Mn^{2+} cations, which are located at the interface of adjacent subunits within a ring, and by a tight domain swapping phenomenon involving the exchange of C-terminal helices and β-strands between the two hexameric rings (Figure 2.2C). In contrast, the N-terminal helix lies above and below the dodecamer pointing towards the exterior of the ring (Eisenberg et al., 2000). This arrangement makes GS an ideal template for symmetrizing proteins using the helix-based connection strategy (§ 2.1). For practical reasons, our experiments used GS from *E.coli,* which shares 97% sequence identity with *S. typhimurium* GS and is therefore expected to have the same structure. GS was produced in *E.coli* and analyzed by SEC and negative stain EM (experimental details are provided in § 4.4.1 and § 4.5.1). The results confirmed that GS was highly homogeneous in conformation and suitable for use as a symmetrization template (Figure 2.2B).

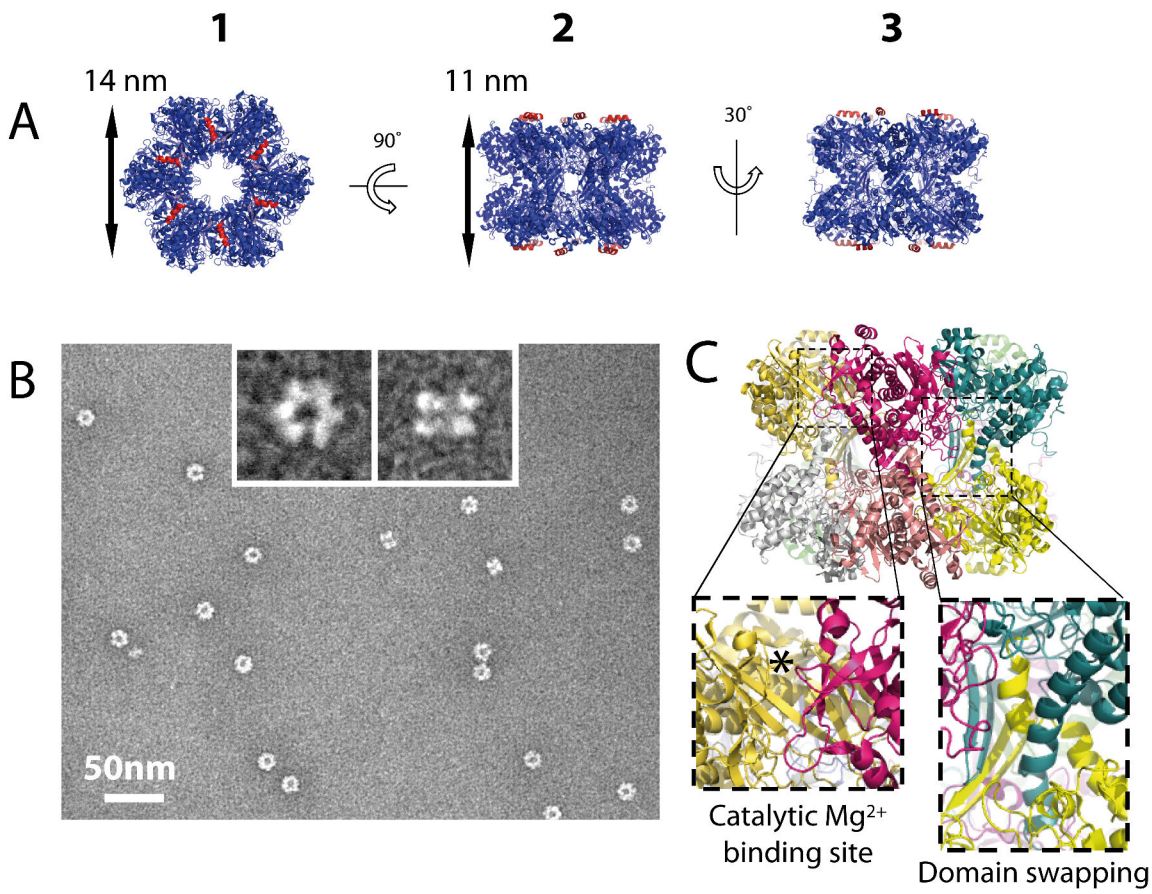


Figure 2.2: Quaternary structure of glutamine synthetase : A) Ribbon diagram of GS showing three views of the dodecamer. The N-terminal solvent accessible helices are in red. B) Electron micrograph of GS negatively stained with SST (Sodium Silico Tungstate) 2% w/v, acquired at a nominal magnification of 22000x and at 120 kV. EM grids were prepared as described (§ 4.5.1) C) Ribbon diagram of GS showing separately coloured monomers (PDB ID 1F52). The boxes highlight the metal binding site and the domain swapping of the C-terminal helices and β sheets belonging to opposite subunits, responsible for stabilizing the quaternary structure.

Lumazine synthase (LS) is a widespread and well conserved enzyme involved in the penultimate step of the riboflavin-biosynthesis pathway (Volk and Bacher, 1991). Relatively small α-β monomers (~16 kDa) assemble as pentamers in which the catalytic sites are located between adjacent subunits (Figure 2.3A-C). In several bacteria, 12 pentameric units are arranged edge-to-edge to form an icosahedral, hollow capsid of ~960 kDa with a diameter of 16 nm (Kumar et al., 2011; Zhang et al., 2001). The crystal structure of the icosahedral LS capsid from *S. typhimurium* (PDB ID code 3mk3) shows that the C-terminal helix is solvent accessible and points outward, making LS suitable for use as a template in the helix-based fusion strategy. We expressed and purified LS as described by Kumar et al., 2011 and pursued a first analysis by negative stain EM (Figure 2.3D;

experimental details are given in § 4.5.1). LS particles appeared as donuts with a central cavity stained by the heavy metal solution, consistent with the hollow spherical crystal structure. However, particles appeared inhomogeneous in their size and shape, suggesting that in solution LS may not be as symmetric as in the crystal. This variability in size and shape was apparently not a staining artifact, as it was also observed under native conditions by cryoEM (Figure 2.3E). Indeed, it has been reported that icosahedral LS from other bacterial species presents a much greater variability of the capsid radius, due to weak interactions among the pentameric subunits (Zhang et al., 2006). The deformation that we observed for *S. typhimurium* LS capsids might derive from a similar structural instability. Because our analysis suggested that the LS oligomer lacked sufficient rigidity and conformational homogeneity for use as a possible symmetrization template, we decided not to further pursue studies on this protein.

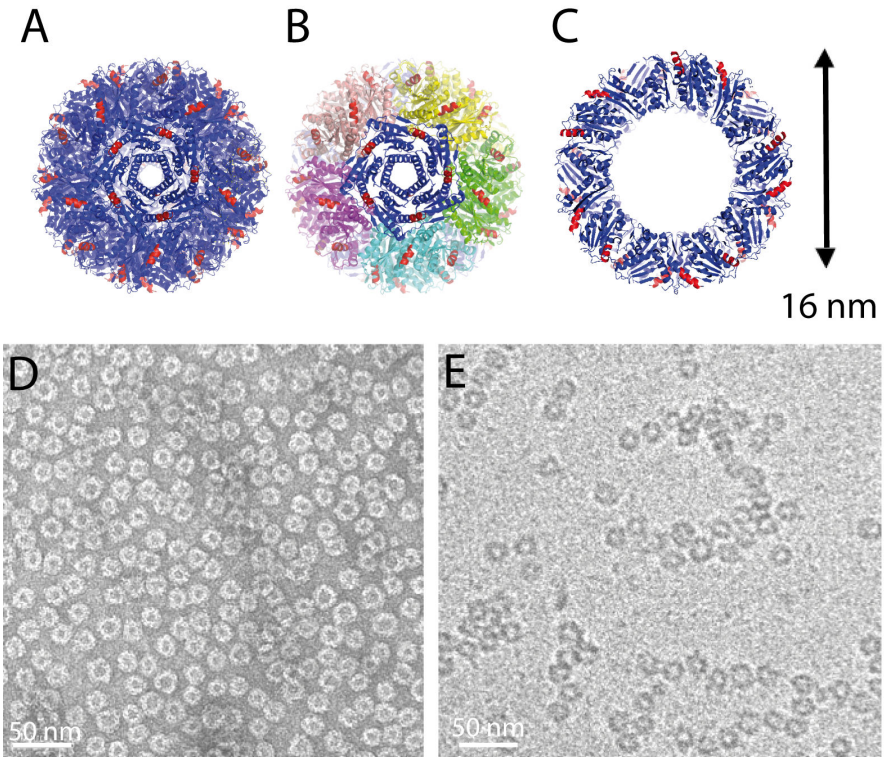


Figure 2.3: Structure of icosahedral Lumazine synthase (LS) A) Ribbon diagram of LS showing the C-terminal helices in red (PDB ID code 3MK3). B) Same view showing adjacent pentameric subunits forming the vertex of an icosahedron C) Cross-section of LS showing the hollow capsid. D) Electron micrograph of LS negatively stained with SST (Sodium Silico Tungstate) 2% w/v acquired at 120kV at a nominal magnification of 22000x E) CryoEM image of LS embedded in vitreous ice, recorded at 200 kV at a nominal magnification of 22000X. EM grids were prepared as described (§ 4.5.1)

Pyruvate dehydrogenase is a very large (5 to 10 MDa), ubiquitous multi-enzyme complex that catalyses the decarboxylation of pyruvate and the acetylation of coenzyme A (CoA). Although the overall architecture of the PDH complex differs across species, in all organisms studied the structural core of the complex is formed by the **transacetylase E2** subunit (Izard et al., 1999; Mattevi et al., 1992). In mammals and Gram-positive bacteria, 60 copies of E2 associate to form the icosahedral core (monomer MW ~28 kDa and total MW of ~1.8 MDa). The icosahedral E2 core from *Bacillus stearothermophilus* has been characterized by EM (Henderson et al., 1979; Milne et al., 2006) and its structure was determined at 4.4 Å resolution by X-ray crystallography (PDB ID code 1B5S)(Izard et al., 1999). The E2 monomer consists of a solvent-accessible N-terminal hook-like region, which adopts a random coil structure and is followed by a 4 turn α-helix and a globular α/β domain (Figure 2.4C). Each monomer associates tightly with two other subunits related by rotation about the three-fold axis. Twenty such trimers assemble to form a hollow icosahedral cage, whose inner and outer diameters measure 12 and 24 nm, respectively. The inner cavity of this cage is accessible via large solvent channels between the subunits (Figure 2.4B,D). We expressed and purified E2 (as described in § 4.3.1) and imaged it by negative stain EM (Figure 2.4A). Particles appeared to be highly homogeneous in size and shape, which were consistent with previous EM and crystallographic studies. These features and the high degree of symmetry make E2 a promising template for symmetrizing target proteins according to the unstructured linker strategy (§ 2.1)

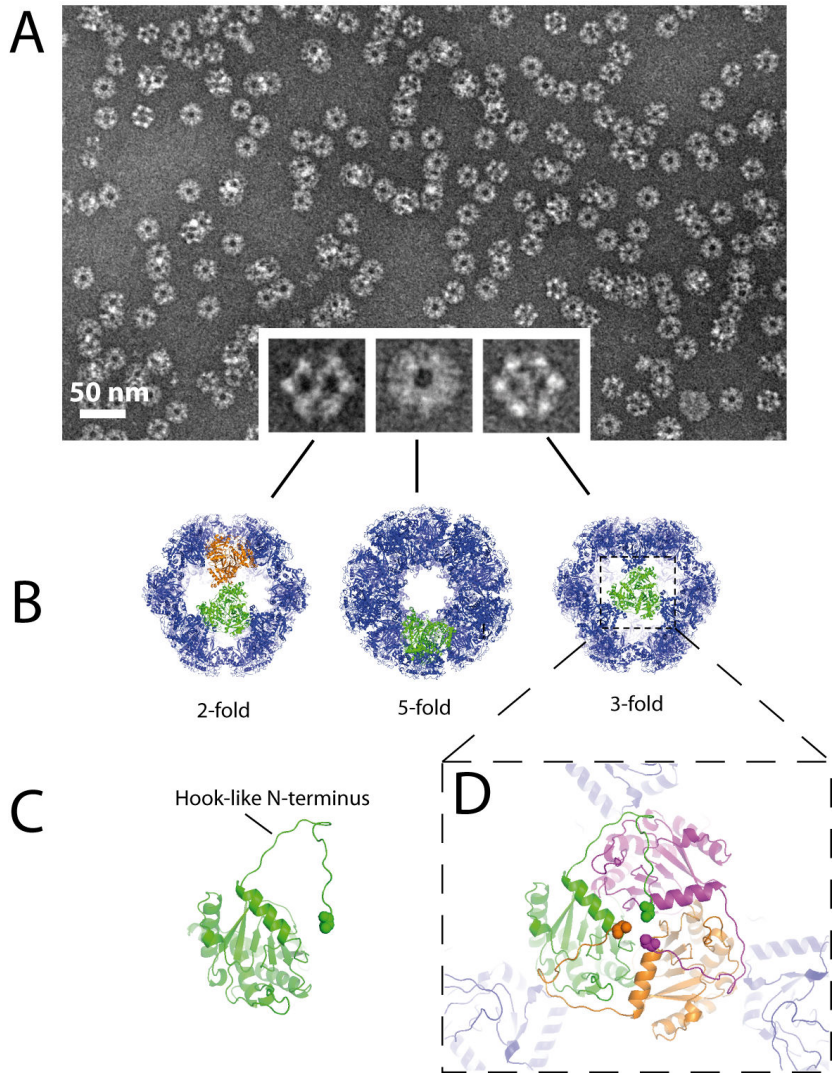


Figure 2.4: Structure of the icosahedral E2 core of the pyruvate dehydrogenase complex. A) Negative-stain electron micrograph of E2, recorded at 120 kV at a nominal magnification of 22000X. Insets show particles viewed along the 2, 5 and 3-fold axes. EM grids were prepared as described (§ 4.5.1) B) Ribbon diagram of E2 viewed along the three rotation axes. Trimeric subunits are shown in green and orange. C) Ribbon diagram of the E2 monomer (PDB ID code 1B5S). The N-terminal residue at the beginning of the hook-like region is shown as spheres. D) View of the E2 trimer. Subunits are coloured differently and the N-terminal residues (where the target will be linked) are shown as spheres.

2.1.2 Targets:

We next considered various proteins for use as the target to be symmetrized. As mentioned in § 1.1, a protein of great interest to our lab is Importin β (Impβ), which mediates the nuclear import of HIV Tat and Rev, essential for viral replication (Truant and Cullen, 1999). Since Impβ is a flexible superhelical protein, the structures of Impβ/Tat and Impβ/Rev complexes are challenging to determine, as they are difficult to isolate in large

amounts for crystallization and too small to be analyzed by cryoEM (~100 kDa). Hence, it would be useful to construct a symmetrized version of Impβ via fusion to a homo-oligomeric template that would allow structural characterization of the Impβ/Tat and Impβ/Rev complexes by cryoEM (Figure 2.5B). Since Impβ is entirely helical and bears a solvent accessible C-terminal helix, its symmetrisation could be accomplished by using the helix-based fusion with GS. Furthermore, as α-helices can be detected in cryoEM maps at 8-10 Å resolution, such a resolution would be sufficient to yield a pseudo-atomic model of these complexes (§ 1.3). Hence, a proof-of-concept of the scaffolding method combined with an interesting biological insight could be provided by using Impβ as a target protein (Figure 2.5).

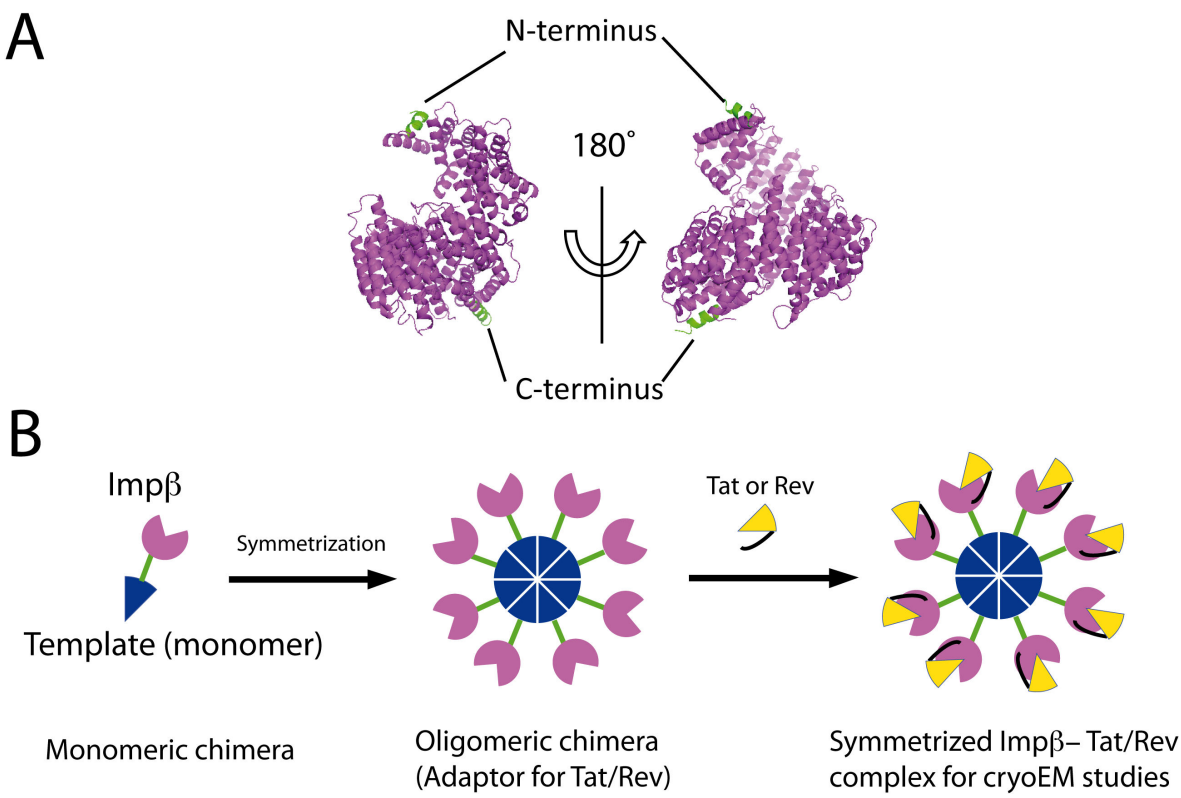
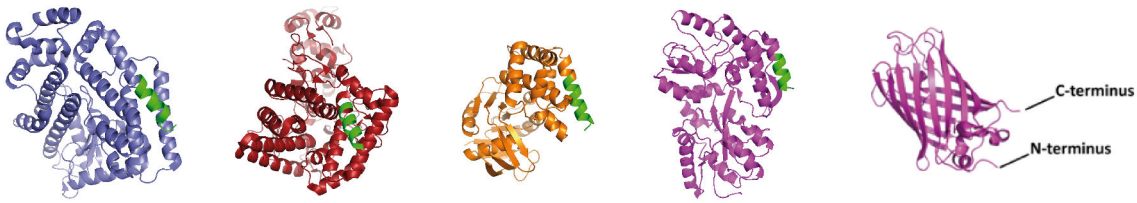


Figure 2.5: Importinβ as a potential target for symmetrization studies A) Ribbon diagram of Impβ (all helical) solenoid shown in two different orientations (PDB ID code 1QGK). C- and N- terminal helices are highlighted in green. B)The idea of symmetrization applied to Impβ to allow the structural study of Impβ/Tat and Impβ/Rev complexes. The oligomeric chimera Impβ-template could serve as an adaptor to study its complexes by cryoEM.

Apart from Impβ, we examined the PDB for globular proteins that would be suitable targets for symmetrization. To facilitate recombinant production in *E.coli* and purification of the resulting fusions, we searched for structures that: (1) were E.*coli* proteins; (2) were monomeric; (3) had a solvent accessible helical N-terminus; and (4) had an isoelectric point (pI) between 5 and 6 to match the pI values of the template proteins. This identified four potentially suitable target proteins: trehalase (Trea), ketopanthoate reductase (Kpr), glutamine synthetase adenylyltransferase (Gsat) and maltose binding protein (Mbp) (Figure 2.6).



Name	**Trea**	**Gsat**	**Kpr**	**Mbp**	**Gfp**
PDB ID	2jf4	1v4a	2ofp	1dmb	4as8
MW (kDa)	60	50.3	36.7	40	27

Figure 2.6: Crystal structures of globular targets for protein symmetrization. From left to the right: ribbon diagram of trehalase, ketopantoate reductase, glutamine synthetase adenylyltransferase, maltose binding protein and green fluorescent protein. In the first four models the N-terminal helix is highlighted in green. In GFP the positions of the non-helical termini are indicated.

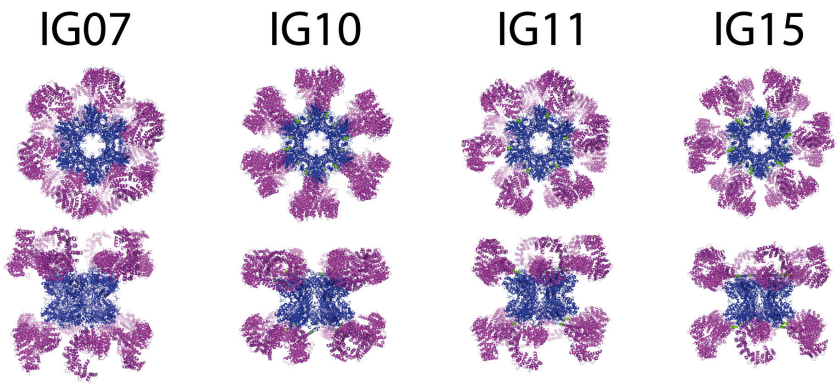
As a target protein having no defined secondary structure at its N or C terminus we chose green fluorescent protein (GFP, ~27 KDa) in its monomeric form (PDB ID code 4as8). Both termini of GFP are solvent exposed and adopt a random coil structure (Figure 2.6)

2.2 SCREENING OF DIFFERENT TARGET-TEMPLATE COMBINATIONS

2.2.1 IMPORTIN-β - GS FUSIONS

Following the helix-based fusion strategy we connected the C-terminal helix of Impβ to the N-terminal helix of GS, using the 18-residue L9 helical linker sequence. The latter sequence has been previously used to accomplish similar protein connections in literature, as described in § 1.4 (Padilla et al., 2001). *In silico* helical alignment (computational

procedure described in § 4.1) allowed us to produce approximate rigid body models of the series of Impβ-GS fusion proteins in which the linker was sequentially truncated from 18 to 0 residues. These constructs were denoted IGN, where I and G stand for Impβ and GS, respectively, and N is the number of linker residues. As explained in § 4.1, the fusions producing sterically forbidden models, i.e., those in which target and template moieties clash, might cause a distorsion of the helical connection or hamper proper folding. Conversely, the sterically allowed models are more likely to fold properly without distortion of the helical connection. Relying on this principle, we inspected the Impβ-GS structural models and identified four sterically allowed constructs, IG15, IG11, IG10, IG07, which we subsequently expressed and purified (Figure 2.7).



Junction sequence	Construct	Linker length
ELRKLKALEAQKQKEQRQAAEHVLT	IG**15**	15
ELRKLKALEAQKQKEQRQAEHVLT	IG**11**	11
ELRKLKALEAQKQKEQEHVLT	IG**10**	10
ELRKLKALEAQKEHVLT	IG**07**	7

Figure 2.7: Illustration of Impβ-GS constructs selected for recombinant production. Predicted structures of "sterically allowed" IGN constructs. In the first line top views of model fusions between Impβ (magenta) and GS (blue) and in the second line corresponding side views are displayed. The table below reports the junction sequence (Impβ in magenta, linker in green, GS in blue) and the corresponding nomenclature and linker length.

The four selected Impβ-GS chimeras were cloned and expressed in *E.coli* and purified by binding to an Impα-affinity resin followed by SEC as described in § 4.4.1. As an example, the results obtained for the best expressed fusion construct, IG07, are shown in

Figure 2.8. SDS-PAGE analysis shows that the construct migrates as expected for its MW (~150 kDa) and is quite pure (Figure 2.8A-B), despite a low expression yield (1 mg of protein for 12 L of culture). Although the expected MW of the oligomeric protein (1.8 MDa) is below the separation limit of the SEC column (5 MDa), the protein elutes at the void volume (Figure 2.8D), suggesting the presence of soluble aggregates. Moreover, the elution peak is not sharp and symmetrical, but presents a long tail extending over several fractions, suggesting that there may be a mixture of oligomeric states or that the protein may be improperly folded. Indeed, native PAGE analysis showed that most of the IG07 protein did not enter the gel matrix (consistent with aggregation), while the fraction that did enter migrated as a diffuse band (consistent with sample heterogeneity), in contrast to the sharp band observed for Impβ alone (Figure 2.8E). The other three fusion constructs gave similar results (data not shown).

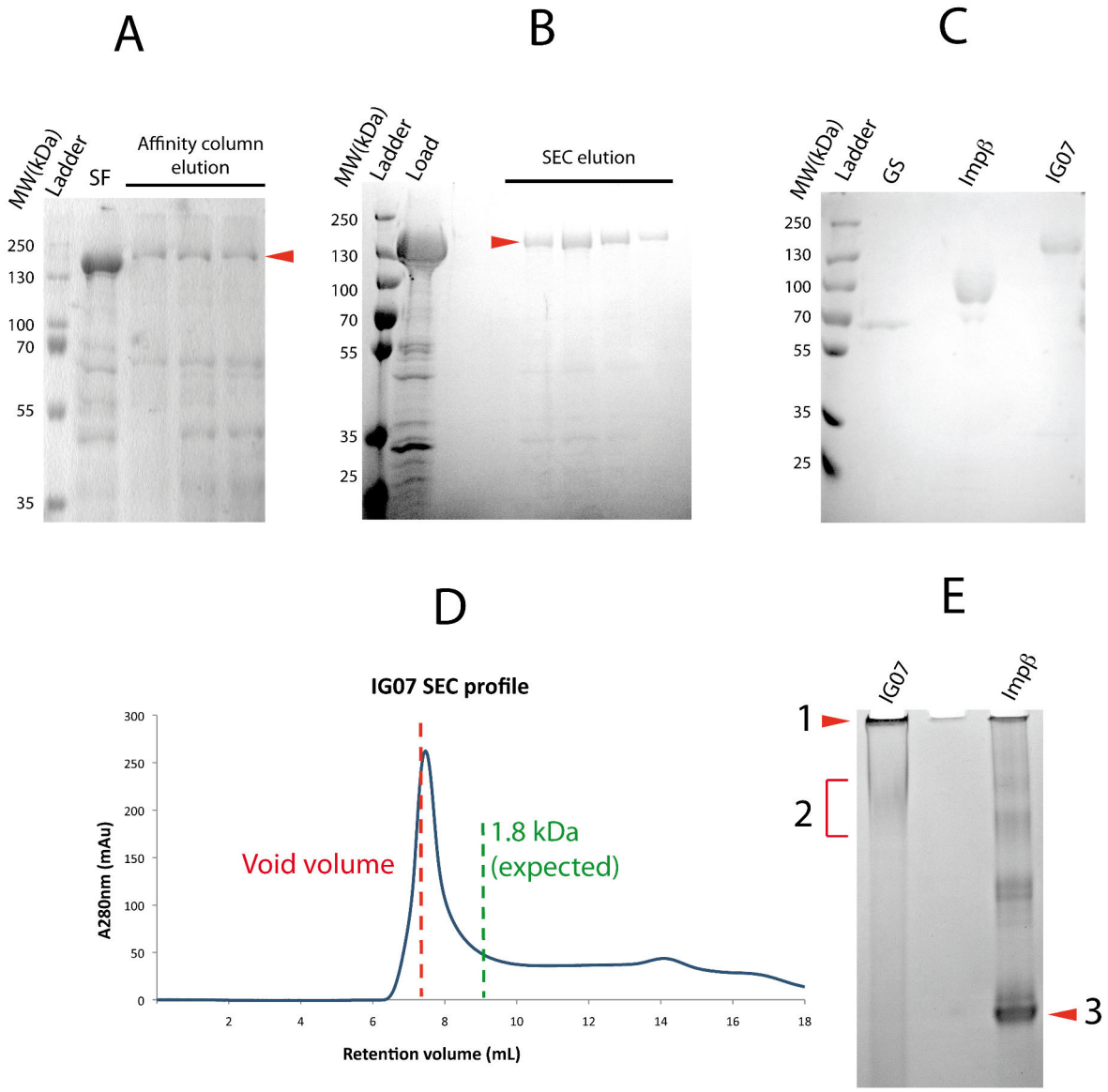


Figure 2.8: Purification and preliminary characterization of IG07. A) SDS-PAGE following Impα affinity purification B) SDS-PAGE following SEC purification and corresponding chromatogram. C) SDS-PAGE analysis showing different migration rates of target, template and fusion. D) SEC profile of IG07 fusion eluting at the void volume instead of the elution volume estimated relying on the calibration curve and mass of the complex (§ 4.4.1). E) Native PAGE analysis of chimera and target confirming the presence of aggregates (1) and sample heterogeneity (smeared band, 2). Conversely in Impβ (3) the migration band is quite sharp (upper bands are impurities).

To further characterize the selected IGN constructs, they were analyzed by negative stain EM (Figure 2.9). In all four cases, micrographs revealed the presence of a mixture of oligomeric species with aggregates. The most promising construct was IG07, for which it was possible to recognize side and top views of the dodecameric template that appeared "decorated" with one to four additional features putatively ascribed to Impβ. However, it

was not possible to find particles that appeared completely decorated with the expected number of Impβ moieties.

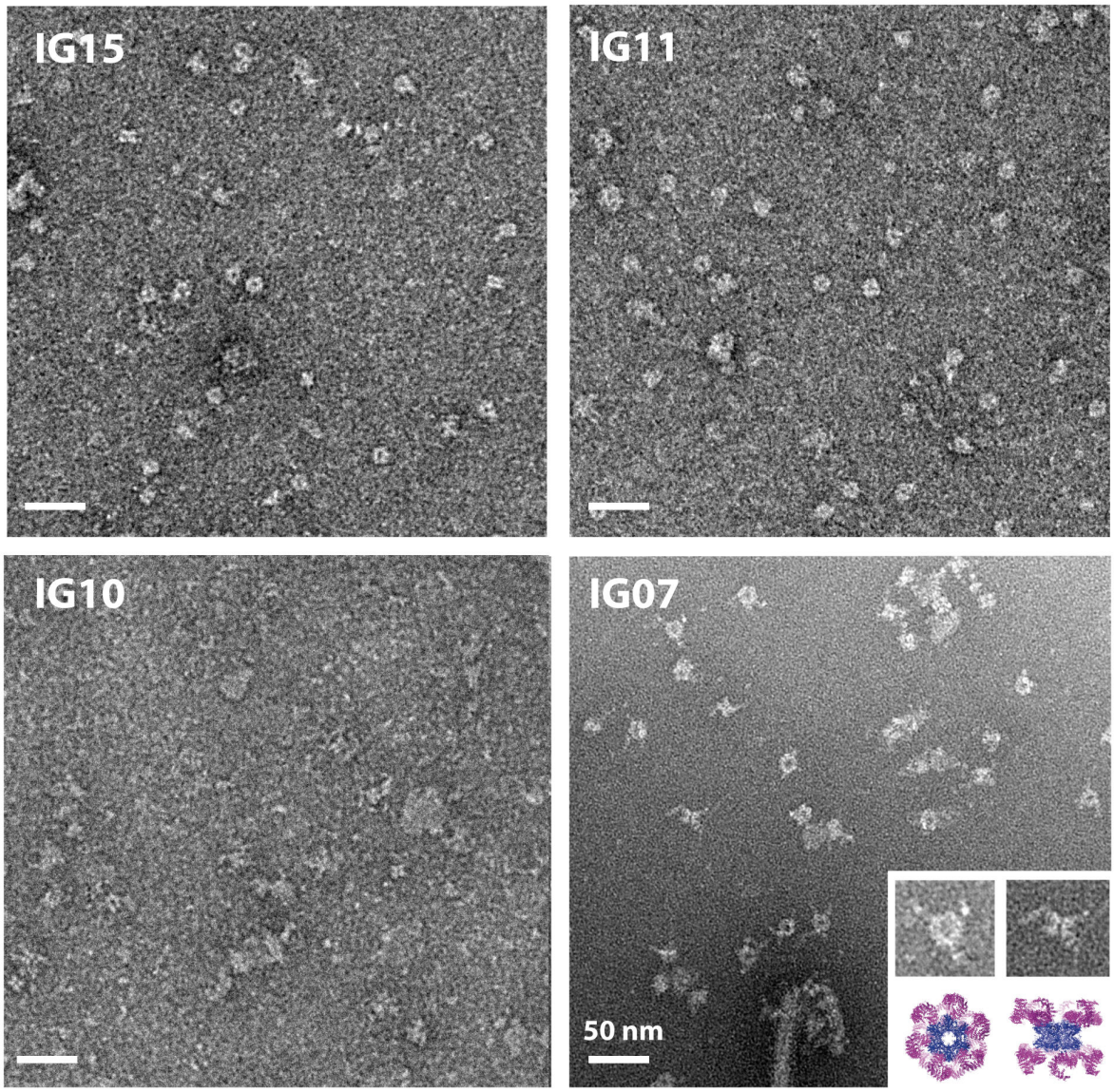


Figure 2.9: Negative stain analysis of four Impβ-GS fusions. EM grids were prepared as described (§ 4.5.1) using SST (Sodium Silico Tungstate) 2% w/v as the stain. Images were recorded at a nominal magnification of 22000x under 120kV voltage. The inset in IG07 shows close-ups of putative top and side views of the particle, as well as the corresponding ribbon diagram of the predicted structure, in which Impβ is shown in magenta and GS in blue. The results reveal the apparent preservation of the dodecameric GS ring structure and a conspicuous absence of Impβ-like features.

We excluded that this phenomenon was due to the loss of Impβ due to proteolysis of the fusion, because SDS-PAGE analysis revealed a single prominent band corresponding to the MW expected for the fusion protein, with no signs of degradation (Figure 2.8C). This

was confirmed by mass spectrometry analysis, which detected the presence of a single species at 149414.6 kDa (Figure 2.10).

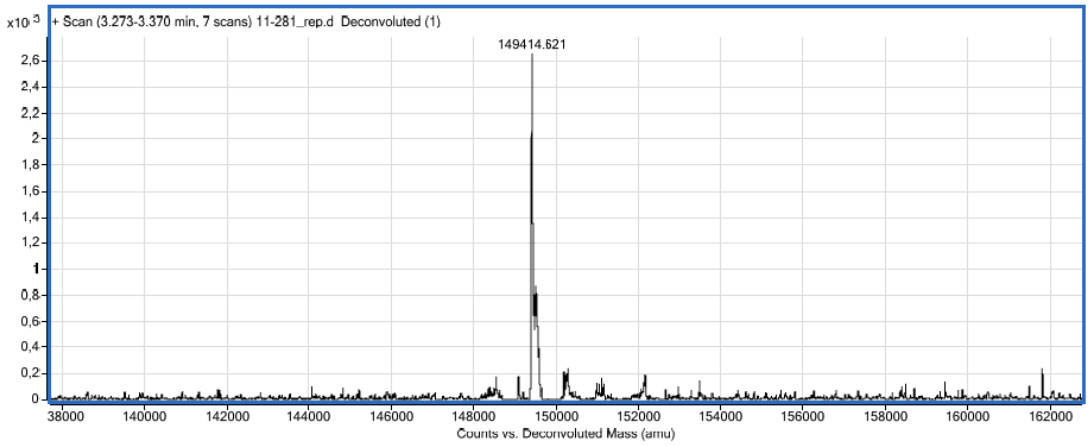


Figure 2.10: Deconvoluted electrospray-ionization (ESI) mass spectrum of IG07. The peak indicates that the most abundant species corresponds to the molecular weight of the intact chimera IG07.

Possible explanations for failing to observe complete decoration of the GS template in IG07 include i) artifacts due to the negative stain used; ii) suboptimal buffer conditions; and iii) improper folding of the target. To explore these possibilities we proceeded as follows.

First, we imaged IG07 embedded in vitreous ice at 200 kV using a large defocus of 3.5 µm (to enhance contrast) and compared images to those taken for the GS template in the same conditions (Figure 2.11). Strikingly, the fusion appeared almost indistinguishable from the template alone. Indeed, Impβ appeared even less visible in cryo-conditions than when negatively stained, suggesting that the poor electron density is an inherent property of the target molecule, rather than a staining artifact.

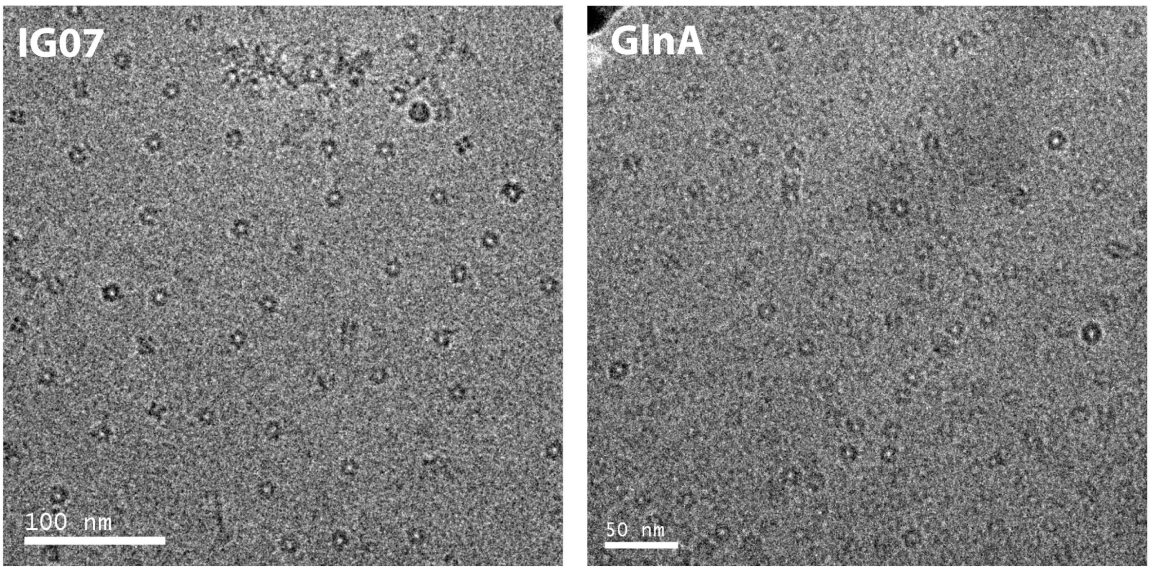


Figure 2.11: Electron micrograph of IG07 and GS embedded in vitreous ice. Images were recorded at 200kV on a PHILIPS CM200 microscope at a defocus of ~ 3.5 µm at a nominal magnification of 22000x (see § 4.5.1 for more details about image acquisition).

Subsequently, we tried to identify improved buffer conditions by studying the thermal stability of the chimera, target and template proteins using different buffers in a thermal shift assay (TSA; Figure 2.12A; the detailed protocol is described in § 4.4.3). The best buffer was identified as TRIS pH 8, the buffer already used in the previous EM analysis. Thermal stability profiles of the template, target and fusion proteins in this buffer are shown in Figure 2.12B. Clearly, Impβ is inherently unstable, with a melting temperature of 40 ˚C. GS presents a two-step profile with melting temperatures of 46˚C and 61˚C. In contrast, the profile for IG07 exhibits only a single transition corresponding to the lower T_m of GS and lacks the higher transition observed for GS, as well as any other transition to which one might attribute the melting of Impβ. This raises the possibility that the Impβ moiety may already be unfolded at the lowest temperature tested in this assay. If true, this would imply that fusing Impβ to GS considerably hampers the ability of Impβ to fold properly.

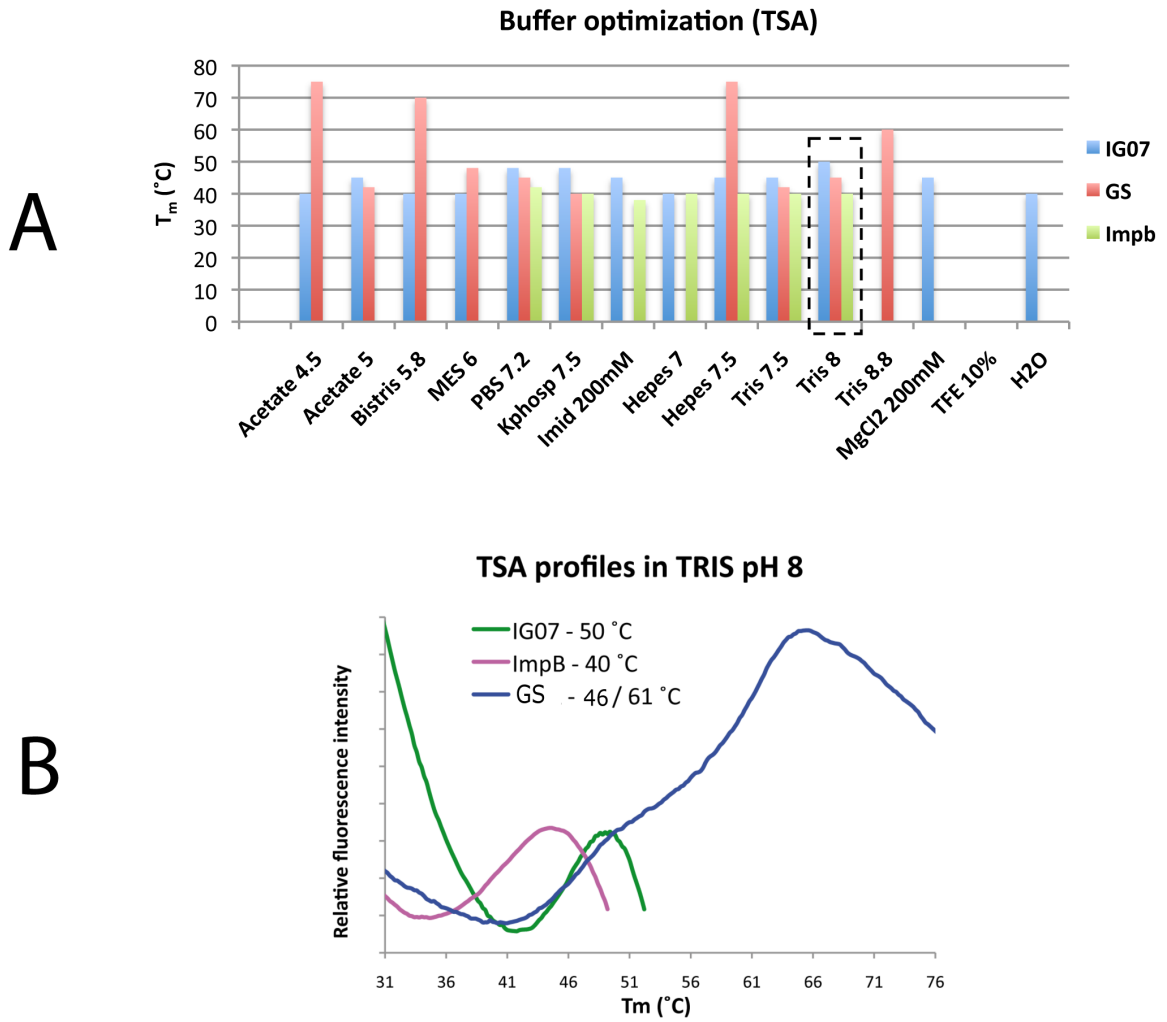


Figure 2.12: Thermal shift analysis Buffer optimization screening of IG07, GS and Impβ. A) Melting temperatures (T_m) of template, target and chimeric proteins (GS, Impβ, IG07) are shown in different buffers, keeping the NaCl concentration constant at 150mM. In the case of GS, only the lower T_m is plotted. The highest thermal stability for IG07 is obtained in TRIS pH 8 (dotted line box). B) Comparison of unfolding profiles of GS, Impβ and IG07 in TRIS pH 8.

We next attempted to stabilize the conformation of Impβ in the IG07 fusion by binding it to a 19-residue peptide derived from HIV Tat (residues 40-58), which had been shown by others in the lab to significantly increase the T_m of Impβ. However, no significant improvement in the negative stain EM images was observed (data not shown). A fluorescently labeled version of the same peptide (TAMRA-Tat 40-58) was used in a fluorescence polarization (FP) assay to compare the Tat-binding affinity of IG07 with that of free Impβ (§ 4.4.5). This assay yielded apparent dissociation constants (K_d) of 55 nM and 800 nM for free and fused Impβ, respectively (Figure 2.13). An equivalent result would be obtained if less than 1/10 of the Impβ moieties in IG07 were able to properly

bind the Tat peptide. This finding is thus consistent with improper folding of Impβ within the fusion construct.

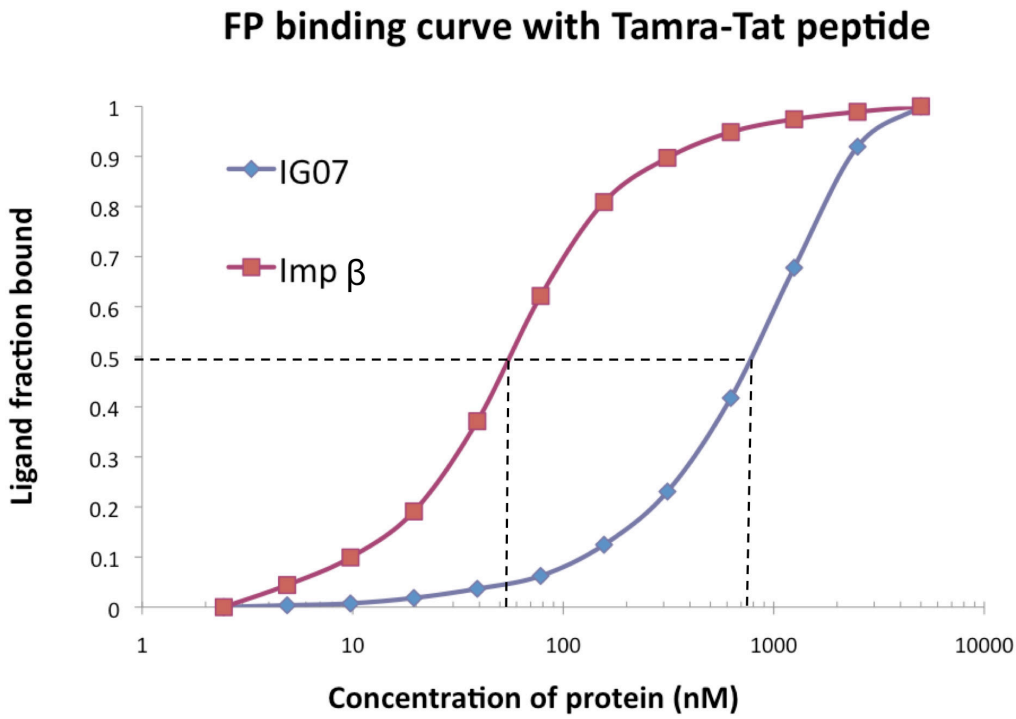


Figure 2.13: Binding of TAMRA-Tat 40-58 peptide to Impβ and IG07 as measured in a fluorescence polarization assay. The dotted line indicates half-maximal binding activity corresponding to the Kd. Kd=55 nM for Impβ and 800 nM for IG07.

Finally, in order to visually assess the degree of Impβ unfolding within the fusion we inserted a hexa-histidine tag at the N-terminus of the chimera (His-IG07) and labeled the protein with gold atom clusters functionalized with Ni^{2+}-NTA (5nm Ni-NTA-Nanogold® - Nanoprobes). Negative stain EM analysis of the labeled chimera thus allows one to localize the N-termini of Impβ within the particle (Figure 2.14). As controls, we performed the same experiment on untagged IG07 and on His-tagged GS. Micrographs of untagged IG07 indicate that gold beads bind non-specifically to aggregates, but rarely bind to isolated particles (Figure 2.14A, bottom panel). Micrographs of gold-labeled GS reveal top views of isolated particles decorated with one to five gold beads, which appear in direct contact with the template (Figure 2.14A, top panel). In contrast, beads appear significantly detached from the template in His-IG07, presumably due to the presence of Impβ (Figure 2.14A, middle panel). For 40-50 well isolated particles we measured the distance from the center of the bead to the center of the nearest GS ring and constructed a histogram

(Figure 2.14). The distance distribution in GS is quite sharp and centered at 13 nm, comparable to the radius of the GS crystal structure, added to the 2.5 nm gold bead radius (7+2.5 nm). In contrast, the distance distribution in His-IG07 is much broader, varying from 7.5 to 40 nm and exhibiting a maximum at ~22 nm, whereas the radius of the predicted *in silico* structure is approximately 12 nm. These observations are consistent with the hypothesis that Impβ is improperly folded when fused to GS. As a result of these findings, we decided not pursue efforts to symmetrize Impβ and to focus instead on other protein targets.

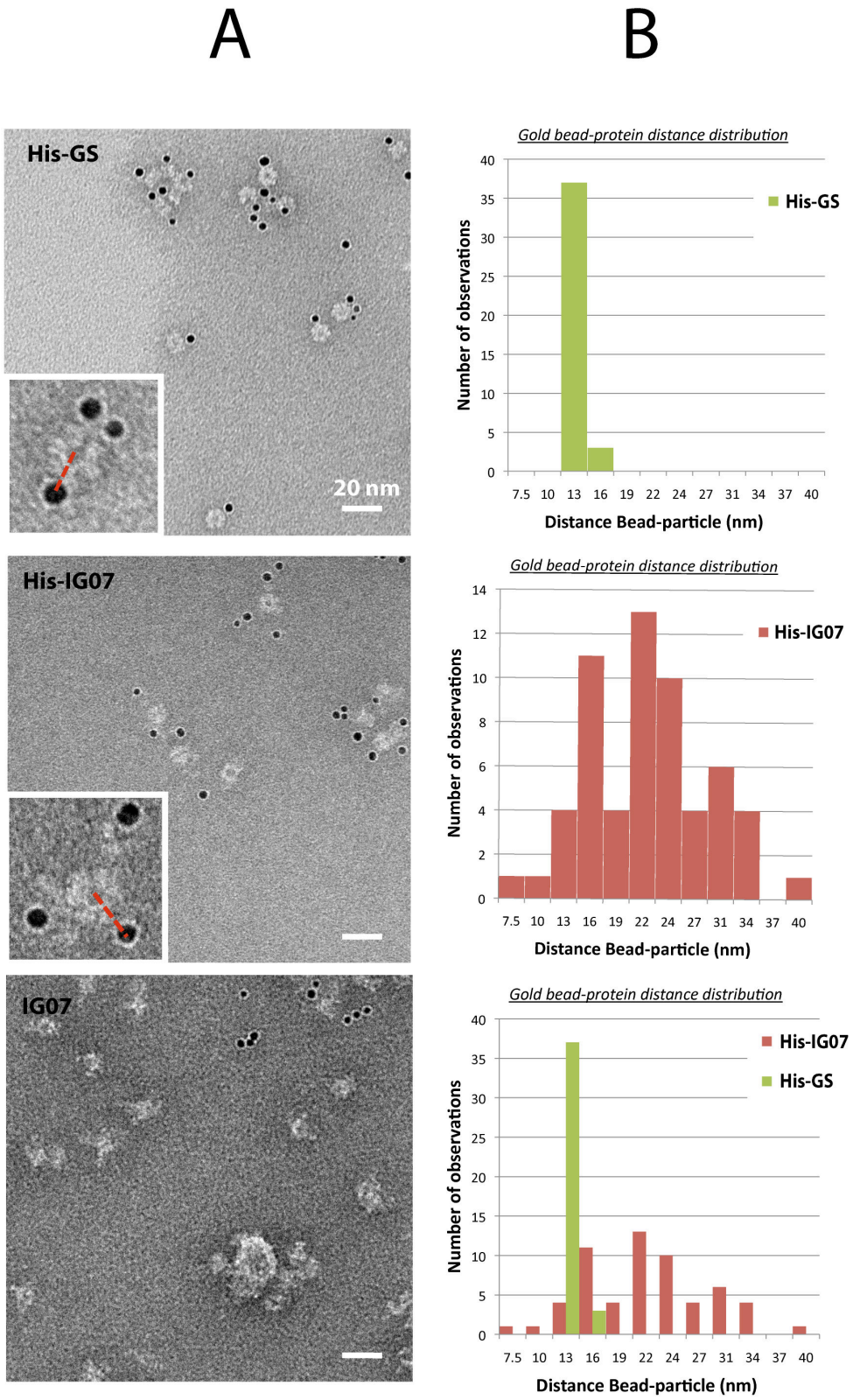
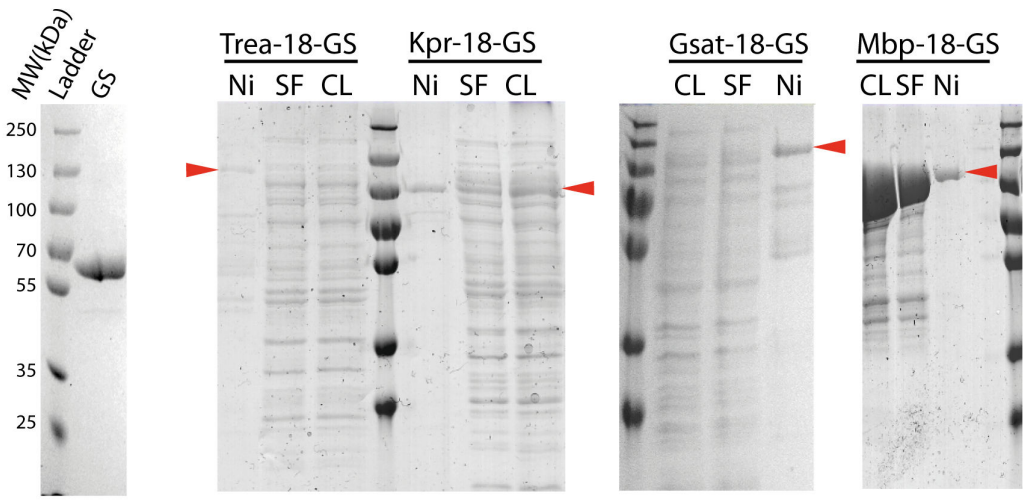


Figure 2.14: Gold labeling experiments. A) Electron micrographs of His-tagged template (GS), His-tagged Impβ-GS fusion (His-IG07) and untagged Impβ-GS fusion (IG07) mixed with 5 nm gold beads functionalized with Ni-NTA (black spots in the image). Insets: Close-up views of labeled particles. The red dotted lines indicate the distance measured to build the histogram. B) Bead-particle distance distribution obtained for 40 to 50 particles per construct. The scale bar indicates 20nm. EM grids were prepared as described (§ 4.5.1)

2.2.2 GS FUSIONS WITH GLOBULAR TARGETS

The four bacterial proteins mentioned in § 2.1.2 (trehalase, ketopanthoate reductase, glutamine synthetase adenylyltransferase and maltose-binding protein) each bear a solvent-accessible helical C-terminus. These were fused to the N-terminal helix of GS via the 18-residue L9 helical linker. These proteins were recombinantly expressed, purified by affinity and size-exclusion chromatography (SEC) (§ 4.4.1) and analysed by negative stain EM (§ 4.5.1). The aim of these efforts was to find at least one well-behaved fusion protein to pursue a proof-of-concept of protein symmetrization. The results of the first purification step are shown in Figure 2.15.



Junction sequence	Name	MW (kDa)
KMLDLAMAKALEAQKQKEQRQAASAEHVLTMLNEHEVKF	Trea-18-GS	~120
KESEYAMAKALEAQKQKEQRQAASAEHVLTMLNEHEVKF	Kpr-18-GS	~100
IGDDEAMAKALEAQKQKEQRQAASAEHVLTMLNEHEVKF	Gsat-18-GS	~110
TRITKAMAKALEAQKQKEQRQAASAEHVLTMLNEHEVKF	Mbp-18-GS	~100

Figure 2.15: SDS-PAGE of bacterial proteins fused to GS expressed in *E.coli*. Crude lysate (CL), soluble fraction (SF) and main elution fraction from Ni-NTA affinity column (Ni) were analyzed. For comparison, the GS template alone is shown on the left. Molecular weights of marker proteins are indicated at left. The table below the gel lists the expected molecular weights of target and fusion proteins in their monomeric form and the junction sequence between target (magenta) linker (green) and template (blue).

In all cases the apparent MW of the band observed in the elution fraction corresponded to the MW expected for the fusion protein. We did not detect proteolytic cleavage of the chimeras in the linker region, since bands corresponding to the template alone were not observed. As judged by SDS-PAGE, the Mbp-18-GS fusion was expressed and purified with the highest yield (50 mg per litre of bacterial culture), while the Trea-18-GS fusion was hardly expressed. Mbp-18-GS and Kpr-18-GS were concentrated and further analyzed by SEC and negative stain EM (Figure 2.16). In contrast, Trea-18-GS and Gsat-18-GS were insoluble after concentration and it was impossible to pursue SEC purification; therefore, we imaged them by negative stain EM after the first purification step (Figure 2.16). In both Mbp-18-GS and Kpr-18-GS SEC chromatograms (Figure 2.16 A) we identified 3 main peaks. The first elution peak corresponds to the void volume of the column, consistent with the presence of large soluble aggregates. Peak 3 corresponds to a species with a hydration radius (R_h) of ~2 nm, (according to the calibration curve of the column; § 4.4.1), much smaller than that of the template alone (R_h ~12nm). SDS-PAGE analysis (data not shown) suggests that this peak is due to the presence of contaminants. Elution peak 2 identifies a species of R_h ~20 nm, which when analyzed by SDS-PAGE reveals a band consistent with the expected MW of the monomeric chimeras. These features are consistent with an oligomeric species comprising 12 copies of Mbp-18-GS and Kpr-18-GS, as also gauged by structural predictions.

Fractions from peak 2 were analyzed by negative stain EM (Figure 2.16B). Visual inspection of Mbp-18-GS and Kpr-18-GS micrographs revealed top and side views of the GS template decorated with additional density attributable to the Mbp and Kpr target proteins. Although conformationally still too heterogeneous for our purposes, the Mbp-18-GS and Kpr-18-GS fusions showed promise as a starting point for subsequent linker optimization. In contrast, the Trea-18-GS and Gsat-18-GS samples looked extremely heterogeneous and only a few undecorated template top views were detected. By analogy with the IG07 results, this suggested that fusing Trea and Gsat to GS might hinder the proper folding and oligomerization of the particle, possibly also explaining the extremely low expression levels. Therefore, these constructs were not further pursued.

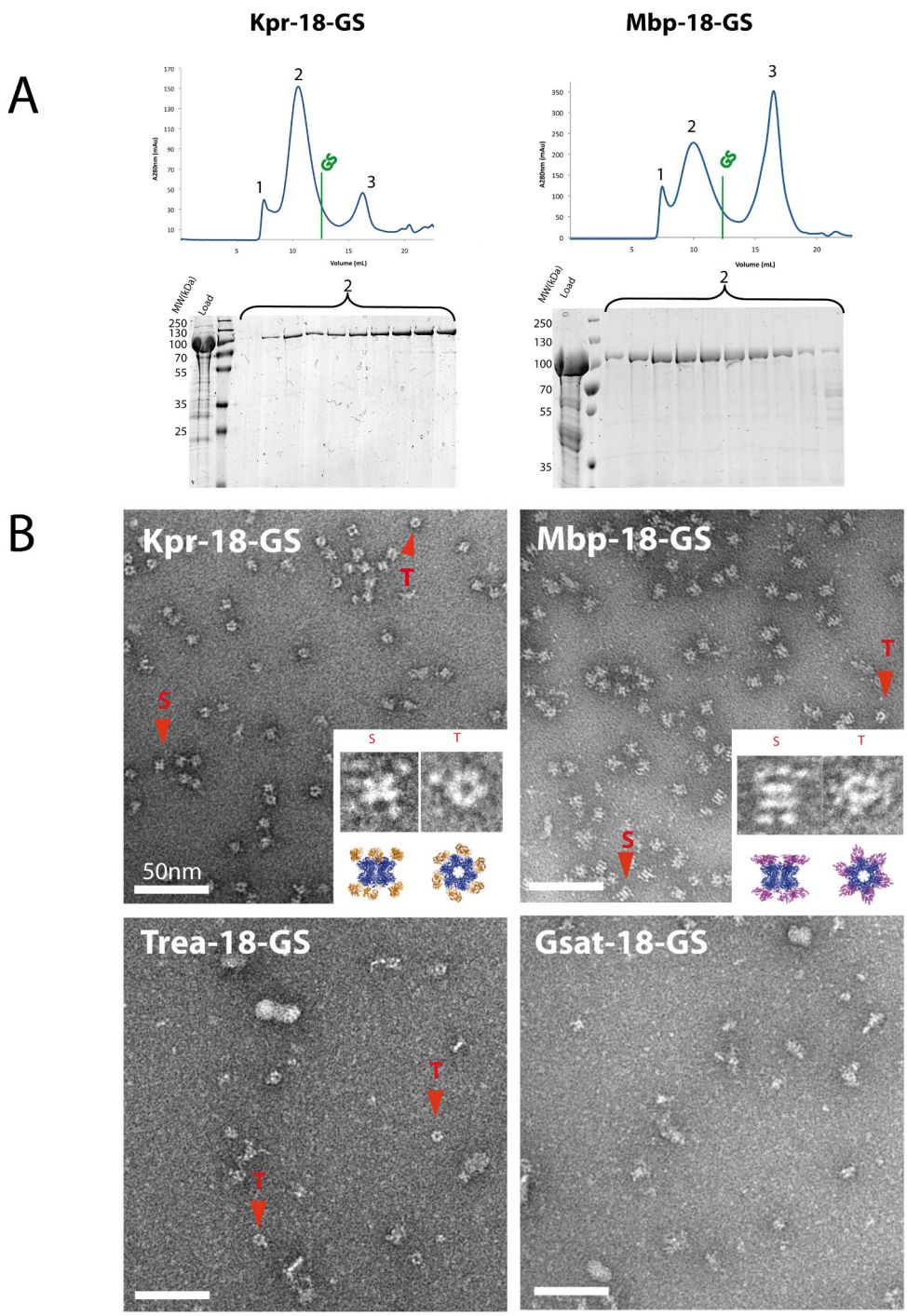


Figure 2.16: Analysis of bacterial proteins fused to GS. A) SEC profile of Mbp-18-GS and Kpr-18-GS: the green line marks the elution volume of the GS template, peak 1 falls within the void volume, peak 2 contains the oligomeric fusion protein, while peak 3 contains low-MW contaminants. B) Preliminary negative stain analysis of bacterial proteins fused with GS. The scale bar corresponds to 50 nm. Red triangles indicate clearly recognizable top (T) and side (S) views of oligomeric particles. Such views are nearly absent in Trea-18-GS and Gsat-18-GS, while they are abundant and decorated with target extra densities in Mbp-18-GS and Kpr-18-GS. Insets: Close-up of top and side views of Mbp-18-GS and Kpr-18-GS. The corresponding predicted structures are shown below at the same scale. EM grids were prepared as described (§ 4.5.1)

Because Mbp-18-GS was highly expressed and well-behaved by negative stain EM, we also attempted a fusion using a shorter linker of only three alanine residues (Mbp-3A-GS) to assess whether the reduced target-to-template distance would hinder correct dodecameric assembly. As mentioned above (§ 1.3), Mbp is often used as a carrier protein in fusion with proteins (P) that in their isolated form are resistant to crystallization. In this approach AAA has been identified as the best linker to produce rigid and well crystallizing Mbp-P fusions, as judged by the diffraction resolution (Moon et al., 2010). Even though AAA is compatible with different secondary structures, placing it between the terminal helices of the target and template increases its chances of adopting an α-helical secondary structure. Under this assumption one can predict the structure of the Mbp-3A-GS fusion particle as was done for Mbp-18-GS.

The Mbp-3A-GS construct was equally well expressed in *E.coli* and following purification was analyzed by SEC and negative stain EM. The comparison of Mbp-GS fusions linked with 3 or 18 residues is shown in Figure 2.17. Based only on the small difference in molecular mass (1%) one would expect the two molecules to elute similarly. However, Mbp-3A-GS has a much larger elution volume yielding an estimated R_h of ~17 nm, compared with an estimated R_h of ~21 nm for Mbp-18-GS (Figure 2.17A). Negative stain electron micrographs reveal that particles of Mbp-3A-GS are more homogeneous than those of Mbp-18-GS. One can identify both top and side views of the template decorated with at least 6 extra densities, suggesting the presence of fully decorated dodecameric species consistent with the structure predicted by *in silico* modelling (Figure 2.17B).

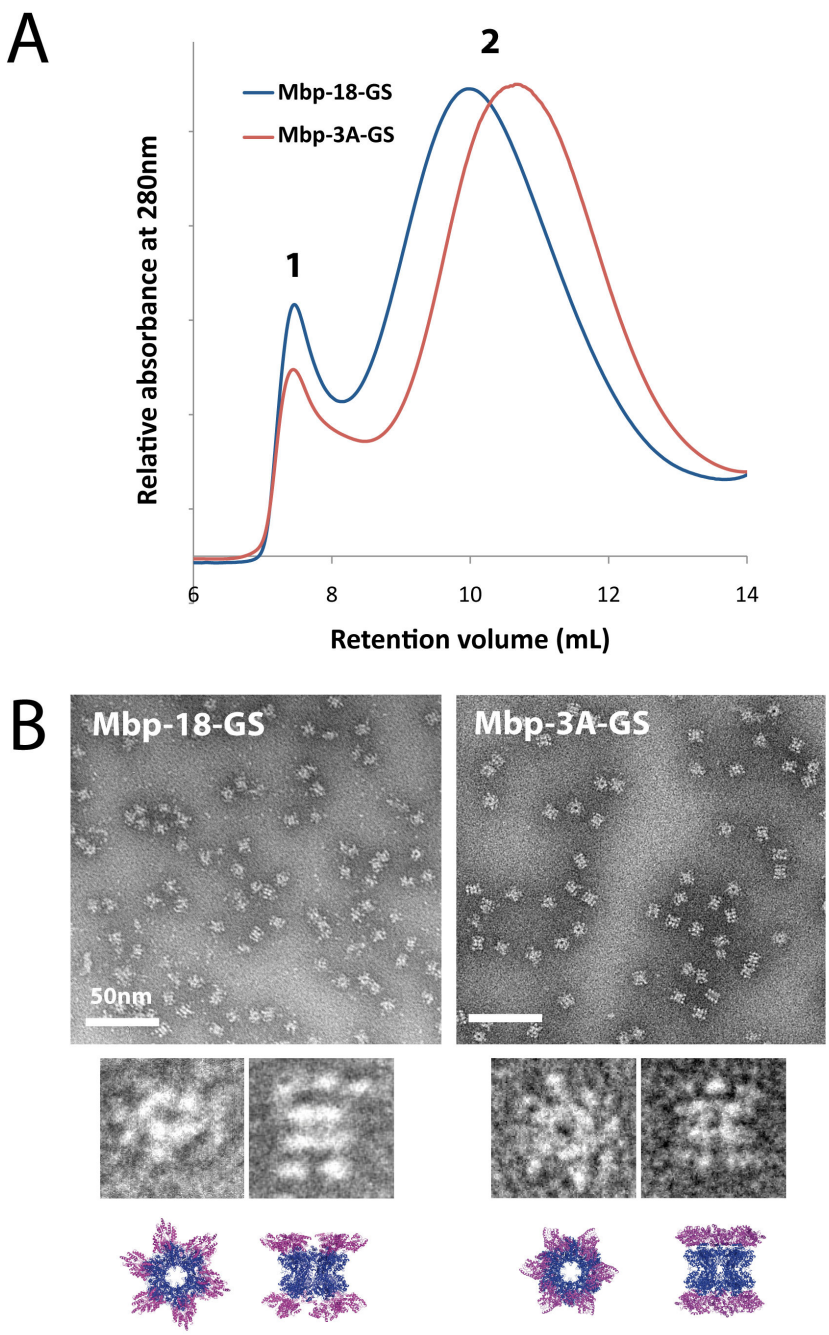


Figure 2.17: Comparison between Mbp-GS fusion proteins with two different linker lengths. A) SEC profiles of Mbp-18-GS and Mbp-3A-GS. Peak 1 corresponds to the void volume while peak 2 contains the oligomeric fusion proteins eluting at different volumes, consistent with Mbp-3A-GS being more compact. B) Electron micrographs of Mbp-3A-GS and Mbp-18-GS negatively stained with SST (Sodium Silico Tungstate) 2% w/v, recorded at 120kV, at a nominal magnification of 22000x. The scale bar indicates 50m. EM grids were prepared as described (§ 4.5.1). The images show a higher degree of homogeneity and a more compact appearance of Mbp-3A-GS compared to Mbp-18-GS. Insets: close-up of top and side views of the chimeras compared to the predicted structures.

Using ~100 selected particles we obtained a D6 restrained *ab initio* model using the webserver RIco http://rico.ibs.fr/RIcoWebServer/, developed by our collaborator Dr. Leandro Estrozi (§ 4.5.2). Reprojections of this model were used to perform alignment and angular assignments of ~4000 Mbp-3A-GS particles selected by hand (the projection matching method was used; see § 4.5.2 for more details). The overall shape of the negative stain volume agrees well with the predicted structure, although the densities corresponding to Mbp are slightly smaller than expected, suggesting the presence of local flexibility (Figure 2.18).

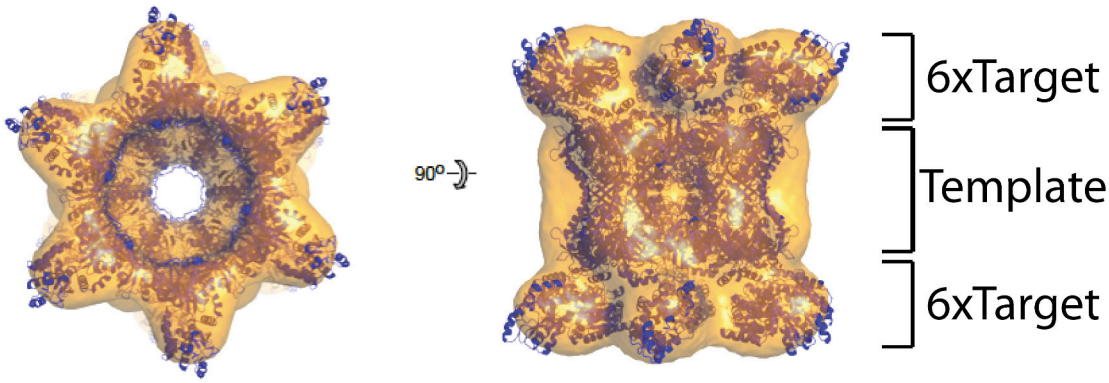


Figure 2.18: Fit of Mbp-3A-GS in the negative stain volume. The map (orange) is shown in the two preferential orientations adopted by particles on the grid (top and side view). The ribbon diagram of the predicted structure (blue) is well covered by the electron density, except for peripheral regions of Mbp. This suggests that Mbp may be flexibly attached to GS and hence its density is partly averaged out when enforcing the D6 symmetry.

Finally, we performed the same gold-labeling experiment as described above for Impβ-GS to investigate eventual folding problems (Figure 2.19). The gold-particle distance distribution observed for His-tagged Mbp-3A-GS is much sharper than for Impβ-GS and peaks at 16 nm, consistent with the presence of a well folded Mbp domain attached to GS.

All these experiments suggested that Mbp-3A-GS was a better behaved fusion protein than previously screened dodecamers and represented a good starting point for linker optimization and detailed structural analysis.

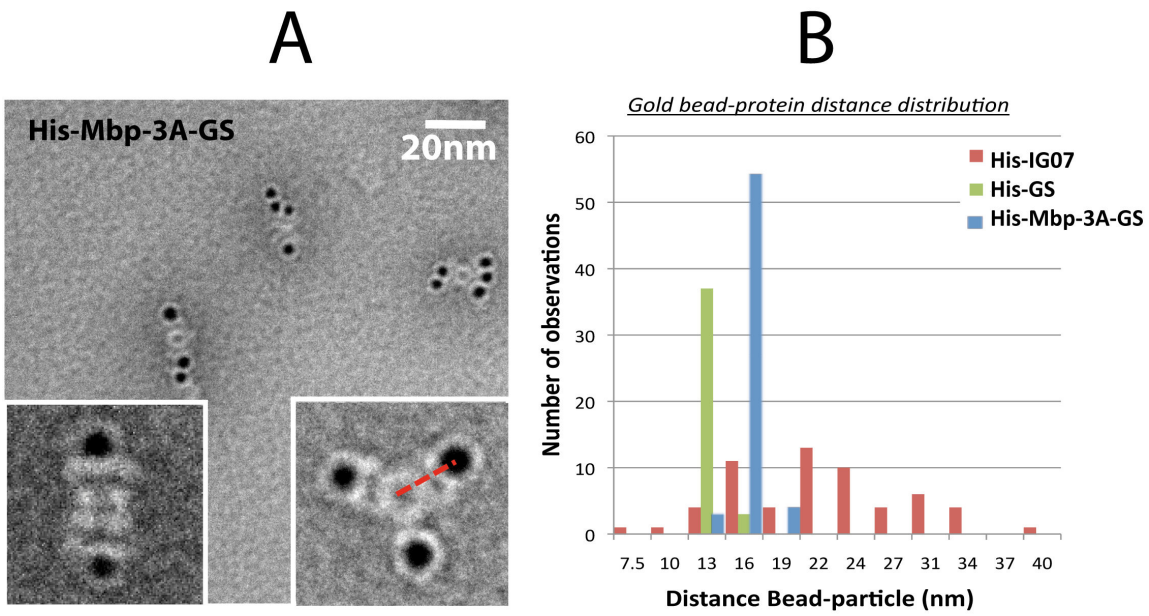


Figure 2.19: Gold labeling of Mbp-3A-GS. A) Negative stain micrographs of Mbp-3A-GS labeled with 5 nm gold beads and negatively stained with SST (Sodium Silico Tungstate) 2% w/v (120kV, magnification 22000x). EM grids were prepared as described (§ 4.5.1). Insets: side and top view close-ups. A particle-bead distance is shown in red. B) Comparison between gold-particle distance distributions of Impβ-GS (IG07), GS and Mbp-3A-GS, showing a sharper distribution for the latter compared to IG07.

In parallel, because a general target protein might have a non-helical C-terminus, we sought to test whether three alanines could still work as a starting linker for connecting such a target to GS. To achieve this, we recombinantly expressed and purified the green fluorescent protein eGFP fused to GS (Gfp-3A-GS) and analyzed it by SEC and negative stain EM (Figure 2.20). The monomer MW estimated by SDS-PAGE was consistent with the expected size of the fusion (90 kDa) and the elution volume corresponded to a species with an R_h of ~15 nm, consistent with a dodecameric species. Negative stain EM analysis confirmed that the particle was likely dodecameric and revealed that the template was decorated by up to 5 target densities. Despite a certain level of heterogeneity of the sample, the result is promising, as it broadens the range of potential applications of the symmetrization approach to proteins that do not possess helical termini.

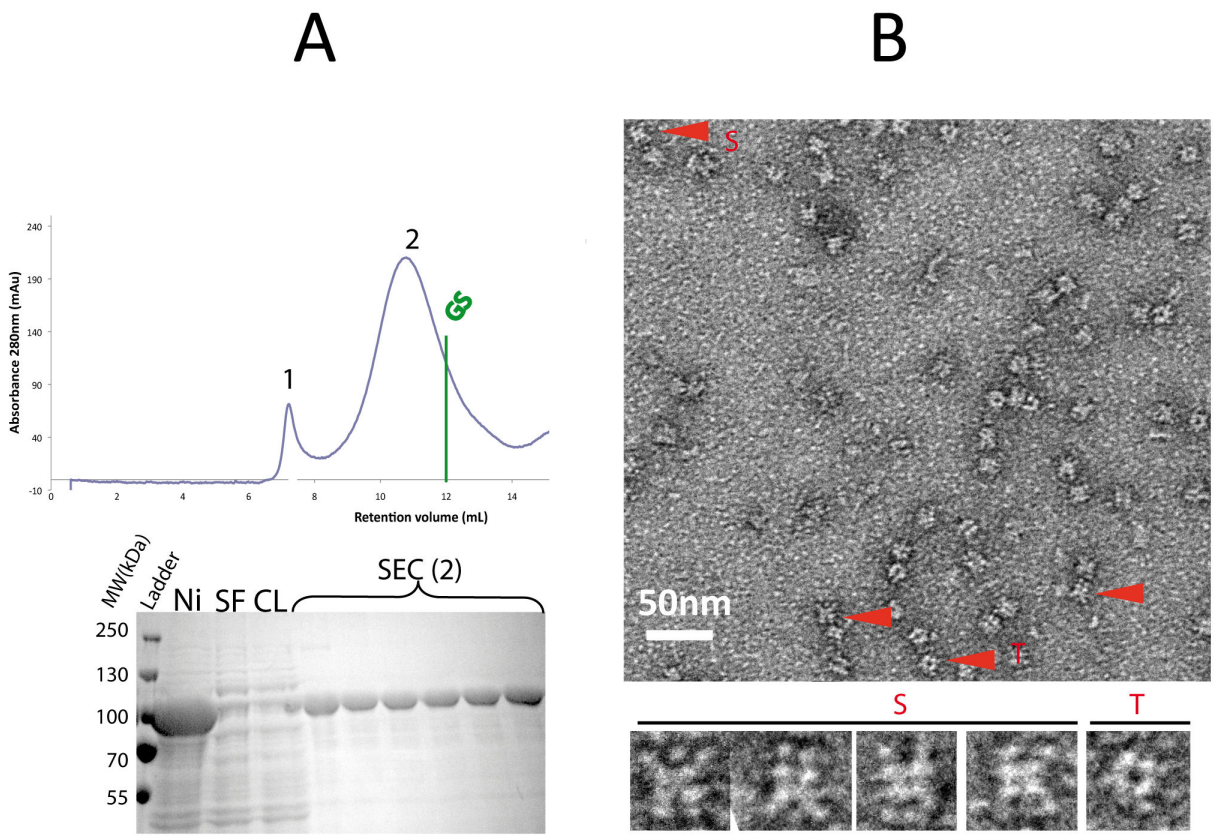


Figure 2.20: Purification and preliminary negative stain analysis of Gfp-3A-GS fusion A) SEC chromatogram: the green line indicates the elution volume of the isolated GS template. Two main peaks are visible: 1, corresponds to aggregates and 2 corresponds to the desidered fusion, as verified by the SDS-PAGE gel below. Ni, SF and CL indicate the elution fraction after Ni-affinity purification, soluble fraction and crude *E.coli* lysate after expression. B) Electron micrograph of Gfp-3A-GS negatively stained with SST (Sodium Silico Tungstate) 2% w/v, recorded at a nominal magnification of 22000x, and a 120kV voltage. EM grids were prepared as described (§ 4.5.1). S and T indicate side and top views of the molecule. Inset : close-ups of oligomeric species detected on the grid.

2.2.3 MBP-E2 FUSIONS

As introduced in § 1.2, the presence of symmetry in a particle considerably facilitates the single particle 3D reconstruction process, providing useful constraints for accurate alignment and high resolution determination. The most favourable case concerns icosahedral particles, for which the reconstruction steps are readily automated and have provided near atomic resolution structures (Zhou, 2008). We sought to obtain an icosahedral fusion by fusing Mbp to the (non-helical) N-terminus of the 60-meric icosahedral template protein E2 of PDH. As a starting linker we used a stretch of three

alanines (successful in GS fusions) and produced two additional shorter constructs by completely removing the linker and by further removing 2 residues from the template. For convenience we refer to these Mbp-E2 fusions as Mae0, MaeΔ3 and MaeΔ5, where the numbers specify how many residues were deleted from the starting linker, as summarized in Figure 2.21A.

We recombinantly produced the 3 chimeras as described in § 4.3.1, and analyzed them by several biophysical techniques and by negative stain EM. As judged by SDS-PAGE (Figure 2.21B), the three Mae constructs present a high degree of purity after the first affinity Ni-NTA step and migrate as bands consistent with the MW expected for the fusion protein (~68 kDa). Despite the expected MW of the oligomer (4.2 MDa) falling within the separation range of the column, the samples elutes at the void volume, hampering estimation of the hydrodynamic radius. Similarly, it was impossible to achieve migration in native PAGE experiments, even at very low acrylamide concentrations (3.5%) (data not shown). As judged by preliminary negative stain analysis (Figure 2.21D), the three constructs consist of a mixture of aggregates, quasi-spherical particles likely corresponding to the decorated template, and smaller species resembling broken oligomers. Compared to dodecameric GS constructs, it is less straightforward to visually estimate how many target densities decorate the template. Moreover, we did not observe a large change in sample quality on varying the linker length, although the shortest construct (MaeΔ5) seems slightly more homogeneous. These observations are in agreement with the size distribution of particles estimated by dynamic light scattering (DLS; § 4.4.4), as shown in Figure 2.21 C. In all the three samples two main species of R_h ~20 and ~100 nm were detected with a rather high polydispersity index (20-25%), consistent with the expected oligomer and higher MW aggregates, respectively. Because the aggregate peak appears less prominent in the shortest construct, MaeΔ5, we focused on this sample for further studies.

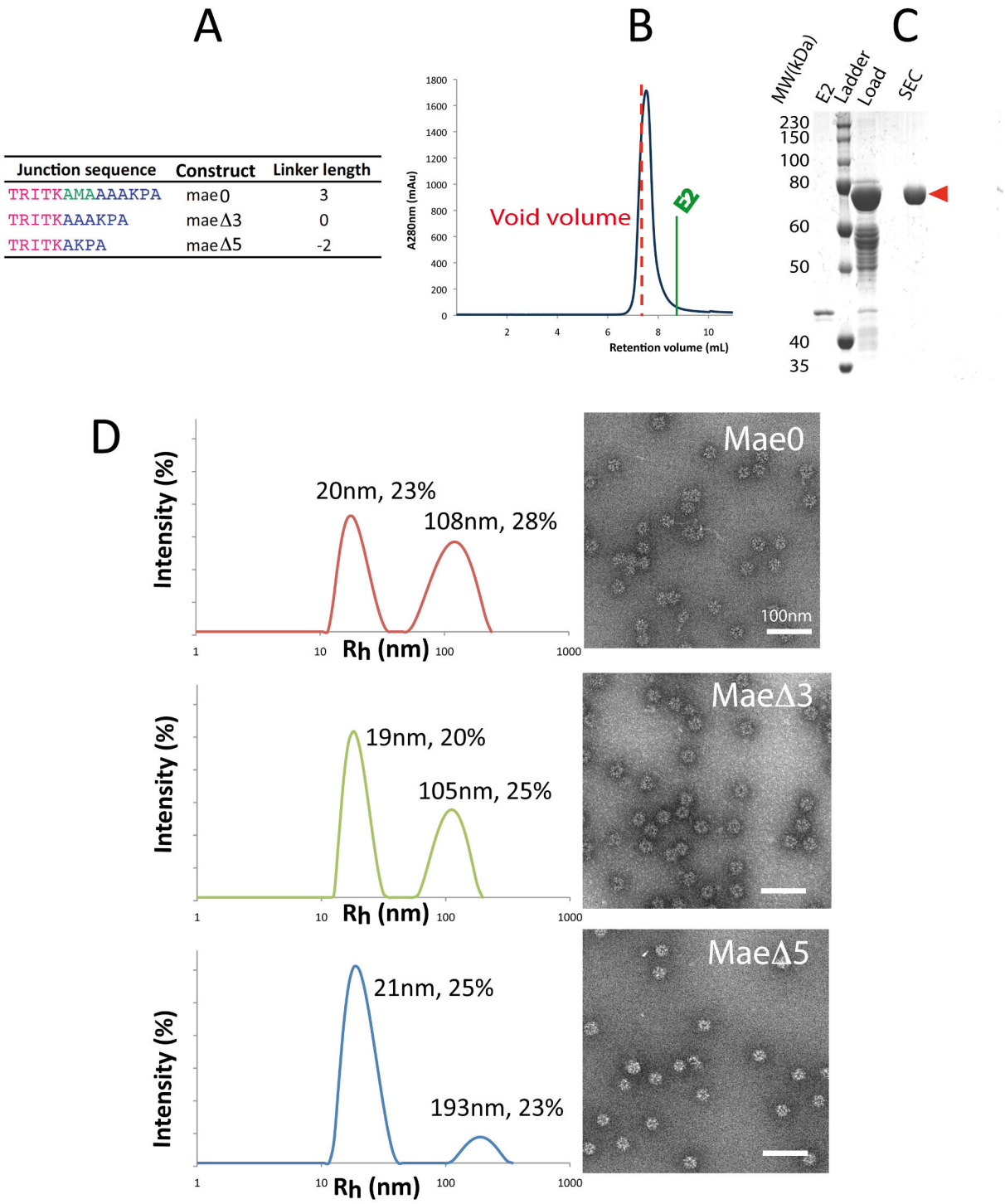


Figure 2.21: Preliminary characterization of Mbp-E2 (Mae) constructs. A) Junction sequence of the three fusions. B) SEC profile of Mae0 (representative of the three fusions). The void volume and elution volume of the isolated E2 template are indicated. C) Corresponding SDS-PAGE analysis, revealing a high degree of protein purity (red arrow) and the absence of proteolytic cleavage of the linker, as judged by comparison with the isolated template. D) *Left,* Size distribution of particles in terms of hydrodynamic radii (R_h) estimated from DLS measurements. The main species have R_h values of 19-21 nm. *Right,* corresponding negative stain micrographs. MaeΔ5 appears to be less contaminated by aggregates (R_h >100 nm) both by DLS and negative stain EM.

First we attempted to improve sample homogeneity by ultracentrifugation in a sucrose gradient (0-50%, 30 krpm for 10'). The majority of the protein was in the fraction containing ~30% sucrose, and as judged by negative stain analysis was devoid of large aggregates. However, particles still appeared inhomogeneous in size (Figure 2.22 A). Despite the heterogeneity, by taking advantage of the high degree of symmetry it was possible to build a symmetry restrained 3D map *ab initio* with only ten well-shaped particles (600 asymmetric units) using the program RIco (http://rico.ibs.fr/RIcoWebServer/) in collaboration with the author, Dr. Estrozi (Figure 2.22C, Estrozi and Navaza, 2010). This revealed the negative stain volume of MaeΔ5 to comprise a double shell of density: an internal layer corresponding to the template, and an external one corresponding to the target moieties. The outer shell is made up of 20 triads compatible with the presence of three Mbp molecules connected to the three N-termini of E2 monomers that cluster at the 3-fold axis (§ 2.1.1, Figure 2.4D). This starting model was used as reference to select ~1000 frozen hydrated particles and to generate a low resolution native 3D reconstruction, by using methods based on symmetry adapted functions (Estrozi and Navaza 2008, § 4.5.2). The MaeΔ5 cryoEM map overall resembles the negative stain model (Figure 2.22D). The E2 crystal structure fits well to the inner shell of density, whereas the density attributed to the Mbp target can accommodate only ~50% of the Mbp crystal structure. One possible explanation is that the Mbp triad, which connects to 3 closely-spaced E2 N-termini near the 3- fold axis, is rotated by ~60° with respect to the underlying E2 trimer, such that individual Mbp subunits make few interactions with the template. This "staggered" arrangement could favour flexibility of the Mbp subunits, whose densities would be averaged out in the symmetry restrained map. This preliminary study demonstrated that it is possible to obtain an icosahedral fusion using E2 as a template. However, the starting MaeΔ5 fusion requires further optimization in terms of homogeneity and rigidity. This can only be done by a trial and error process, because the non-helical linker makes it difficult to rationalize how to improve fusion constructs by changing the linker length. To expedite progress on the thesis project, it was therefore decided to focus on a target-template fusion protein having a helical linker.

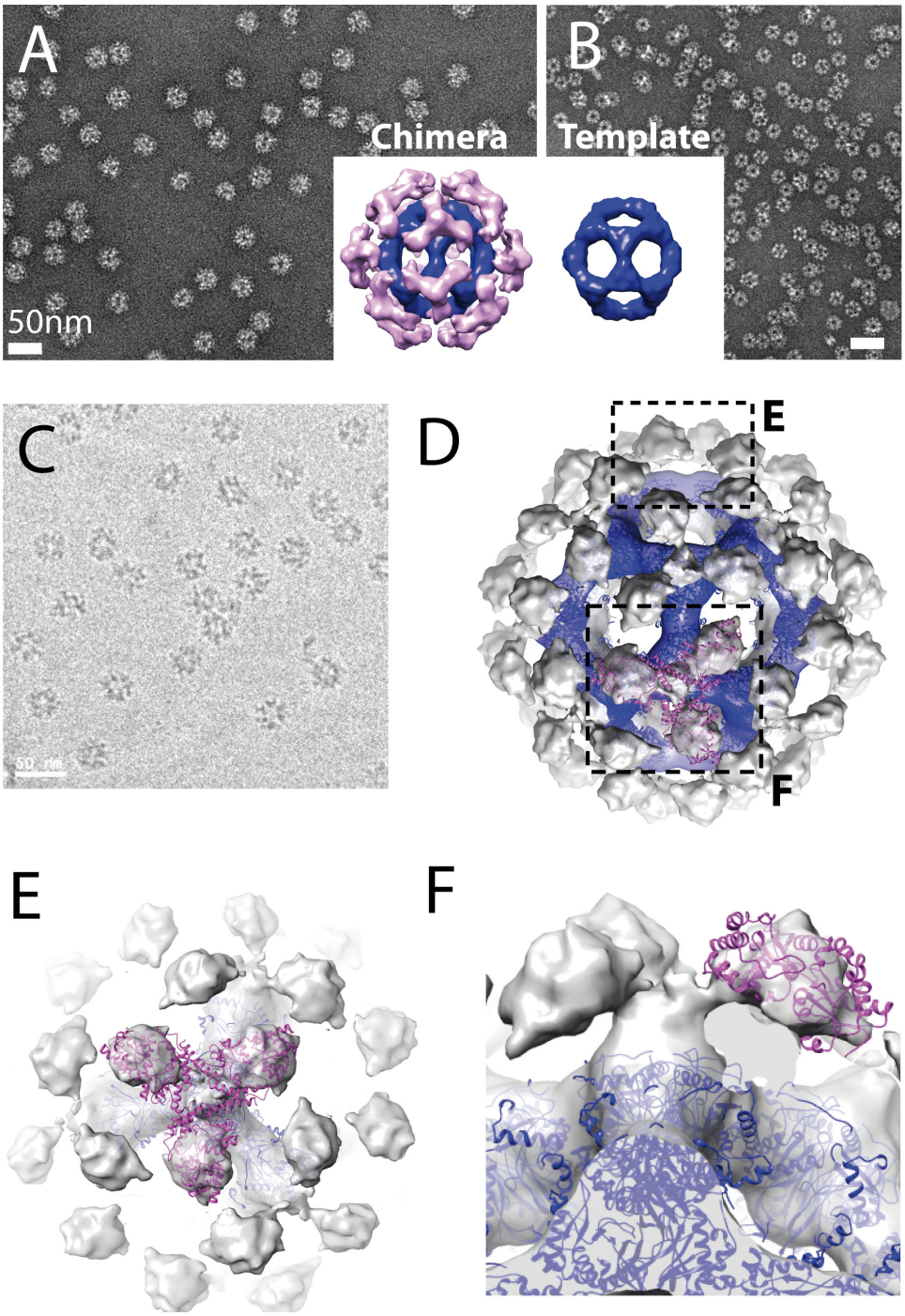
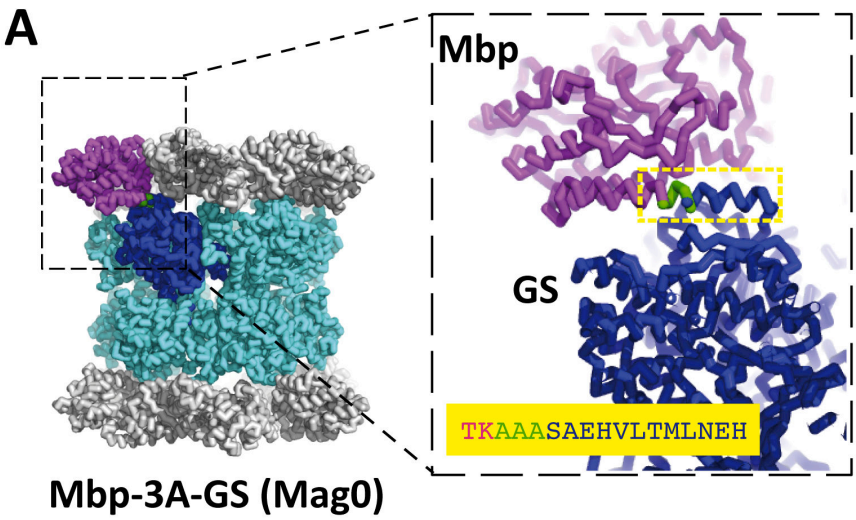


Figure 2.22: Negative stain and cryoEM analysis of MaeΔ5. A,B) Negative stain micrographs of A) MaeΔ5 and B) the E2 template. Both samples were stained with SST (Sodium Silico Tungstate) 2% w/v and imaged at 120kV and at a nominal magnification of 22000x. Insets: corresponding 3D reconstruction, in which the E2 template is shown in (blue) and additional density is coloured pink. C) MaeΔ5 embedded in vitreous ice and imaged at 300kV on a POLARA microscope, at a 23K nominal magnification and with a -3 µm defocus. D) CryoEM reconstruction of MaeΔ5, where the template map is coloured blue. Mbp trimeric unit decorating the template (shown in magenta) was fitted into the presumed target density. E) The view of the central Mbp triad (magenta) shows how Mbp subunits are staggered relative to the underlying E2 subunits. F) The fit of a single Mbp molecule shows that the map volume is smaller than the target size, suggesting a certain degree of flexibility.

2.3 OPTIMIZATION OF THE LINKER LENGTH IN MBP-GS FUSIONS

Among the screened target-template combinations, the Mbp-GS fusion bearing a three-alanine linker (Mbp-3A-GS) was the most promising construct for structural analysis. As suggested by the negative stain volume analysis (§2.2.2, Figure 2.18), this construct appeared to have local flexibility that would hamper high resolution structure determination of the Mbp target. By examining the junction region in the predicted structure, we noticed that the last 2 residues of Mbp and the entire N-terminal helix of GS do not appear critical for the integrity of the globular fold of these proteins (Figure 2.23A). In fact, the two domains could be brought closer by shortening this stretch of 17 residues, leading to more compact symmetrical particles potentially better suited for cryoEM analysis. On the other hand, the deletions might adversely affect the rigidity of the linker or the ability of the template to oligomerize.

Therefore, we produced a panel of 18 Mbp-GS constructs by sequentially truncating the junction sequence (Figure 2.23 B) and screened these by various biophysical techniques and by negative stain EM to identify constructs amenable to high resolution cryoEM analysis, as described below. Henceforth, for simplicity, Mbp-3A-GS will be referred to as Mag0, and the deletion constructs as MagΔN, where N is the number of residues removed from the original AAA linker sequence.



B

Junction sequence	Fusion name	Linker length
TRITKAAASAEHVLTMLNEHEVKF	Mag 0	3
TRITKAASAEHVLTMLNEHEVKF	Mag Δ1	2
TRITKASAEHVLTMLNEHEVKF	Mag Δ2	1
TRITKSAEHVLTMLNEHEVKF	Mag Δ3	0
TRITKAEHVLTMLNEHEVKF	Mag Δ4	-1
TRITKEHVLTMLNEHEVKF	Mag Δ5	-2
TRITKHVLTMLNEHEVKF	Mag Δ6	-3
TRITKVLTMLNEHEVKF	Mag Δ7	-4
TRITKLTMLNEHEVKF	Mag Δ8	-5
TRITKTMLNEHEVKF	Mag Δ9	-6
TRITKMLNEHEVKF	Mag Δ10	-7
TRITKLNEHEVKF	Mag Δ11	-8
TRITKNEHEVKF	Mag Δ12	-9
TRITKEHEVKF	Mag Δ13	-10
TRITKHEVKF	Mag Δ14	-11
TRITKEVKF	Mag Δ15	-12
TRITEVKF	Mag Δ16	-13
TRIEVKF	Mag Δ17	-14

Figure 2.23: Optimization of Mbp-GS linker length A) Predicted structure of Mbp-3A-GS (Mag0) and close-up of the three alanine (green) junction, revealing the presence of 17 effective linker residues (yellow box) between the target (magenta) and template (blue) subunits. B) Table describing the junction sequence of MagΔN constructs obtained by deletion of linker residues. Mbp, tri-alanine and GS residues are shown in magenta, green and blue, respectively.

2.3.1 Biophysical characterization

To obtain an accurate structure of the target protein by cryoEM it is crucial to have a conformationally homogenous sample composed of highly symmetrical particles

exhibiting minimal flexibility of the target relative to the template. This condition is more likely to be achieved for fusion constructs yielding highly compact oligomeric particles than for those yielding more extended particles. Conformational homogeneity and compactness can be associated with the presence of a single sharp band by native PAGE, a small hydrodynamic radius (R_h) and low polydispersity index (PD) by SEC and DLS, and a single cooperative unfolding step and high melting temperature (T_m) by TSA (these methods are detailed in § 4.4). To compare the 18 Mag fusions and select the best construct(s) for cryoEM, we analyzed them using the four biophysical assays mentioned above. Figure 2.24 summarizes the results obtained as a function of linker length. As examples, we show the raw data for 3 constructs that are illustrative of the range of results observed (form the best to the worst behaved construct). The fusion constructs Mag0 to MagΔ11 migrated as a single band on a native gel, whereas shorter linkers led to the appearance of multiple bands (Figure 2.24A). Therefore, constructs MagΔ12-MagΔ17 were excluded from further analysis, being deemed unsuitable for cryoEM analysis. The remaining 12 constructs, despite have very similar molecular weights, exhibited a range of SEC retention volumes which corresponded to R_h values between 13.2 and 17.5 nm, as estimated from the calibration (§ 4.4.1) (Figure 2.24B). In this case, the best behaved construct was MagΔ11, which exhibited the largest elution volume (corresponding to an R_h of 14 nm). In DLS experiments the estimated R_h values varied between 9.3 and 10.8 nm with polydispersity indices varying from 7% to 25% (Figure 2.24C). In this case the best behaved constructs were MagΔ7 and MagΔ8. In the thermal shift assays, both thermal stability (T_m) and the cooperativity of the transition were evaluated. The higher the stability, the less accessible the hydrophobic core is to the fluorescent probe. A single step is interpretable as a compact particle in which the target and template subunits unfold cooperatively, whereas multiple steps suggest independent unfolding events and a low degree of particle compactness. The thermal stability was almost constant (T_m ~65 ˚C) for constructs Mag0-MagΔ5 and was lower and rather variable for shorter linker lengths (Figure 2.24D). Similarly, almost all Mag0-MagΔ5 constructs showed a cooperative transition, while shorter constructs presented double or multiple transitions. The best behaving construct was MagΔ1 with a T_m of 66 ˚C and a single transition.

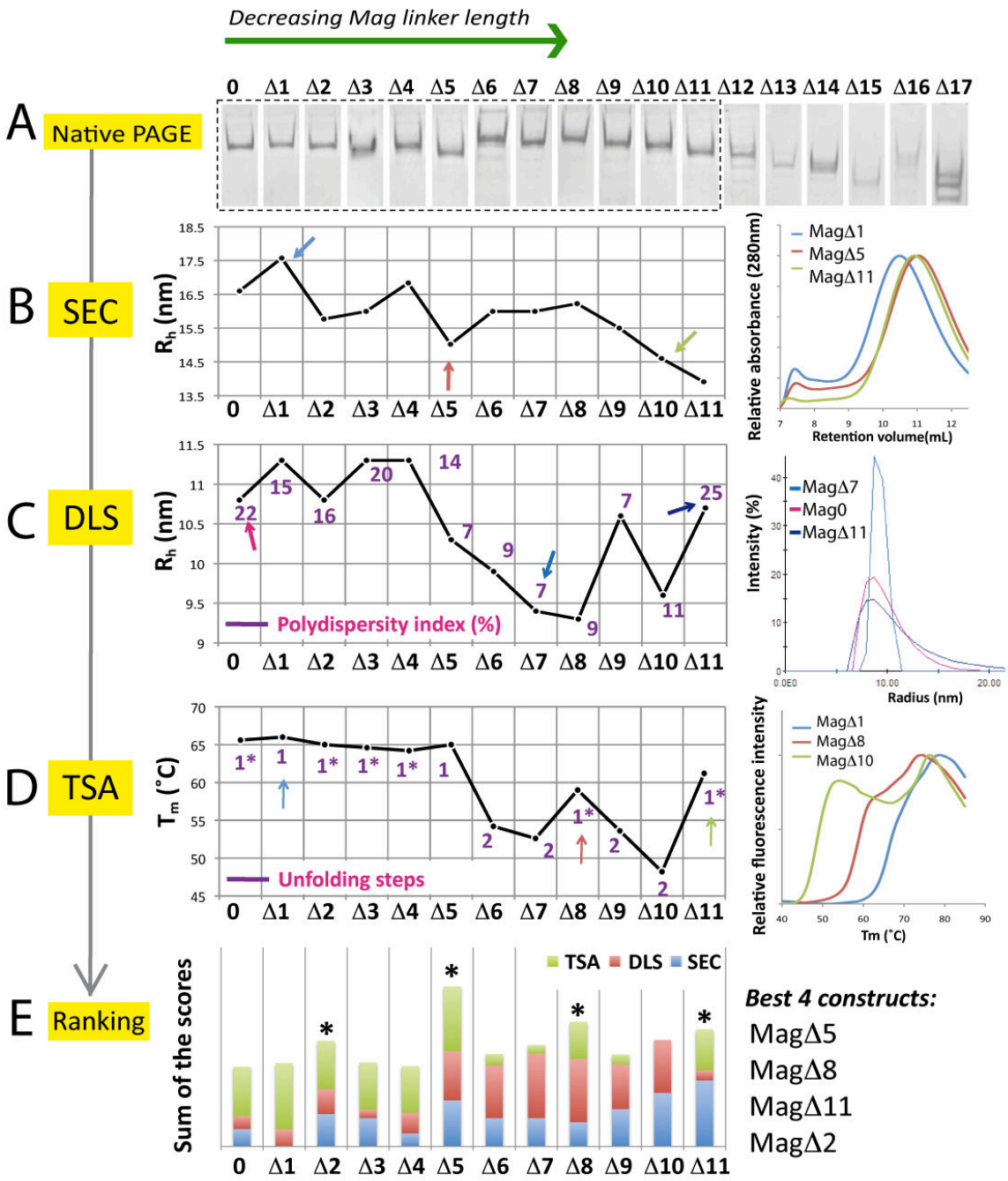


Figure 2.24: Biophysical studies of MagΔN constructs. The parameters retrieved from each technique are plotted as a function of linker length. The arrows indicate the three constructs representative of the range of observed parameters and whose raw data are shown at the right. A) Native PAGE analysis: the dotted-line box delimits constructs migrating as a single band and retained for further analysis. B) *Left,* Plot of hydrodynamic radii (R_h) determined by SEC. *Right,* Chromatograms illustrating different elution profiles. C) *Left,* Plot of hydrodynamic radii (R_h) determined by DLS and corresponding polydispersity index (%) values (in purple). *Right*: Profiles illustrating broad (high polydispersity) and narrow (low polydispersity) size distributions. D) *Left,* Plot of melting temperatures (T_m) determined by TSA and corresponding number of transitions observed (in purple). 1: single step, 1*: quasi-two steps, 2: multiple steps. *Right,* Profiles illustrating different melting behaviour of constructs). E) Ranking: sum of individual scores identifying the four best constructs (*)

Interestingly, by comparing the MagΔ1 TSA profile with those of the template alone (presenting a quasi-two step profile with melting temperatures of 42˚C and 61˚) and of the target alone (which unfolds cooperatively at a T_m= 55˚C (Soon et al., 2012)) we noticed a mutual stabilization between the target and template in the fusion. This phenomenon occurs in all three constructs showing a single denaturation step: MagΔ1, MagΔ5 and MagΔ11.

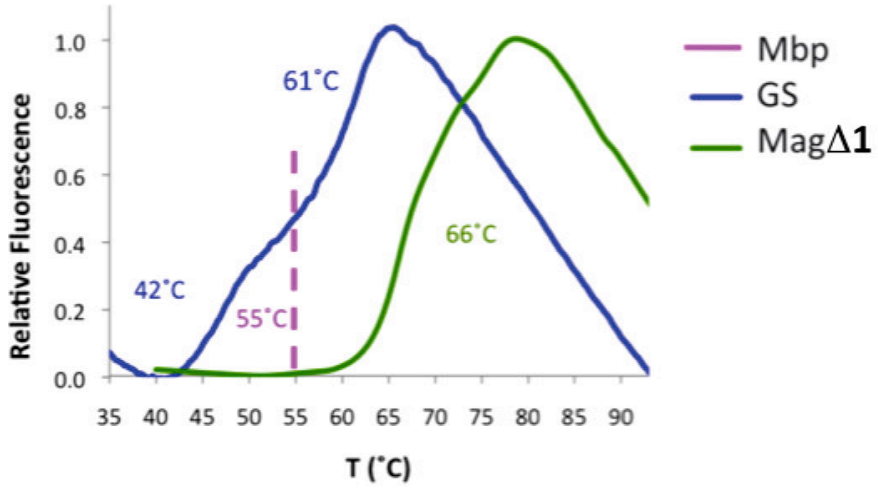


Figure 2.25: Comparison of TSA profiles of MagΔ1 with free template (GS) and target (Mbp), revealing their mutual stabilization in the fusion.

To combine the results observed for each fusion construct, we assigned a score between 1 and 0 for each technique based on the retrieved parameters (R_h, PD and T_m and number of unfolding transitions). In each case a normalized value (X_N) was calculated as: $X_N = (X - X_{min}) / (X_{max} - X_{min})$, where X is the value measured for a given construct and X_{min} and X_{max} are the minimum and maximum values across all constructs, respectively. The resulting scores between 1 (best) and 0 (worst) were calculated as follows:

1. SEC Score = $1- R_{hN}$
2. DLS Score = $[(1 - R_{hN}) + (1-PD_N)]/2$
3. TSA Score = $(T_{mN} + S_T) / 2$

where R_{hN}, PD_N, and T_{mN} are the normalized values of R_h, PD and T_m, respectively. (For TSA experiments, where more than one thermal transition was observed, the T_m used was that of the first transition.) S_T is an additional parameter that reflects whether proteins

exhibited a single unfolding transition (S_T=1) or multiple transitions (S_T=0). When a quasi-two step transition (similar to the GS profile) was observed, an intermediate value of 0.5 was assigned. The sum of the three scores plotted on a histogram permits a global comparison of the different fusion constructs and allows the four best behaved constructs to be identified: MagΔ5, MagΔ8, MagΔ11 and MagΔ2 (Figure 2.24E).

2.3.2 Negative stain EM analysis

Next, the 18 Mag constructs with different linker lengths were all analyzed by negative stain EM (Figure 2.26). The MagΔ12-MagΔ17 fusions that showed multiple bands in native PAGE were all extremely heterogeneous. As an example, the negative stain micrograph of MagΔ12 is shown. Constructs migrating as a single band in native PAGE presented a large variation of conformational homogeneity: some constructs appeared dodecameric and quite homogeneous, others showed partial oligomerization and the presence of aggregates. MagΔ1, MagΔ4, MagΔ5, MagΔ8 and MagΔ11 seemed well behaved; however it was difficult to establish which of these constructs was most suitable for cryoEM analysis by simple visual inspection. Of the four constructs which gave the best scores by biophysical characterization (indicated by a star in Figure 2.26 and presented again in Figure 2.27), MagΔ2 seems to adopt different oligomerization states while MagΔ11 appears dodecameric but exhibits a certain degree of flexibility. Conversely, MagΔ5 and MagΔ8 appear much more homogeneous. By applying multivariate statistical analysis (§4.5.2) to a small (400 particles) dataset, 2D class averages of side and top views were obtained (Figure 2.27 insets). Side views measure approximately 20 nm along the longest axis and exhibit four distinct layers, consistent with a central double layer of GS subunits flanked on both sides by a layer of Mbp subunits: in MagΔ5 and MagΔ8 we observe 2x2 and 3x2 Mbp extra densities, respectively. Top views have a donut-like appearance with an outer diameter of ~15 nm and have a 6-fold symmetry for MagΔ5, but the symmetry is less pronounced in MagΔ8. These initial observations suggest that both MagΔ5 and MagΔ8 are dodecameric, consistent with the D6 symmetry of the template and with Mbp adopting either an eclipsed or staggered position relative to GS, respectively. As both fusions ranked well in the biophysical screening and looked promising by negative stain EM, they were both pursued for cryoEM analysis.

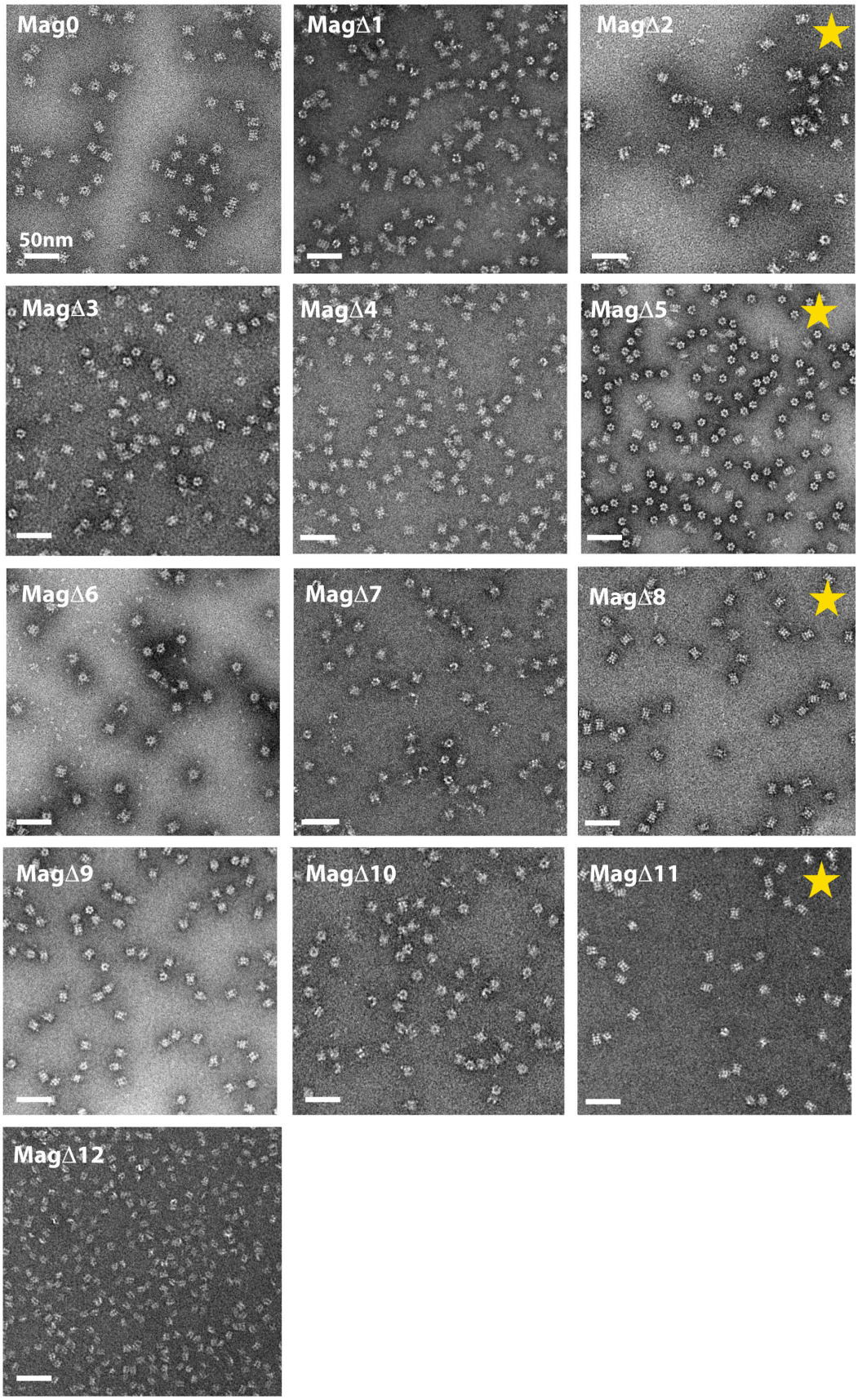


Figure 2.26: Negative analysis of Mag constructs with different linker lengths. Staining solution SST (Sodium Silico Tungstate) 2% w/v, micrographs collected at 120kV and a nominal magnification of 22000x (§ 4.5.1). Yellow stars indicate the samples that are gauged as most compact by biophysical characterization.

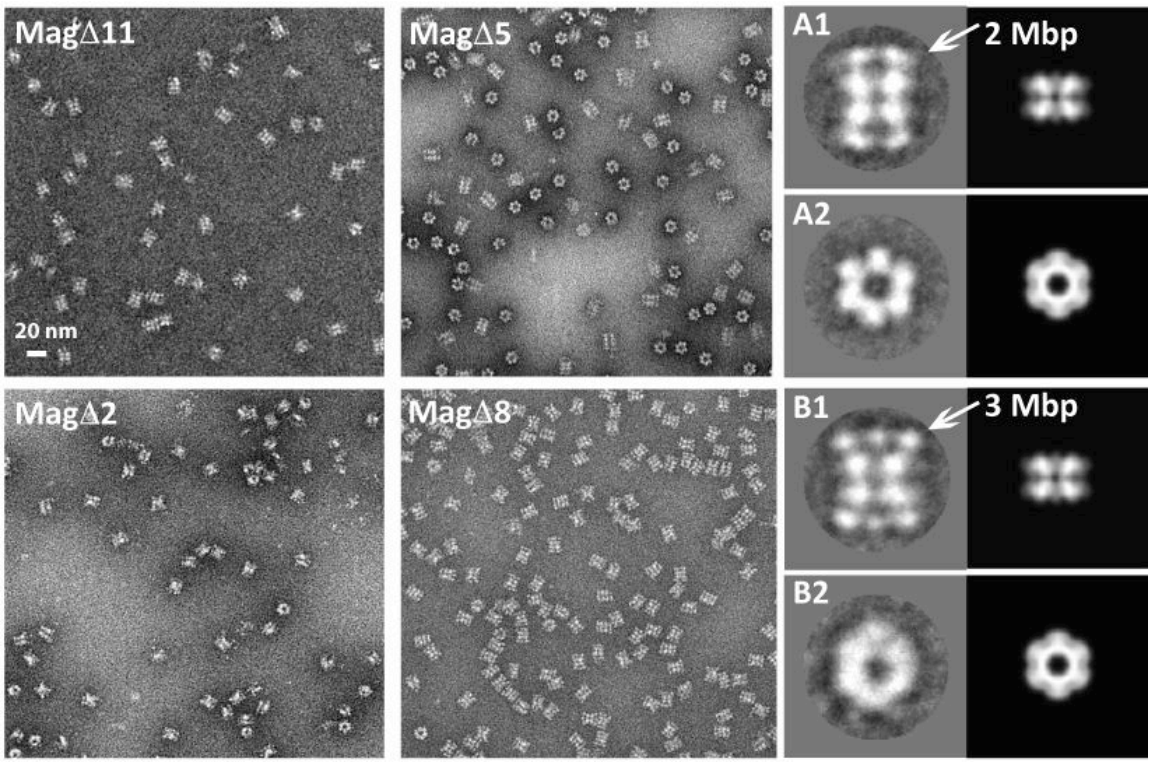


Figure 2.27: Negative stain analysis of the Mag constructs selected by biophysical characterization. Electron micrographs of MagΔ11, MagΔ2, MagΔ5, MagΔ8 negatively stained with SST (Sodium Silico Tungstate) 2% w/v, recorded at 120kV at a 22000x magnification and shown on the same scale. EM grids were prepared as described (§ 4.5.1) By visual appearance MagΔ8 and MagΔ5 look more homogeneous and show characteristic side and top views consistent with a D6 dodecameric structure. Mag 08 and MagΔ5 side views resemble the GS template decorated with either six (3 x 2) or four (2 x 2) extra densities, respectively. Top views of MagΔ5 exhibit a pronounced 6-fold symmetry, whereas those of MagΔ8 appear as a relatively smooth donut. This difference suggests that when viewed along the 6-fold, the Mbp and GS moieties are aligned in MagΔ5, whereas in MagΔ8 they are staggered.

2.4 CRYOEM ANALYSIS OF MAGΔ5 AND MAGΔ8

2.4.1 CRYOEM SAMPLE OPTIMIZATION

A preliminary analysis of vitrified MagΔ8 and MagΔ5 specimens (prepared as described in § 4.5.1) was carried out on a CM200 microscope at 200 kV. The raw images of both constructs show particles with a shape and features consistent with those observed by negative stain EM. In MagΔ8 the ice was quite thin and homogeneous (Figure 2.28C3), the particles are well spread over the hole and mostly side views are observed. In MagΔ5 a concave meniscus was formed, thereby making the ice uneven (Figure 2.28A1). This

problem was overcome by reducing the blot time from 3 to 2 s and increasing the blot force during the freezing procedure from 2 to 2.5 force units on the Vitrobot (Iancu et al., 2006) (Figure 2.28A2). The molecules appeared to form fibers of 100-500 nm in length, thereby hampering single particle analysis (Figure 2.28A3). In fact, fiber formation was also visible by negative stain EM a few hours after purification, and appeared to be due adjacent dodecameric rings stacking one another (Figure 2.28B1). Since the interaction between oligomers within the fiber appeared to occur where the Histidine tags were located (six heptahistidine tags per oligomeric face), we hypothesized that interactions between tags, possibly mediated by traces of Nickel ions from the first purification step, might be the cause of fiber formation. While the addition of 10 mM EDTA did not disrupt the fibers (Figure 2.28B2), and addition of 1 mM $NiCl_2$ only caused further aggregation and affected the dodecameric assembly (Figure 2.28B3), adding 20 mM imidazole was effective at breaking the filaments without compromising sample integrity (Figure 2.28B4 and 2.28C1). This came, however, at a hefty price: the resulting particles almost exclusively oriented themselves on the grid by adopting top views, making the 3D reconstruction problematic. We tried to revert this tendency by adding a mild concentration of detergent (BOG 0.08%), which had solved a similar problem reported previously (Schoehn et al., 2000), but were unsuccessful. Consequently, we removed the His-tag and purified the protein by amylose affinity chromatography, exploiting the presence of Mbp in the construct. CryoEM micrographs of untagged MagΔ5 show that the sample is primarily composed of isolated particles, with only the occasional dimer or trimer of oligomers visible (Figure 2.28C2). Hence this sample is more suited for single particle structural analysis.

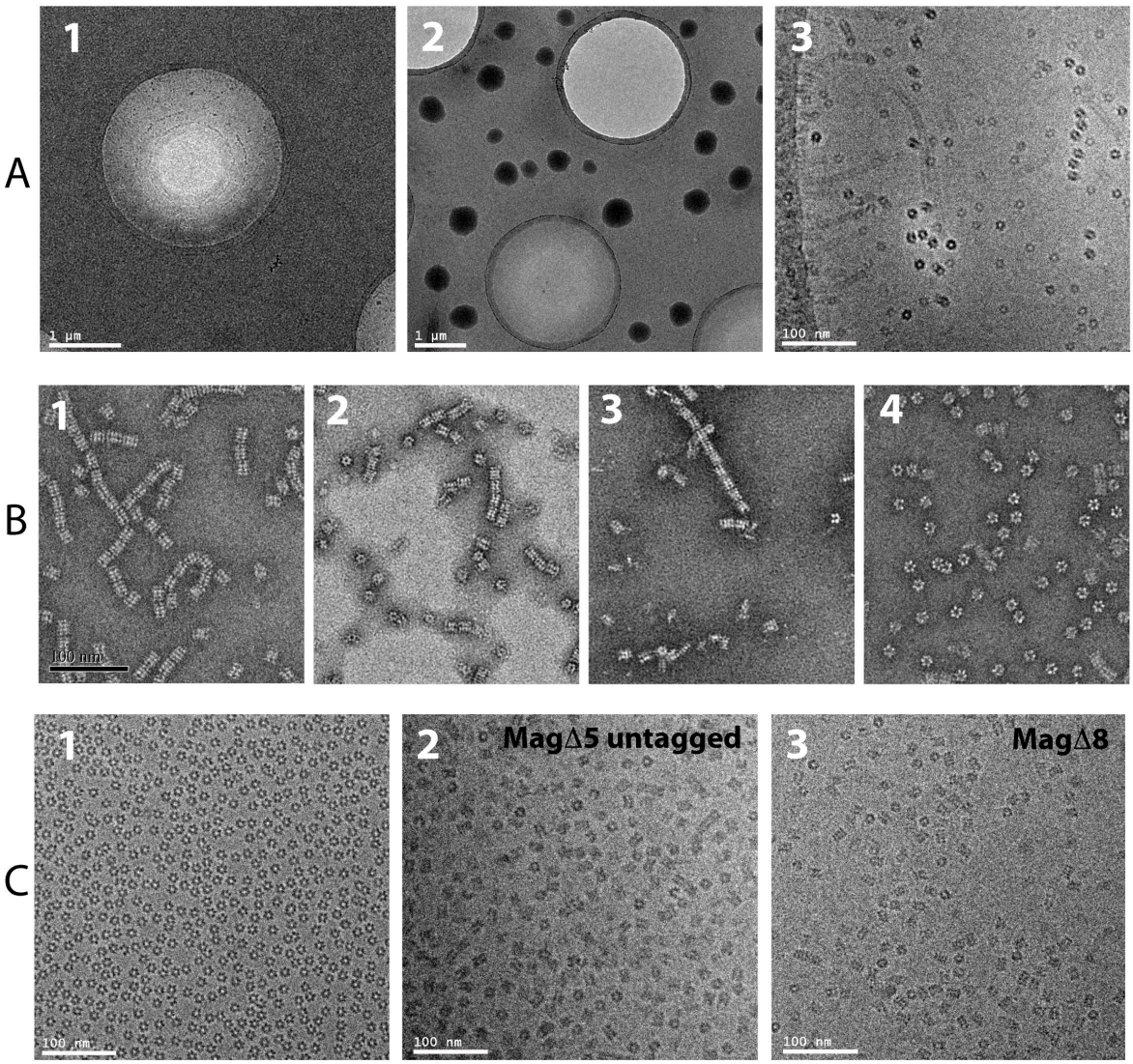


Figure 2.28: Optimization of frozen hydrated samples. A) Vitrified specimens of MagΔ5. A1. Concave meniscus formed in holey carbon support in cryo conditions. A2. Even ice obtained by decreasing blot time. A3. Fibers of MagΔ5 in cryo conditions. **B)** Negative stain screening: B1. MagΔ5 fibers. B2 MagΔ5 + 10mM EDTA. B3. MagΔ5 + 1mM $NiCl_2$. B4. MagΔ5 + 20 mM imidazole pH 8. **C)** Optimized preparation of vitrified specimens. C1. MagΔ5 + 20 mM imidazole pH 8. C2. Untagged MagΔ5. C3 .MagΔ8.

2.4.2 CRYOEM RECONSTRUCTION OF MBP-GS CHIMERAS

The MagΔ5 and MagΔ8 chimeras were frozen according to previously optimized conditions (§ 2.4.1) and imaged at 300 kV on a FEI POLARA microscope recording images on photographic films. Micrographs were scanned to yield digital images with a final pixel size of 1.8 Å, as described in § 4.5.1. Images were corrected for the contrast transfer function (CTF) and datasets of 8797 (MagΔ8) and 16025 (MagΔ5) particles were normalized and utilized for image processing, in order to achieve a 3D map of the dodecameric chimeras. For both MagΔ5 and MagΔ8 an *ab initio* starting model was

produced via the angular reconstitution method (§ 4.5.2). The ensemble of raw individual images were subjected to reference-free alignment, multivariate statistical analysis and classified as described in § 4.5.2. A starting model was then produced from the best quality class averages (~300 particles in total for each constructs) enforcing D6 symmetry (§ 4.5.2). In Figures 2.29 and 2.30 the image processing results are presented for MagΔ5 and MagΔ8, respectively. In the Figures sections **A** the raw cryoEM micrographs are shown, in **B** a few eigenimages, 2D class averages and the corresponding reprojections of the angular reconstitution models are reported. Finally, in Figure 2.29C and 2.30C the 3D reconstructions of the two chimeras are illustrated in three different orientations. For both chimeras the 2D class averages match quite well the reprojections of the model.

The cryoEM volume of **MagΔ5** (Figure 2.29C) shows that Mbp moiety is almost aligned with the GS subunit when viewed along the longest molecular axis, accounting for the 6-fold "flower-like" appearance of the top view and the 2 extra densities decorating the template in the side view. The top view 6-fold symmetry conceivably corresponds to the symmetric variations observed in the first eigenimages. Conversely, in **MagΔ8** (Figure 2.30C) each Mbp moiety is positioned between two GS subunits when viewed along the 6-fold axis, accounting for the smoother donut-like appearance of this view (lacking a clear 6-fold symmetry) and the three extra densities observed in the side view. Indeed, among the eigenimages a clear 6-fold variation is not apparent. Hence, this analysis suggests that in MagΔ5 and MagΔ8 each ring of Mbp subunits adopts an eclipsed and staggered arrangement, respectively, relative to adjacent ring of GS subunits. This result, obtained by cryoEM, is in agreement with previous observations of the samples by negative stain EM (Figure 2.27).

In the case of MagΔ5, a second *ab-initio* model was independently obtained using the RIco webserver (http://rico.ibs.fr/RIcoWebServer/), as described in § 4.5.2. The overall shape of the model and the arrangement of Mbp subunits with respect to GS ones correlate well with the MagΔ5 model obtained by angular reconstitution (Figure 2.31).

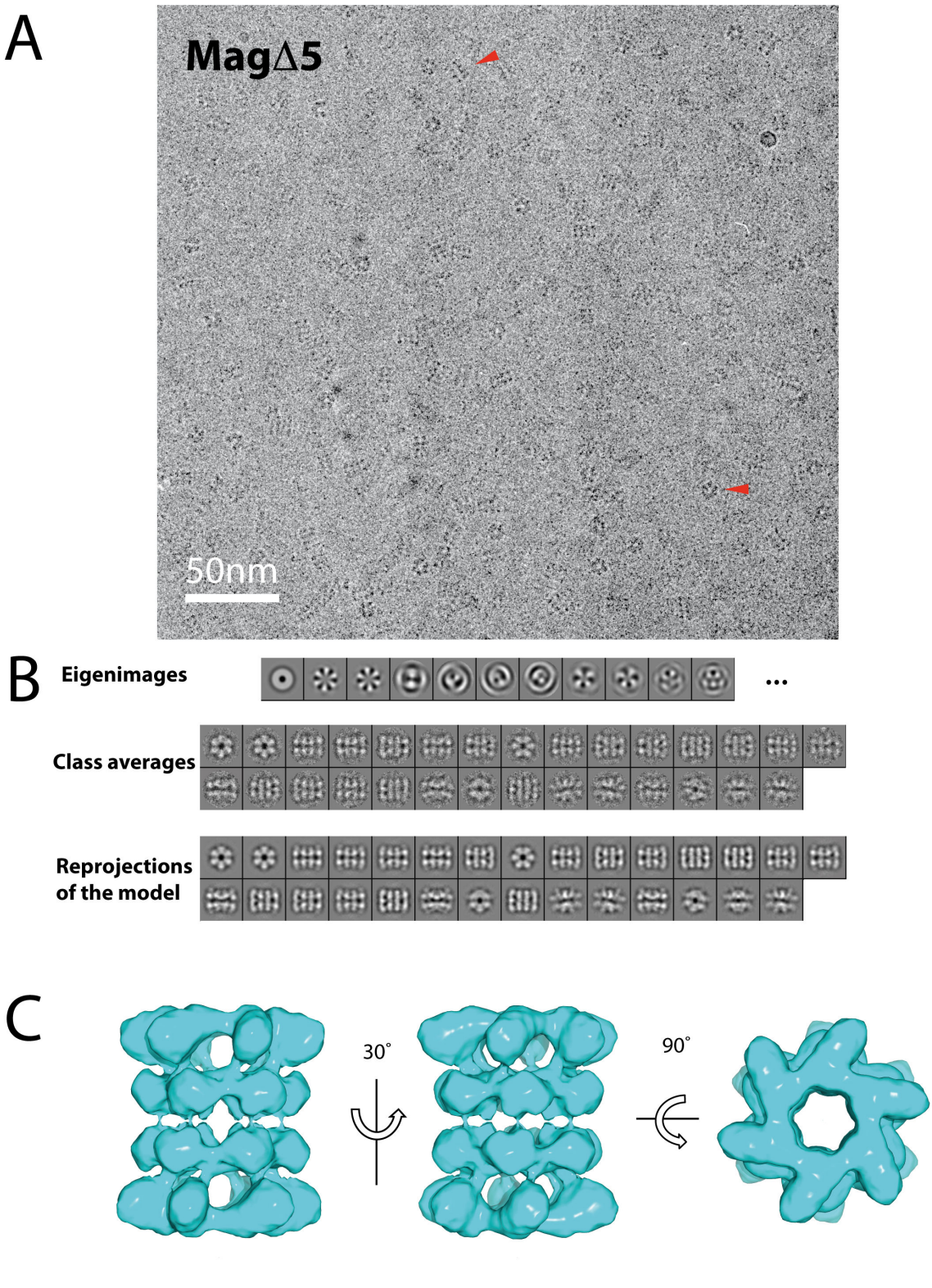


Figure 2.29: MagΔ5 reference cryoEM map obtained by angular reconstitution A) CryoEM raw micrograph of MagΔ5. EM grids were prepared as described (§ 4.5.1) B) Eigenimages, class averages and reprojections of the model obtained by angular reconstitution image processing. C) MagΔ5 final volume obtained by angular reconstitution displayed with 6-fold axis parallel (side views) and perpendicular (top view) to the plane of the page. Mbp and GS are eclipsed resulting in a "flower-like" top view, probably corresponding to the 6-fold main variations in the first two eigenimages.

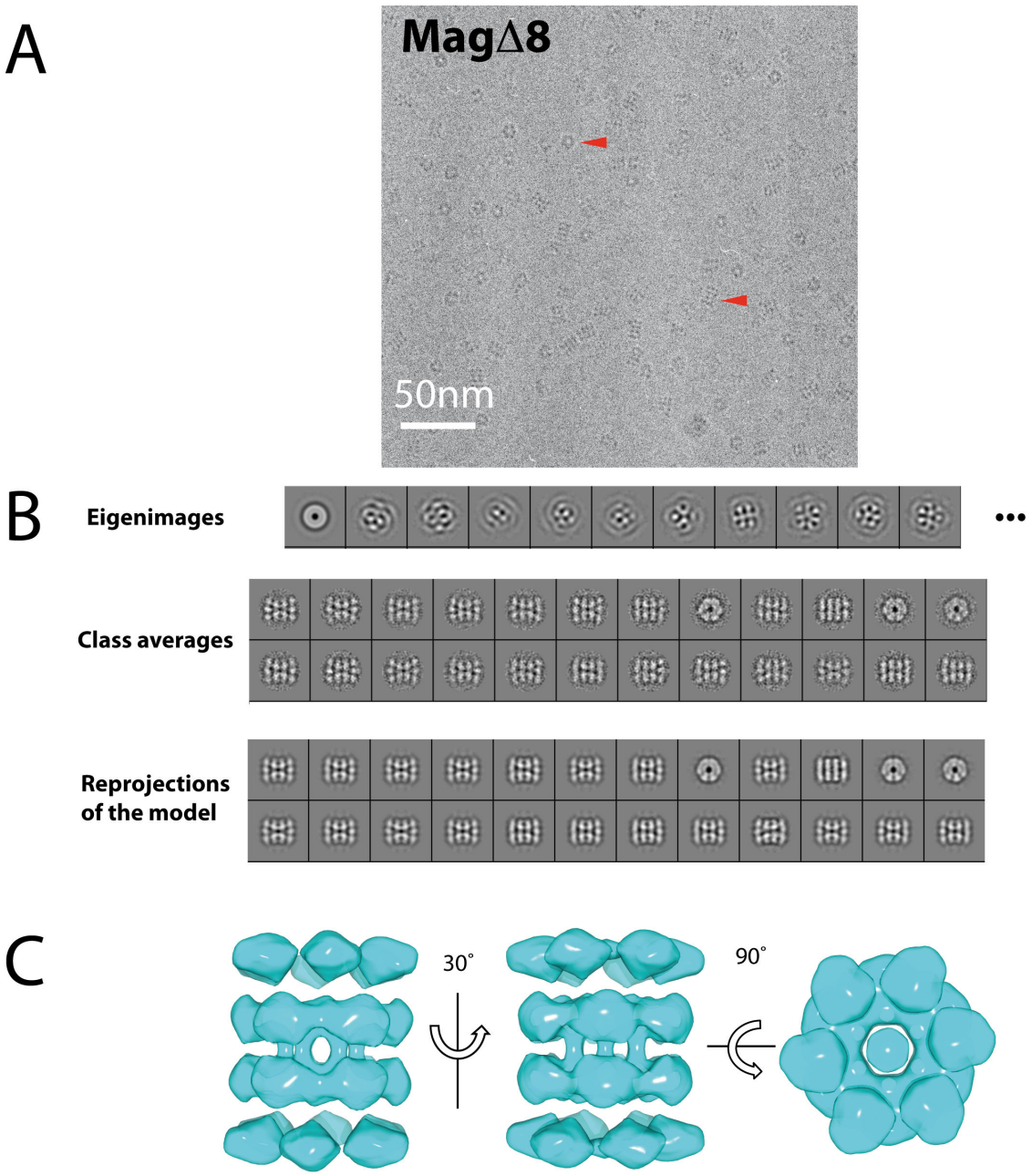
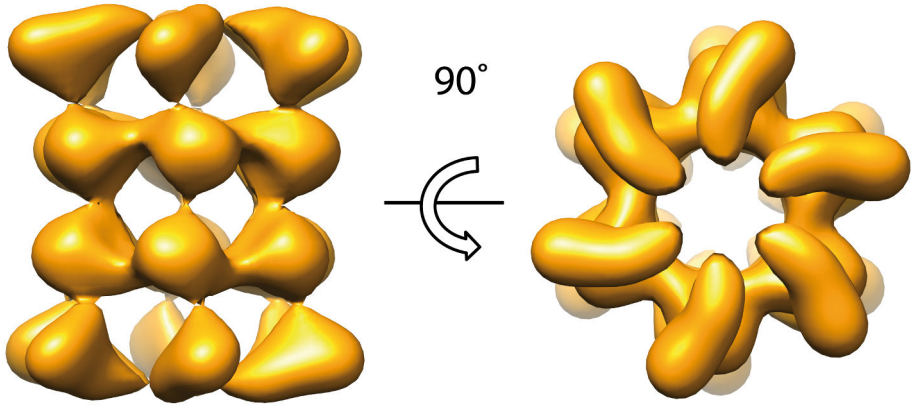


Figure 2.30: MagΔ8 reference cryoEM map obtained by angular reconstitution A) CryoEM raw micrograph of MagΔ8. EM grids were prepared as described (§ 4.5.1) B) Eigenimages, class averages and reprojections of the model obtained by angular reconstitution image processing. C) MagΔ8 final volume obtained by angular reconstitution displayed with 6-fold axis parallel (side views) and perpendicular (top view) to the plane of the page. Mbp and GS are staggered resulting in a smooth donut as top view.



2.31: *Ab initio* model of MagΔ5 obtained using RIco.

2.4.3 Refinement of the cryoEM volumes

Refinement of both the MagΔ5 and MagΔ8 volumes was performed through a projection matching procedure by enforcing D6 symmetry, using the reference structures obtained by angular reconstitution and rejecting 20% of the particles by correlation coefficient (§ 4.5.2). As evident from the angular coverage diagrams in Figure 2.32, ~40% of projection directions correspond to top views for both datasets. This tendency has been reported for other dihedral molecules (Schoehn et al., 2000) and is often due to a preferential interaction of one ring of subunits with the water-air interface. The remaining projection directions correspond primarily to side views. In MagΔ8 the different possible side views are not equally represented; instead, one predominantly observes a view in which the lateral 2-fold axis faces the water-air interface. In MagΔ5 this is also one of the most populated views, but a slightly broader angular coverage is achieved (Figure 2.32).

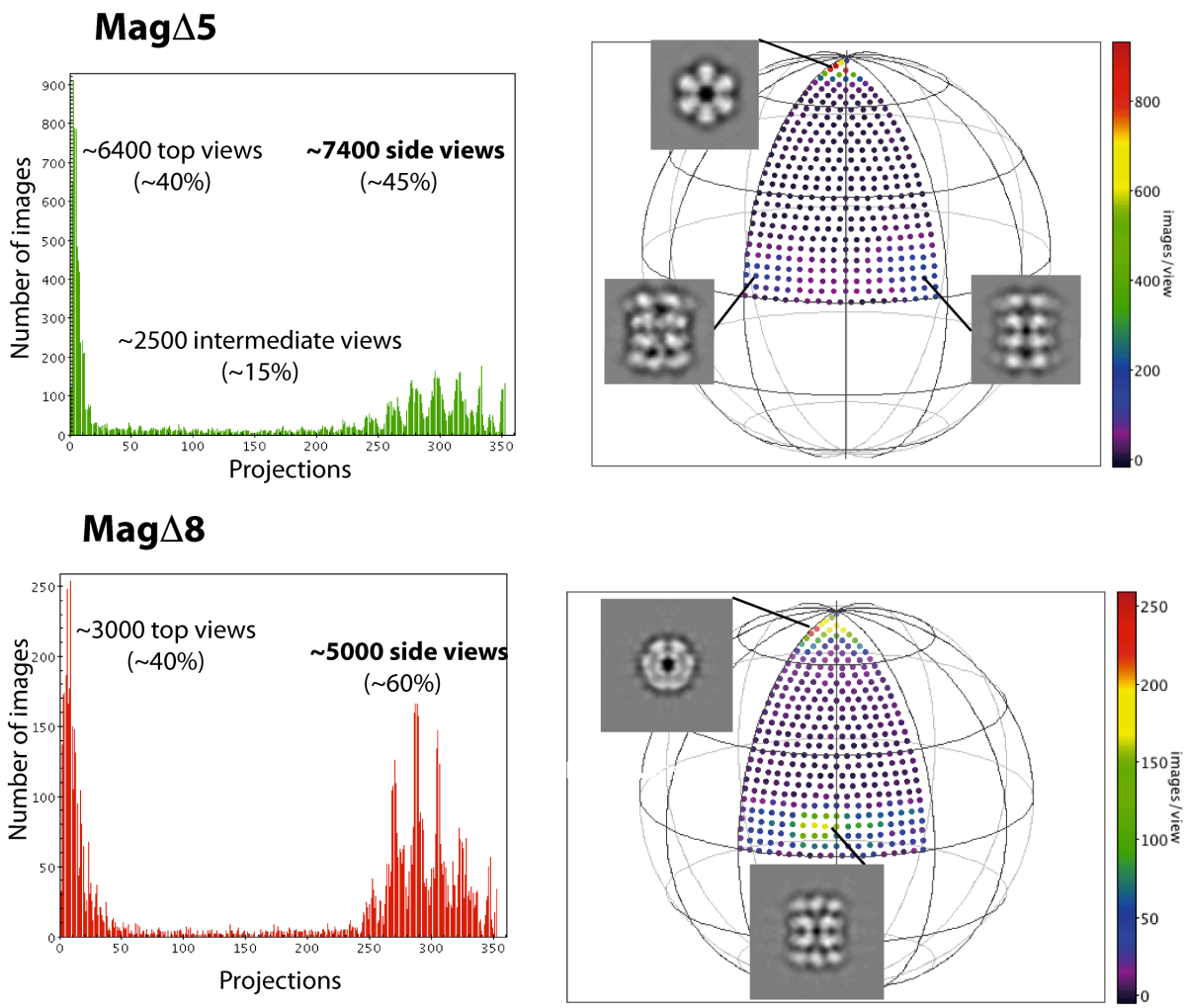


Figure 2.32: Angular coverage in MagΔ5 and MagΔ8 datasets. Number of images aligned by maximizing the correlation coefficient (CC) with 352 projections of the reference model, visualized on a histogram (left) and in spherical coordinates (right). The most populated views are the top and side views shown in the insets.

In order to enrich the poorly populated orientations, we attempted to collect data by tilting the CompuStage by 10˚ and 20˚. However, for the collected micrographs it was impossible to accurately determine and correct the CTF, due to a strong drift phenomenon in the experimental setup (Figure 2.34).

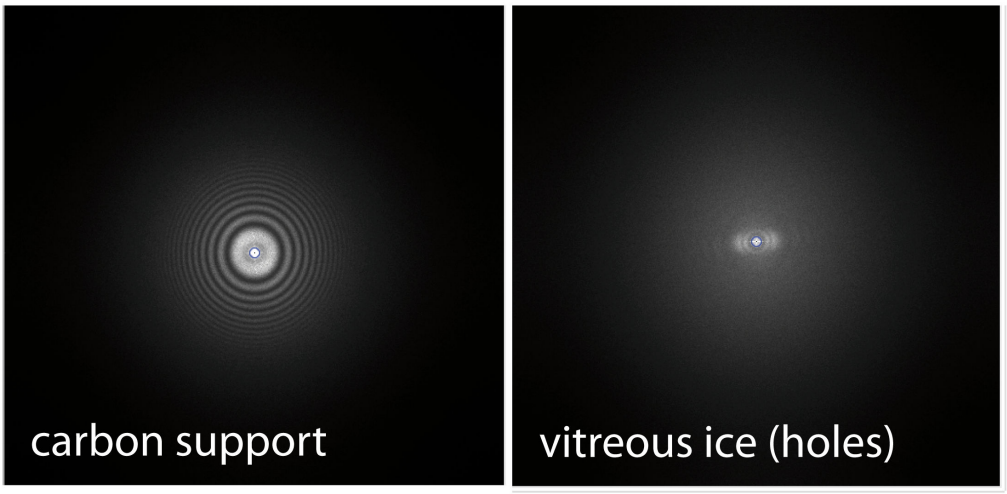


Figure 2.33: Power spectra calculated for MagΔ5 cryoEM micrograph. *Left,* power spectrum of a micrograph in which carbon support is imaged. Thon rings are clearly visible to a resolution > 10 Å. *Right*, power spectrum of a micrograph corresponding to vitreous ice (holes) in the same grid. Here, a clear phenomenon of drift is observed, which hampers CTF determination and correction.

Hence, only the zero tilt dataset could be utilized to produce final maps of both dodecameric chimeras MagΔ8 and MagΔ5, represented in Figures 2.34 and 2.35, respectively. By using the *fitmap* routine in Chimera (Pettersen et al., 2004), we fitted the crystal structure of Mbp and that of the GS dodecamer into the EM maps to evaluate the quality of the final reconstruction. Whereas the GS dodecamer showed a good fit to the EM maps of both chimeras, important differences were observed at the level of Mbp density.

In the **MagΔ8** map (Figure 2.34) the density for Mbp is detached from the template and appears as an elongated volume much smaller than the Mbp crystal structure (Figure 2.34E). This made it impossible to establish a unique position and orientation of the target and suggests a possible movement of Mbp subunits relative to GS. The presence of a round extra density in the middle of the target ring (on the 6-fold axis, Figure 2.34A) seems to be an artefact of the D6 enforcement and is likely due to flexible Mbp subunits partly occupying positions close to, or overlapping with, the 6-fold. As estimated from the FSC 0.5 criterion the overall resolution of the map is about 15 Å.

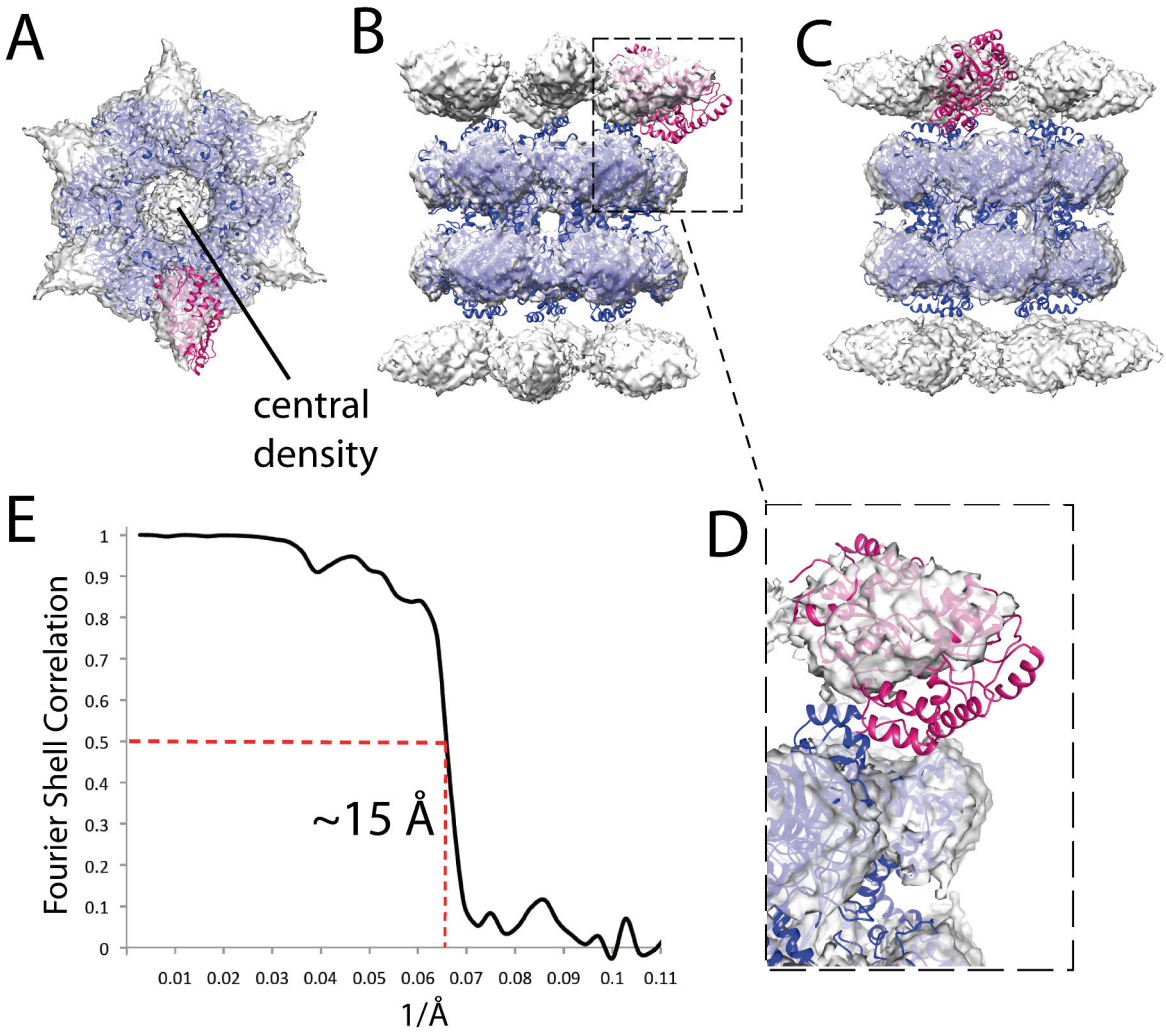


Figure 2.34 CryoEM reconstruction of the MagΔ8 chimera. The density map is depicted in gray at a contour level of ~ 0.1 sigma. The fitted crystal structures of the GS template and Mbp target are in blue and magenta, respectively. A) Top view revealing the presence of an uninterpreted round density at the level of the Mbp ring. B,C) Two different side views of the same map. D) Close-up of one Mbp-GS subunit. The Mbp density does not cover the whole X-ray structure, suggesting partial flexibility of Mbp relative to GS. E) FSC curve. The dotted red line indicates that the estimated resolution at FSC= 0.5 is ~15Å.

In **MagΔ5** (Figure 2.35) no central densities are present on the 6-fold axis. Continuous density is observed between the outer and inner rings, and the outer density presents a volume and shape that nicely accommodates six copies of the Mbp crystal structure. It is therefore possible to confidently position the target with respect to the template. Indeed, the two lobes enclosing the active site of Mbp are clearly visible in the map (Figure 2.35D). The Mbp subunits appear quite distant from one other and do not seem to share a large interface (Figure 2.35A). The FSC presents a regular fall off and the resolution estimated

by the 0.5 criterion (10 Å) is coherent with the overall features of the map (Figure 2.35E). If higher frequency terms are boosted by applying a B-factor correction [B= -300 Å^2, applied with *embfactor* (Fernandez et al., 2008)], some α-helical elements appear better resolved in the junction region between Mbp and GS and on the template (Figure 2.35F). As a control to check for possible model bias, angular refinement based reconstructions for MagΔ5 were also performed using the RIco model as a reference model (Figure 2.31). The reconstructions starting from two independent (IMAGIC and RIco) reference structures converged to extremely similar volumes.

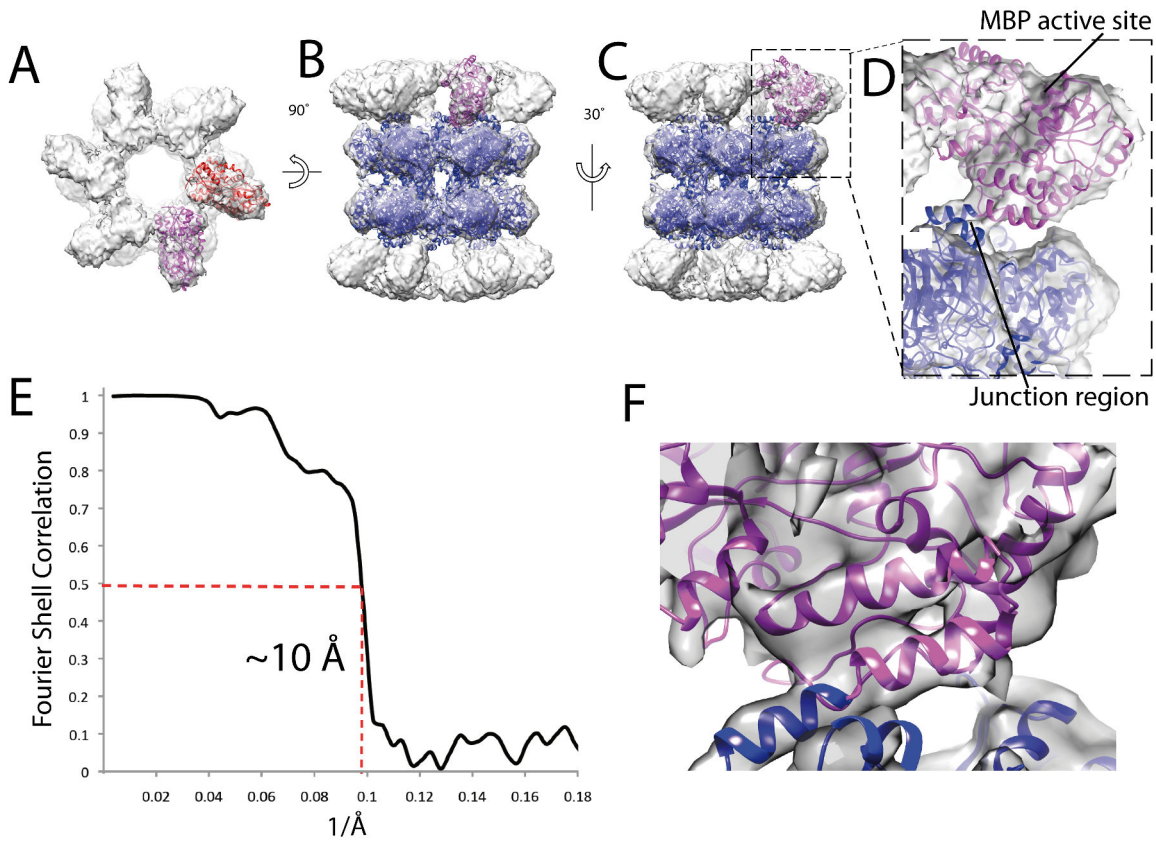


Figure 2.35: CryoEM reconstruction of the MagΔ5 chimera. The density map is depicted in gray at a contour level of 0.1 sigma. The fitted crystal structures of the GS template and Mbp target are in blue and magenta, respectively. A) Top view of the cryoEM map in which two Mbp monomers have been fitted. B,C) Two different side views of the same map. D) Close-up of one Mbp-GS monomer, showing good coverage of the fitted structures. E) Close up of the linker region and C-terminal region of Mbp following B factor sharpening. Tube-like features in the sharpened map agree well with the expected location of α helices. F) FSC curve, the dotted red line indicates that the resolution estimated at FSC= 0.5 is ~10 Å.

2.3.4 LOCAL RESOLUTION ESTIMATION OF CRYOEM MAPS

The FSC 0.5 criterion (van Heel and Schatz, 2005) gave an overall estimate of the 3D map resolution of ~10 Å and ~15 Å for MagΔ5 and MagΔ8, respectevely. However, the local resolution can vary significantly across different regions of the map (Cardone et. al, 2013). Indeed, the MagΔ5 and MagΔ8 maps exhibit similar density for the GS template but strikingly different densities for the Mbp target. We calculated the local resolution for the MagΔ5 and MagΔ8 both maps utilizing the *blocres* routine (Cardone et. al, 2013) in the Bsoft program suite (Heymann, 2001) and coloured both maps according to the resulting local FSC values (Figure 2.36). The range of local resolutions observed match the overall FSC estimation, varying between 9 and 12 Å for MagΔ5 and between 9 and 16 Å for MagΔ8. In both MagΔ8 and MagΔ5, the GS template exhibits a resolution ranging from 9 to 11 Å, with the highest resolution observed at the innermost diameter (Figure 2.36B). However, in MagΔ5 the GS template is better resolved in the junction and contact regions with the Mbp target (where the subunits eclipse each other). Regarding Mbp, in MagΔ8 the corresponding density exhibits very poor resolution (11-16 Å), consistent with the volume of this density being smaller than expected. The peripheral regions and the central density inside the target ring appear less well resolved, presumably because these regions show the greatest variation in density as Mbp subunits fluctuate about some average position. In MagΔ5 the Mbp density presents a local resolution that ranges from 9 Å near the target-template linker region, where one can indeed distinguish a few helices visually, to 12 Å at the top of the dodecamer, where the protein is conceivably more flexible. Taken together, the above findings demonstrate that it is feasible to obtain an accurate structure of Mbp at a resolution of 10-12 Å by cryoEM, firmly establishing a proof-of-concept of the protein symmetrization approach. Moreover, the higher quality of the MagΔ5 map with respect to MagΔ8 mirrors the observation that MagΔ5 was the better behaved construct in the various biophysical assays (Figure 2.24E), highlighting the usefulness of performing such studies prior to embarking on cryoEM analysis.

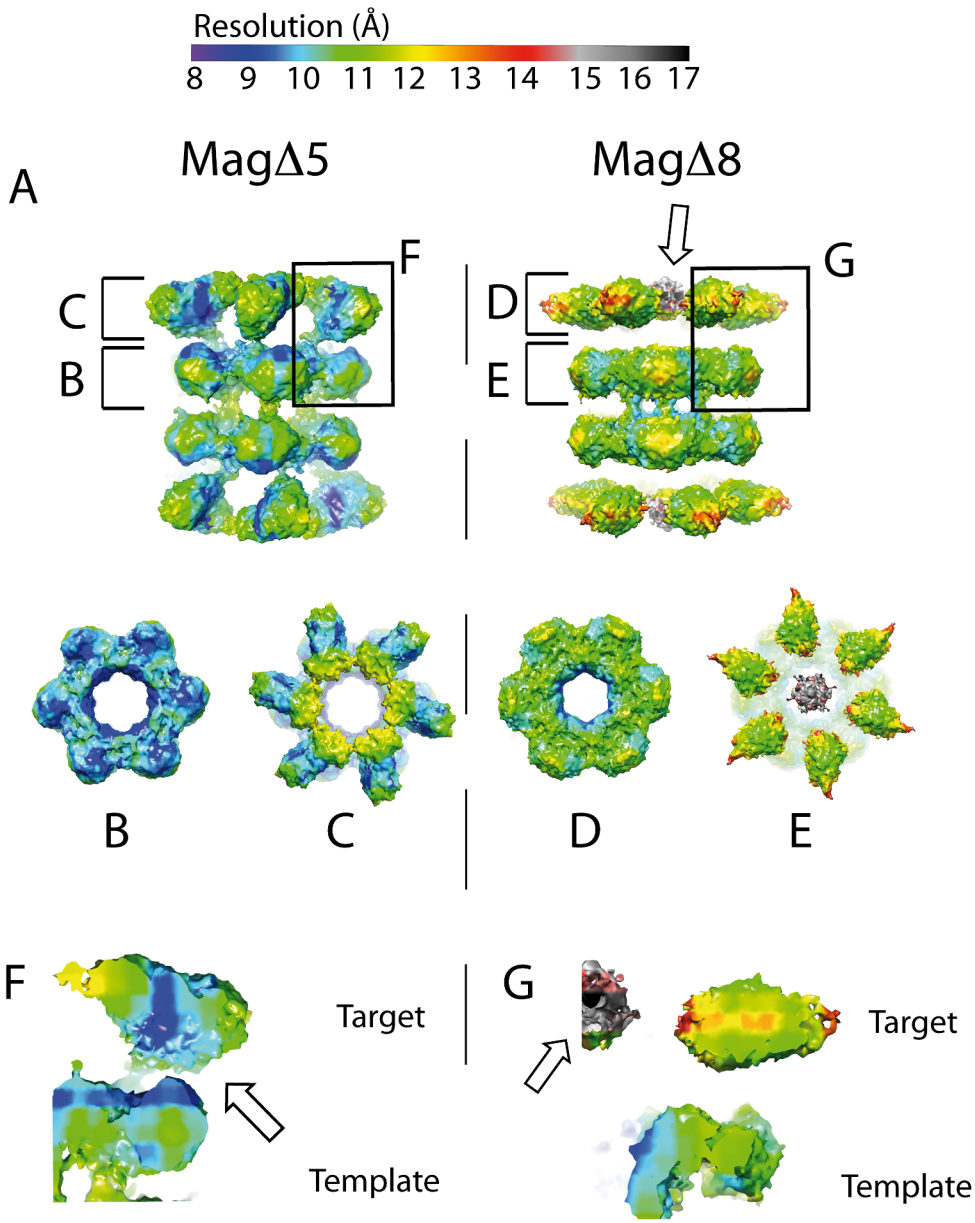


Figure 2.36: Illustration of local resolution in MagΔ5 and MagΔ8 cryoEM maps. A) Surface rendering of cryo-EM MagΔ5 and MagΔ8 reconstructions, coloured according to local resolution (palette above). For MagΔ5, the resolution varies from 9-12 Å and for MagΔ8 between 9-16 Å. The lower resolution region in MagΔ8 is the central density on the 6-fold axis, indicated by the arrow. This is probably generated by D6 enforcement and due to flexibility of the target moieties. B,D) *Template local resolution*. Maps viewed along the 6-fold axis from the plane separating target and template. In both Mag constructs the resolution on the template is ~11 Å (green), with the inner region best resolved (9 Å, blue). At the Mbp-GS interface the local resolution on the template is higher for MagΔ5 (blue regions) and lower for MagΔ8 (green-yellow). C,E) *Target local resolution*. Maps viewed along the 6-fold axis from the plane above the target moieties. MagΔ5 is less resolved in the central region (12 Å), far from the GS juction. In MagΔ8 the local resolution is poorest in peripheral parts (14 Å) and on the central density (16 Å). F,G) Vertical central sections showing the target-template junction region. In MagΔ5 the target density is of the expected size and is adjacent to the template, unlike in MagΔ8, where these densities are disconnected.

The ~10 Å resolution reached for MagΔ5 map does not allow *de novo* tracing of the protein backbone, either within the target or the template. One reason for the limited resolution might be the narrow angular distribution of particle orientations in the cryo dataset, i.e. the prevalence of a top view and one particular side view. In order to investigate this possibility we simulated an electron density map at 10 Å resolution from the atomic coordinates of our *in silico* MagΔ5 model. This calculation was performed using the *proc3d* routine in the EMAN program suite (Ludtke, 2010). Subsequently, in SPIDER (Frank et al., 1996), we reprojected the simulated map along evenly distributed directions in the Euler space, using the same angular increment as used for determination of the MagΔ5 experimental map. Then, we reconstructed two maps by backprojection in Fourier space (see projection matching procedure, § 4.5.2):

i) A simulated map with ***ideal angular coverage,*** using equally populated and the evenly spaced projections

ii) A simulated map with ***real angular coverage***, using an uneven distribution of projections similar to that obtained experimentally.

Subsequently, we visually inspected the simulated and experimental MagΔ5 maps in the helical regions of the template (Figure 2.37). In the "ideal" simulated map α-helices are only slightly better resolved than in the "real" simulated map, indicating a minor effect of view distribution on map quality (Figure 2.37 A,B). In contrast, helices are much more poorly resolved in the experimental map than in the "real" predicted map (Figure 2.37 C). Therefore, incomplete angular coverage at most accounts only partly for the loss of resolution in the MagΔ5 map. A more important factor is likely to be the low signal-to-noise ratio (SNR), as the MagΔ5 reconstruction was obtained from relatively few particles (~7500 side views).

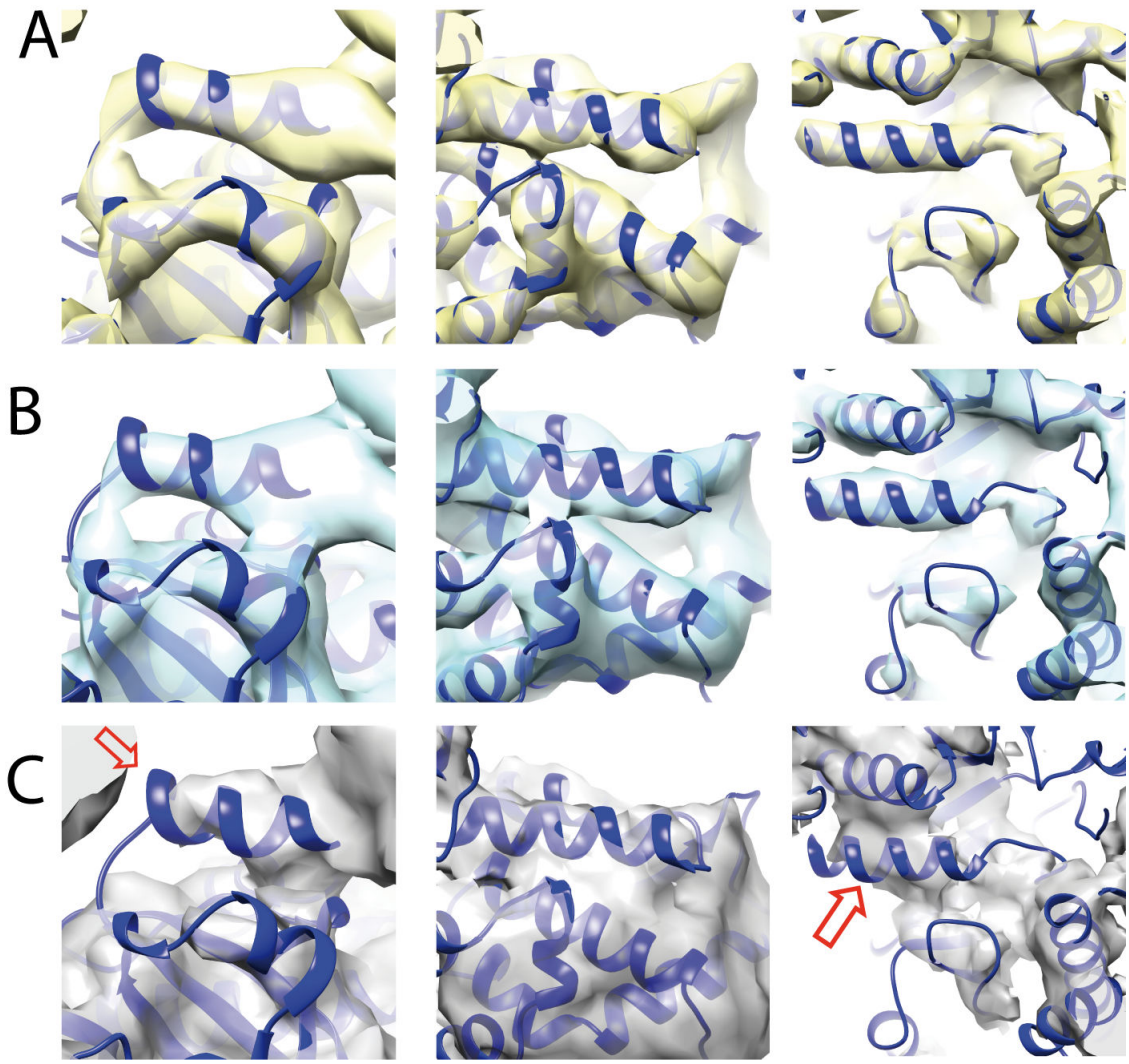


Figure 2.37: Effect of angular coverage on cryoEM MagΔ5 map resolution at 10 Å resolution. Fit of GS atomic model (PDB ID code 1F52A; blue ribbon diagram) in three MagΔ5 maps, visualized in selected helical regions of the template: A) Simulated MagΔ5 map calculated at 10 Å from the *in silico* model using ideal homogeneous angular coverage (yellow) B) Simulated MagΔ5 map calculated at 10 Å from the *in silico* model using experimental uneven angular distribution, i.e. mainly one top and one side view (cyan) C) Experimental MagΔ5 cryoEM (grey), whose estimated resolution is 10 Å. The definition of α-helices in map A is only slightly better than in map B, suggesting that angular distribution has only a small effect on the map resolution. In contrast, the experimental map is overall much less defined than map B, derived from the same angular distribution, and lacks density in several helical regions, indicated by the red arrows. This suggests that the limited resolution of experimental map C is predominantly due to factors other than an inhomogeneous view distribution.

Another factor limiting the resolution of the MagΔ5 map could be inherent flexibility of the chimera. Indeed, the presence of multiple conformations (i.e. loss of symmetry) within the fusion, could result in a less defined map when enforcing D6 symmetry in the cryoEM reconstruction. In order to study this possibility we pursued a crystallographic study of MagΔ5.

2.5 CRYSTALLOGRAPHIC STUDIES OF MAGΔ5

To verify the structure of MagΔ5 obtained by cryoEM, we attempted to determine its structure by X-ray crystallography. We used the purified protein (at concentrations between 5 and 10 mg/mL) to screen for crystallization conditions at 20 ˚C using the high-throughput crystallization facility at the EMBL-Grenoble. The screen, performed on a small scale (200 nL initial drop volume) by the method of sitting drop vapour diffusion, yielded crystals which were shaped as hexagonal prisms measuring ~200 µm in length (Figure 2.38 A; see figure legend for crystallization conditions). We manually reproduced this condition by the method of hanging drop vapour diffusion (2 µL initial drop volume). Crystals were harvested in loops, transferred to a cryoprotectant solution and flash-cooled in liquid nitrogen. Depending on the cryoprotectant used, crystals diffracted at a resolution between 7 and 15 Å, with the best diffraction observed using 8% (v/v) 1,4-butanediol as the cryoprotectant. The fact that crystals did not diffract to higher resolution suggests that there is some degree of disorder or flexibility within the MagΔ5 construct, possibly explaining the limited resolution of the cryoEM map. A complete diffraction data set was collected at ~7 Å resolution at ESRF beamline ID23-1. Data were processed using XDS (Kabsch, 2010) and programs of the CCP4 suite (Winn et al., 2011). Data collection statistics are summarized in Table 2.38.

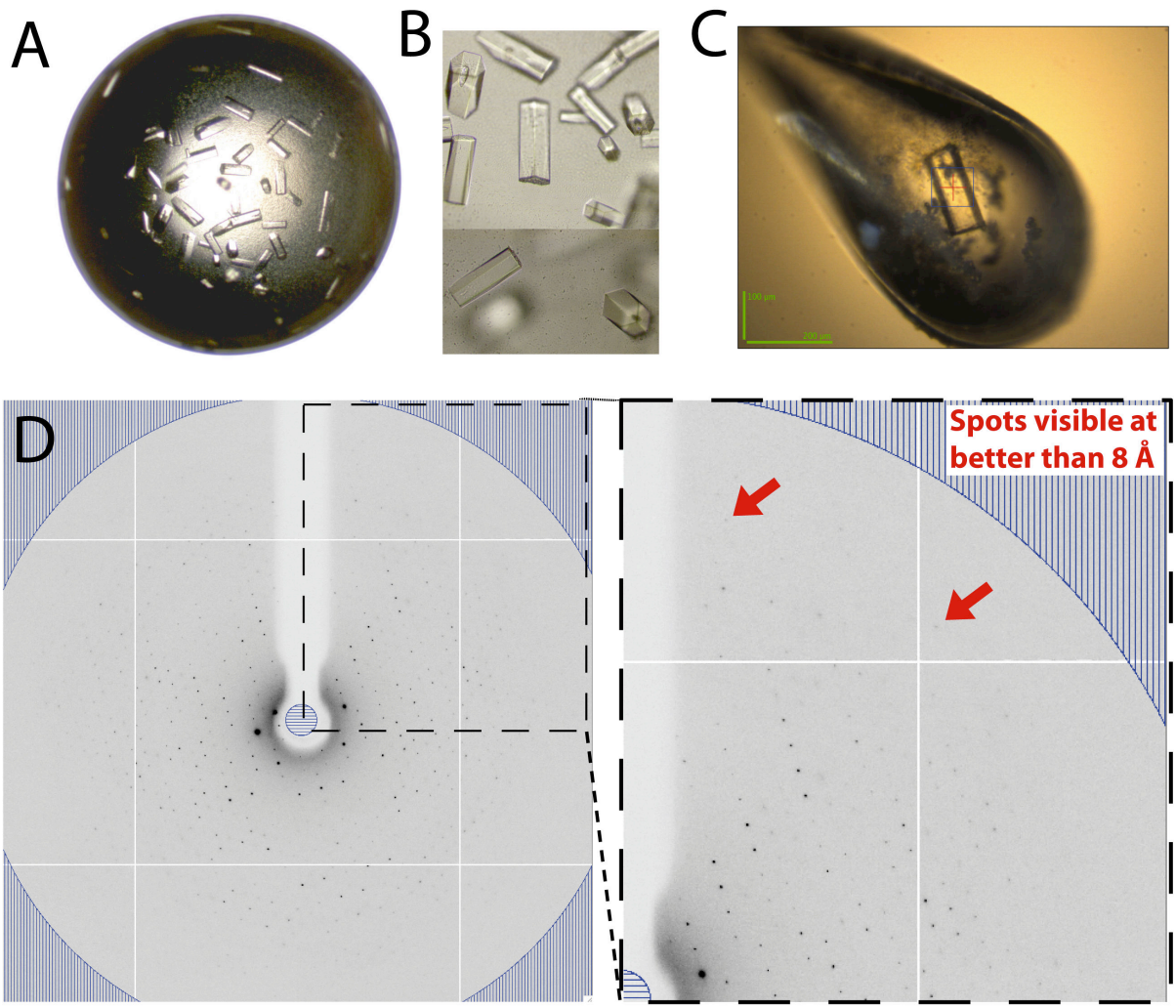


Figure 2.38: X-ray diffraction data of MagΔ5 crystals. A-B) Crystals of MagΔ5, appearing as hexagonal prisms. Crystals were grown by hanging drop vapour diffusion by mixing 1 μL of protein (10 g/L) with 1 μL of a solution containing 10% PEG 3350, 0.1 M HEPES pH 7.5 and 0.2 M L-proline. C) Image of the frozen crystal inside the loop prior data collection at ESRF beamlines ID23-1 D) An example of diffraction frame where spots are visible till ~ 7 Å maximum resolution.

Table 2.38. Crystallographic data collection and refinement statistics.

Data Collection	
Space group	$P3_221$
Cell dimensions (Å)	*a*= *b*=181.0, *c*=430.9, α=β=90°, γ=120°
Solvent content (%)	66.5
ESRF Beamline	ID23-1
Wavelength (Å)	0.97487
Resolution (Å)	49-7.06 (7.25-7.06)[a]
R_{merge} (%)[b]	10.9 (122.3)
$CC_{1/2}$	0.999 (0.608)
$\langle I/\sigma(I)\rangle$	14.5 (1.5)
No. observed reflections	141006 (11376)
No. unique reflections	13182 (1030)
Completeness (%)	99.8 (100.0)
Multiplicity	10.7 (11.0)
Refinement	
Resolution (Å)	49 – 7.06
No. reflections (test)	10669 (661)
R_{work}/R_{free} (%)[c]	29.6 / 31.8
Rmsd bond lengths (Å)	0.010
Rmsd bond angles (°)	2.34
Molprobity Score	
Overall	3.1
Clashscore	19.6
Ramachandran Analysis	
Favored (%)	91.8
Allowed (%)	7.4
Outliers (%)	0.8

[a] Values in parentheses are for the highest resolution shell.

[b] $R_{merge} = \Sigma_{hkl}\Sigma_i |I_i - \langle I\rangle| / |\langle I\rangle|$, where I_i is the intensity for the *i*-th measurement of an equivalent reflection with indices h, k and l.

[c] $R = \Sigma|F_o-F_c| / \Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.

The crystal structure of MagΔ5 was determined by Dr. Carlo Petosa. Crystals belong to the trigonal space group $P3_221$, with six MagΔ5 monomers per asymmetric unit. The six monomers comprise three monomers (A, B, C) from one hexameric ring and three (G, H, I) from the other (Figure 2.39). A crystallographic dyad generates the remaining six monomers to reconstitute the complete dodecamer. Molecular replacement using program Phaser (McCoy, 2007) was initially used to locate GS subunits A, B and C within the asymmetric unit. The resulting $2F_o$-F_c map revealed clear density for the three remaining GS subunits, confirming that the space group assignment and molecular replacement solution were correct (Figure 2.39B).

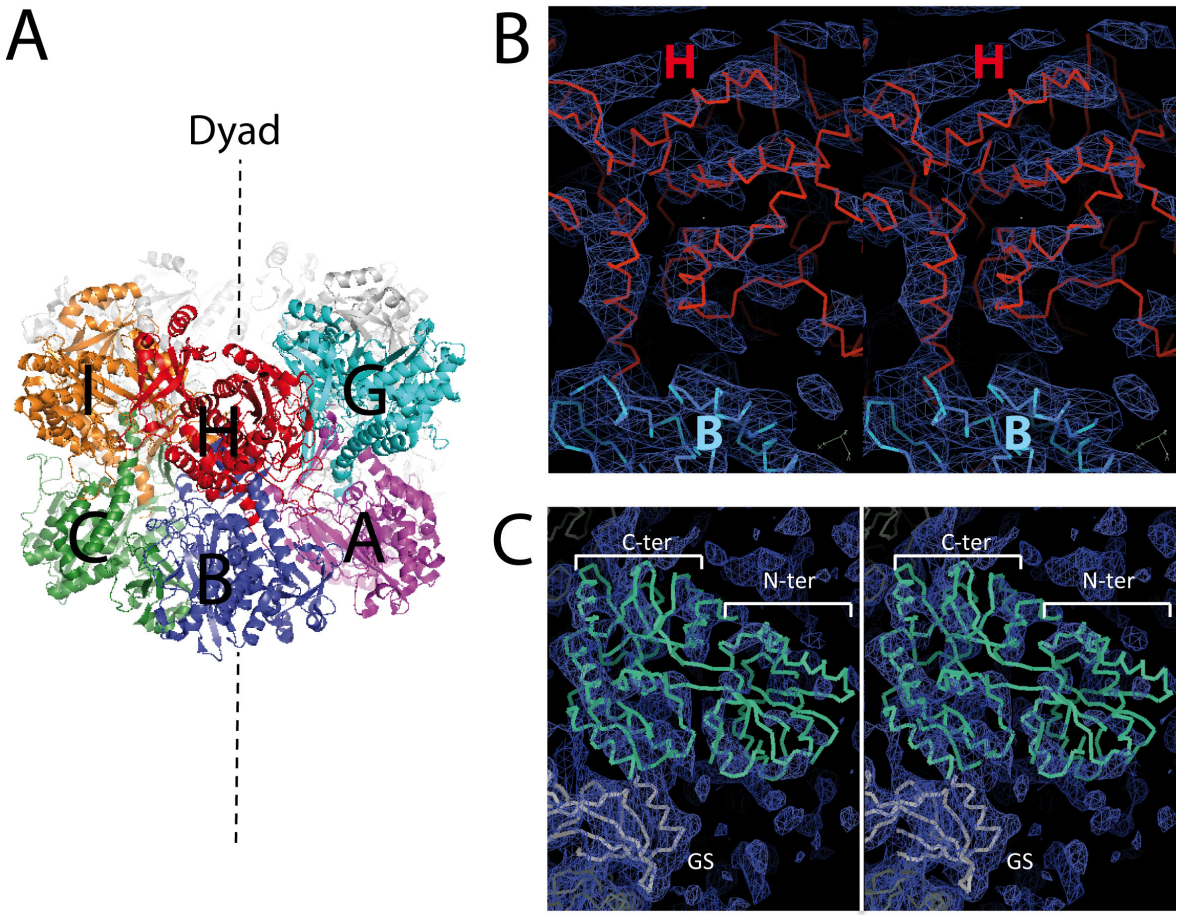


Figure 2.39: Crystal structure of MagΔ5. A) The asymmetric unit contains half a dodecamer. B) Cross-eyed stereo view of a 2Fo-Fc map calculated at 7.4 A resolution using phases from the GS subunits A-C. The map shows clear electron density for GS subunit H (red ribbon), confirming the correctness of the molecular replacement solution. C) 2Fo-Fc map phased on all six GS subunits showing density for Mbp subunit C. The density for the C-terminal domain is better defined than that for the N-terminal domain, which is more distal to the GS subunits.

The six GS subunits were then placed into density and refined as rigid bodies using program PHENIX (Adams et al., 2011). The resulting electron density map revealed density for all six Mbp subunits. In all cases, the density was better defined for the Mbp C-terminal domain (which is proximal to GS) than for the N-terminal domain (Figure 2.39C). Moreover, the density was best defined for Mbp subunits H and I, somewhat noisier for subunits A-C, and poorest for subunit G (Figure 2.40).

The six Mbp subunits were manually fitted into density and subsequently refined as individual rigid bodies, followed by a round of TLS refinement in which the GS subunits and the N- and C-terminal domains of each Mbp subunit within the asymmetric unit were defined as (a total of 18) separate TLS groups. The final structure refined at 7.1 Å yielded

a crystallographic R-factor of 0.296 (Rfree=0.318), which is a typical value for this resolution.

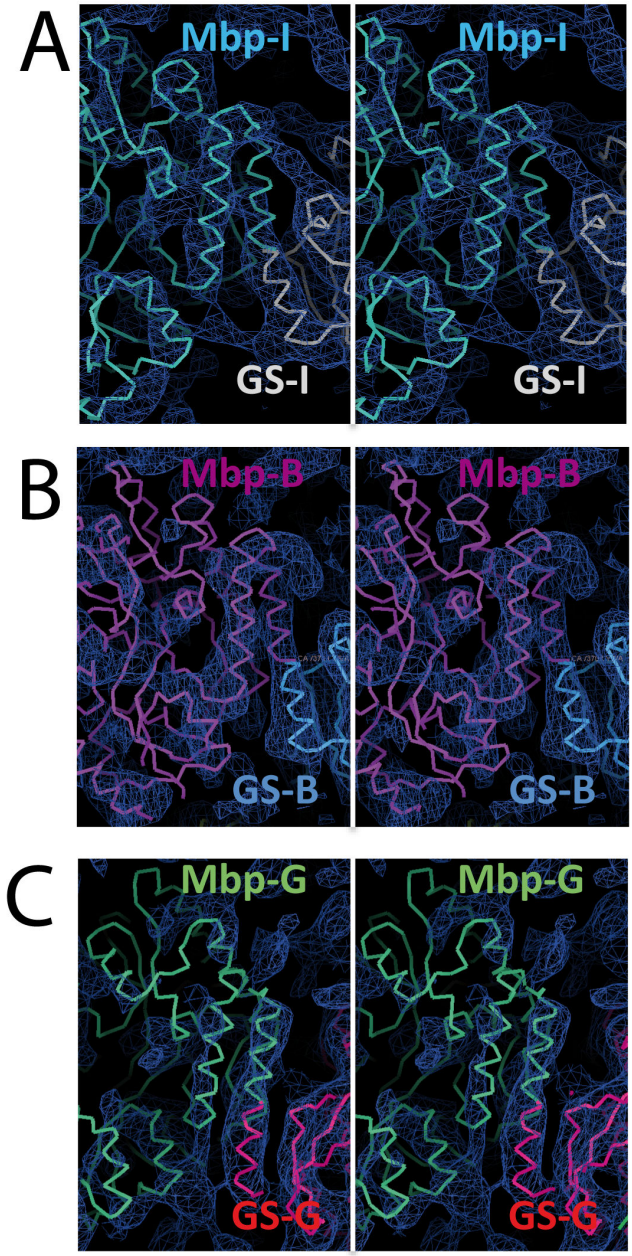


Figure 2.40: MagΔ5 crystallographic map. Cross-eyed stereo view of 2Fo-Fc map calculated at 7 A phased on all six GS subunits showing density for the Mbp C-terminal domain of A) subunit I, B) subunit Band C) subunit G. The density is best defined for subunit I and worst for subunit G.

As expected, the structure of the template within the MagΔ5 dodecamer is identical to the crystal structure of the isolated GS dodecamer present in the Protein Databank, confirming that the tertiary and quaternary structures of GS are unperturbed by the fusion

to Mbp. Interestingly, the different Mbp subunits adopt different orientations relative to the GS subunit to which each is fused (Figure 2.40H). Whereas Mbp subunits A, B and C all share the same relative oriention, chains H and I are tilted closer towards the GS subunit by ~13°, while chain G is tilted by ~9° in the opposite direction. Thus, while the hexameric ring formed by Mbp subunits A-C and the crystallographically related subunits A'-C' [the $(ABC)_2$ ring] is highly symmetrical, that formed by Mbp subunits G-I and G'-I' deviates considerably from 6-fold symmetry. These deviations appear to be due to differences in the crystal packing environment of the two Mbp rings.

These crystal packing interactions can best be understood as arising from three levels of structure. First, MagΔ5 dodecamers stack head-tail to form fibres (Figure 2.41A), highly reminiscent of those observed by negative stain and cryoEM analysis (Figure 2.28). Second, the fibres line up in parallel (with a translational offset of ½ dodecamer) to form sheets in the x-y plane (Figure 2.41A). Finally, the x-y sheets are layered on one another with a 120° rotation along the z direction (the crystallographic 3_2 screw axis) to form the 3D crystal (Figure 2.41B,C,D).

Monomers A, B, C and G mediate interactions between neighbouring dodecamers within the fiber, but do not participate in inter-fiber or inter-sheet contacts (Figure 2.40D,E). In contrast, in addition to mediating intra-fiber contacts, monomer I mediates inter-fibre interactions within an x-y sheet, while monomer H mediates inter-sheet interactions. (Figure 2.41). The Mbp H and I subunits must tilt away from the 6-fold axis to mediate these interactions, explaining the observed loss of symmetry for the $(GHI)_2$ ring of Mbp subunits. Moreover, whereas the Mbp subunits within the $(ABC)_2$ ring interact with one another, the outward tilting of Mbp subunits H and I leaves subunit G unbuttressed by lateral contacts. This looser packing probably explains why subunit G exhibits such poorly defined electron density. Taken together, these observations suggest that the $(GHI)_2$ ring of Mbp subunits is distorted by crystal packing interactions, while the more symmetric $(ABC)_2$ ring is more likely to be representative of the MAGΔ5 conformation in solution.

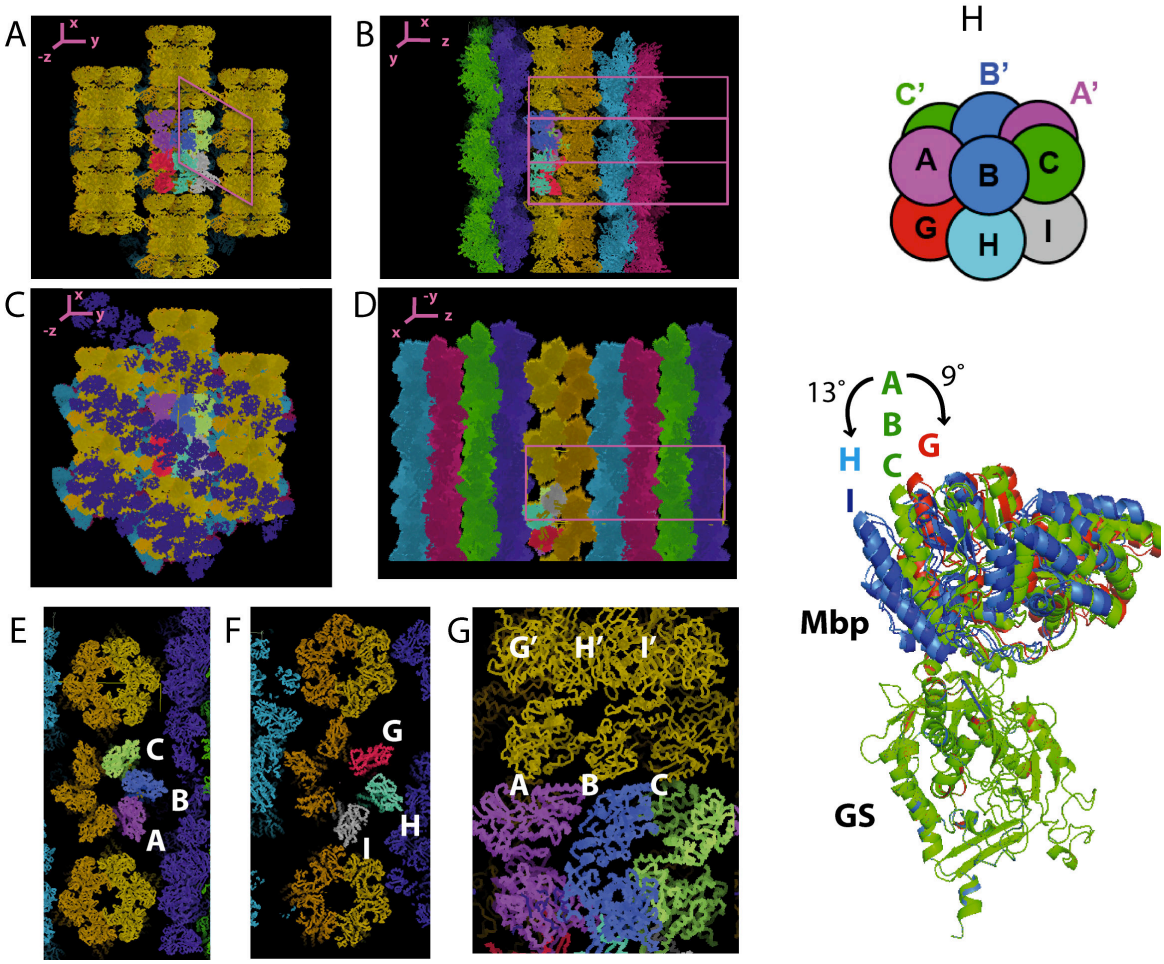


Figure 2.41: Crystal packing and Mbp orientation in MagΔ5. Packing interactions observed in crystals of MagΔ5. The crystallographic unit cell is outlined in pink in panels A-C, which show different views, as indicated by the x,y,z axes. Panel C is in the same viewing orientation as panel A but shows a larger slice along the z-direction to reveal interactions occurring between sheets of parallel fibers.

To confirm this hypothesis, we generated four perfectly symmetric models of MagΔ5 dodecamers by applying the D6 symmetry to the conformations observed for crystallographic subunits A, G, H and I. The $(A)_{12}$ model showed excellent agreement with the cryoEM density map and yielded a relative Mbp orientation closely resembling that obtained by independently fitting the Mbp crystal structure into the EM map (Figure 2.42). In contrast, the $(G)_{12}$, $(H)_{12}$ and $(I)_{12}$ models gave a significantly poorer fit, with a greater proportion of residues (~2000 atoms) outside the map, as estimated by the *fitmap* routine in CHIMERA (Pettersen et al., 2004). This result suggests that the MagΔ5 conformation in the $(ABC)_2$ ring of subunits most likely represents the solution conformation. By the same token, it confirms the validity of determining the structure of Mbp by cryoEM via the symmetrisation method. The fact that different monomer conformations are observed in

the crystal indicates that the MagΔ5 fusion constructs allows a certain degree of flexibility between the template and target domains, which may at least partly explain the limited resolution of the cryoEM map.

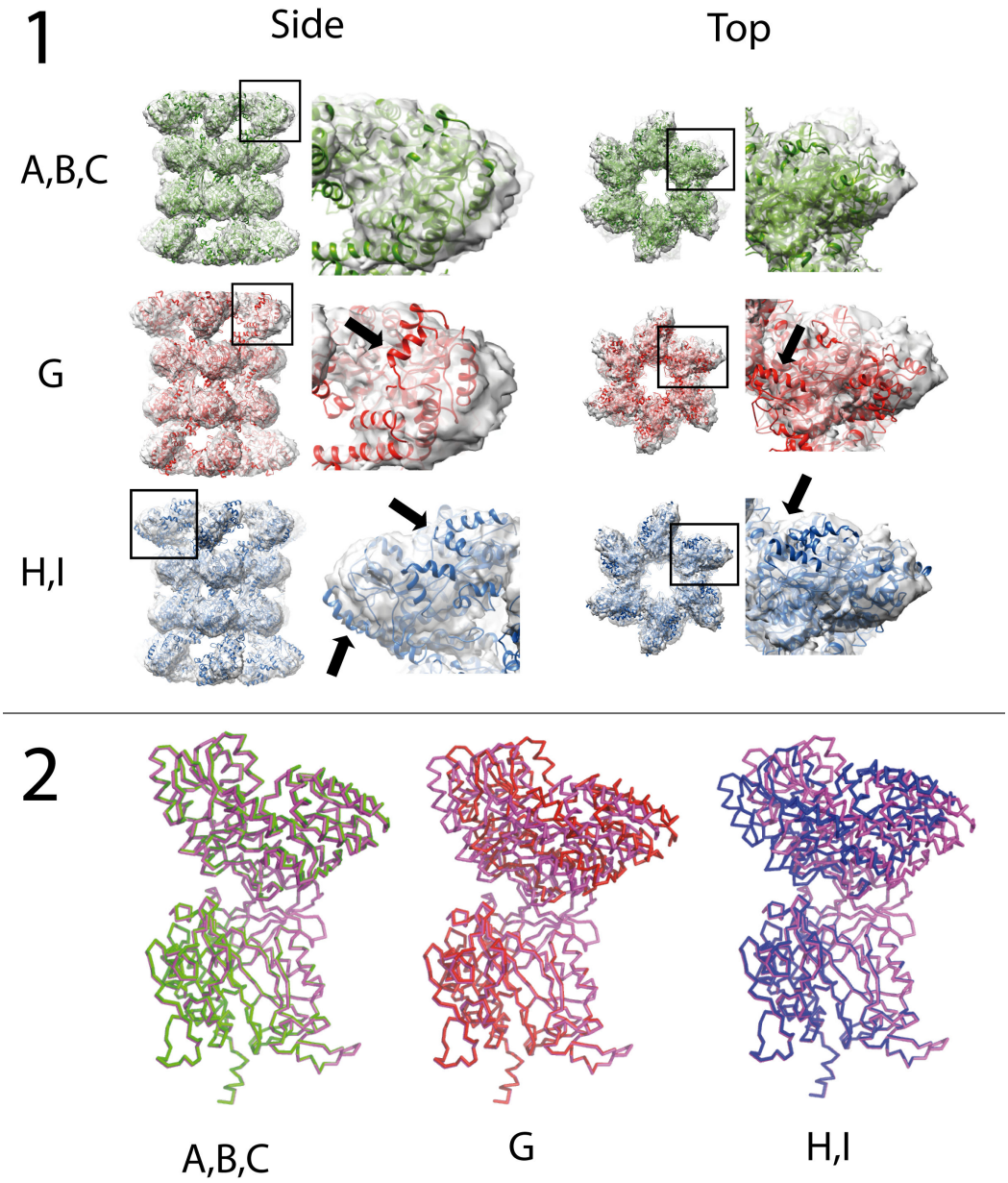


Figure 2.42: Fit of crystal structure in cryoEM map. 1) Fit of different A, G and H conformation observed in MagΔ5 crystal structure into cryoEM map. The best fit of the cryoEM map is obtained with the A conformation, whereas in the other two fits ~2000 more atoms lie outside the density (pointed by black arrows) 2) Comparison of Mbp-GS cryoEM model (magenta) with different conformations of Mbp-GS adopted in the crystal (ribbon diagrams superimposed via GS moiety). The A,B,C conformation is in good agreement with the one obtained by individually fitting a Mbp monomer into the cryoEM density. Conversely, G, H, I conformations diverge significantly from the cryoEM model.

3. DISCUSSION AND CONCLUDING REMARKS

ABSTRACT

In this thesis the proof-of-concept of protein symmetrization was demonstrated by solving the structure of Mbp fused to GS via a helical junction. Comparing this combination with other constructs studied in this work suggests that the presence of a large buried surface area and favorable target-template interactions considerably limit the flexibility of the chimera and improve the resolution of the corresponding cryoEM map. In E2 and Impβ fusion constructs these features have to be optimized. For the symmetrization of a target of unknown structure we envisage proceeding by fusing it to a panel of helical linker templates with different surface properties. This screening would maximize the likelihood of favorable template-target interactions resulting in increased rigidity of the final chimera. A three-alanine peptide could be used as the starting linker, and the degree of compactness of chimeras generated by varying the template or linker sequence could be monitored by using the biophysical ranking scheme that we applied to Mbp-GS..

RÉSUMÉ

La preuve de concept de la méthode de symétrisation de protéines a été démontrée à travers la structure cryoME de Mbp fusionné à GS via un peptide de liaison de structure hélicoïdale (Mag). La présence d'interactions favorables entre la cible et la matrice limite considérablement la flexibilité de la chimère et améliore la résolution de la reconstruction 3D obtenue par cryoEM. Pour les autres constructions de E2 et Impβ, ces caractéristiques doivent être optimisées. Dans le cas de la symétrisation d'une cible inconnue, nous envisageons de la fusionner à un ensemble de matrices ayant différentes propriétés de surface. Cette approche permettrait de maximiser la probabilité de générer des interactions favorables et d'augmenter la stabilité de la chimère. Le contrôle du degré de compaction des chimères avec diverses matrices et liaisons pourrait être effectué en utilisant un système de score attribué à un ensemble de méthodes biophysiques tel que nous l'avions appliqué à Mbp-GS.

3.1 PROTEIN SYMMETRIZATION FEASIBILITY STUDIES: EXPLORATORY SCREENING

Recent advances in cryoEM now permit structures to be determined at near-atomic resolution for large (> 300 kDa) proteins from frozen-hydrated solutions (Kuhlbrandt, 2014; Liao et al., 2013; Smith, 2014). However, because many biomedically relevant proteins are monomeric and < 100 kDa in mass, they remain unsuitable for cryoEM analysis. We envisaged circumventing this obstacle by fusing a monomeric target protein to a homo-oligomeric protein (template), thereby generating a self-assembling particle whose large size and symmetry would facilitate cryoEM analysis. The goal of the present thesis was to demonstrate the proof-of-concept of this approach, by solving the cryoEM structure of a "symmetrized" target and comparing it to its known atomic structure.

To accomplish this idea, we had to design a fusion protein in which both the target and template moieties are properly folded and their interaction is sufficient to produce a rigid particle amenable to cryoEM analysis. However, even knowing the structure of the building blocks (target, template and linker sequence) it is uncertain that a given fusion protein will fold correctly into the desired structure. Therefore, we tested symmetrized versions of more than one known protein target considering two linker strategies (described in § 2.1): ***i) a helix-based connection***, in which an α-helical linker continuous with helices in both the template and target is used, such that the target's orientation relative to the template is sampled in a discrete fashion dependent on the linker length; and ***ii) an unconstrained connection,*** in which the linker has no defined secondary structure, allowing the target to explore a larger space of orientations relative to the template.

A helix-based strategy was used to fuse the N-terminal helix of the dodecameric template GS to five targets bearing a C-terminal helix: Impβ, Mbp, Trea, Gsat, Kpr. The unstructured linker strategy was used to fuse the N-terminal helix of the E2 60-meric template to Mbp, and that of GS to Gfp (§ 2.2). Intuitively, for both strategies one expects that a sufficiently long linker should not significantly affect the individual structures of the fused moieties, which should be able to fold independently. Conversely, when the linker is short, both its sequence and length may affect the relative position of the connected domains and becomes crucial in defining particle shape. Following this

idea, we first connected our proteins with relatively long linkers, and screened for correct template-mediated oligomerization and degree of decoration, using SEC and negative stain electron microscopy. Subsequently, the most promising constructs were subjected to linker optimization to maximize particle compactness for structural studies. This procedure is summarized in Figure 3.1.

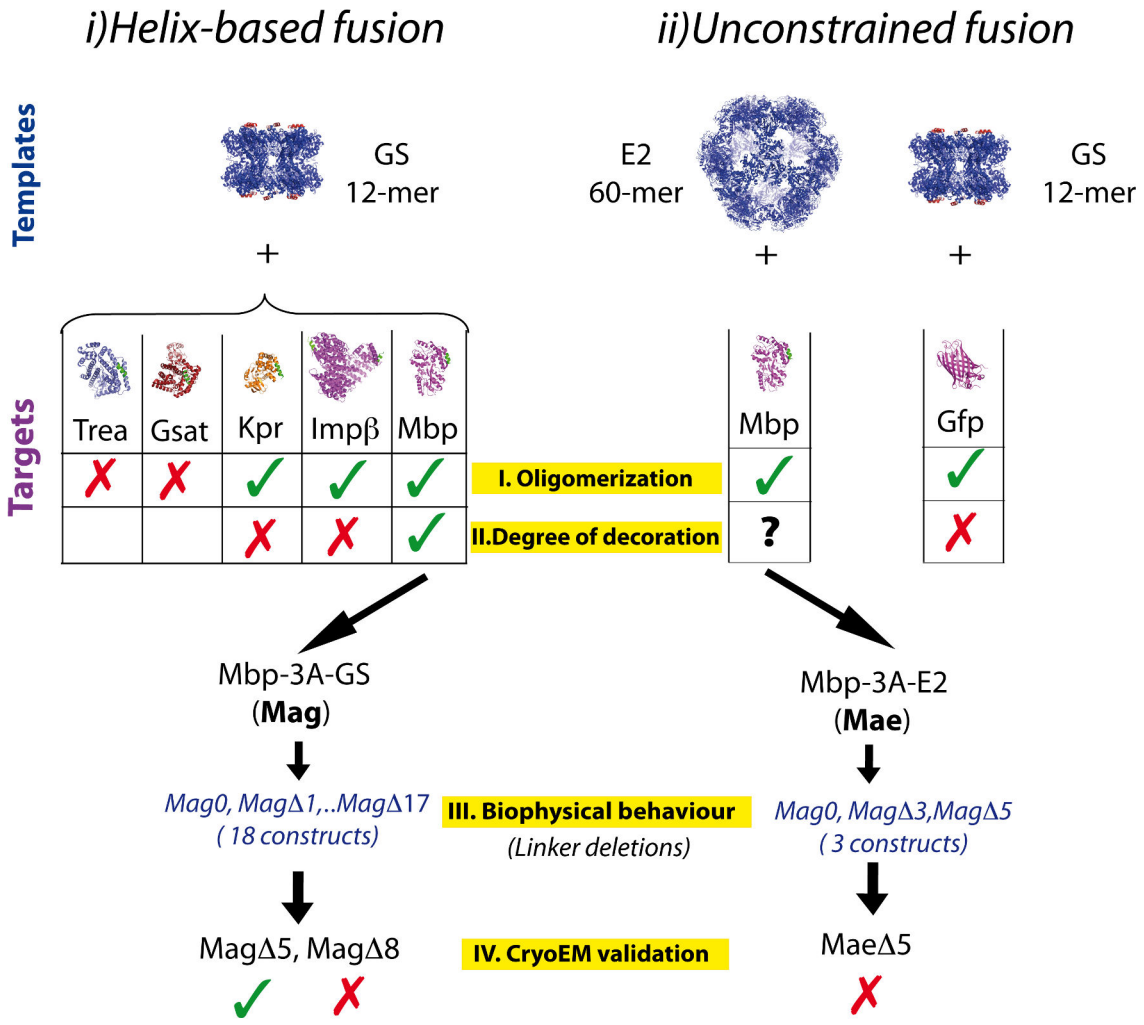


Figure 3.1: Selection procedure for target-template fusions used in this study.

In **GS fusions** the prevalence of top and side views and the peripheral position of the N-terminal junction made it easy to verify correct oligomerization and number of visible target moieties ("degree of decoration"), because one could visually recognize the template core. Indeed, negative stain micrographs indicated that fusions of Impβ, Kpr, Mbp, and Gfp assembled correctly as dodecamers, whereas those of Trea and Gsat did not. Mbp yielded the best expressed fusion protein, which also gave particles that were visually the most homogeneous in conformation. By replacing the long L9 linker with a

three alanine linker we obtained a quite promising Mbp-GS starting chimera for structural analysis, as judged by the negative stain model (§ 2.2.2, Figure 2.18).

Since **Impβ** has nearly twice the mass of the template, it should have been even easier to identify this target on GS. However, the target density was poorly visible (§ 2.2.1). Moreover, the fusion protein was highly prone to aggregation. These findings were independent of the linker length or staining solution used. A highly informative experiment was that in which gold beads were used to label the target moiety's N-terminal His-tag and determine its distance from the center of the particle: the long distances and high variability observed confirmed that Impβ was improperly folded in these particles. Conversely, the corresponding distances measured for the Mbp-GS fusion presented a narrow distribution centered at a value consistent with expectation (§ 2.2.2). Other structural and biophysical studies reported in the literature have shown that, in the absence of partner proteins, Impβ is a highly flexible solenoid protein, whose flexibility is central to its ability to bind and transport a large range of substrates diverse in size and shape. Indeed, Impβ has been described as a "molecular spring", whose fold is intermediate between that of a globular protein and of an intrinsically disordered protein (Fukuhara et al., 2004; Kappel et al., 2010; Zachariae and Grubmuller, 2008). Hence, it is not surprising that tethering such an elastic protein to GS resulted in misfolding. Alternative strategies to symmetrize Impβ without compromising folding might be more successful, such as fusing the solenoid's N-terminus to a template with a helical C-terminus, or using a different linker.

E2 fusions to Mbp appeared as promising round capsids of the expected size. However, because the outer shell of target moieties hides the inner template core, it is difficult (visually) to confirm that the template has correctly oligomerized and how many target copies decorate the template.

In summary, from this first exploratory screen we selected two well behaved quasi-symmetric chimeras, both bearing a three alanine linker, one made using the helix-based strategy, Mbp-GS (Mag), and the other using the unstructured linker strategy, Mbp-E2 (Mae).

3.2 LINKER OPTIMIZATION IN THE HELIX-BASED STRATEGY

3.2.1 BIOPHYSICAL ANALYSIS OF MAGΔN CONSTRUCTS

The advantage of the helix-based strategy is that it allows one to predict the structure of the fusion protein by a rigid body alignment of the target and template subunits via the linker connection (§ 2.1). Visual inspection of the Mag0 *in silico* model suggested that up to 17 residues in the linker region of Mag0 could be deleted without compromising the folding of GS (§ 2.3, Figure 2.23). Therefore we deleted the linker residues one by one and characterized their heterogeneity by biophysical techniques, seeking the most suitable construct for cryoEM analysis. Selection criteria included migration as a single band on a native gel, a compact hydration radius, high thermal stability, cooperative unfolding, and monodispersity (§ 2.3.1). By examining the contribution of each assay to the overall score calculated, it seems that no single parameter is prevalent in defining the ideal construct (Figure 2.24). For instance, MagΔ1 and MagΔ11 have the best scores in TSA and SEC, but relatively low scores in the other techniques. Conversely, MagΔ5 ranked best in none of the assays, but it presents consistently high scores that made it stand out in the overall statistics. While a single biophysical parameter was not sufficiently informative to pinpoint the best construct, the combined results identified the most suitable constructs for subsequent cryoEM analysis: MagΔ5 and MagΔ8. These two constructs also exhibited a favorable appearance by negative stain EM analysis, confirming the validity of our ranking method. In principle, the selection of the best candidate over a panel of constructs with different linker lengths could have been done exclusively by negative stain EM. However, it is difficult to establish the relative degree of compactness and rigidity by visual appearance. Conversely, the biophysical parameters provide more quantitative information and are highly suited to automation in the case of large screens. For instance, MagΔ8 and MagΔ5 appeared equally homogeneous by negative stain EM, but MagΔ5 behaved better in the biophysical assays. Indeed, MagΔ5 also gave the better cryoEM reconstruction, highlighting the usefulness of performing biophysical assays prior to cryoEM analysis. Moreover, this selection analysis is independent of the nature of linker used and can readily be applied to fusions involving target proteins of unknown structure.

3.2.2 VALIDATION OF PROTEIN SYMMETRISATION

The cryoEM D6 restrained map of the best dodecameric chimera MagΔ5 was obtained at a resolution of 10 Å, as estimated by FSC 0.5 criterion (§ 2.4.3). Features of the map indicate that the cryoEM structure corresponds to the crystallographic model of Mbp. This result proves that symmetric scaffolding is a valid approach for solving protein structures by cryoEM that are below the acknowledged molecular size limit of 100 kDa (Henderson, 1995). To our knowledge, the 40 kDa Mbp structure we obtained represents the lowest protein molecular weight limit reached so far by cryoEM. In principle smaller symmetrised protein structures could be solved by a similar approach. It would have been desirable to reach higher resolution for MagΔ5, as the attained 10 Å resolution does not allow *de novo* tracing of the target backbone. The limited resolution may be due to technical reasons, such as the narrow angular coverage and limited size of the dataset. Therefore, it would be interesting to collect a larger dataset using modern detector systems in order to repopulate low abundance views and increase the SNR, and assess the effect of these on map quality. Flexibility of Mbp relative to the template could also limit resolution when enforcing D6 symmetry. In our 7 Å crystal structure of MagΔ5, six crystallographically independent Mbp copies exist within the asymmetric unit (§ 2.5). Of these, three are strongly affected by crystal packing and present different orientations. In contrast, the other three, which do not participate in crystal contacts, adopt the same orientation relative to the template as observed in our cryoEM map. This suggests that in solution the Mbp subunits probably tend to be symmetrically disposed around the template, but that some heterogeneity in conformation likely exists. Even though side chain conformations are not visible at 7 Å resolution, rigid body fitting of Mbp and GS atomic models into the crystallographic density map suggests that two salt bridges are likely present at each Mbp-GS interface (Figure 3.2). These target-template interactions might explain the mutual thermal stabilization of target and template observed by TSA (§ 2.3.1, Figure 2.25) and the higher local resolution at their interface, estimated by FSC (§ 2.4.4, Figure 2.36). Conceivably, introducing point mutations to additionally stabilize this interface could further increase the overall rigidity of the particle and improve the final resolution.

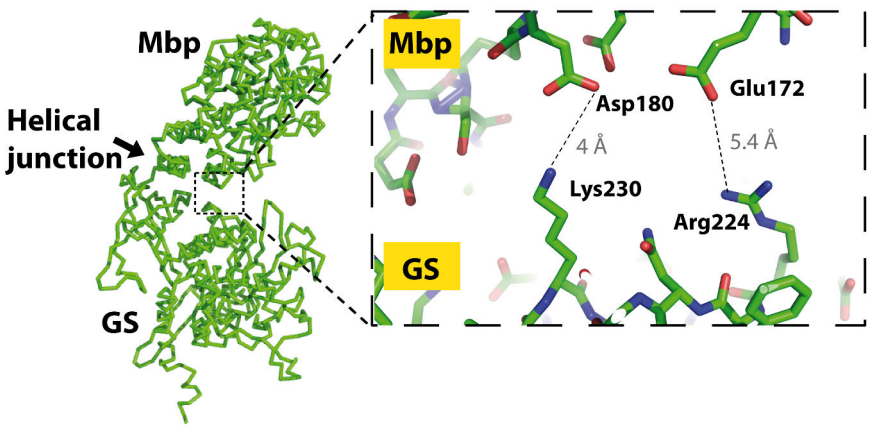


Figure 3.2: Polar contacts at Mbp-GS interface in MagΔ5. *Left*: ribbon diagram of the crystal structure of MagΔ5 monomer A, which agrees well with the cryoEM structure. *Right*: Close-up of the lateral region where the target and the template subunits eclipse each other. Two salt bridges involving GS residues Lys230 and Arg224 and MBP residues Asp180 and Glu172 are likely to be formed. Such interactions would contribute to the overall stability of the MagΔ5 dodecamer.

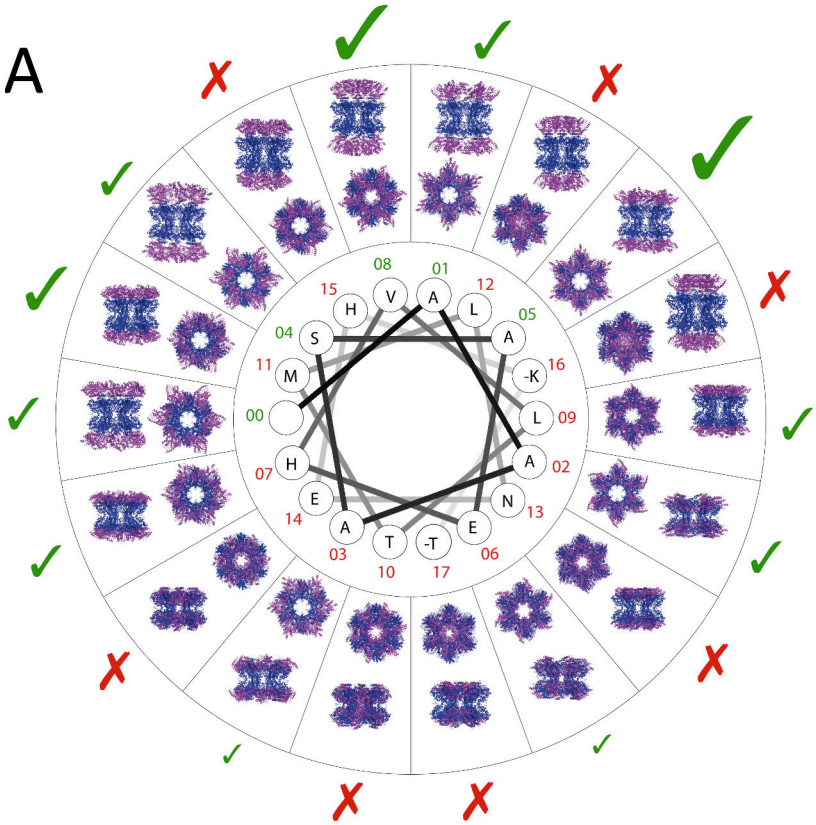
MagΔ8, the second best behaved construct according to our ranking method, was analyzed by cryoEM using identical data collection parameters and image processing methods as those used for MagΔ5. However, the overall resolution achieved for MagΔ8 was substantially lower than that for MagΔ5 (15 Å vs 10 Å). Moreover, the Mbp density is smaller than expected and detached from GS. This finding correlates well with local FSC calculations, which reveals a similar (9-11 Å) resolution for the GS moieties of both constructs, but much higher resolution for Mbp in MagΔ5 than in MagΔ8. Interestingly, in MagΔ8 each ring of 6 Mbp subunits is staggered with respect to the underlying ring of GS subunits, such that, apart from the linker region, the target lacks contact points with the template. Conversely, in MagΔ5 the Mbp ring is eclipsed with respect to the GS subunits, providing a larger contact interface (Compare panels C and E in Figure 2.36). This finding suggests that the higher resolution of MagΔ5 probably reflects a larger target-template interface (buried surface area). In MagΔ8, although the linker is 3 residues shorter, the helical geometry orients the Mbp far off the template, resulting in a less rigid particle.

3.2.3 EXPERIMENTAL DATA RATIONALIZED BY *IN SILICO* MODELLING

We generated *in silico* models of all the oligomeric MagΔN structures assuming perfect continuity between the terminal helices of the template and target proteins (for details see § 4.1). By virtue of the helical nature of the linker, deletion of one residue

results in a 100˚ rotation of Mbp relative to GS about the linker helical axis, yielding a total rotation of 1700˚ when all 17 residues are deleted. A further deletion would cover five complete turns (1800˚ = 5x360˚) to produce a model with the template and target proteins in the same relative orientation as Mag0 (except for the translational offset). This means that all possible relative template-target rotational orientations about the helix axis were mapped with a 20° sampling interval.

The Mag0-MagΔ17 models, displayed on a helical wheel diagram (Figure 3.3), were inspected and classified as sterically "forbidden" or "allowed". In the fusions located in the lower part of the diagram Mbp is oriented towards and overlaps with GS, thereby always causing "sterically forbidden" arrangements. Conversely, in the constructs at the top of the wheel, Mbp points away from GS, generating more favourable configurations. However, when the linker is further reduced (MagΔ11 and shorter constructs), neighbouring Mbp subunits clash with each other, producing additionally "forbidden" assemblies. Comparing these findings with the experimental data (§ 2.3.1, §2.3.2) we observe that chimeras corresponding to "allowed" models (Mag0, MagΔ1, MagΔ4, MagΔ5, MagΔ8) scored well in biophysical studies, showed a high degree of homogeneity by negative stain EM, and appeared to adopt an overall structure resembling the predicted structure, consistent with helical integrity of the linker residues. Concerning constructs corresponding to "forbidden" models, those with sufficiently long linkers still appeared to assemble correctly as dodecamers, but were less rigid and stable (e.g. Mag2, Mag9 and MagΔ11), consistent with a loss of helical linker conformation to accommodate folding of the target and template moieties. Conversely, "forbidden" constructs with shorter linkers (MagΔ12-MagΔ17) produced a highly heterogeneous sample that displayed multiple bands when analysed by native PAGE and multiple unfolding transitions by TSA. Presumably, this happens because the lack of conformational freedom due to the short linker results in unfavourable target-template or target-target interactions, leading to incorrect assembly of the dodecamer and/or partial unfolding of Mbp or GS.



B

Junction sequence	Fusion name	Linker length
TRITKAAASAEHVLTMLNEHEVKF	Mag 0	3
TRITKAASAEHVLTMLNEHEVKF	Mag Δ1	2
TRITKASAEHVLTMLNEHEVKF	Mag Δ2	1
TRITKSAEHVLTMLNEHEVKF	Mag Δ3	0
TRITKAEHVLTMLNEHEVKF	Mag Δ4	-1
TRITKEHVLTMLNEHEVKF	Mag Δ5	-2
TRITKHVLTMLNEHEVKF	Mag Δ6	-3
TRITKVLTMLNEHEVKF	Mag Δ7	-4
TRITKLTMLNEHEVKF	Mag Δ8	-5
TRITKTMLNEHEVKF	Mag Δ9	-6
TRITKMLNEHEVKF	Mag Δ10	-7
TRITKLNEHEVKF	Mag Δ11	-8
TRITKNEHEVKF	Mag Δ12	-9
TRITKEHEVKF	Mag Δ13	-10
TRITKHEVKF	Mag Δ14	-11
TRITKEVKF	Mag Δ15	-12
TRITEVKF	Mag Δ16	-13
TRIEVKF	Mag Δ17	-14

Figure 3.3: *In silico* prediction of Mag constructs. Top and side views of structural models for fusions between Mbp (magenta) and GS (blue) are displayed on a helical wheel and show the sampling of different target-template relative orientations. The residue inside each circle indicates the linker residue deleted from the previous longest construct (–T and –K indicate removal of the two C-terminal residues of Mbp). Sterically "forbidden" and "allowed" constructs are indicated by red and green numbers, respectively. The ✗ symbol outside the circumference indicate constructs that either migrated as multiple bands on a native gel and/or do not assemble properly , as judged by negative stain EM (Figure 2.25). Constructs that migrate as a single band in native gel are indicated by a ✓ symbol where size is scaled to the overall biophysical score (Figure 2.24)

The quasi-periodicity of sterically “allowed” and “forbidden” *in silico* alignments correlates well with the overall behaviour of constructs as judged by biophysical assays and negative-stain EM. However, superimposing the MagΔ5 experimental structure with the *in silico* model shows that in the former Mbp is rotated by ~30˚ away from GS, thereby resulting in a less compact dodecamer than that predicted (Figure 3.4). Inspection of the *in silico* model reveals a small number of minor sterical clashes involving bulky side chains at the Mbp-Mbp and Mbp-GS interfaces (Figure 3.4). The experimental structure avoids such clashes by tilting Mbp away from GS, which implies a slight distorting of the helical connection. In principle, the interfaces could be engineered to reduce these clashes and generate a more compact particle. Thus, *in silico* modeling based solely on secondary structure provides a useful first approximation of the overall shape of chimeras. However the effective compactness of the oligomeric assembly can be affected by the chemical nature of the components. This often requires a further engineering step to be optimized, as also previously reported for the design of protein nanohedra (Lai et al., 2013).

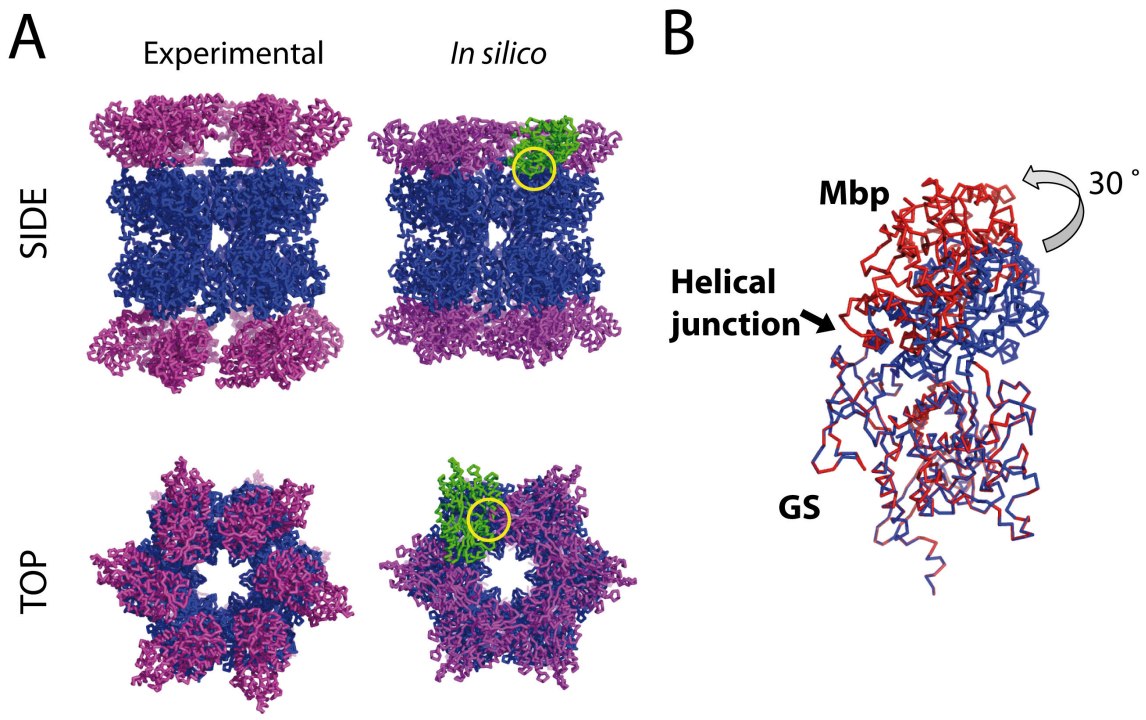


Figure 3.4: Comparison between *in silico* and experimental MagΔ5. A) Ribbon diagrams of MagΔ5 models displayed in side and top views. The idealized model appears more compact than the real one (by ~5 nm in the vertical direction for the side view). This can be attributed to minor steric clashes in the predicted model involving side chains in the region highlighted by the yellow circle. Mbp-GS monomers probably adopt a more elongated structure to avoid such clashes. B) Superimposition of *in silico* MagΔ5 model (blue) and experimental structure (red). In the latter the target is rotated further away from the template by ~30˚.

3.3 LINKER OPTIMIZATION IN UNCONSTRAINED CONNECTION STRATEGY

Since Mae0 does not present a linker with defined secondary structure *in silico* structural modelling is less reliable than for helix-based fusions. Therefore, to identify the effective number of linker residues that could be deleted without affecting protein folding we considered only the crystal structure of E2. In E2 the hook-like N-terminus lies on the surface of the icosahedron; however, its electron density is poorly defined due to the low resolution (4.4 Å, Izard et al., 1999). In the crystallographic study of a homologous 24-meric enzyme, the whole hook-like region was postulated as being essential for oligomer stability (Mattevi et al., 1992). On the other hand, a cryoEM structure (at 25 Å) of E2 bound to its partners E1/E3 revealed the whole hook-like region to be flexible and detached from the core surface, putatively not involved in oligomerization (Milne et al., 2006). To minimize the risk of oligomer instability, we deleted at most only two residues

from the E2 N-terminus (MaeΔ5 construct). In MaeΔ5, if the hook-like region were bound to the core (as in the Mattevi structure), the expected distance between Mbp and the E2 N-terminus would be less than 1 nm. In contrast, we measured an approximate distance of 4 nm, consistent with the structure of Milne and collaborators (Figure 3.5).

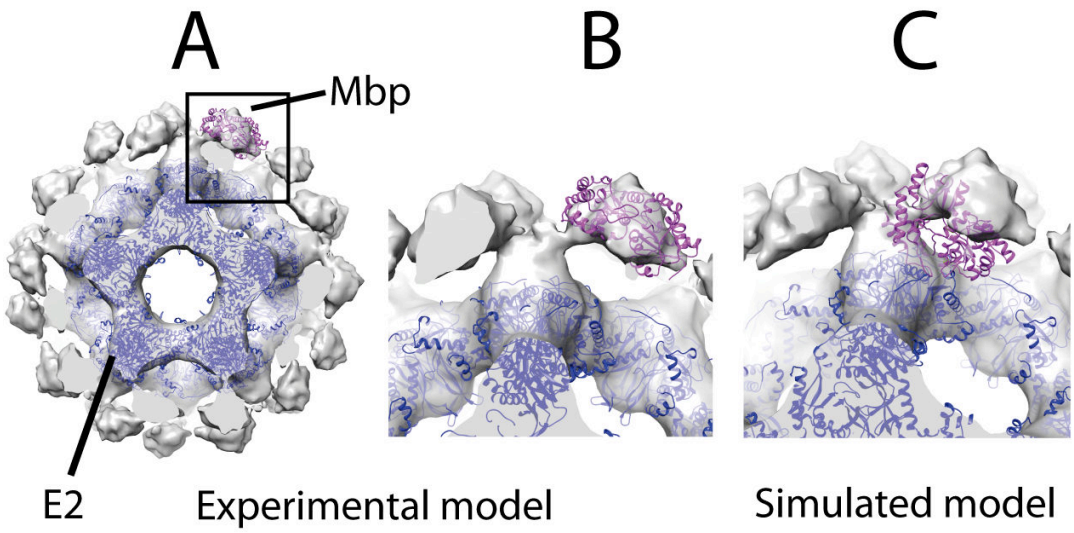


Figure 3.5: Target-template distance in MaeΔ5 cryoEM. A) CryoEM map of MaeΔ5, fitted with atomic models of Mbp (magenta) and of E2 (in blue), visualized along the 5-fold axis of the icosahedron. B) Close-up of the Mbp-E2 junction region. The distance between these two moieties is ~4 nm, suggesting the presence of a flexible E2 N-terminus not bound to the core surface, consistent with the structure of Milne at al., 2006. C) Simulated model in which the N-terminal E2 region is bound to the core surface as in the structure of Mattevi et al, 1992. This model is clearly less consistent with the experimental map.

N-terminal flexibility of E2 would yield poor tethering of Mbp and variability in its position, causing its density to be averaged out in the symmetry-restrained cryoEM map. This, and a lack of template-target interactions, could explain why the volume of Mbp density in our cryoEM reconstruction of MaeΔ5 is much smaller than expected. These observations suggest that the effective linker region between Mbp and E2 is much longer than we estimated relying on the crystal structure. Hence, the rigidity of E2 icosahedral fusions might be substantially improved by partly or completely deleting the N-terminal hook-like region. An interesting future study would be to determine the optimal deletion by structurally and biophysically characterizing several N-terminally truncated E2 cores.

In this work we biophysically characterized three E2 fusions (Mae0, Δ3 and Δ5) and 18 GS fusions (Mag00-MagΔ17) to screen for constructs suitable for cryoEM analysis. Concerning Mae fusions, only the analysis by DLS, TSA and negative stain EM were

informative. Due to their large size (~50 nm) we were unable to characterize these constructs by native PAGE and SEC. Analytical ultracentrifugation (AUC) would be a better technique to study such large particles. By analogy with MagΔ5, one might expect to improve the situation by engineering an eclipsed configuration between template and target trimers. However, this arrangement (simulated by rotating the Mbp trimers with respect to the template) would likely cause a sterical clash between target subunits. I.e., the desired eclipsed configuration might not be achievable because Mbp is too big compared to E2 (Figure 3.3). (This might also partly explain the inefficient assembly and partial aggregation observed for Mbp-E2 constructs).

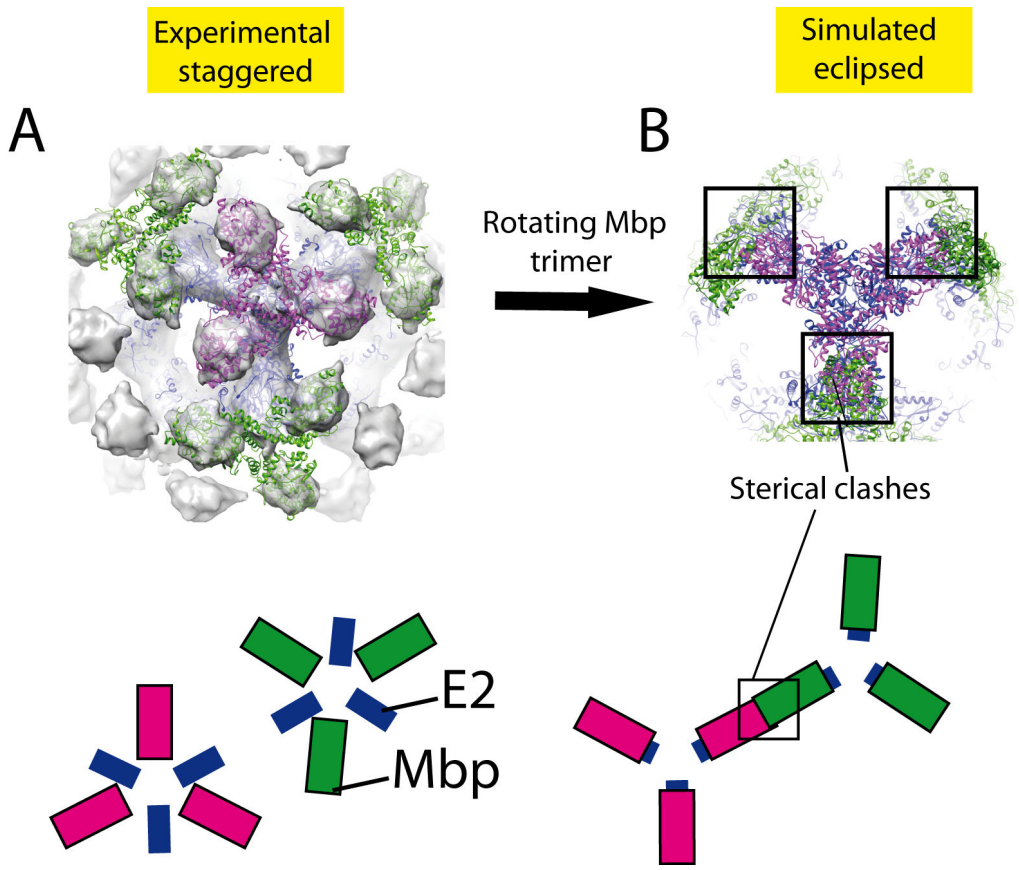


Figure 3.3: Target-template interactions in the Mbp-E2 fusion A) CryoEM density map of MaeΔ5, fitted with atomic model of the template (in blue) and a trimeric unit of the target, staggered with respect to the template. B) Same map with trimeric target units surrounding central one underlined in green. Below, a schematic diagram representing the staggered configuration of two adjacent trimeric units (In blue the template, in magenta and green the target). C) Simulated eclipsed configuration of Mbp-E2, generated by rotating the target trimers to align them with E2. This configuration is predicted to cause sterical clashes between adjacent Mbp trimers, as shown in the schematic diagram below.

As E2 is a well behaved high-symmetry template, it would be interesting to attempt fusing smaller targets to it.

3.4 PERSPECTIVES FOR FUTURE APPLICATIONS

Our studies of the Mbp-GS constructs reveal important factors for achieving a template-target fusion protein suitable for cryoEM analysis. The minimal requirements are the proper folding of the template and target and correct assembly of the oligomer. An important factor affecting the resolution of the target density is the target's ability to form stable interactions with the template and with other copies of itself (this makes MagΔ5 a better construct than MagΔ8). Such interactions depend on the complementarity of shape and surface properties between the target and template, which cannot be controlled when dealing with a target of unknown structure. Naively, however, one could imagine that the lower the target-to-template mass ratio, the greater are the chances of having a large target-template interface and of reducing target mobility. Such reasoning might explain the outcomes observed for the various target-template combinations studied in this work. Indeed, in the case of the most successful construct, MagΔ5, the target-to-template mass ratio is ~0.8, whereas for the less successful Impβ-GS and Mbp-E2 constructs it is 1.7 and 1.4, respectively. Accordingly, in the case of Impβ it would be interesting to test a larger template subunit to increase the chances of forming a stable target-template interface and potentially reduce problems related to improper folding. In the case of Mbp-E2 we achieved a low resolution reconstruction in which the Mbp trimers are staggered with respect to the underlying E2 trimers (§ 2.2.3, Figure 2.22). As with MagΔ8, the absence of contacts with the template surface (other than in the linker region) might be responsible for the poor resolution attained.

Experience gained during this thesis project suggests general guidelines for future efforts aimed at symmetrizing proteins, including those of unknown structure. The main parameters to choose are the template and the linker. Concerning the linker, we confirmed that a three alanine linker is a good starting point for both a helical and a non-helical connection (Mae, Mag and Gfp-GS). In cases where secondary structure predictions indicate that the target protein has a helical N- or C-terminus, a good strategy

is to design a fusion construct that maximizes the likelihood of forming a continuous helix between the template and target. Once a promising construct is identified, the linker should be shortened to select the most compact construct using a panel of biophysical techniques (ranking method we developed), ideally implemented in a high-throughput fashion. Well behaved constructs should then be screened by negative stain EM, prior to selecting a limited number for cryoEM analysis.

Another important feature we pinpointed for the achievement of a compact fusion is target-template surface complementarity and effective interactions. To maximize potential buried surface area, a series of templates with comparable or higher MW than the target should be tested. In order to favour stable interactions it would be interesting to produce templates with different surface properties. For instance, a panel of GS templates could be designed so as to include molecules whose target-proximal surface was enhanced for i) short, polar side chains, ii) acidic or iii) basic residues, or iv) hydrophobic and aromatic side chains. In this perspective, the series of protein cages designed for biotechnological applications could be a useful source of templates with tunable size, symmetry and surface properties (King et al., 2014; Lai et al., 2012b; Yeates and Padilla, 2002)

3.5 CONCLUSIONS

In the last few years, the resolution of 3D structures attainable by electron microscopy has evolved enormously, moving from a morphological description of macromolecules (<20 Å) to the near atomic resolution (~3 Å) of large complexes such as ribosomes. However, as the size and symmetry of the molecule decreases, cryoEM analysis becomes increasingly difficult, and currently is practically impossible for monomeric proteins below ~100 kDa in mass.

The aim of this work was to develop a new approach that would lower this molecular weight limit by genetically fusing the protein of interest to a homo-oligomeric template. We engineered fusion proteins comprising a small number of target and template proteins, characterized these in biophysical assays and by negative stain EM, and identified a promising candidate for further study, Mbp-GS. We investigated a panel of Mbp-GS chimeras having different linker lengths and selected the most compact member

as judged by various biophysical parameters. The 10 Å cryoEM map of the symmetrized 40 kDa Mbp protein presents shape and features that correlate well with the crystal structure. This result establishes the proof-of-concept that protein symmetrization can be used for the structure determination of monomeric “small” protein targets by cryoEM.

A factor determining higher resolution of this construct with respect to other tested combinations was a stable target-template interface. The helical linker probably also contributed importantly to particle rigidity, although additional studies with other constructs are required to confirm this. We envisage that the protein symmetrization of an unknown target could be accomplished via a trial and error approach, by initially fusing it to a large panel of templates having different sizes and surface properties. The future of protein symmetrization is therefore in the production of a broader “toolkit” of templates. Biophysical assays and negative-stain and cryoEM would then be used to screen for the best chimera. This approach would be comparable to the multi-factorial approach used to search for initial hits in a crystallization experiment, followed by evaluation of different crystal forms by diffraction experiments. Using this approach, even if a near atomic resolution map is not achieved for a particular target by cryoEM, useful information about the overall shape and size of domains could be anyway obtained from the electron density maps.

In conclusion, the PhD work described above explored the concept of protein symmetrization to harness the enormous potential of cryoEM for analyzing low molecular weight targets. Based on a limited number of bacterial fusion proteins, we demonstrated that structural characterization of a 40 kDa protein is feasible using such an approach. Our findings suggest avenues to explore for future improvement of the methodology and pave the way for the analysis of more challenging protein targets.

4. EXPERIMENTAL AND COMPUTATIONAL PROCEDURES

ABSTRACT

In order to set up protein symmetrization we utilized a series of experimental and computational methods for steps ranging from the design to the structural determination of symmetric chimeras. First, *in silico* modelling of helix-based fusion constructs was performed using PyMOL and CCP4 routines. We developed a versatile cloning procedure to express the histidine-tagged fusion constructs in *E.coli* and purify them by Nickel affinity chromatography. After purification the proteins were evaluated by negative stain EM and biophysical techniques (TSA, DLS, SEC, native PAGE) that gave information about the homogeneity, stability and hydrodynamic radius of the chimeras. The best ranked chimeras were imaged by cryoEM at 300 kV. The starting model was mainly generated by angular reconstitution and the map refined by projection matching. Finally, to check flexibility of the best chimera we pursued crystallographic analysis and solved the structure by molecular replacement.

RÉSUMÉ

Afin de mettre en place la symétrisation de protéines, une série de méthodes expérimentales et informatiques a été utilisées La modélisation *in silico* des constructions à base de fusion hélicoïdale a été réalisée en utilisant PyMOL et CCP4. Nous avons développé une procédure de clonage pour exprimer en même temps toutes les constructions dans *E. coli* et les purifier par chromatographie d'affinité. Après purification, les protéines ont été évaluées par microscopie (coloration négative) et par des techniques biophysiques (TSA, DLS, la SEC, gel natif) qui donnent des informations sur l'homogénéité, la stabilité et le rayon hydrodynamique des chimères. Les mieux classées ont été imagées par cryoME à 300 kV. Les modèles 3D ont été générés, principalement par reconstitution angulaire, et ensuite raffinées. Enfin, pour vérifier la flexibilité de la meilleure chimère, nous avons résolu la structure cristallographique par remplacement moléculaire.

4.1 *IN SILICO* MODELLING OF HELIX-BASED FUSIONS

The helix-based fusion strategyused to link target and template proteins was inspired by the work of Todd Yeates and colleagues, who produced protein nanohedra by connecting the helical C- and a N-termini of a pair of proteins via a linker sequence with strong helical propensity (Padilla et al., 2001). By doing so, the whole junction is likely to fold as a continuous α-helix, spanning from one domain to the other. In this assumption, by superimposing the C-terminal, linker and N terminal helices it is possible to construct a rough structural model *in silico*, only taking into account the secondary structural elements of the components (Figure 4.1A). In the present thesis work we applied this principle to generate *in silico* models for symmetric fusions of GS with Impβ, Kpr and Mbp. As an example, in this section we describe the computational procedure for the Mbp-GS (Mag) fusions (Figure 4.1).

First, we generated an ideal 100-residue polyalanine α-helix named helix.pdb using PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC). Then, by using the *lsqkab* routine present in the CCP4 package (Winn et al., 2011) we performed two types of alignments, both over 8 residues (~ two helical turns). First, we overlapped the C-terminal portion of helix.pdb onto the GS N-terminal helix (gs.pdb), thereby generating helix_on_gs.pdb (Figure 4.1C1). Subsequently, we overlapped the Mbp C-terminal helix onto helix_on_gs.pdb. The last overlap range defines the net number of linker residues between Mbp and GS. By “sliding” the Mbp C-terminus over the helix_on_gs.pdb in the direction of GS, we reduce the effective number of residues between target and template (Figure 4.1C2). Afterwards, new Mbp coordinates, linker and GS were concatenated, and monomeric fusions overlapped to the template, by generating the oligomeric chimera *in silico*. Keeping the template coordinates fixed, by virtue of the α-helix geometry, the removal of each linker residue implies a rotation of 100˚ and a shift along the helical axis of 1.5 Å of the target. Hence, gradual shortening of the linker generates a quasi-periodical change in the orientation of the target, by affecting the overall shape of the monomeric fusion and in turn of the the oligomeric particle (Figure 4.1B and Figure 4.1D). By inspecting the obtained models and checking for sterical clashes, we can classify the structures as “forbidden” or “allowed”.

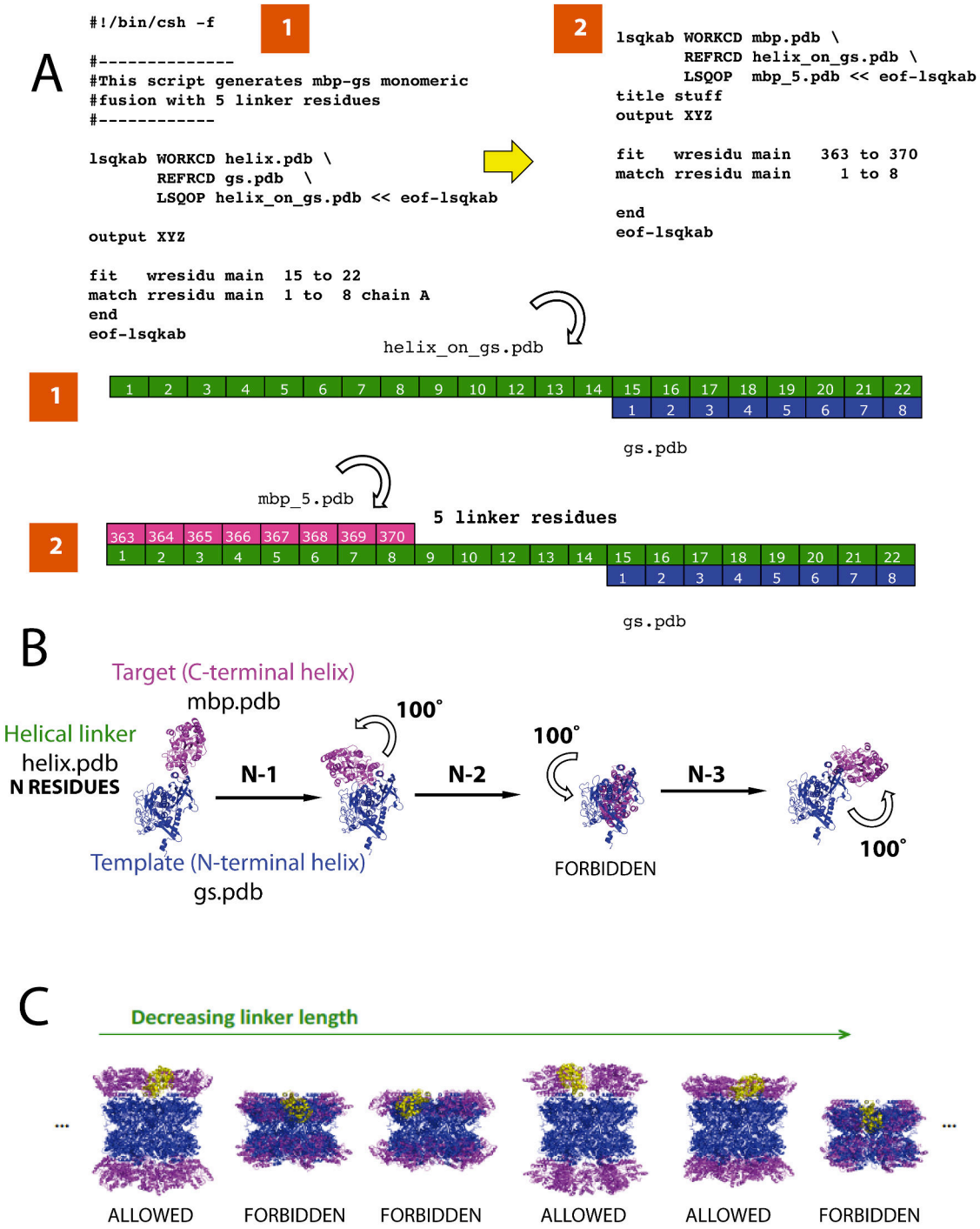


Figure 4.1: *In silico* helical alignment. A)Computational procedure to generate Mbp-GS fusion with five linker residues: 1) alignment of linker helix on GS 2) alignment of Mbp onto linker helix, leaving 5 residues between target and template. B) Ribbon diagram of Mbp-GS monomeric fusions having from N to N-3 linker residue, showing the change in orientation of Mbp with respect to GS by 100˚ every linker residue deletion. C) Ribbon diagram of dodecameric Mbp-GS fusions where the central Mbp copy is highlighted in yellow to underline the change in orientation of Mbp with respect to GS that generates sterically allowed or forbidden structures.

In the latter models, it is possible (but not guaranteed) that the linker helix is helical and that the actual structure resembles the predicted model; the severely "forbidden" models, are incompatible with preservation of the linker helical structure and we have no prediction for the true assembly. Therefore we can assume that either the linker helix becomes distorted, or the proteins do not fold or assemble properly.

4.2 CLONING

4.2.1 CLONING STRATEGIES AND RECOMBINANT EXPRESSION IN *E.COLI*

A pETM-11 expression vector (EMBL) was already available in the lab for the expression of a His-tagged L9-GS protein bearing the 18-residue linker sequence AMAKALEAQKQKEQRQAA, derived from the central portion of a long solvent-exposed helix in the ribosomal protein L9 (PDB ID code 1div). The gene was cloned between NcoI and KpnI restriction sites in frame with an N-terminal hexa-histidine tag (His_6-tag) coding sequence followed by a TEV cleavage sequence. The vector was modified to facilitate the generation of N-terminally His-tagged fusion constructs by simple restriction ligation (Figure 4.2A).

The CCC codon (encoding proline) following the His-tag was mutated to CAT (encoding histidine) so as to generate an NdeI restriction site (Figure 4.2 B). The resulting vector, pETM-11*, presents a His_7-tag coding sequence followed by NdeI, NcoI and KpnI sites, thereby allowing for the insertion of one coding sequence between the NdeI and NcoI sites and a second coding sequence between the NcoI and KpnI sites. The pETM-11* vector thus allowed for the rapid exchange of either the target or the template sequence (Figure 4.2C). Following generation of the initial template-target fusion vector, the linker sequence was modified by sequential deletion of residues to yield fusions with different linker lengths (Figure 4.2D).

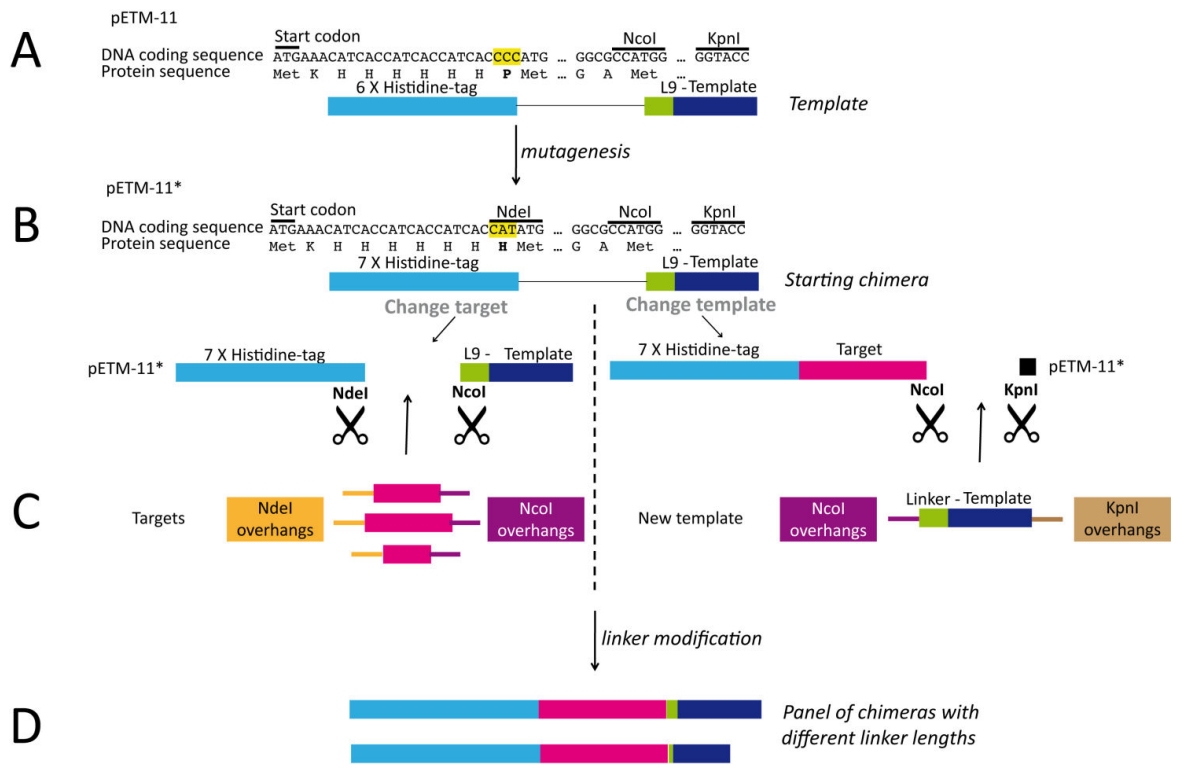


Figure 4.2: Strategy developed to allow rapid cloning of a given target-template fusions gene by restriction-ligation from the same plasmid pETM11*. A) Mutation of CCC to CAT codon introduces a NdeI site, and corresponds to a mutation from P to H (vector pETM11*). B,C) restriction ligation with either NdeI/NcoI or NcoI/KpnI allows replacement either of the target or the template. D) The original linker was shortened by deletion.

<u>Cloning of multiple targets in fusion with L9-GS:</u> the genes of the chosen targets (§ 2.1.2) Trea, Kpr, Mbp, Gsta, Gfp were amplified from the *E.coli* genome with specific primers bearing extensions harbouring NdeI and NcoI sites. The PCR fragments were subcloned in a pJET1.2/blunt cloning vector (Thermo-scientific). The pJET-insert plasmids were purified at high yield and purity (NucleoSpin® Plasmid, Macherey Nagel) and digested with NcoI/NdeI enzymes (New England Biolabs). Restriction fragments bearing the desired coding sequence were then ligated into the pETM-11* vector (linearized with the same enzymes) with a T4 ligase (Fermentas) following the manufacturers' procedures.

<u>Cloning of Mbp-E2:</u> the E2 gene was PCR-amplified with extensions bearing NcoI-linker/KpnI from a pETE2disp vector, provided by Dr. De Berardinis (Institute of Protein Biochemistry, Naples, Italy). The PCR fragment was subcloned in a pJET1.2/blunt cloning

vector (Thermo-scientific). The pJET-E2 plasmid was purified at high yield and purity (NucleoSpin® Plasmid, Macherey Nagel), digested with NcoI/KpnI enzymes (New England Biolabs) and the E2-coding fragment ligated into a pETM-11*-Mbp-18-GS vector (linearized with the same enzymes) with a T4 ligase (Fermentas) following the manufacturers' procedures. The same E2-coding fragment was analogously subcloned in the unmodified pETM-11 vector to produce a His-tagged template for EM studies of the unfused template.

4.2.2 IMPORTIN-β – GS CLONING

An efficient expression and purification protocol for Impβ, expressed from a pQE30 vector (Qiagen), had been previously established in the laboratory. Our goal was to fuse Impβ to GS via the L9 linker. Previous tests performed in our laboratory showed that moving the Impβ-coding sequence to other expression vectors negatively affected the expression yield of the protein. Therefore, rather than applying the cloning strategy involving pETM-11 described in § 4.2.1, we decided to modify the original vector bearing the Impβ gene. The L9-GS DNA sequence was first PCR amplified from the *E.coli* K12 genome using appropriate primers, purified by agarose gel extraction followed by an affinity column step (NucleoSpin® Plasmid, Machery Nagel) and inserted downstream of the Impβ gene by restriction free (RF) cloning (van den Ent and Lowe, 2006). The correct DNA insert was then verified by sequencing. The starting chimera cloned in a pQE30 vector was used as a template for subsequent deletion of codons within the linker region. The choice of linker length was guided by *in silico* modeling (§ 4.1) In order to perform gold-labeling experiments by EM, a His_6-tag was inserted by RF cloning at the N-terminus of the best Impβ-GS construc, IG07 (§ 2.2.1).

4.3 PROTEIN EXPRESSION AND PURIFICATION

4.3.1 RECOMBINANT EXPRESSION AND PURIFICATION OF TEMPLATES AND OF GLOBULAR PROTEIN FUSIONS

The pET28c LS (provided by Prof. Subramanian Karthikeyan, Institute of Microbial Technology, Chandigarh, India) and the pETM-11* bearing the genes of the designed chimeras and of the free templates (described in § 4.2) all code for the corresponding

proteins, with a C or N terminal Histidine-tag. In order to allow their simultaneous and straightforward production in *E.coli* we developed a common expression and purification procedure, described as follows.

The expression vectors were used to transform *E.coli* BL21 DE3 cells. The transformed cells were grown in lysogeny broth (LB) supplemented with 50 µg/mL kanamycin at 37°C and protein expression was induced with 1 mM IPTG overnight (O/N) at 20˚C. The cells were centrifuged at 5000 g for 20 min, resuspended in lysis buffer (50 mM Tris pH 8, 200 mM NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol) in the presence of lysozyme (1g/L) and protease inhibitors, lysed by sonication at 4˚C and then centrifuged at 40,000 g for 20 min. The recovered soluble fraction was loaded onto a column filled with Ni-NTA resin (500 µL/L culture) to perform affinity chromatography purification. Washes were performed with lysis buffer and proteins were eluted with the same buffer containing 500 mM imidazole. All the buffers used for Mbp fusion proteins were supplemented with 10 mM maltose to stabilize the conformation of the target. All the buffer solutions used for GS fusions were supplemented with 10 mM $MgCl_2$ to stabilize the dodecameric state of GS (Eisenberg et al., 2000). The eluted fractions were collected, concentrated by ultrafiltration and purified by size exclution chromatography (SEC) on a Superose 6, 10/300 GL column (GE Healthcare) pre-equilibrated with 50 mM TRIS pH 8, 150 mM NaCl, using an AKTA Prime system (Amersham Biosciences). The purity and yield of protein at each step of the procedure were gauged by discontinuous polyacrylamide gel electrophoresis in denaturing and reducing conditions (SDS-PAGE) (Laemmli, 1970).

4.3.2 IMPORTIN-β – GS EXPRESSION AND PURIFICATION IN *E.COLI*

After transformation of *E. coli* strain Tg1 (Lucigen), fusion proteins were expressed in LB medium supplemented with 100 µg/ml ampicillin at 20°C for 16 hours. Cells were harvested at 7000 g for 20 minutes, resuspended in lysis buffer (20 mM TRIS pH 8, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol) in the presence of 1g/L DNase, 1g/L lysozyme and protease inhibitors, lysed by sonication at 4˚C and then centrifuged at 40,000 g for 20 minutes. To purify the untagged Impβ-GS chimeras we referred to the Impβ purification protocol reported in literature (Weis et al., 1996). The first step of the Impβ purification consists of an affinity chromatography on a resin functionalized with Impα, a well known Impβ protein partner (Cingolani et al., 1999). In principle, only the

well folded Impβ in the soluble fraction are able to bind the IBB Impα domain, via tight and specific electrostatic interactions. Subsequently, Impβ is eluted from the resin by disrupting the complex with Impα, increasing the ionic strength of the solution. This protocol allows one to obtain highly homogeneous and active Impβ sample. In order to maximize the conformational homogeneity of Impβ within the fusion to GS, we adopted the same purification protocol reported in literature for Impβ. First, we loaded the soluble fraction at 4˚C onto a CNBr-activated Sepharose™ 4B (GE Healthcare) resin, pre-functionalized with Impα according to the manufacturer's procedure. Subsequently, we eluted the bound Impβ-GS fusion protein, with the lysis buffer supplemented with 500 mM $MgCl_2$. The eluted fractions were collected, concentrated by ultrafiltration and purified by SEC on a Superose 6, 10/300 GL column (GE Healthcare) pre-equilibrated with 50 mM TRIS pH 8, 150 mM NaCl, 5mM $MgCl_2$, using an AKTA Prime system (Amersham Biosciences). The purity and yield of protein at each step of the procedure were assessed by SDS-PAGE (Laemmli, 1970).

4.4 BIOPHYSICAL CHARACTERIZATION OF CHIMERIC CONSTRUCTS

4.4.1 SIZE EXCLUSION CHROMATOGRAPHY

After affinity purification all the chimeras were purified by Size exclution chromatography (SEC). This technique allows separation of a mixture of species according to size and shape and yields an approximate estimation of the hydrodynamic radius (R_h). In SEC the analyte is sieved through a porous inert stationary phase (column) under the constant flow of a mobile phase. A detector (for proteins usually a 280 nm spectophotometer) at the end of the column detects the eluting molecules producing a chromatogram (absorbance as a function of the time or eluted volume). The larger the molecule-pore size ratio the less the molecules are retained by the column, causing an elution delay, thereby permitting the separation of different sized molecules in the loaded mixture. At the limit, species larger than the maximal pore volume (e.g., aggregates) elute at a volume excluded by the matrix, named the void volume (V_0), which is characteristic of each column. After this volume the separation range (sieving effect) starts. A linear relationship exists between the elution volume and the logarithm of the hydrodynamic radius, which in the case of globular proteins is correlated with molecular

weight (MW). Hence, by calibrating the column with proteins of known hydrodynamic radius, the elution volume can be used to estimate the hydrodynamic radius (hence MW) of the analyte. Here, SEC was used to investigate the change in hydrodynamic radius among chimeras having nearly identical masses and different linker lengths. The analysis was performed using the column with the largest pore size commercially available (Superose 6, 10/300 GL; GE Healthcare) pre-equilibrated with 50 mM TRIS pH 8, 150 mM NaCl, 10mM $MgCl_2$, using an AKTA Prime system (Amersham Biosciences). The protein elution profile was monitored at 280 nm and 260 nm to allow for the detection of nucleic acid contamination. Calibration was performed using available commercial standard proteins. The hydrodynamic radius of the proteins was estimated from its elution volume by interpolation of the calibration curve (Figure 4.3). Despite the broad fractionation range (5 MDa-5 kDa) of the column reported by manifacturers, it was impossible to obtain an accurate R_h estimate for the Impβ-GS and Mbp-E2 constructs (expected particle sizes of 1.8 and 4.3 MDa, respectively), since all the fusions eluted at the void volume, outside the fractionation range. In contrast, reliable estimates were obtained for the Mbp-GS, GFP-GS and Kpr-GS of ~ 1.2 MDa in MW.

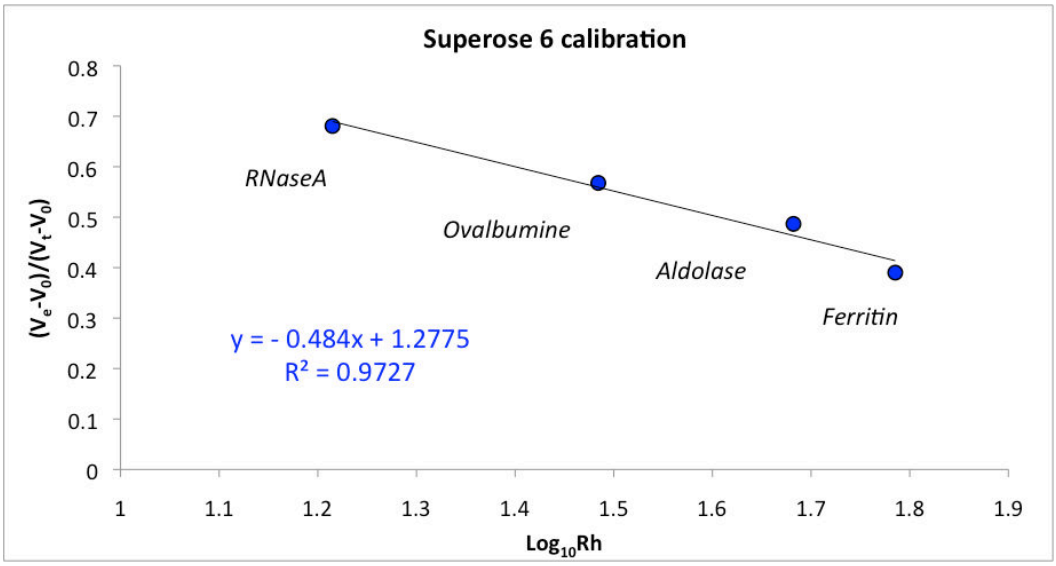


Figure 4.3: Calibration curve of Superose 6, 10/300 GL (GE Healthcare) using standard proteins with known molecular weight and hydrodynamic radius (R_h).

4.4.2 NATIVE POLYACRYLAMIDE GEL ELECTROPHORESIS

The quaternary structure and homogeneity of chimeric constructs having different linker lengths was gauged by native polyacrylamide gel electrophoresis (Native PAGE). Here, a differential migration under an electric field through a sieving matrix is exploited to identify species with different mass/charge ratios in a mixture (Wittig and Schagger, 2005). 5 µg of each protein solution were loaded on a 4% continuous polyacrylamide gel and a voltage of 150 V was applied for 7 h at 4˚C using 150 mM TRIS/glycine pH 8.8 as running buffer. After migration, proteins were visualized by Coomassie blue staining. The native PAGE experiments were unsuccessful for both Impβ and E2 constructs, since the samples remained stuck at the bottom of the wells and did not enter the gel. However, native PAGE analysis was crucial for selecting homogeneous Mbp-GS constructs. Proteins exibiting a single band were presumed to be composed of a single oligomeric species (or more than one species in rapid equilibrium). On the contrary, proteins showing multiple bands were judged to be composed of multiple oligomerization states, and hence were likely to be incorrectly assembled and unsuitable for EM analysis.

4.4.3 THERMAL SHIFT ASSAY

A thermal shift assay (TSA) or Differential Scanning Fluorimetry (DSF) consists of the thermal denaturation of a protein in the presence of a fluorescent dye whose fluorescence intensity is quenched by water and is enhanced when in a hydrophobic environment. As the protein unfolds with increasing temperature, the hydrophobic core becomes more accessible and the dye binds to it, generating a thermal denaturation profile, whose midpoint corresponds to the melting temperature (T_m) of the molecule. For globular, single-domain proteins, denaturation is highly cooperative and the melting curve displays a single transition. DSF is commonly used as a high-throughput technique to select compact and homogeneous proteins for crystallization studies in different buffer conditions or to study ligand binding (Ericsson et al., 2006; Niesen et al., 2007). For oligomeric or multi-domain proteins there may be multiple thermal transitions, and comparison between TSA curves can be used to assess formation and relative compactness of particles, suggesting good leads for EM analysis. Here, the technique was used for two purposes: optimizing buffer conditions for Impβ-GS constructs and investigating the compactness of Mbp chimeras having different linker lengths in the same buffer. Assays were carried out in white 96-well plates in an RT-PCR machine (Bio

Rad CFX96). Each well (20 µl) contained SyproOrange Dye (Sigma-Aldrich) diluted 5000x and 5 µM final protein concentration in the presence of a 100 mM buffer. The plate temperature was ramped from 20 to 99°C with a 0.5˚C temperature increment. SyproOrange dye (Sigma-Aldrich) was excited at 483 nm and fluorescence intensity detected at 568 nm. T_m values were calculated as the temperature at which the first derivative of thermograms (−dF/dT) displayed a minimum using the integrated software Biorad CFX-manager 2.1.

4.4.4 DYNAMIC LIGHT SCATTERING

Dynamic light scattering (DLS) is a non-invasive technique for determining protein size distribution in solution by exploiting the scattering of visible light (Wilson, 2003). According to Rayleigh scattering theory, if the scattering objects are much smaller than the wavelength of the incident beam ($d \leq \lambda/10$) then the scattering intensity is proportional to the sixth power of the object's diameter (Lorber et al., 2012). DLS experiments were carried out in a DynaPro Nanostar machine (Wyatt) equipped with a thermostat set at 25 ˚C. Protein solutions at a concentration of approximately 40 µM were placed in 50 µL plastic cuvettes. A laser beam in the visible range (He-Ne laser at λ = 632.8 nm) illuminated the solution and a detector placed at 90˚ measured the time-dependent (every 5 µs) fluctuations in scattering intensity due to particles undergoing Brownian motion. An autocorrelation function of the signal was calculated over time as $G(t) = e^{-\Gamma t}$, where Γ is proportional to the diffusion coefficient D. The autocorrelation is maximal at time zero and tends to zero as t increases. Large particles move slowly and exhibit a delayed decay (lower D), while small particles move more rapidly, causing the signal to become uncorrelated more quickly (high D). The calculated diffusion coefficient allows one to derive the Stokes radius or hydrodynamic radius R_h (the radius of a hard sphere that diffuses at the same rate as the protein). In the ideal noise-free, single-species (monodisperse) case the autocorrelation curve decays exponentially and can be transformed to yield a size distribution function (I vs R_h), where a single radius is detected. In practice, the sample is always polydisperse: the correlation curve must be deconvoluted according to cumulant analysis (Koppel, 1972) and the size distribution function consists of multiple Gaussian peaks with a standard deviation, named the polydispersity index (PD). The latter indicates how broad is the size distribution in solution and represents an indication of heterogeneity and interface stability of

multimeric proteins (Marion et al., 2010; Shiba et al., 2010) that can be used prior to structural analysis (Wilson, 2003). In this work the radius and polydispersity index were calculated from the DLS data using the integrated Dynamics 7 software (Wyatt technology), and used to assess the compactness of chimeric proteins as a function of linker length.

4.4.5 FLUORESCENCE POLARIZATION BINDING ASSAY

Fluorescence polarization (FP) spectroscopy is based on the following phenomenon: when a small fluorescent molecule is excited with polarized light of a certain wavelength, the light emitted is usually not polarized. This happens because the molecules rotate rapidly in solution during the life time of fluorescence (i.e. the time elapsing between excitation and emission). However, if the fluorescent molecule (ligand) is bound to a large molecule the rotation of the ligand is slowed, so that a greater proportion of the light emitted is in the same plane of polarization of the excitation light. Therefore, free and bound ligand have an intrinsic low and high value of polarization, respectively. The measured polarization allows one to estimate the fraction of ligand bound. As FP is often linearly proportional to the fraction of bound ligand, we can estimate the dissociation constant K_d, as the concentration of protein which yields a half-maximal change in FP signal (Jameson and Seifried, 1999). The technique of FP, in this study, was used to calculate the apparent affinity constants between the free Impβ and Impβ in fusion with GS with a fluorescently labelled NLS peptide from the HIV Tat protein (Tamra-TKALGISYGRKKRRQRRRA). In the assay, the peptide concentration is kept constant at 20 μM and the protein concentration is progressively increased (2 nM – 20 μM). The samples were loaded in a multi-well plate (for a total of 80 μL of solution for each dilution) and the values of FP read on a Synergy HT Multi 47 Mode Microplate Reader (BioTek). We then plotted the value of fluorescence anisotropy versus the concentration of protein. Both sets of data points reached a plateau corresponding to saturation and could be well fitted assuming a 1:1 binding interaction.

4.5 ELECTRON MICROSCOPY

4.5.1 SAMPLE PREPARATION AND DATA COLLECTION

The preparation of a thin aqueous bio-macromolecular sample resistant to evaporation and tolerant to radiation damage for optimal EM analysis can be achieved mainly through two techniques: negative stain EM and cryoEM.

In **negative stain EM** the molecule of interest is bound to a support (typically carbon) and stained with an electron-dense heavy atom salt solution replacing the solvent surrounding the molecule. Therefore, a negative stain micrograph provides a high contrast, low resolution image of the strongly scattering stain background (black) surrounding the analyte surface (white). This method can be described as generating a footprint of the molecule in the heavy atom salt medium. Due to the interaction of the stain with the molecule and to the surface and variations of the stain meniscus, partial staining and shape artefacts as well as flattening phenomena can occur (Harris, 1991). On the other hand, sample preparation is rather fast and allows detection of small molecules (~100 kDa) at low concentrations (of the order of 50 nM or ~ 0.05 g/L) and EM analysis at RT at relatively low voltages (120 kV down to 60 kV).

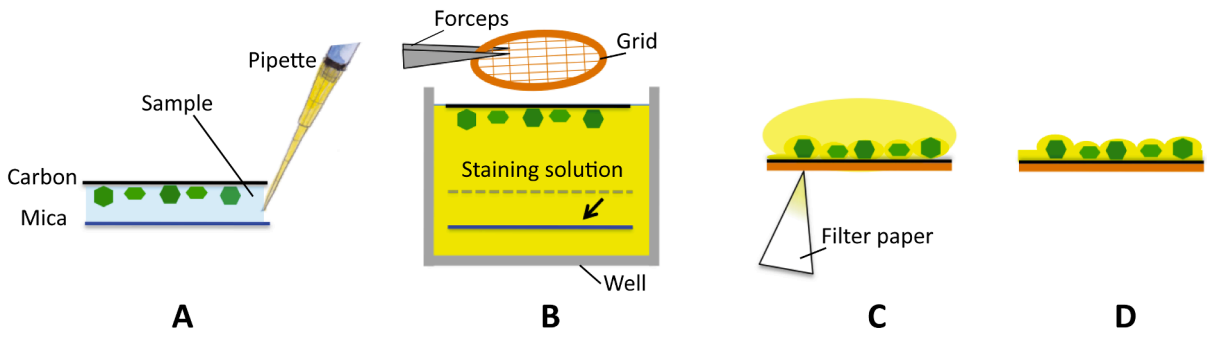


Figure 4.4 Negative stain EM grid preparation by negative staining-carbon technique. A-B): First, a freshly cleaved mica slice is coated with a carbon film produced by resistance evaporation of carbon rods in vacuum. Second, the protein solution (4 µL at 0.01-0.05 g/L) is applied at the interface between the mica-carbon bilayer, causing the absorption of the protein onto carbon. C) The wet bilayer is introduced in a 2% w/v pH 7.2 Sodium Silico Tungstate (SST) solution. The Mica is allowed to detach and fall to the bottom of the well, while the floating carbon absorbed with proteins is fished with a 400 mesh (lines/inch) copper grid support D) The grid is turned over, air-dried on filter paper few minutes and ready to be inserted into the microscope.

Therefore, it provides a useful tool to screen sample quality and study overall quaternary structure or binding of a ligand (Jinek et al., 2014). In this thesis work, negative stain EM was used to investigate the homogeneity and oligomerization state of the recombinantly expressed chimeras, prior to cryoEM analysis. The negative staining-carbon technique used for this purpose is described in Figure 4.4.

Negative stain micrographs were recorded under low-dose conditions with a Philips CM12 microscope at 120 kV, a magnification of 22000X, with a defocus range between -3 and -1.5 µm on a Gatan Orius 1000 CCD camera (Å/pix= 3.24), exposing for 1 s at an electron dose of 30 $e^-/Å^2$. Different staining solutions were tested at the beginning of the EM work for His-tagged GS : Ammonium molybdate 2% w/v pH 7.3, Uranyl acetate 2% w/v pH 4.5 and SST 2% w/v pH 7.2. The latter gave the best compromise between contrast, meniscus level and detection detail of solvent exposed cavities (Figure 4.5). SST was used for all the negative stain EM experiments.

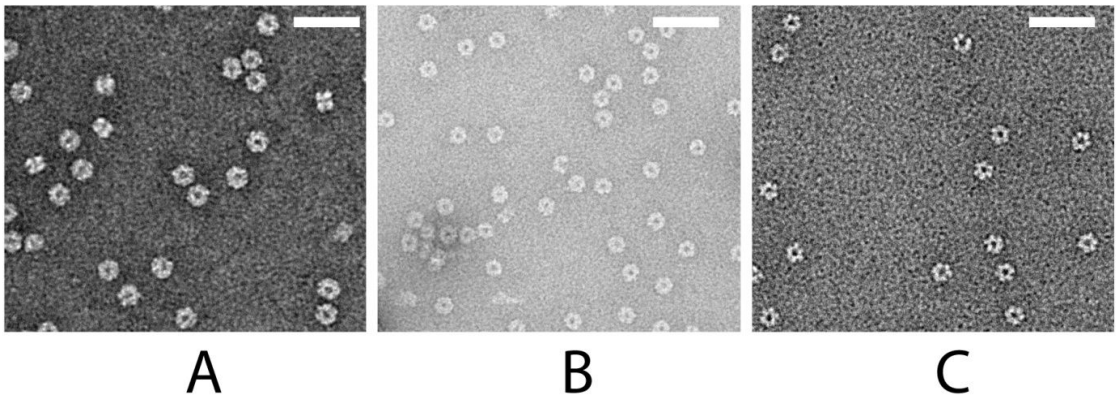


Figure 4.5 Electron micrographs of GS negatively stained with different heavy metal salts solutions A) Uranyl acetate 2% w/v pH 4.5 B) Ammonium molybdate 2% w/v pH 7.3, and C) SST 2% w/v pH 7.2 (The best staining solution). Images recorded on a CM12 operating at 120 kV, magnification 45000x, defocus ~ 3µm. The scale bar corresponds to 50 nm.

The CryoEM method was first developed during the 1980s (Adrian et al., 1984; Dubochet et al., 1988) and consists of flash freezing a thin layer of protein solution (usually in liquid ethane, T = -182 ˚C), thereby generating amorphous ice instead of crystalline ice, which can be damaging to the sample. The low temperature prevents evaporation inside the microscope, and at the same time reduces radiation damage induced by the electron beam. In this condition the molecule ideally adopts random orientations (as in solution at RT) and its image consists of a collection of density

projections of the molecule in its native hydration state, surrounded by the lower density (transparent) buffer. By averaging these noisy individual particle 2D densities (image processing, § 4.5.2) it is possible to reconstruct the proper 3D electron density map of the sample, as described in the next section. In the present work, the symmetric chimeras presenting a high degree of compactness by biophysical analysis and negative staining EM were imaged in native conditions by cryoEM. The general procedure of cryoEM sample preparation is summarized in Figure 4.6.

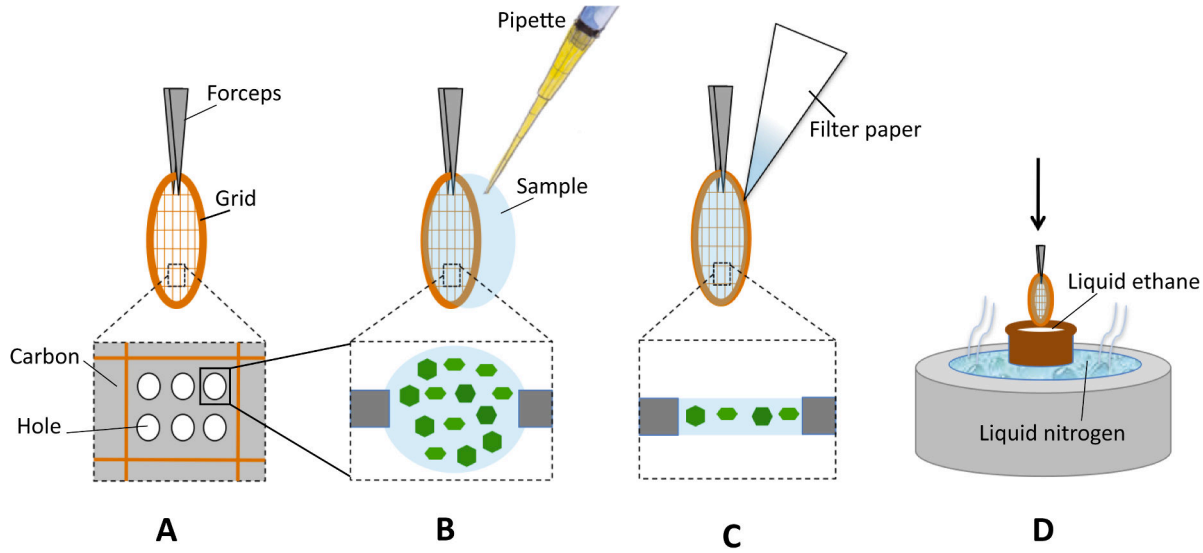


Figure 4.6: CryoEM sample preparation A-C) The protein solution (4 µL at ~ 0.5 g/L) is suspended over holey carbon coated grids (Cu/Rh 400mesh Quantifoil grids), at 100% humidity and 20˚C, held by forceps. The liquid in excess is blotted onto filter paper. D) The sample is then plunged in liquid ethane (-182˚C) cooled by liquid nitrogen, for a rapid heat transfer. Cooling by plunging into liquid ethane is much faster than plunging directly into liquid nitrogen because liquid ethane is used close to its freezing point rather than at its boiling point, so it does not evaporate to produce an insulating gas layer (Orlova and Saibil, 2011). The suspension over the holes improves the contrast of the image by reducing the electron scattering caused by the carbon film and also avoids possible distortions caused by interaction with the carbon film. The sample once frozen was kept at close to liquid nitrogen temperature (-196˚C) over the whole experiment. In practice the freezing procedure was pursued using an automated freezing system (Vitrobot –FEI) (Iancu et al., 2006).

Grids frozen in different conditions (blotting time, temperature, blotting force, draining time, etc.) were imaged on a Philips CM200 microscope operating at 200 kV at a magnification of 30000X, by recording the micrographs on a Gatan slow scan 1K CCD camera using quite a high defocus (~4 µm), to increase contrast and better detect the

frozen particles. This way, we could identify the best freezing conditions in terms of ice thickness, presence of a considerable number of non-empty holes and concentration for the three chimera samples Mae5, MagΔ8 and MaeΔ5. The grids frozen under optimal conditions were analyzed with a FEI-POLARA microscope operating at 300 kV, at a magnification of 39000X, collecting the images on photographic films. The samples were exposed for 1 s for a total dose of 15-20 $e^-/Å^2$. Selected negatives were then digitalized on a Zeiss scanner (Photoscan TD) at a step size of 7 µm, giving a digital micrograph pixel size of 1.8 Å. All measurements were performed in collaboration with Dr. Guy Schoehn, at the IBS EM facility.

4.5.2 IMAGE PROCESSING

A brief overview of image formation by TEM was given in § 1.2 and § 4.5.1. The aim of image processing is to average the inherently noisy projected density of single particles so as to enhance common features and remove the random noise, ultimately allowing the achievement of an accurate 3D electron density map of the molecule of interest (Orlova and Saibil, 2011; Saibil, 2000; van Heel et al., 2000). In the present work, single particle cryoEM image processing was performed to reconstruct 3D maps of the symmetric chimeras and consisted of mainly three steps :

i) Preprocessing: restoring of image information due to distortions of the optical system, particle picking, and normalization

ii) Achievement of a reference model: sorting and combination of similar projections to obtain a first *ab initio* 3D model

iii) Model refinement

i) Preprocessing. As introduced in § 1.2, raw electron micrographs of native biological thin samples (embedded in vitreous water) are extremely noisy because of the inherently weak scattering power of low molecular weight components and the low dose conditions imposed by radiation sensitivity. In other words, cryo biological specimen images present both low amplitude (<10%) and low phase contrast, and hence are nearly electron transparent (weak phase objects). However, by applying a certain defocus (i.e. virtually moving the object plane) the electron path and interference can be changed to enhance the resultant phase contrast. The transfer of contrast by the optical system as a function

of the spatial frequency (resolution) is described by the contrast transfer function (CTF). CTF is a sine-like function whose zeros depend on the defocus, spherical aberration of the microscope and wavelength (voltage) used (Figure 4.7). CTF modulates the sign and intensity of the Fourier components of the image and, by enhancing lower spatial frequencies, allows the better detection of particles on the micrograph. However, at the same time the CTF causes loss of information at its zeros (sign inversion points). Therefore, CTF determination and correction of EM images is necessary to restore image information prior to further image processing. Moreover, the lack of information at its zeros have to be compensated by merging several micrographs acquired with different defoci. On the power spectrum of each micrograph we can visually detect Thon rings (determined by the CTF), whose intensity gradually decreases from low to high spatial frequencies (Figure 4.7A). This dampening is caused by partial coherence, chromatic aberrations, stage or beam-induced motions, etc. and can be approximated as a Gaussian decay (envelope function; Figure 4.7B).

In this thesis work, the CTF of the digitalized micrographs was determined by fitting the experimental oscillations (Thon rings) with a model CTF, using the CTFFIND3 software (Mindell and Grigorieff, 2003). By visually comparing the experimental oscillations with calculated ones (Figure 4.7C-D), we rejected micrographs presenting either strong astigmatism (different defocus value in the plane, with elliptical Thon rings), or drift (missing of Thon rings in one direction or their total absence). In cryoEM micrographs acquired at low defocus value (-1.5 µm) CTF oscillations were visible until ~ 10 Å. After CTF determination, we corrected the images with the *bctf* routine from the BSOFT package, flipping the phases in every second Thon ring (Heymann, 2001).

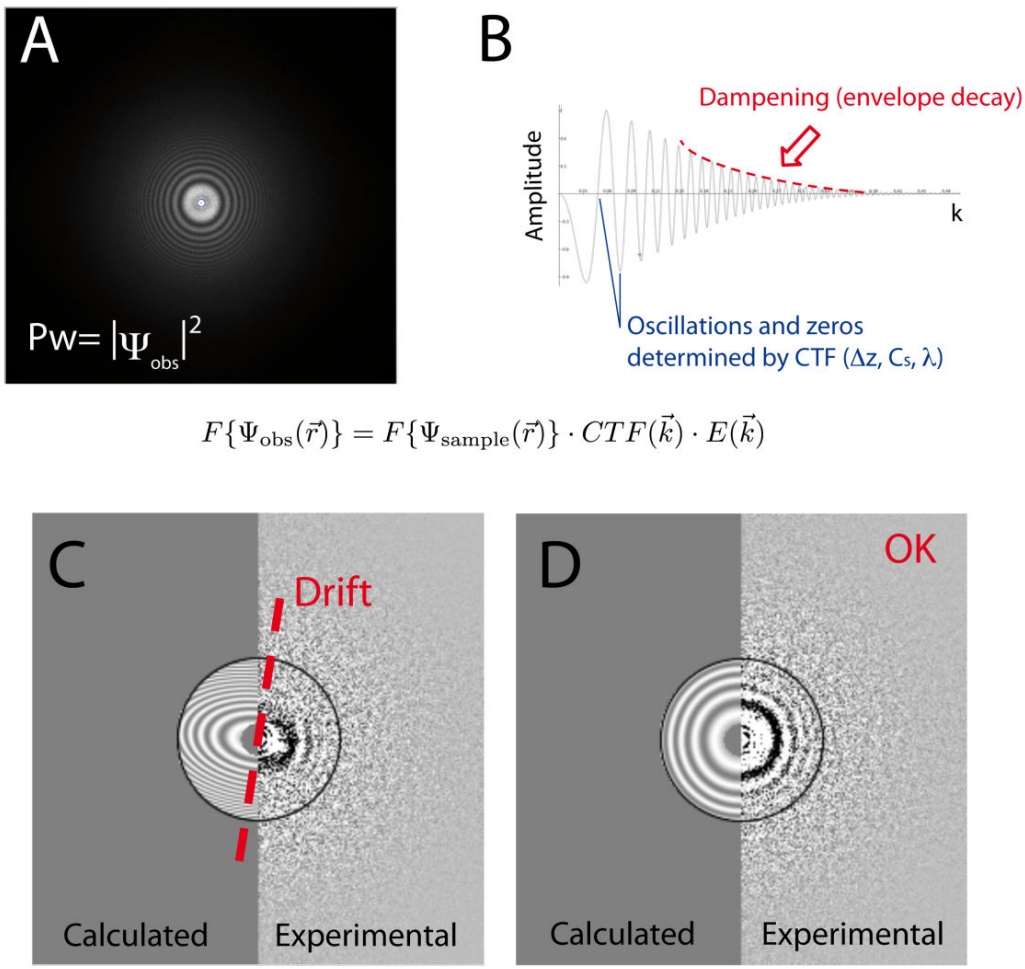


Figure 4.7 CTF determination from the power spectrum of an image. A) The observed Fourier transform of the image $F(\Psi_{obs})$ corresponds to the Fourier transform of the projection of the sample $F(\Psi_{sample})$ multiplied by the contrast transfer function (CTF) and the Envelope function. The power spectrum of a recorded raw image is the intensity (squared amplitude) of the Fourier components of the image and can be used to determine the CTF oscillations prior to its correction. B) The rotationally averaged 1D spectrum (Amplitude vs spatial frequency k) is an oscillating function whose zero values depend on defocus Δz, wavelength λ, and spherical aberration of the microscope C_s. This is dampened by the Gaussian decay E, due to partial radiation coherence, radiation damage and other instabilities of the optical system. In CTFFIND3, by fitting the Thon ring oscillations with a model CTF (calculated till 10 Å, indicated by the black circle) it is possible to establish the optical parameters relative to the micrographs. Moreover, by comparing the calculated and experimental spectrum it is possible to select good micrographs for image analysis C) CTFFIND3 output comparison reveals a strong drift in the direction of the dotted line, therefore the image is not usable for image processing. D) An example of a good match between calculated and experimental regular oscillations, which allow an accurate determination of optical parameters.

After CTF correction, micrographs were binned twice (pixel size 3.6 Å) and filtered to enhance contrast. In particular, a smoothed band pass filter was applied between 300 Å (about the particle diameter) and 15 Å. Particle centres were selected by hand in *boxer* (EMAN) (Ludtke, 2010) and the dataset (stack of images) normalized (to average = 0 and standard deviation 0.2) in IMAGIC (van Heel et al., 1996). By visual inspection we cleaned up the dataset of bad looking particles (dust, ice or aggregates). In this step the presence of the central template in dodecameric chimeras dramatically helped the task of particle detection and centring.

ii) Achievement of a reference model. After preprocessing we intended to determine a set of three Euler angles defining the orientation and two translational parameters defining the position of every particle. This first assignment allows one to obtain a starting 3D reference model, describing the rough shape of the macromolecule. Two methods were used to generate reference models: angular reconstitution implemented in IMAGIC (van Heel et al., 2000; van Heel et al., 1996) and the use of symmetry adapted functions implemented by one of our collaborators, Dr. Leandro Estrozi on the webserver RIco http://rico.ibs.fr/RIcoWebServer/ (Estrozi and Navaza, 2010; Navaza, 2003).

Angular reconstitution is a reference-free method implemented in IMAGIC (van Heel et al., 2000; van Heel et al., 1996) whose main steps are summarized in Figure 4.8 and described as follows. In the first cycle (i=1), we iteratively centred the images (previously normalised, binned twice and low pass filtered) by translational alignment to the rotationally averaged total sum of the image dataset. Then, by multivariate statistical analysis (MSA) we decomposed the dataset in terms of a limited set of independent eigenimages, describing the local density variations in the image dataset. The eigenimages containing small variations were attributed to noise and excluded. Subsequently, those “main components” were used to sort the dataset in N groups (classes) containing 10-20 images with similar features. The principle of reference-free classification is that images with similar features correspond to similar projections of the molecular density (along similar orientation defined by Euler angles). Therefore, by averaging the images belonging to the same class, the typical views of the molecule will be apparent with higher contrast with respect to individual images. The best defined class averages, as judged by overall contrast and detection of side and top views of GS, were

extracted to create a group of reference class averages. To each one of these, a set of three Euler angles defining their orientation can be attributed according to the common lines method (Van Heel, 1987). In fact, each pair of 2D projections of a 3D object has at least one 1D (line) projection in common (projection theorem). As a consequence, for an asymmetric object three projections are sufficient to define the relative orientation in space and to derive a 3D map. The presence of symmetry (in our case D6) provides many more constraints and results in multiple common lines. These allow more accurate determination of the projection directions (sections in Fourier space) and of the 3D reconstruction of the object (Orlova and Saibil, 2011). Using the *euler* routine in IMAGIC we assigned orientations to the first set of reference class averages, with relative errors. Subsequently, with the *true* routine we obtained a first 3D model. The quality of the angular assignment was gauged on the similarity between the projection of the model and the reference class averages. A few low quality class averages were rejected and new ones were incorporated in the reference set (*anchorset* routine in IMAGIC), to improve the features of the resulting map. This procedure was iterated until no further improvement of the model was observed. Subsequently, the model was re-projected according to equispaced directions (every 10˚) and raw images realigned to the model reprojections by multi reference alignment (i>1). Classification was then repeated as previously, and class quality was improved, since the original images had already been aligned to the model projections. In turn, this implied an improvement of the output model quality. The angular reconstitution procedure was reiterated until convergence to a “stable” model.

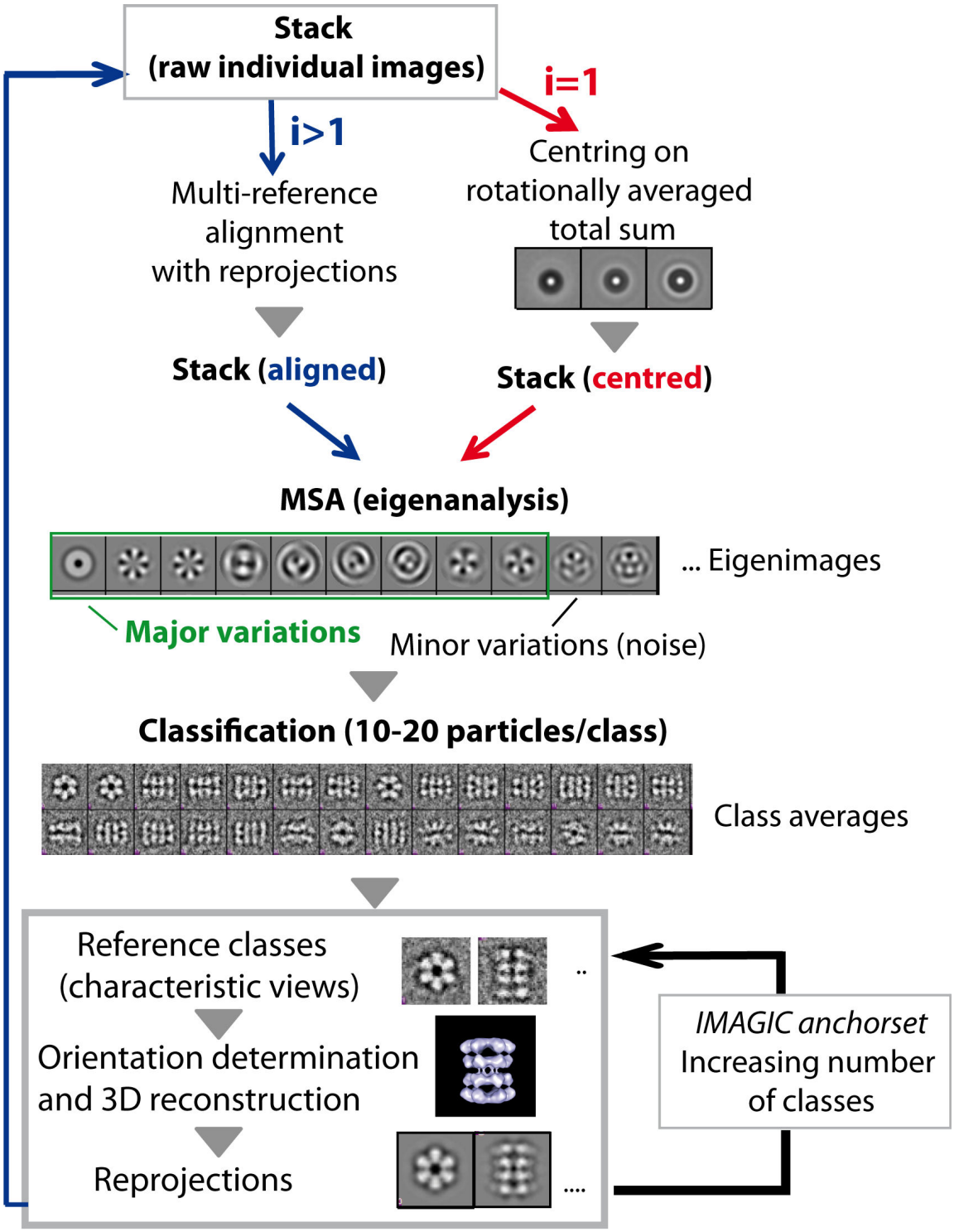


Figure 4.8: Computational procedure for angular reconstitution implemented in IMAGIC and illustrated for the reconstruction of MagΔ5 *ab initio* reference model. Raw images (stack) are aligned and classified based on local variations. Among the class averages, typical views of the object (easily identified due to the presence of the GS template) are used to construct a first model. Model reliability is gauged by comparing re-projections to the reference classes (In the figure as examples a top and a side view are displayed). The model is improved by adding new classes (*Anchorset* routine) till no further changes. The first model is re-projected, used to realign the stack of images and to reiterate the angular reconstitution procedure.

The RIco webserver http://rico.ibs.fr/RIcoWebServer/ developed by Dr. Leandro F. Estrozi allows the reconstruction of a low resolution *ab initio* model from a few raw individual particle images of a symmetric molecule with low computational cost. The presence of point group symmetry (D6 and I) in our chimeras implies that several positions in the 3D object are related by symmetry operations. Therefore their 3D density can be represented in terms of symmetry adapted functions (SAF), i.e. appropriate combinations of spherical harmonics that are invariant with respect to the specific point group symmetries. For MagΔ5 and MaeΔ5 ~ 50 and 10 views were supplied to the server, by specifying the radius and either D6 or I symmetry, respectively. The automated protocol of RIco consists of assigning an orientation to each particle and calculating a 3D reconstruction. Subsequently, the 2D projection of the reconstruction is compared to the image through a correlation coefficient (CC):

$$CC(X,Y) = \frac{\sum_{i=1}^{N\times N} x_i y_i}{\sqrt{\sum_{i=1}^{N\times N} x_i^2 \sum_{i=1}^{N\times N} y_i^2}}$$

where x_i and y_i are the pixel density values for projections and individual images, respectively, and N is the dimension of the pixels matrix.

The view that gives a maximum CC is assigned to the image. In this single-image search the resolution is limited to about 10% of the particle diameter. The image that gives the most reliable view assignment in terms of its CC profile (peak value, uniqueness and contrast) is included as a fixed contribution in another exhaustive re-assessment of views of the remaining images. The last step may be repeated, testing other images as fixed contributions. Eventually, views and centres are checked for consistency and refined.

In the present thesis work we analysed one negative stain dataset (Mag0) and three cryo datasets MagΔ8 and MagΔ5 (D6 symmetry) and MaeΔ5 (icosahedral symmetry). Mag0 and MaeΔ5 initial references were obtained with RIco, whereas MagΔ8 was obtained by angular reconstitution. For MagΔ5 we produced two independent reference models, one obtained with RIco and one by angular reconstitution.

iii) Model refinement. For all datasets the initial volumes were refined by projection matching. The **Projection matching** method is based on finding the projection directions of the individual molecular images by comparing them to the reprojections of a 3D model (Penczek, 2012). The similarity of the images is measured by the value of the cross correlation (CC) which have to be maximized.

The drawback of this approach is that it requires prior knowledge of a molecular model that could be unavailable or incorrect for a completely unknown object, especially an asymmetric one. The choice of a good starting model is decisive for determining the quality and reliability of the final model. Furthermore, it was demonstrated that this approach could lead to a molecular model that reproduces the original reference, a phenomenon known as model-induced bias (Henderson, 2013b).

In the case of Mbp-GS dodecamers, the presence of symmetry, of the template structure as a "fiducial marker" and of the *in silico* structural models, gives a higher degree of control on the reliability of the reconstruction. The reference model (from RIco and/or IMAGIC) was band pass filtered between 300 Å and 50 Å, and its reprojections along equispaced directions were used to align the individual raw images (classification by alignment). The models were initially filtered to perform a rough centring and alignment of the image stack, to avoid artefacts derived by alignment to noise. In order to sample the Fourier space for a maximum theoretical resolution of 3.6 Å (twice the pixel size), we set the angular increment of reprojections to 1˚. After alignment and rejection of 20% of the particles by CC, a backprojection procedure in Fourier space was applied to the class-averages to reconstruct a 3D volume (Penczek, 2012). The whole procedure was reiterated until convergence of the alignment parameters. The same projection matching procedure was used to refine the second model derived from the angular reconstitution method. We tried to adopt larger angular increment values to reduce the computational cost of the projection matching procedure. For all datasets, the largest increment causing no change in map quality and resolution was 3˚ (corresponding to a maximum resolution of 10 Å). The projection matching method was performed by using SPIDER software (Frank et al., 1996), as summarized in Figure 4.9.

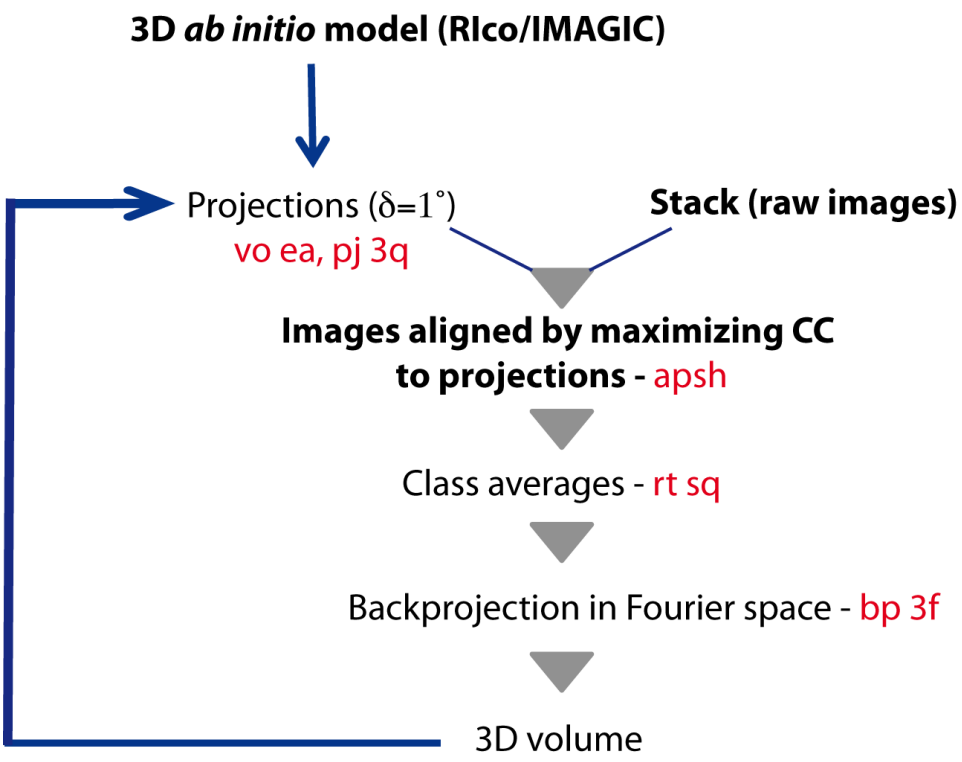


Figure 4.9 Projection matching scheme diagram. In purple are indicated the SPIDER program routines used to carry out the different steps. *vo ea* generates equi-spaced direction projections, along which the reference model is reprojected with *pj 3q*. The CC-based alignment is performed with *ap sh* and applied to the images with *rt sq*. Finally, class averages of images corresponding to the same view are used to reconstruct a 3D volume in Fourier space with *bp 3f*. This model is reprojected and the procedure reiterated by monitoring the two shifts and three Euler angles until parameter convergence.

As an example, the whole MagΔ5 image processing procedure, using two independent reference models is summarized in the diagram in Figure 4.10.

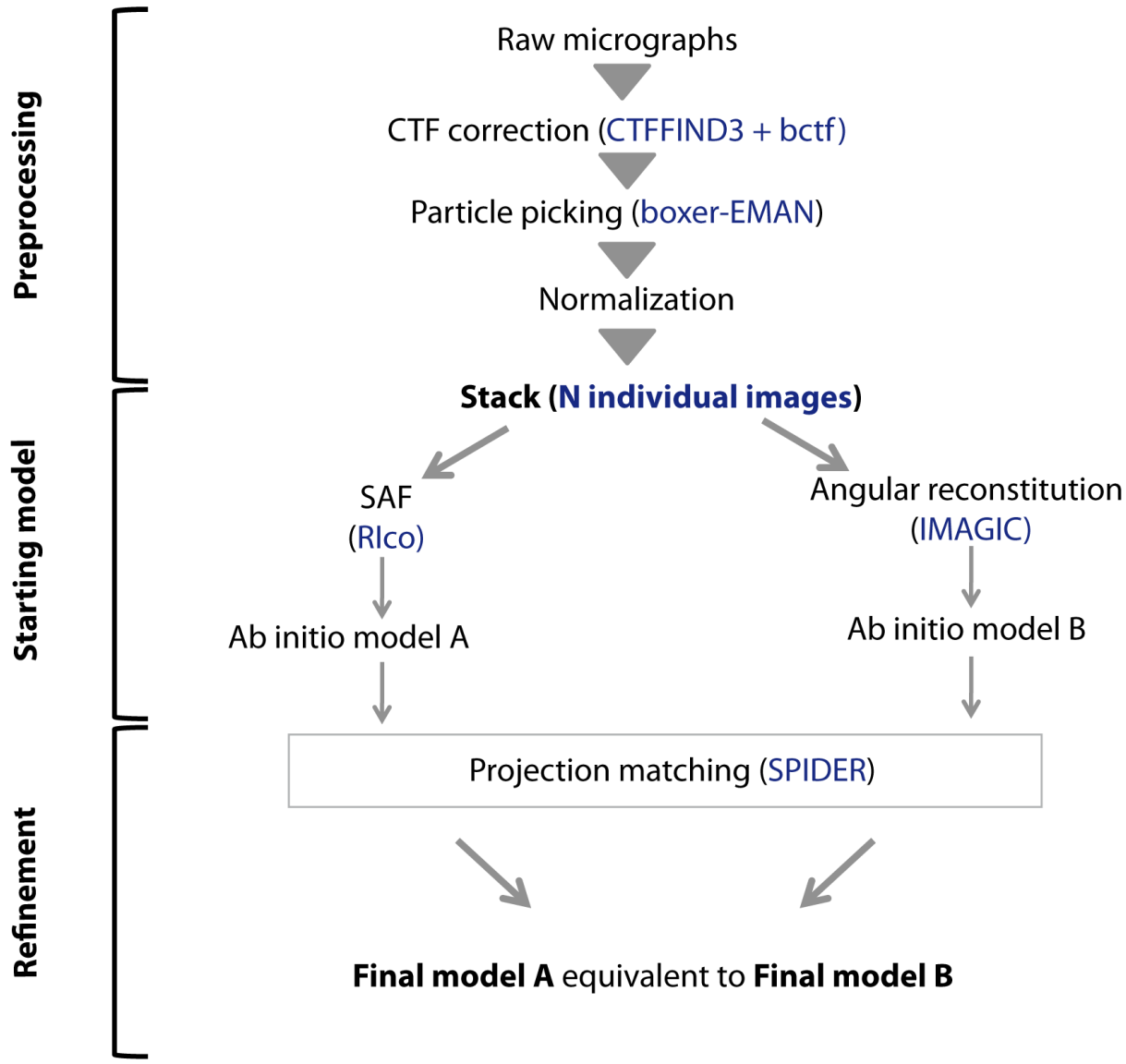


Figure 4.10 : Image processing steps of MagΔ5. After preprocessing the stack of individual images was used to reconstruct two independent 3D models (A and B) by the symmetry adapted function (SAF) method implemented in RIco and angular reconstitution implemented in IMAGIC. The refinement of the maps was performed by projection matching in SPIDER, by converging to the same MagΔ5 volume, as judged by visual appearance and Fourier Shell Correlation estimation (§ 4.5.3). All the procedure was pursued by applying D6 symmetry.

4.5.3 Resolution estimation and map visualization

At the end of the reconstruction procedure, in order to assess the resolution and internal coherence of the map, the dataset was divided into two halves. From these two separate maps (1 and 2) were built and the normalized cross correlation calculated in spatial frequency shells (Fourier Shell Correlation, FSC):

$$FSC = \frac{\sum_{k,\Delta k} F_1(k) F_2^*(k)}{\sqrt{\sum_{k,\Delta k} |F_1(k)|^2 \sum_{k,\Delta k} |F_2(k)|^2}}$$

where F1 is the complex structure factor for map 1, F2* is the complex conjugate of the structure factor for map 2, k is the special frequency and and Δk is the spatial frequency shell. In this work the 0.5 criterion has been used to estimate resolution, although the resolvability of structural elements is a better guide to gauge the effective and local map resolution (van Heel and Schatz, 2005). In theory the gold standard of FSC should be used to estimate the resolution from two independent reconstructions obtained by splitting the dataset directly at the beginning of the image processing (Scheres and Chen, 2012). However, in our case this procedure could not be applied due to the limited size of the dataset. This calculation represents an average over the entire reconstructed volume. In practice, however, substantial local variations in resolution may occur. Blocres, a software program in the Bsoft package, was used to estimate the local resolution of template and target (Cardone et al., 2013). In this procedure, the experimental volumes calculated from the two halves of a dataset are further divided into small subvolumes and the FSC at 0.5 determined for each of them. The size of the subvolumes was set to 20 voxels, following the indications provided by the authors (Cardone et al., 2013). The original map was coloured using the blocres output FSC local values, with the *surface color* routine in Chimera (Pettersen et al., 2004). In this work, the majority of map visualizations and figures as well as the fitting of atomic models was performed with Chimera (REF). The visualization of class averages and projections was carried out with bshow and EMAN (Heymann, 2001; Ludtke, 2010). Micrograph gray scales were adjusted *ad hoc* in the figures present in the text for better visualization.

ACKNOWLEDGEMENTS

First of all, I thank the *Nanoscience fondation*, that gave me the financial support and the opportunity to carry out the present PhD project.

Second, I would like to thank the members of the Thesis Jury, Dr Stephane Bressanelli, Dr Irina Gutsche, Prof Elena Orlova, and Prof Anthony Watts, for reading this manuscript and evaluating my work.

I thank my supervisor Dr Carlo Petosa, for his kindness, constant advice and patience. Moreover, I thank him for his example of critical thinking, for his help in the scientific writing and to have given me freedom during these years of work. I also thank him for the exciting experience at the ESRF beamlines and for solving the structure of one of my proteins.

I thank my co-supervisor Dr Guy Schoehn, for his crucial help, expertise and guidance in electron microscopy studies and for collecting the majority of EM datasets analyzed in this work.

I thank all the members of the EM facility at the IBS and in particular Daphna Fenel for her help and constant smile, Dr Emmanuelle Neumann for training me on the CM12 and T12 microscopes. Thanks to Dr Maria Bacia, for her exceptional support and patience in teaching me not to be scared about the microscope.

From the UVHCI I thank Dr Leandro Farias Estrozi, Dr Helene Malet and Dr Irina Gutsche for the fruitful support and discussion about image processing.

I thank, of course, all the members of my group and in particular: Florent Bernaudat and Cyril Dian who taught me the basics of the wet lab, Dr Marjolaine Noirclerc-Savoye for her help in the last part of the PhD and for her great humanity, Mizar and Didier who shared with me the PhD troubles and doubts. I thank Dr Dimitrios Skoufias who read the manuscript and gave me important tips to ameliorate it. I owe big thanks to Prof Fabienne Hans, who taught me cloning and French language at the same time, and for her encouragement during these years. She has been for me a model of passion and foresight in the lab life.

Among my friends, I thank my “squire”, Pamela, who shared at more than 1000 miles of distance the funny and tragic moments of the PhD and who has always believed in me. I thank Elena, Claudio, Liza, Angela, Bridgette, Agnès and Winnie for their sympathy and true friendship. I also thank Angelo, for the endless weekends of writing and talking about latex and quantum chemistry. I thank Simone, Martin, Fabrizio, Florian and Jonathan for the amusing nights at the café Bayard and to show me how playing jazz can help a lot in the jam sessions of life and of science. Huge thanks to Tommaso, to be at my side in the creative and tough moments of the scientific work, to always encourage me and to have changed the horizons of my life.

I thank very much my parents for their love and assistance during my stay in Grenoble. Thanks also to my sister Flora, for her support and for cheering me up with Neapolitan humour.

Overall Grenoble has been a fertile laboratory where growing up on the scientific and human side. This PhD experience has enormously expanded my field of view and have helped me to look at my future with more consciousness and courage.

REFERENCES

Abrahams, J.P., Leslie, A.G., Lutter, R., and Walker, J.E. (1994). Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. Nature *370*, 621-628.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Echols, N., Headd, J.J., Hung, L.W., Jain, S., Kapral, G.J., Grosse Kunstleve, R.W.*, et al.* (2011). The Phenix software for automated determination of macromolecular structures. Methods *55*, 94-106.

Adrian, M., Dubochet, J., Lepault, J., and McDowall, A.W. (1984). Cryo-electron microscopy of viruses. Nature *308*, 32-36.

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

Banatao, D.R., Cascio, D., Crowley, C.S., Fleissner, M.R., Tienson, H.L., and Yeates, T.O. (2006). An approach to crystallizing proteins by synthetic symmetrization. Proceedings of the National Academy of Sciences of the United States of America *103*, 16230-16235.

Bethea, H.N., Xu, D., Liu, J., and Pedersen, L.C. (2008). Redirecting the substrate specificity of heparan sulfate 2-O-sulfotransferase by structurally guided mutagenesis. Proceedings of the National Academy of Sciences of the United States of America *105*, 18724-18729.

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.

Bottcher, B., Wynne, S.A., and Crowther, R.A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. Nature *386*, 88-91.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D.C., Joachimiak, A., Horwich, A.L., and Sigler, P.B. (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 A. Nature *371*, 578-586.

Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., Henderson, R., and Grigorieff, N. (2012). Beam-induced motion of vitrified specimen on holey carbon film. Journal of structural biology *177*, 630-637.

Caffrey, M., and Cherezov, V. (2009). Crystallizing membrane proteins using lipidic mesophases. Nature protocols *4*, 706-731.

Cardone, G., Heymann, J.B., and Steven, A.C. (2013). One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. Journal of structural biology *184*, 226-236.

Channon, K., Bromley, E.H., and Woolfson, D.N. (2008). Synthetic biology through biomolecular design and engineering. Current opinion in structural biology *18*, 491-498.

Cherezov, V. (2011). Lipidic cubic phase technologies for membrane protein structural studies. Current opinion in structural biology *21*, 559-566.

Cingolani, G., Petosa, C., Weis, K., and Muller, C.W. (1999). Structure of importin-beta bound to the IBB domain of importin-alpha. Nature *399*, 221-229.

Clare, D.K., Vasishtan, D., Stagg, S., Quispe, J., Farr, G.W., Topf, M., Horwich, A.L., and Saibil, H.R. (2012). ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. Cell *149*, 113-123.

Cooper, D.R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., and Derewenda, Z. (2007). Protein crystallization by surface entropy reduction: optimization of the SER strategy. Acta crystallographica Section D, Biological crystallography *63*, 636-645.

Crowther, R.A., Amos, L.A., Finch, J.T., De Rosier, D.J., and Klug, A. (1970). Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs. Nature *226*, 421-425.

Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. (1985). Structure of the protein subunits in the photosynthetic reaction centre of Rhodopseudomonas viridis at 3A resolution. Nature *318*, 618-624.

Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowall, A.W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. Quarterly reviews of biophysics *21*, 129-228.

Eisenberg, D., Gill, H.S., Pfluegl, G.M., and Rotstein, S.H. (2000). Structure-function relationships of glutamine synthetases. Biochimica et biophysica acta *1477*, 122-145.

Ericsson, U.B., Hallberg, B.M., Detitta, G.T., Dekker, N., and Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. Analytical biochemistry *357*, 289-298.

Estrozi, L.F., and Navaza, J. (2010). Ab initio high-resolution single-particle 3D reconstructions: the symmetry adapted functions way. Journal of structural biology *172*, 253-260.

Faruqi, A.R., and Henderson, R. (2007). Electronic detectors for electron microscopy. Current opinion in structural biology *17*, 549-555.

Fernandez, J.J., Luque, D., Caston, J.R., and Carrascosa, J.L. (2008). Sharpening high resolution information in single particle electron cryomicroscopy. Journal of structural biology *164*, 170-175.

Fischlechner, M., and Donath, E. (2007). Viruses as building blocks for materials and devices. Angew Chem Int Ed Engl *46*, 3184-3193.

Frank, J. (2006). Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State (2nd edition).

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. Journal of structural biology *116*, 190-199.

Fukuhara, N., Fernandez, E., Ebert, J., Conti, E., and Svergun, D. (2004). Conformational variability of nucleo-cytoplasmic transport factors. The Journal of biological chemistry *279*, 2176-2181.

Goldschmidt, L., Cooper, D.R., Derewenda, Z.S., and Eisenberg, D. (2007). Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. Protein science : a publication of the Protein Society *16*, 1569-1576.

Goodsell, D.S., and Olson, A.J. (1993). Soluble proteins: size, shape and function. Trends in biochemical sciences *18*, 65-68.

Gradisar, H., Bozic, S., Doles, T., Vengust, D., Hafner-Bratkovic, I., Mertelj, A., Webb, B., Sali, A., Klavzar, S., and Jerala, R. (2013). Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. Nature chemical biology *9*, 362-366.

Gradisar, H., and Jerala, R. (2011). De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. Journal of peptide science : an official publication of the European Peptide Society *17*, 100-106.

Harris, J.R. (1991). Negative staining-carbon film technique: new cellular and molecular applications. Journal of electron microscopy technique *18*, 269-276.

Hecht, H.J., Sobek, H., Haag, T., Pfeifer, O., and van Pee, K.H. (1994). The metal-ion-free oxidoreductase from Streptomyces aureofaciens has an alpha/beta hydrolase fold. Nature structural biology *1*, 532-537.

Henderson, C.E., Perham, R.N., and Finch, J.T. (1979). Structure and symmetry of B. stearothermophilus pyruvate dehydrogenase multienzyme complex and implications for eucaryote evolution. Cell *17*, 85-93.

Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Quarterly reviews of biophysics *28*, 171-193.

Henderson, R. (2004). Realizing the potential of electron cryo-microscopy. Quarterly reviews of biophysics *37*, 3-13.

Henderson, R. (2013a). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. Proceedings of the National Academy of Sciences of the United States of America *110*, 18037-18041.

Henderson, R. (2013b). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. Proceedings of the National Academy of Sciences of the United States of America.

Heymann, J.B. (2001). Bsoft: image and molecular processing in electron microscopy. Journal of structural biology *133*, 156-169.

Hoffman, D.W., Davies, C., Gerchman, S.E., Kycia, J.H., Porter, S.J., White, S.W., and Ramakrishnan, V. (1994). Crystal structure of prokaryotic ribosomal protein L9: a bi-lobed RNA-binding protein. The EMBO journal *13*, 205-212.

Iancu, C.V., Tivol, W.F., Schooler, J.B., Dias, D.P., Henderson, G.P., Murphy, G.E., Wright, E.R., Li, Z., Yu, Z., Briegel, A.*, et al.* (2006). Electron cryotomography sample preparation using the Vitrobot. Nature protocols *1*, 2813-2819.

Izard, T., Aevarsson, A., Allen, M.D., Westphal, A.H., Perham, R.N., de Kok, A., and Hol, W.G. (1999). Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes. Proceedings of the National Academy of Sciences of the United States of America *96*, 1240-1245.

Jameson, D.M., and Seifried, S.E. (1999). Quantification of protein-protein interactions using fluorescence polarization. Methods *19*, 222-233.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S.*, et al.* (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science *343*, 1247997.

Kabsch, W. (2010). Xds. Acta crystallographica Section D, Biological crystallography *66*, 125-132.

Kappel, C., Zachariae, U., Dolker, N., and Grubmuller, H. (2010). An unusual hydrophobic core confers extreme flexibility to HEAT repeat proteins. Biophysical journal *99*, 1596-1603.

Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E.*, et al.* (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. Nature methods *5*, 53-55.

Kendrew, J.C. (1959). Structure and function in myoglobin and other proteins. Federation proceedings *18*, 740-751.

Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., and Shore, V.C. (1960). Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution. Nature *185*, 422-427.

King, N.P., Bale, J.B., Sheffler, W., McNamara, D.E., Gonen, S., Gonen, T., Yeates, T.O., and Baker, D. (2014). Accurate design of co-assembling multi-component protein nanomaterials. Nature *510*, 103-108.

Koppel, D. (1972). Analysis of macromolecular polydispersity in intensity correlation spectroscopy: the method of cumulants. Journal of Chemical Physics *57*, 4814–4820.

Kratz, P.A., Bottcher, B., and Nassal, M. (1999). Native display of complete foreign protein domains on the surface of hepatitis B virus capsids. Proceedings of the National Academy of Sciences of the United States of America *96*, 1915-1920.

Kuhlbrandt, W. (2014). Biochemistry. The resolution revolution. Science *343*, 1443-1444.

Kumar, P., Singh, M., and Karthikeyan, S. (2011). Crystal structure analysis of icosahedral lumazine synthase from Salmonella typhimurium, an antibacterial drug target. Acta crystallographica Section D, Biological crystallography *67*, 131-139.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Lai, Y.T., Cascio, D., and Yeates, T.O. (2012a). Structure of a 16-nm cage designed by using protein oligomers. Science *336*, 1129.

Lai, Y.T., King, N.P., and Yeates, T.O. (2012b). Principles for designing ordered protein assemblies. Trends in cell biology *22*, 653-661.

Lai, Y.T., Tsai, K.L., Sawaya, M.R., Asturias, F.J., and Yeates, T.O. (2013). Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. Journal of the American Chemical Society *135*, 7738-7743.

Lander, G.C., Saibil, H.R., and Nogales, E. (2012). Go hybrid: EM, crystallography, and beyond. Current opinion in structural biology *22*, 627-635.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature methods *10*, 584-590.

Liao, M., Cao, E., Julius, D., and Cheng, Y. (2013). Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature *504*, 107-112.

Lorber, B., Fischer, F., Bailly, M., Roy, H., and Kern, D. (2012). Protein analysis by dynamic light scattering: methods and techniques for students. Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology *40*, 372-382.

Low, H.H., Gubellini, F., Rivera-Calzada, A., Braun, N., Connery, S., Dujeancourt, A., Lu, F., Redzej, A., Fronzes, R., Orlova, E.V.*, et al.* (2014). Structure of a type IV secretion system. Nature *508*, 550-553.

Luca, S., Heise, H., and Baldus, M. (2003). High-resolution solid-state NMR applied to polypeptides and membrane proteins. Accounts of chemical research *36*, 858-865.

Ludtke, S.J. (2010). 3-D structures of macromolecules using single-particle analysis in EMAN. Methods Mol Biol *673*, 157-173.

Marion, J.D., Van, D.N., Bell, J.E., and Bell, J.K. (2010). Measuring the effect of ligand binding on the interface stability of multimeric proteins using dynamic light scattering. Analytical biochemistry *407*, 278-280.

Mattevi, A., Obmolova, G., Schulze, E., Kalk, K.H., Westphal, A.H., de Kok, A., and Hol, W.G. (1992). Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. Science *255*, 1544-1550.

McCoy, A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. Acta crystallographica Section D, Biological crystallography *63*, 32-41.

McPherson, A. (2009). Introduction to Macromolecular Crystallography.

Milne, J.L., Wu, X., Borgnia, M.J., Lengyel, J.S., Brooks, B.R., Shi, D., Perham, R.N., and Subramaniam, S. (2006). Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy. The Journal of biological chemistry *281*, 4364-4370.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. Journal of structural biology *142*, 334-347.

Moon, A.F., Mueller, G.A., Zhong, X., and Pedersen, L.C. (2010). A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. Protein science : a publication of the Protein Society *19*, 901-913.

Navaza, J. (2003). On the three-dimensional reconstruction of icosahedral particles. Journal of structural biology *144*, 13-23.

Nie, Y., Viola, C., Bieniossek, C., Trowitzsch, S., Vijay-Achandran, L.S., Chaillet, M., Garzoni, F., and Berger, I. (2009). Getting a grip on complexes. Current genomics *10*, 558-572.

Niesen, F.H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nature protocols *2*, 2212-2221.

Ogihara, N.L., Weiss, M.S., Degrado, W.F., and Eisenberg, D. (1997). The crystal structure of the designed trimeric coiled coil coil-VaLd: implications for engineering crystals and supramolecular assemblies. Protein science : a publication of the Protein Society *6*, 80-88.

Orlova, E.V., and Saibil, H.R. (2011). Structural analysis of macromolecular assemblies by electron microscopy. Chemical reviews *111*, 7710-7748.

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.

Pellam, J.R., and Harker, D. (1962). Nobel Awards--Physics and Chemistry. Science *138*, 667-669.

Penczek, P.A. (2012). Fundamentals of three-dimensional reconstruction from projections Vol Volume 482.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C. (1960). Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature *185*, 416-422.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry *25*, 1605-1612.

Pille, J., Cardinale, D., Carette, N., Di Primo, C., Besong-Ndika, J., Walter, J., Lecoq, H., van Eldijk, M.B., Smits, F.C., Schoffelen, S.*, et al.* (2013). General strategy for ordered noncovalent protein assembly on well-defined nanoscaffolds. Biomacromolecules *14*, 4351-4359.

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

Rothemund, P.W. (2006). Folding DNA to create nanoscale shapes and patterns. Nature *440*, 297-302.

Saad, A., Ludtke, S.J., Jakana, J., Rixon, F.J., Tsuruta, H., and Chiu, W. (2001). Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination. Journal of structural biology *133*, 32-42.

Saibil, H.R. (2000). Macromolecular structure determination by cryo-electron microscopy. Acta crystallographica Section D, Biological crystallography *56*, 1215-1222.

Sara, M., and Sleytr, U.B. (1996). Biotechnology and biomimetic with crystalline bacterial cell surface layers (S-layers). Micron *27*, 141-156.

Saxton, W.O., and Baumeister, W. (1982). The correlation averaging of a regularly arranged bacterial cell envelope protein. Journal of microscopy *127*, 127-138.

Scheres, S.H., and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. Nature methods *9*, 853-854.

Schoehn, G., Hayes, M., Cliff, M., Clarke, A.R., and Saibil, H.R. (2000). Domain rotations between open, closed and bullet-shaped forms of the thermosome, an archaeal chaperonin. Journal of molecular biology *301*, 323-332.

Seeman, N.C. (1982). Nucleic acid junctions and lattices. Journal of theoretical biology *99*, 237-247.

Selmi, D.N., Adamson, R.J., Attrill, H., Goddard, A.D., Gilbert, R.J., Watts, A., and Turberfield, A.J. (2011). DNA-templated protein arrays for single-molecule imaging. Nano letters *11*, 657-660.

Sennhauser, G., and Grutter, M.G. (2008). Chaperone-assisted crystallography with DARPins. Structure *16*, 1443-1453.

Sha, B., and Luo, M. (1997). Crystallization and preliminary X-ray crystallographic studies of type A influenza virus matrix protein M1. Acta crystallographica Section D, Biological crystallography *53*, 458-460.

Shiba, K., Niidome, T., Katoh, E., Xiang, H., Han, L., Mori, T., and Katayama, Y. (2010). Polydispersity as a parameter for indicating the thermal stability of proteins by dynamic light scattering. Analytical sciences : the international journal of the Japan Society for Analytical Chemistry *26*, 659-663.

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.

Smith, M.T.J.R., J. L. (2014). Beyond blob-ology. Science *Vol. 345 no. 6197* pp. 617-619

Soon, F.F., Suino-Powell, K.M., Li, J., Yong, E.L., Xu, H.E., and Melcher, K. (2012). Abscisic acid signaling: thermal stability shift assays as tool to analyze hormone perception and signal transduction. PloS one *7*, e47857.

Stark, H., Rodnina, M.V., Wieden, H.J., van Heel, M., and Wintermeyer, W. (2000). Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation. Cell *100*, 301-309.

Streicher, S.L., and Tyler, B. (1980). Purification of glutamine synthetase from a variety of bacteria. Journal of bacteriology *142*, 69-78.

Truant, R., and Cullen, B.R. (1999). The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. Molecular and cellular biology *19*, 1210-1217.

Ullah, H., Scappini, E.L., Moon, A.F., Williams, L.V., Armstrong, D.L., and Pedersen, L.C. (2008). Structure of a signal transduction regulator, RACK1, from Arabidopsis thaliana. Protein science : a publication of the Protein Society *17*, 1771-1780.

Valentine, R.C., Shapiro, B.M., and Stadtman, E.R. (1968). Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from Escherichia coli. Biochemistry *7*, 2143-2152.

van den Ent, F., and Lowe, J. (2006). RF cloning: a restriction-free method for inserting target genes into plasmids. Journal of biochemical and biophysical methods *67*, 67-74.

Van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy *21*, 111-123.

van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M.*, et al.* (2000). Single-particle electron cryo-microscopy: towards atomic resolution. Quarterly reviews of biophysics *33*, 307-369.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. Journal of structural biology *116*, 17-24.

van Heel, M., and Schatz, M. (2005). Fourier shell correlation threshold criteria. Journal of structural biology *151*, 250-262.

Volk, R., and Bacher, A. (1991). Biosynthesis of riboflavin. Studies on the mechanism of L-3,4-dihydroxy-2-butanone 4-phosphate synthase. The Journal of biological chemistry *266*, 20610-20618.

Weis, K., Dingwall, C., and Lamond, A.I. (1996). Characterization of the nuclear protein import mechanism using Ran mutants with altered nucleotide binding specificities. The EMBO journal *15*, 7120-7128.

Whitesides, G.M., Mathias, J.P., and Seto, C.T. (1991). Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. Science *254*, 1312-1319.

Wilcox, S.K., Putnam, C.D., Sastry, M., Blankenship, J., Chazin, W.J., McRee, D.E., and Goodin, D.B. (1998). Rational design of a functional metalloenzyme: introduction of a site for manganese binding and oxidation into a heme peroxidase. Biochemistry *37*, 16853-16862.

Wilson, W.W. (2003). Light scattering as a diagnostic for protein crystal growth--a practical approach. Journal of structural biology *142*, 56-65.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta crystallographica Section D, Biological crystallography *67*, 235-242.

Wittig, I., and Schagger, H. (2005). Advantages and limitations of clear-native PAGE. Proteomics *5*, 4338-4346.

Wong, W., Bai, X.C., Brown, A., Fernandez, I.S., Hanssen, E., Condron, M., Tan, Y.H., Baum, J., and Scheres, S.H. (2014). Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. eLife, e03080.

Woolfson, D.N., and Alber, T. (1995). Predicting oligomerization states of coiled coils. Protein science : a publication of the Protein Society *4*, 1596-1607.

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

Wu, S., Avila-Sakar, A., Kim, J., Booth, D.S., Greenberg, C.H., Rossi, A., Liao, M., Li, X., Alian, A., Griner, S.L.*, et al.* (2012). Fabs enable single particle cryoEM studies of small proteins. Structure *20*, 582-592.

Yeates, T.O., and Padilla, J.E. (2002). Designing supramolecular protein assemblies. Current opinion in structural biology *12*, 464-470.

Yu, H. (1999). Extending the size limit of protein nuclear magnetic resonance. Proceedings of the National Academy of Sciences of the United States of America *96*, 332-334.

Yu, X., Jin, L., and Zhou, Z.H. (2008). 3.88 A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. Nature *453*, 415-419.

Zachariae, U., and Grubmuller, H. (2008). Importin-beta: structural and dynamic determinants of a molecular spring. Structure *16*, 906-915.

Zhang, X., Jin, L., Fang, Q., Hui, W.H., and Zhou, Z.H. (2010). 3.3 A cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. Cell *141*, 472-482.

Zhang, X., Konarev, P.V., Petoukhov, M.V., Svergun, D.I., Xing, L., Cheng, R.H., Haase, I., Fischer, M., Bacher, A., Ladenstein, R.*, et al.* (2006). Multiple assembly states of lumazine synthase: a model relating catalytic function and molecular assembly. Journal of molecular biology *362*, 753-770.

Zhang, X., Meining, W., Fischer, M., Bacher, A., and Ladenstein, R. (2001). X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 A resolution: determinants of thermostability revealed from structural comparisons. Journal of molecular biology *306*, 1099-1114.

Zhang, X., Settembre, E., Xu, C., Dormitzer, P.R., Bellamy, R., Harrison, S.C., and Grigorieff, N. (2008). Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. Proceedings of the National Academy of Sciences of the United States of America *105*, 1867-1872.

Zhao, P. (2011). The 2009 Nobel Prize in Chemistry: Thomas A. Steitz and the structure of the ribosome. The Yale journal of biology and medicine *84*, 125-129.

Zheng, H., Hou, J., Zimmerman, M.D., Wlodawer, A., and Minor, W. (2014). The future of crystallography in drug discovery. Expert opinion on drug discovery *9*, 125-137.

Zhou, Z.H. (2008). Towards atomic resolution structural determination by single-particle cryo-electron microscopy. Current opinion in structural biology *18*, 218-228.